The Transfer of Digitisation Knowledge and Practice from Large Organisations to Small: The National Library of Wales to the Centre for Performance Research

Clare Wood-Fisher

Department of Information Studies

Aberystwyth University, UK

Submitted in partial fulfilment of the requirements of the Degree of Master of Philosophy

2012

Abstract

The aim of this research was to investigate how the knowledge of digitisation practices and connected standards may be transferred from a large organisation to a small one through the digitisation of special collections. In this project the large organisation was the National Library of Wales and the small the Centre for Performance Research at Aberystwyth University.

Using the 'learning by doing' paradigm, the techniques required to carry out a digitisation project were taught at the National Library of Wales where two collections were digitised and their metadata written. Learning was then demonstrated by independently digitising a collection in CPR.

To facilitate the transfer of the learned information two guides were written. One addressed the requirements of project organisers by describing all the phases of a project at a generic level. The other gave the specific directions needed to use the equipment in CPR. The guides were evaluated by staff from the Centre for Performance Research for their usability.

Research sponsorship included an allowance to attend selected outside events. This allowed additional information to be gathered and added to that gained from the NLW and by talking to selected individuals within the information profession.

The research showed that digitisation practices were transferred, and specific issues on the use of metadata and digital objects discussed. Recommendations on the next developmental steps for CPR were outlined.

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed .... (candidate)

Date ....

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where ***correction services** have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references.

A bibliography is appended.

Signed .... (candidate)

Date ....

[*this refers to the extent to which the text has been corrected by others]

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .... (candidate)

Date ....

Acknowledgements

I wish to thank all of the staff at the National Library of Wales and Aberystwyth University for the support they have given me while undertaking this research.

I particularly wish to thank my supervisors, Dr. Menna Morgan at the National Library of Wales, Professor Richard Gough, Sarah Higgins, Dr Amy Staniforth and Lucy Tedd at Aberystwyth University for giving freely of their help, encouragement and support. I would like to give special thanks to Dr Staniforth as it was due to her that this research was instigated.

I wish to thank all the staff at the National Library of Wales who gave me instruction in the digitisation skills I needed to learn to carry out this research. Among the many Julia Davies, Galen Jones, Will Troughton and Scott Wabey in particular patiently answered my many questions and helped me correct my mistakes. I also wish to thank all the staff in CPR. They made me feel very welcome and supported my research with their particular expertise.

Last, but not least, I want to thank the participants who consented to be interviewed and freely shared their knowledge with me.

This research was funded by the European Social Fund.

Contents

List of Figures

List of Tables

Acronyms and Initials Used in the Research

INITIALS	WHAT THEY STAND FOR	WEB ADDRESS (checked 16/02/2012)
AACR2	Anglo-American Cataloguing Rules	http://www.aacr2.org/
AHDS	Arts and Humanities Data Service	http://www.ahds.ac.uk/
AHRC	Arts and Humanities Research Council	http://www.ahrc.ac.uk/Pages/default.aspx
ANDS	Australian National Data Service	http://www.ands.org.au/
AONS	Automatic Obsolescence Notification System	
BL	British Library	http://www.bl.uk/
CADAIR	Aberystwyth University Repository name	http://cadair.aber.ac.uk/
CASIS	Centre for Advanced Software and Intelligent Systems	http://www.aber.ac.uk/casis/
CCD	Charge-Coupled Device	
CLT	Cardiff Laboratory Theatre	http://www.thecpr.org.uk/archive/ gallery_details.php?galleryID=25
CMYK	Cyan, Magenta, Yellow, Key (Black)	
CPR	Centre for Performance Research	http://www.thecpr.org.uk/
DCC	Digital Curation Centre	http://www.dcc.ac.uk/
DCMES	Dublin Core Metadata Element Set	http://dublincore.org/documents/dces/
DPC	Digital Preservation Coalition	http://www.dpconline.org/
dpi	Dots per inch	
DROID	Digital Record Object Identification	http://www.nationalarchives.gov.uk /information-management/our-services/dc-file-profiling-tool.htm
DSpace	Open source repository software (used by CADAIR)	http://www.dspace.org/
ECLAP	European Collected Library of Artistic Performance	http://bpnet.eclap.eu/drupal/?q=en-US/home
FOSS	Free and Open Source Software	
Greenstone	Open source repository software	http://www.greenstone.org/

INITIALS	WHAT THEY STAND FOR	WEB ADDRESS (checked 16/02/2012)
HEIs	Higher Education Institutions	
ICA Atom	Open source repository software	http://ica-atom.org/
IC-ININFO	International Conference on Integrated Information	http://history.icininfo.net/2011/
ICT	Information Communication Technology	
InterPARES	The International Research on Permanent Authentic Records in Electronic Systems	http://www.interpares.org/
JISC	Joint Information Systems Committee	http://www.jisc.ac.uk/
LC	Library of Congress	http://www.loc.gov/index.html
LCSH	Library of Congress Subject Headings	
Mac or iMac	Apple Mac computer	
MADS	Metadata Authority Description Schema	http://www.loc.gov/standards/mads/
MARC	Machine-Readable Cataloguing	http://www.loc.gov/marc/status.html
METS	Metadata Encoding and Transmission Standard	http://www.loc.gov/standards/mets/ mets-home.html
MODS	Metadata Object Description Schema	http://www.loc.gov/standards/mods/
NLS	National Library of Scotland	http://www.nls.uk/
NLW	National Library of Wales	http://www.llgc.org.uk/
NOF-digitise	New Opportunities Fund-Digitise project. Digitisation projects funded by the Heritage National Lottery	http://www.nof-digitise.org/launch.htm
OCR	Optical Character Recognition	
OPF	Open Planets Foundation	http://www.openplanetsfoundation.org/
PC	Personal computer	
PLANETS	Preservation and Long Term Access Through Networked Services	http://www.planets-project.eu
PLATO	Tool to implement PLANETS preservation suite.	http://www.planets-project.eu/software/

INITIALS	WHAT THEY STAND FOR	WEB ADDRESS (checked 16/02/2012)
ppi	Pixels per inch	
PRONOM	Database of file formats and software products	http://www.nationalarchives.gov.uk/PRONOM/Default.aspx
RCAHMW	Royal Commission for Ancient and Historical Monuments Wales	http://www.rcahmw.gov.uk/
RDF	Resource Description Framework	http://www.w3.org/RDF/
RGB	Red, Green, Blue	
TFTS	Theatre Film and Television Studies Department	http://www.aber.ac.uk/en/tfts/
TIDSR	Toolkit for the Impact of Scholarly Resources	http://microsites.oii.ox.ac.uk/tidsr/
TNA	The National Archives	http://www.nationalarchives.gov.uk/
TRAC	Trusted Repositories Audit & Certification	http://www.dcc.ac.uk/resources/tools-and-applications/trustworthy-repositories
URI	Uniform Resource Identifier	
Wombat	Software used by NLW to manage digitisation workflow	
XML	eXtensible Markup Language	http://www.w3.org/XML/

Table 1: Table of Acronyms and Initials Used in the Research

Terms and Phrases and How they are Used in the Research

This list includes terms used in the thesis and the written guides produced as part of the research. The terms from the written guides have been included. The written guides can be found in Appendix B: Organiser's Guide and Appendix C: CPR Quick Start Pages.

Term	**Definition**
Additive Mixing	Where colours (light) are added together to obtain other colours (Langford & Bilissi, 2011, p. 445)
Archival Digital Objects	The digital object that has the largest file size and is used to make other copies from.
Archival Material	What you start with e.g. Photographs or video tape.
Colour Temperature	Is the colour of light given off by a black body radiator, an ideal source, at a particular temperature in Kelvin (Bermingham, 1989, p. 11)
Curation	Preserving digital objects for active access (Digital Curation Centre, 2010c)
Data	Raw unprocessed facts about something (BBC, 2012)
Digital Object(s)	Are what you get when you have scanned the archival material.
Digitisation	The process which changes archival materials into digital objects (Digital Preservation Coalition, 2009b)
Dublin Core Metadata Element Set (DCMES)	The metadata scheme used for this collection. It has 15 parts to be filled in. Each of which can be repeated or left out, unless it is made mandatory locally.
Format	The physical manifestation of a resource. This could be paper, a DVD, a digital file etc.
Information	Facts stored in a coherent manner about something (BBC, 2012)
Interpolation	Where pixels without information surrounding an image are completed by the software making 'a best guess' (Langford & Bilissi, 2011, p. 459)
Knowledge	Information re-accessed and re-processed for use in an appropriate form (BBC, 2012)
Metadata	Data compiled about archival materials and/or digital objects and used to manage it (Rowley & Hartley, 2008, p. 42).
Native Resolution	The fixed resolution the scanner is designed to display images at (Langford & Bilissi, 2011, p. 187).
Open Source Software (OSS)	Applied to software resources whose source code is freely available. Sometimes this is available at no cost, sometimes not (Open Source Initiative, n.d.).
Original Source	Same as archival material
Preservation	An act designed to allow archival materials or digital objects to be kept for a long time (InterPARES, 2011, p. 39).

Term	Definition
Quality Assurance (QA)	The process of checking that a digital object and its metadata meet minimum standards
Representational Digital objects	The group name for the viewing and thumbnail copies of the digital objects.
Resolution	How well an imaging system distinguishes between adjacent points. The better it does this, or the closer together the points the better the resolution (Langford & Bilissi, 2011, p. 465)
Resource	An item, or a group of items that are a unit, in the collection.
Subtractive Mixing	Where adding a colour (cyan, magenta & yellow) subtracts those tones from white light (Langford & Bilissi, 2011, p. 466)
Thumbnail digital object	A low quality, small file size digital object created to give an impression of the viewing digital object
Viewing Digital object	The digital object created for viewing the resource

Table 2: Table of Terms Used in the Research

Chapter 1: Introduction

1.1 Project Background

The aim of this research was to investigate how the knowledge of digitisation practices and connected standards may be transferred from a large organisation to a small one through the digitisation of special collections. This was done by using the acknowledged expertise of the National Library of Wales (NLW) and investigating ways in which it could be transferred to smaller, often isolated, collections. The collection to be used to test this transfer of skills knowledge was part of the Centre for Performance Research (CPR) Archive. The transfer was to be achieved through a combination of technical and theoretical training.

The person specification in the advert for this research project indicated that the Supervisors were looking for someone who had either information, or performance related experience. Having just completed an MSc Econ Records and Information Management and previously a Stage Manager for amateur theatre companies; I had a foot in both camps. I had developed an interest in the management and preservation of digital objects when I worked in a Further Education college learning resource centre. This interest was prompted by the discovery that a colleague had previously worked on a number of Joint Information Systems Committee (JISC) digitisation projects.

1.1.1 KESS

The research was funded under the Knowledge Economy Skills Scholarship (KESS) programme. KESS is part-funded by the European Social Fund through the European Union's Convergence Programme. Selection is limited to those students who either live, or work, in specific parts of Wales. As part of the scholarship, students undertake the study of 30 credits of postgraduate skills courses and spend a period in a partner institution. Any type of business can be a KESS partner, but an on-going benefit to the community from the work must be demonstrated and the research focussed on the partner needs. The partner institution and the University both contribute financially to the scheme. There are strict time limits imposed so that the aims of the study and the sponsorship may be achieved (Cardiff University, n.d.).

For this project funding from KESS included:

- £3,000 total equipment budget with a single item limit of £1,300
- £1,000 for travel to conferences and other similar opportunities related to the research
- A stipend of £9,000, and
- All University fees.

1.1.2 Management of Project

The structure of the supervisory side of this project was a little different to the usual structure at Aberystwyth University. Because KESS projects have a link with industry I had a supervisor at the NLW in addition to academic supervisors and the Archive Collections Director in CPR. The academic supervisors came from the two different University academic departments involved in the project: the Department of Theatre, Film and Television Studies (TFTS) and the Department of Information Studies. Each Supervisor also had their own specialist area of interest relevant to the research. For this research the Director of CPR and the two Supervisors from the Department of Information Studies oversaw my academic supervision and the production of this thesis. There were formal meetings with my Supervisors on a monthly basis, and less formal contact at least weekly. The monthly meetings discussed academic topics for two months, and every third month was the required quarterly meeting. Between these meetings, timesheets were required from everybody involved.

For my course, the 30 study credits to achieve the Postgraduate Skills Development Award were achieved by attending a four day postgraduate research training course, and a two day residential course. The four day course was assessed via a presentation and a written assignment. The two day course was not formally assessed as the aim was to improve personal skills in a range of communication based skills. In addition I asked to attend the three day University Postgraduate Writing course.

1.2 Aim

The aim of this research was to investigate if the elements that comprise digitisation practice, and their associated standards, may be transferred between large and small institutions. The transposed techniques were then to be tested by applying them to the digitisation of special collections.

1.3 Objectives

The main steps completed to achieve the results in this research process were:

- Acquire the practical knowledge to carry out a digitisation project by digitising two collections at National Library of Wales (NLW)
- Research the available literature on digitisation and other relevant areas, e.g. photography
- Interview selected professionals to gain an insight into their views on the past, present and possible future of digitisation
- Transfer the acquired learning to Centre for Performance Research (CPR) via a written guide
- Test the written guide by digitising a pre-selected collection at CPR
- Evaluate the use of the written guides.

1.4 Reasons for Research

The literature review shows that there have been academic papers that describe how institutions large and small carried out their respective digitisation projects; but none describe the transfer of techniques directly between themselves and other bodies. There are also few writers who have described the transfer of the fundamental principles the area of information takes for granted. It was also observed that many of these projects were completed on archival materials under the direct control of the information sector. The archive in CPR stands outside the direct control of the University libraries and archives although they collaborate on many projects. This project set out to discover if the techniques and associated values of good digitisation practice could be transferred with a system that could continue to be utilised.

1.5 Limitations on Research

This research had limitations placed upon it by:

- the time allowed,
- the budget available,
- the boundaries of the subject area,
- what could be digitised within the allowed time, and
- Ethical considerations as described in Section 4.2.

1.6 Definitions

The nature of this type of research means that some specific technical terms are used. There are explanations of some of these in Appendix A: Technical Background. However, in his paper, Reid (2000, pp. 143-144) suggests that terms should have agreed meanings to avoid confusion. Taking this idea forward, these are some of the key ones used in this research.

1.6.1 Knowledge

Knowledge is generally accepted as being the active use, or re-use, of information in an appropriate manner. The BBC BiteSize web pages define knowledge as (BBC, 2012):

> *the ability to understand information and to then form judgements, opinions, make predictions and decisions based on that understanding*

This definition is a generally accepted definition of knowledge. This definition was important to this research because receiving data, changing it into information, and passing on that knowledge was central to the aim of developing and testing a model to transfer digitisation practices.

1.6.2 Archive

The International Research on Permanent Authentic Records in Electronic Systems (InterPARES) is a large, international, collaborative project that aims to ensure the long term preservation of authentic records. It is currently working on its third series of topics and is an important project in the archives and records management fields (InterPARES, 2010). It concentrated on the individual digital record as the object and defined an archive as:

> *v., To save digital data, documents, and records, typically those that are not current, offline. [Computer and Information Sciences] (InterPARES, 2011, p. 6)*.

The Open Archival Information System (OAIS) was a project that described a high level model of the elements that comprise a system to preserve information for the long term, whether in a digital or analogue format (British Standards Institute, 2003, p. 13).

Their definition was a high level holistic view and defined an archive as:

> *An organization that intends to preserve information for access and use by a Designated Community (British Standards Institute, 2003, p. 20)*.

The definition of the word ‘archiving’ has been under scrutiny since the ICT industry coined the word to mean “*storing back-up copies of computer files*” (Reid, 2000, p. 144). Reid, an archivist in Powys, Wales, when the paper referenced was written, goes on to explain that in his view, within the archive profession the term archiving means:

> *the process of evaluating records worthy of permanent preservation and taking measures for their long-term storage*.

InterPARES, in comparison, defined archives three different ways depending upon the point of view being expressed by its use:

> *n., The whole of the records of a creator. or*

> *n., An agency or institution responsible for the preservation and communication of records selected for permanent preservation. Or*

> *n., A place in which records selected for permanent preservation are kept*.

In this research it was used to represent the place where archival materials selected for preservation were kept. It is important that this term was defined to eliminate confusion that may be caused by adopting other definitions.

1.6.3 Archival Material

The International Council on Archives (ICA) exists to promote good archival practice by representing its members in high level international situations such as at UNESCO (International Council on Archives, 2009). In its introduction to the General International Standard Archival Description (ISAD(G)) ICA use the term archival materials twice, but do not explain it in the glossary (International Council on Archives Ad Hoc Commission on Descriptive Standards, 2010, p. 7). The way the word was used implies strongly that the term was applied to the objects that are described in an archive.

This term was used throughout this research to refer to the material that is being digitised whatever the original format. This was important as it avoids suggesting that the processes described apply only to particular formats of materials in the archive.

1.6.4 Digitisation

In her book published in 2003 Hughes (p. 4) stated that

> *digitization is the process by which analogue material is converted into a sequence of 1s and 0s and put into binary code to be readable by a computer.*

And the Digital Preservation Coalition (DPC) as (2009b):

> ***Digitisation:*** *The process of creating digital files by scanning or otherwise converting analogue materials. The resulting digital copy, or digital surrogate, would then be classed as digital material and then subject to the same broad challenges involved in preserving access to it, as "born digital" materials*.

Both of the above definitions concentrated upon the practical processes associated with digitisation. However, the definition by Hughes drew attention to the fact that digitised material needs additional hardware to allow access which the analogue original does not.

In this research, digitisation was used to describe the process of turning archival materials into digital objects. It is an important definition because it limits the range of the word to the scanning or photographing process.

1.6.5 Digital Object

Digital object was defined in the Joint Information Systems Committee (JISC) *Beginners Guide to Digital Preservation* (JISC, n.d.) as encompassing all digital materials. This definition also distinguished between born digital materials and digital files created by the transformation of archival materials. This definition means that digital objects created through digitisation are the digital surrogates of archival materials.

In the InterPARES Project Glossary digital object is given as (InterPARES, 2011, p. 17):

> *n., A discrete aggregation of one or more bit streams and the metadata about the properties of the object and, if applicable, methods of performing operations on the object. [General Dictionaries].*

Using this definition indicates that the creation and care of digital objects follows the same principles without reference to the format. It was used in the same manner in this project.

1.6.6 Digital Preservation

Although this research did not include preservation activities explicitly, its definition is included here to avoid the confusion that sometimes arises between this and digitisation. The DPC define digital preservation as (2009b):

> *Refers[ing] to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. Digital preservation ....refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technological change. Those materials*

> *may be records created during the day-to-day business of an organisation;"born-digital" materials created for a specific purpose (e.g. teaching resources); or the products of digitisation projects.*

This definition encompasses the holistic nature of the management of digital files beyond creation.

1.7 Large UK Organisations Involved in Digitisation

Particular organisations have had a large effect on digitisation practice in UK education through their size and the work they oversaw and funded. An outline of some of those whose resources were consulted during the research are in this section. These resources were used to research applicable standards and the advice given to small organisations. I then extrapolated from this general advice to apply to the research situation.

The strap line on the first page of the JISC web site introducing their functions states that:

> *JISC is the UK's expert on information and digital technologies for education and research (JISC, 2012a)*

On the page describing their programmes they described how their digitisation strategy funds projects that fall into three categories. Creating resources that:

- would not otherwise be made
- would otherwise be lost to education
- contribute to a critical mass of resources for an area

They go on to list that they have sponsored 15 programmes that have created 151 projects. Funding of nearly £30 million was listed for these programmes in the yearly highlights from 2004 (JISC, 2012b).

JISC Digital Media is a section of JISC whose resources were consulted extensively throughout this research. Formerly known as TASI it aims to help Higher Education (HE) and Further Education (FE) institutions use digital resources to the fullest degree by achieving effective solutions (JISC Digital Media, 2012).

The Arts and Humanities Data Service (AHDS) was a body funded by JISC and the Arts and Humanities Research Council (AHRC). It existed to

> *collect, preserve and promote the electronic resources which result from research and teaching in the arts and humanities (Arts and Humanities Data Service, 2005)*.

Its main focus was on the promotion of shared standards and integrated access. The web site is now maintained by UKOLN because the AHDS ceased to exist in 2008. They have a portfolio of 42 projects listed by completion dates and 36 by topic, with dates running from 2003 to 2008. Some of the contents of the two lists overlap and some are not strictly digitisation projects, although all relate to the general area of the digitisation and management of digital resources (Arts and Humanities Data Service, 2008).

The Digital Curation Centre (DCC) describes itself as:

> *the UK's leading hub of expertise in curating digital research data* (Digital Curation Centre, 2010d).

The DCC was established in 2004 in response to a recommendation by JISC. It was part of a consortium designed to solve digital curation issues (Digital Curation Centre, 2010a). Its third strand of activities are designed to increase the resources available and the skills base in higher education establishments. It is due to end in 2013 and concentrates on research data management (Digital Curation Centre, 2010a).

UKOLN is "*a centre of expertise in digital information management, providing advice and services*" (UKOLN, 2010) for the education and information community. It is funded by JISC with funding for specific projects raised from various sources. Support for its infrastructure requirements is received from Bath University (UKOLN, 2011).

The Digital Preservation Coalition (DPC) describes itself as an *"advocate and catalyst for digital preservation"*. Their main emphasis is on training and supporting practice development in the public and private sector (Digital Preservation Coalition, 2009c).

The relationship between these organisations is illustrated in Figure 1 where a line signals an interchange of ideas.

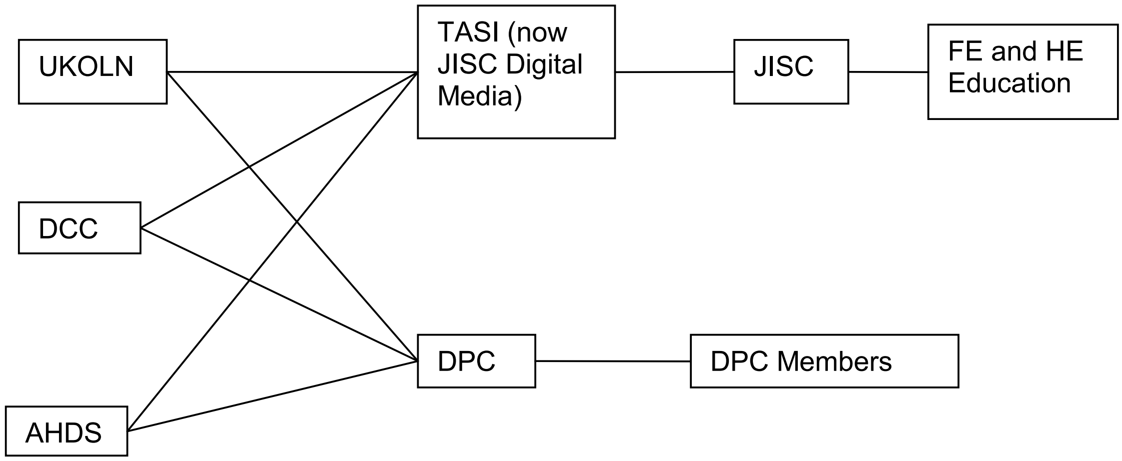


Figure 1: Diagram of the Relationship between Organisations Involved in Digitisation

1.8 Structure of Research

Explanation of the background to the partner organisations involved in this research situation is given in Chapter two of this thesis. Following this, Chapter three is a review of the literature, Chapter four describes the Research Design and Chapter five the research results. The discussion of the results is in Chapter six and the conclusions drawn from them are in Chapter seven. A Bibliography is followed by the Appendices. The topics in the Appendices are presented in the order in which they are first mentioned in the thesis. Appendix B: Generic and Organiser's Guide and Appendix C: CPR Quick Start Pages are the written guides referred to throughout the thesis. They are written so that there is a minimum of technical language included in them to make them as accessible as possible. The page numbers continue without any breaks throughout the thesis. Chapters and sections of chapters in the thesis are numbered sequentially whilst those in the appendices have the appendix letter in front. Figures and tables appearing in the appendices also have the relevant appendix letter before the number.

1.9 Chapter Summary

The research, which has been sponsored by KESS, aims to discover if digitisation knowledge can be transferred from a large institution to a small. The process involved the researcher being embedded in the NLW to learn the applicable skills and to transfer them to CPR. The research gains its validity from its emphasis on the transfer of the underpinning values of digitisation practice.

Chapter 2: Background to Project Organisations

The two partner institutions in this project were the NLW and Aberystwyth University. At Aberystwyth University the main departments involved were CPR, which is located in the Department of Theatre, Film and Television Studies (TFTS) and the Department of Information Studies. Support was also received from the Information Services Department.

The European Collected Library of Performance (ECLAP) is described here because for CPR it had an influence upon the research outcomes.

2.1 National Library of Wales

The first Royal Charter for the NLW was granted on the 19th March 1907. Sir John Williams is credited with being the moving force behind its location in Aberystwyth in a building overlooking the town. The foundation stone of the building designed by Sidney Kyffin Greenslade to echo the Classical style (BBC, 2007), was laid in 1911 (National Library of Wales, 2011a).

Among its required duties the NLW houses the Welsh National Archive as well as being one of the national legal deposit libraries. It has a large staff whose working lives are devoted to managing, preserving and maintaining access to the many collections, both large and small, that constitute the archive and library.

The NLW collections are focussed on Welsh cultures with its role defined "*as a collector and guardian of the intellectual record of the life of Wales and the Welsh people*" (National Library of Wales, 2011c, p. 3). The original collection was extended considerably by Sir John Williams who promised to leave his large collection of books and manuscripts to the library if it was situated in Aberystwyth. This Welsh Collection was previously kept by, what was then, the University College of Wales (National Library of Wales, 2011a). As well as books and manuscripts, the Library houses a large collection of paintings and the first book printed in Welsh. In 2008 (Jones) the complete collections were described as comprising 5,000,000 printed items plus 15km of archives.

The National Screen and Sound Archive was established in the Library in 2001 when the Wales Film and Television Archive was merged with the National Library of Wales' Sound and Moving Image Collection. The collections included feature films, TV and educational productions as well as oral history recordings, radio broadcasts and music recordings (National Screen and Sound Archive Wales, 2011).

NLW has had a separate digitisation strategy since 2001 and a digital preservation policy and strategy since 2003 (Jones, 2008). Figure 2 shows that digitisation enclosed or interacted with most Library systems, except for its deposit library and statutory records management functions. For users the diagram shows that there are two initial points of contact with collections; via the catalogue or via Web exhibitions.

In this systems map the parts of the Library that were drawn completely inside the boundary of another, operated within that part of the library. Sections that were intersected by the boundaries of other parts of the Library interacted with that part of the Library. The amount of overlap was not an indicator of the extent of the interaction in these diagrams. So in Figure 2, the creation of digital object metadata and digitisation are completely within the digitisation workflow element.

2.1.1 Equipment Available in NLW During Research

The equipment available in the NLW reflects its commitment to digitisation. Within their digitisation suite I can remember there being seven scanners of differing types in the room in which I worked. There was a second room separate from my base which contained more scanners and their operators.

In his article, Jones cites the number of posts in the digital developments department as 41 in 2007. Of these he indicates that seven are directly involved in digitisation of archival materials and 15 creating metadata (2008, p. 104). However he goes on to state that 11 people were employed on the Welsh Journals Online project when he was writing the article (2008, p. 112).

The Welsh Journals Online project was funded by JISC and the Welsh Assembly Government. It was one of the largest major, high quality, digitisation projects in

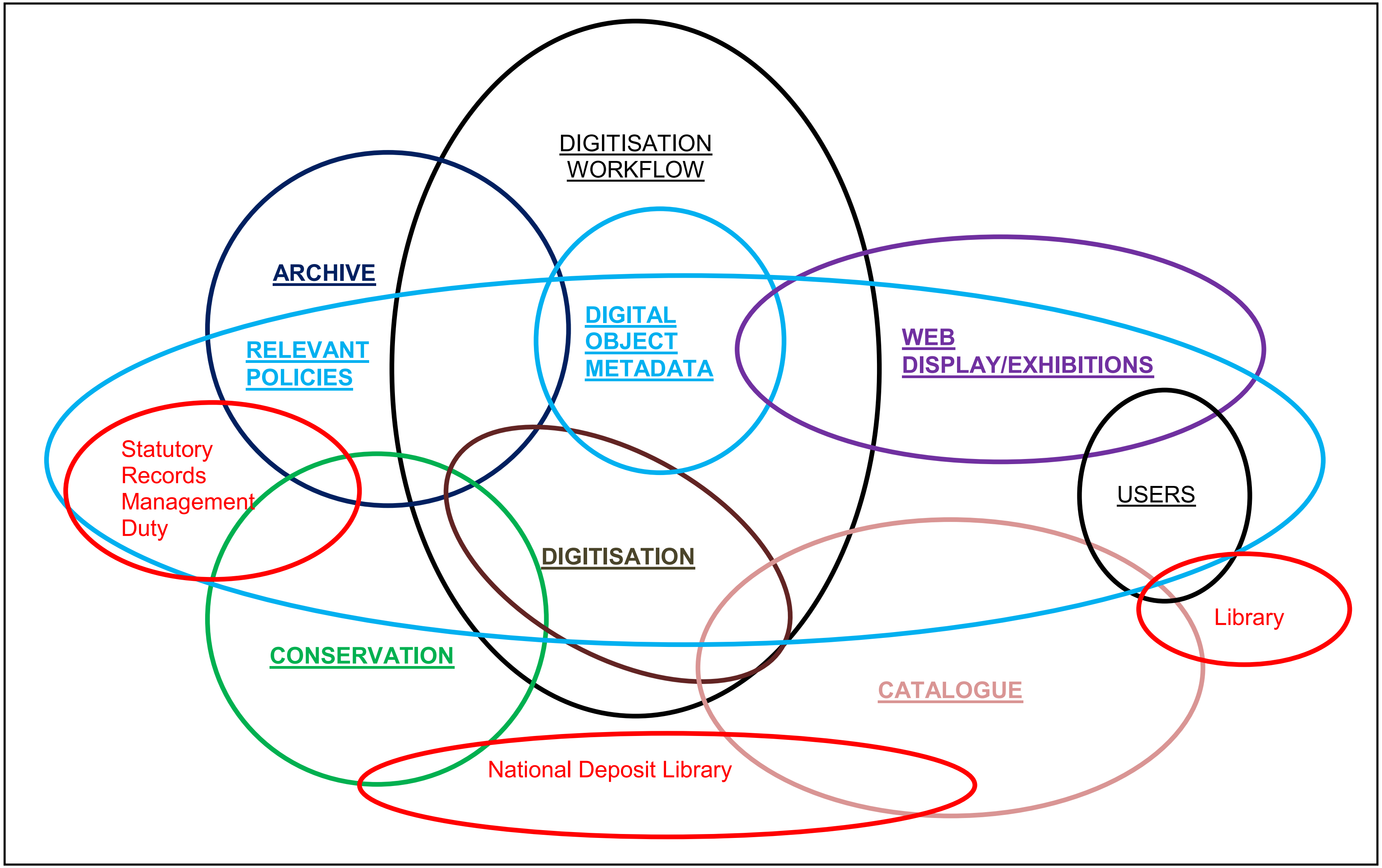


Figure 2: Systems Map of NLW

which the Library has taken part. It has involved scanning and writing the metadata for 50 journals, including their back copies. The project does not include born digital journals. The digital objects created were made fully searchable by utilising Optical Character Recognition (OCR) (National Library of Wales, 2011b).

Another recently completed project is the digitisation of all the newspapers and journals published in Wales and out of copyright up to 1911. It is expected that the resources created from this project will be available in 2012 (National Library of Wales, 2011b).

Since the first digitisation strategy at the NLW in 2001 it has taken part in a number of digitisation projects that include:

- Welsh Biography Online
- Early English Books Online
- The Probate Project
- St Asaph Notitiae
- The Black Book of Carmarthen
- The White Book of Rhydderch
- The Book of Llandaf
- Gathering the Jewels with Culturenet Cymru
- The Ohio Project
- The Portread (Welsh Portraits) Project.

Some of these, like Gathering the Jewels, have their own web site. Many are featured in the Digital Mirror area of the NLW web site. The Digital Mirror is an area of the web site that showcases chosen digital objects from selected projects.

2.1.2 Collections Digitised

The two collections digitised by the researcher for the study at the NLW were:

- The HW Lloyd Collection, which comprised around 400 glass negatives dating from roughly 1900 – 1920. The collection of glass plates contained images of people and events in Bala, North Wales, and prisoners of war held in a nearby camp.

- The more contemporary Dwynwen Belsey Collection which contained in the region of 180 postcards, black and white photographs and colour slides dated approximately 1955 – 1965. The Dwynwen Belsey Collection had images from trips to the Welsh colony in Patagonia, South America, and dated from the 1950s and 1960s.

2.2 Centre for Performance Research (CPR)

The CPR described itself on its web site as "*a multi-faceted theatre organisation located and rooted in Wales, working nationally and internationally*" (Centre for Performance Research, 2011). It was established in 1988 as the successor to Cardiff Laboratory Theatre (CLT). At the time of the research it had three full time posts occupied by four staff: the Centre Director, the Collections Director who is a qualified archivist, and two administration posts. One administration post focused upon production of the journal *Performance Research* and the other, a wider range of tasks that included selling books produced by the department.

CPR specialised in "*contemporary visual and experimental performance*" based collections (Centre for Performance Research, 2009, p. 1) which were organised into four main areas. These areas are:

- The International Theatre Collection
- Performance Research
- Cardiff Laboratory Theatre
- Giving Voice Archive

The CPR collections were housed within TFTS at Aberystwyth University. Unlike some University collections, they are not controlled by the University Information Services or Archives departments, but had their own staff and building.

CPR has an entry in the National Register of Archives, the Archon Directory. There are more than 44,000 entries for collections in the UK and abroad in the index. Each entry contains outline details supplied by the location and is updated whenever new data is received. Most collections list the nature of their resources as well as their locations and opening times. Many locations also supply a copy of their catalogue.

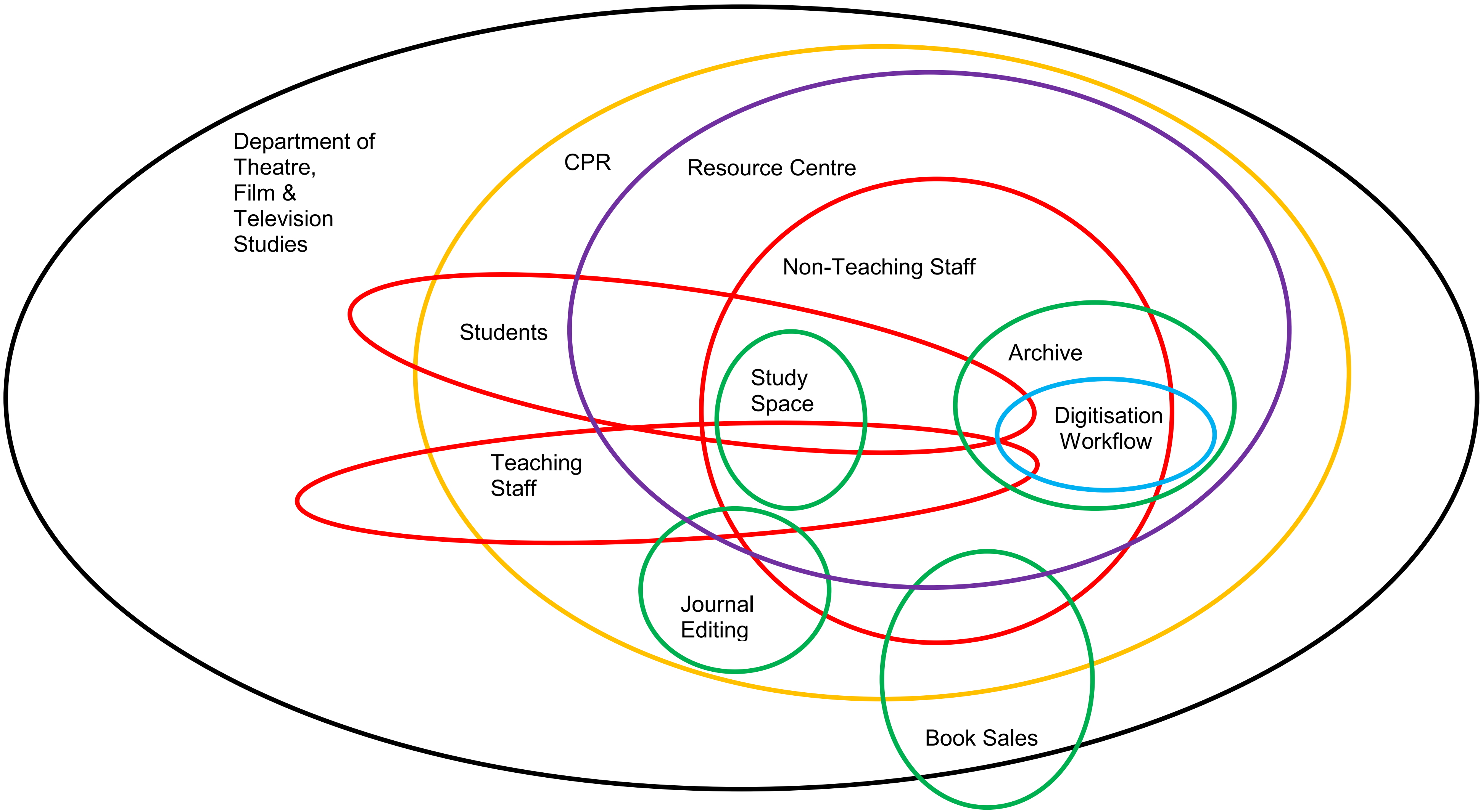


Figure 3: Systems Map of CPR

In addition to that supplied by a site, supplementary information is taken from referenced printed sources (The National Archives, n.d.c).

Each Archon entry is given a number, for CPR this is GB3261. To be included in the register there is a requirement to lodge a collection preservation policy with the TNA.

As the systems map shows (Figure 3) completing a digitisation workflow is only one of the functions of the Centre. It has not undertaken a digitisation project before.

2.2.1 Equipment Available in CPR at Beginning of Research

At the start of the research CPR had a flat bed scanner, a Bookeye, attached to a PC. It was expected that this equipment could be used as a starting point for the research. There was also a Mac in an adjacent office with software suitable for editing digital objects. I could use this computer when it was not being used by department staff for their work.

2.2.2 The International Theatre Collection

The International Theatre Collection contains items relating mainly to musical theatre and world theatre, historic and contemporary theatre. This includes items in their original language. These archival materials and their variety of formats significantly add to the identity of CPR (Centre for Performance Research, 2011).

its unique value is in the breadth of subject and the diversity of subject is matched by the languages and cultures represented, (Gough, 2009, p. 1)

and houses the life's collections of John Cavanagh. It consists of 22,000 printed texts and 8,000 pieces of theatre ephemera (Staniforth, 2010, p. 2).

2.2.3 Performance Research

Performance Research is an international journal whose emphasis is on contemporary performance. It is now published six times a year. The archives contain all the published papers, images and related correspondence, including the research for themed issues since its formation in 1993 (Centre for Performance Research, 2011).

2.2.4 Cardiff Laboratory Theatre

This collection contains all the documentation relating to the productions of Cardiff Laboratory Theatre and its successor CPR. There are recordings of workshops, festivals and conferences and documented research into themes for the outputs. The resources date back to 1974 and have continued with the resources contributed by CPR (Centre for Performance Research, 2011).

2.2.5 Giving Voice Archive

Giving Voice is a collection created from the biennial workshops. The collection has a wide audience appeal, as it shows the performances, workshops and talks that have taken place at the ten festivals.

2.2.6 Collection Digitised

The title of the collection digitised at CPR was The Blaenau Ffestiniog Residency 1981, also known as BF1981. This collection sub-divided into four areas that were given the names:

- The Dragon Procession
- The Dragons Wedding
- The Widow’s Dream, and
- Cabaret and Workshops

The contents of the collections were photographic slides and photographs. Some were duplicates, in both colour and black and white, with the photographic prints taken from slides. As far as I am aware, this was the total extent of the collection.

The residency in Blaenau Ffestiniog took place over three weeks in September and October 1981. While there, CLT co-ordinated a series of community based projects that had three recorded production outputs. Professionals taking part were from the CLT and four members of the Odin Teatret companies. Table 3 shows a summary of the productions and their contribution to the archive holdings.

Collection Section	Contents	Story
The Dragon Procession	19 slides & 8 photographs	Cardiff Laboratory Theatre took The Dragon Procession to a variety of Welsh Towns during the summer of 1981. Many of the images in the collection were taken at the performance in Llandysul. In general, the processions started on the edge of a town, and processed to the centre, gathering followers on the way. Once they had reached the centre, there was a short performance that involved the crowd that had gathered around them
The Dragons Wedding	64 slides & 24 photographs	This performance was a collaboration that particularly involved the children from the town and the Professionals. There were some preparations, but most of the production took place over the course of one day. In the morning a New Dragon came down from the mountain above the town and made its way to the Miners Institute. There, a ceremony took place where the New Dragon proved itself worthy of the Town Dragon to the characters on stilts. At the end of the ceremony the stilted characters paraded the New Dragon around the town. The Town Dragon, with eight children inside, then processed around the town in the afternoon on the way to the 'wedding' ceremony. The 'wedding' took place on the town football pitch, after which there was a celebration outside the town hall and a meal.
The Widow's Dream	19 slides & 6 photographs	This production took place in Llechwedd Slate Caverns in Blaenau Ffestiniog. This is an attraction that takes tourists on a route exploring the inside of the mountain. The owners of the caverns cooperated with the production to allow it to take place. The story line cue was the realisation that due to accidents, most families in Blaenau Ffestiniog had, at some time in the past, lost a relative in mining accidents. The production moved from above ground, through the caverns, using the tourist routes, and back. During this the lighting from the tourist displays was augmented with specials to pick out specific features, and props used to centre the action. The performance journeyed down through the caverns stopping to tell the tale at each location. The images tell the tale of a woman widowed by a mining accident. She journeyed into the mine, and struck a deal to free the body of her husband so that she could ensure he

		received a fitting funeral.
Cabaret and Workshops	12 slides & 21 photographs	As part of the residency, a series of workshops that explored many skills associated with the Performing Arts were organised. At the end of the residency there were two performances; one that involved mainly children, and another for adults. The Dragons' Wedding was the children's production. The adults requested a cabaret. The images in this section showed some of the workshops, preparations for the cabaret and the performance.

Table 3: Blaenau Ffestiniog Residency Production Summary

2.3 Department of Information Studies

The Department of Information Studies delivers courses across the information spectrum, both on Campus and through distance learning. Academic support for this research was given in the form of access to lectures with relevance to the research and office space. There was also the opportunity to attend an on-going series of lectures arranged for all Postgraduate students; the Aberystwyth Information Research Seminars (AIRS).

I had two academic supervisors based in this department. As well as overseeing the complete research process and thesis production, their advice on their individual specialisms' included the fields of digital curation, digital lifecycle management, digital information sources and digitisation.

2.4 Information Services Department

The Information Services Department at Aberystwyth University provide services across the information spectrum to the University. These include IT, library and printing services. In this research they helped with advice on the options and costs of storage services they provide and the purchase of equipment. They also discussed the uploading of the digital objects produced to the institutional repository, CADAIR, with me.

2.5 European Collected Library of Artistic Performance

The European Collected Library of Performance (ECLAP) is an organisation that aims to create an online archive of performing arts materials. It aims to do this by building a network of contributors who use best practice methods and tools to build their digital content. This content can either be uploaded to the repository or its metadata can be added and links back to the host institution provided. ECLAP also provide aggregated metadata to the Europeana project, thereby widening knowledge of its resources (ECLAP, 2011).

Chapter 3: Literature Review

3.1 Introduction

This literature review presents a topically organised view of digitisation through the research of writings presented on both paper and Web based media. The broad topic areas covered include the practical implications of choices made during the digitisation process via case studies and technical manuals, and the history of digitisation. More detailed information on some technical topics is in Appendix A: Technical Background.

3.2 Literature Review Method

The function of a literature review is to present a view of the current situation in the chosen topic; in this case the topic is the transfer of digitisation techniques and knowledge. In particular the transfer of these techniques in a manner that can be used and understood by those whose main function is not a cultural heritage facility. Specifically sought to discover sources which explained:

- how digitisation practices have developed to provide background information
- case studies that provided examples of practice to learn from
- practical techniques for the written guides

The initial research involved looking at known sources such as Web sites belonging to JISC, the DCC, the AHDS and DPC. This was followed by a keyword search of journal databases hosted on Aberystwyth University's library catalogue. Books were sourced using keyword searches of the University library catalogue and from previous knowledge of key works in specific related areas of information practice.

To ensure that the resources discovered were comparable, the databases in Table 4 were all searched on 4th August 2011 using the terms “digitisation” and “digitization” with no further filtering applied to the results.

A more targeted search was conducted on web sites and journals, identified from previous experience, such as the *International Journal of Digital Curation, Program* and *Ariadne*. Specific databases such those in Table 4, were also searched more thoroughly for potentially useful items throughout the research period. Having retrieved a large number of items using broad search terms these were whittled down using narrower terms appropriate to the particular topic area being researched e.g. by combining the terms digitisation and metadata.

Database	Number of Hits
Web of Science	463
LISTA	143
Emerald	830
JSTOR	228
LISA	613

Table 4: Table of Search Terms and Sources

Over time, the area of investigation moved from a largely theoretical basis to include operational tools and methods of practice that could be encountered during a digitisation project. These are shown in the decision tree, Figure 4. The tree is used by following the branches to select items to read and categorise for inclusion by deciding whether they fit into the area indicated in a box.

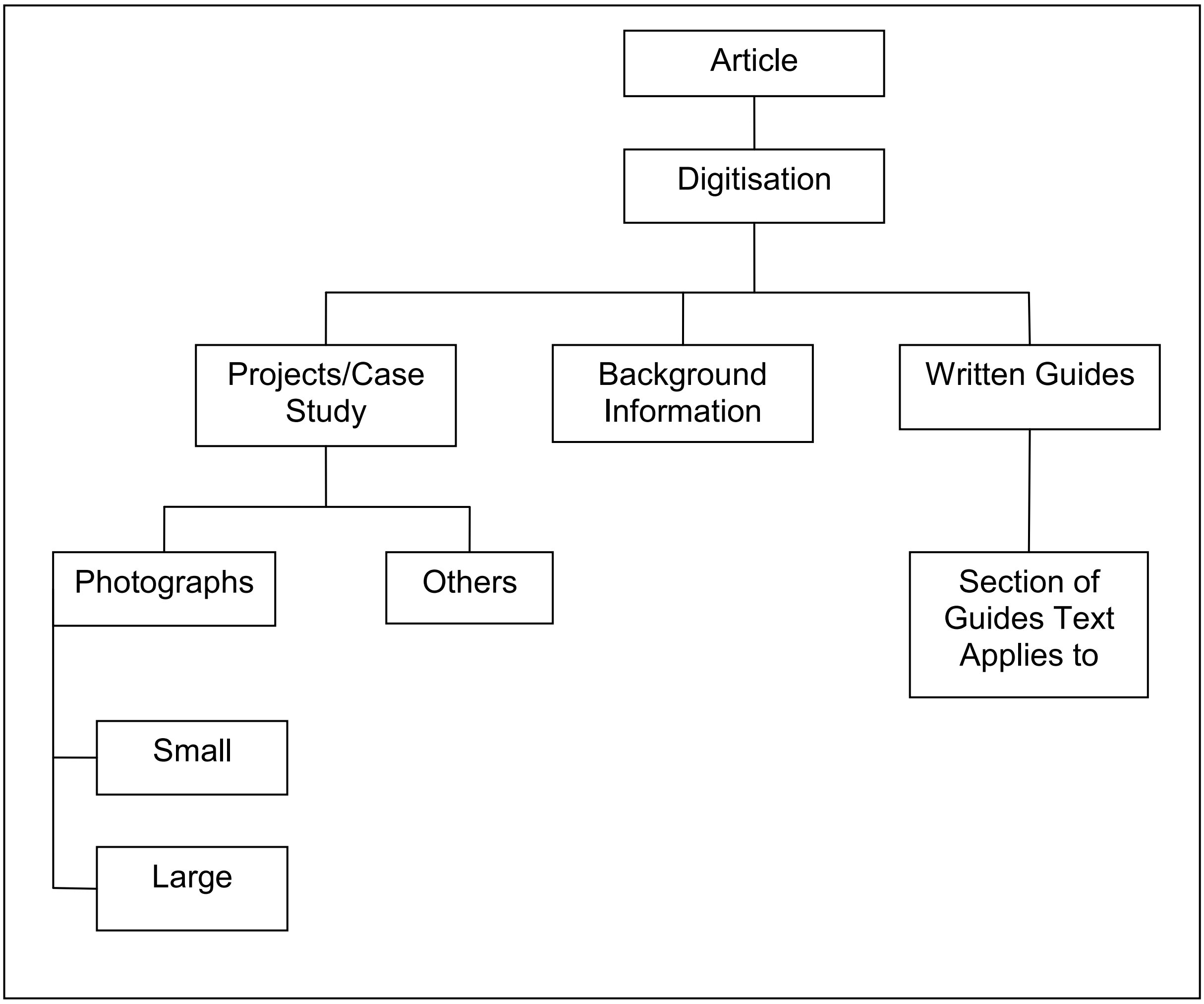


Figure 4: Decision Tree for Choosing Literature

Items relating to the technical aspects of digitisation were sourced using knowledge gained from previous professional experiences and in known related fields. These fields included digital photography, the performance industry, media studies and teaching manuals and were written by authors such as Berger (1972), Canadian Ski Instructors Alliance (2000), Chandler (2000), Elkins (2011), Giles (2010), Mitchell (1984) and Sheddon (1993). Some of the topics researched, but not included here are listed in Appendix A: Technical Background.

The search for information utilised journals and Web publications extensively because this is the chosen method of dissemination for some areas of information practice. For example JISC publishes all of its material on its Web pages. This allows the users to choose the articles they are most interested in and download those. To JISC the benefit is that they can update their Web pages regularly to reflect the most current view on a topic. This allows for greater contemporary control of the content. The DPC also regularly publish articles but as a continuing series of updates to their web pages rather than as a discrete journal. Along with the DCC, these sources publish articles and reviews from differing perspectives but, if reviewed together, cover the topic area thoroughly from a technical point of view.

As can be seen from the decision tree (Figure 4) this review has two longer branches and a short one; one long one concentrated on projects and case studies and the other, text applicable to the written guides.

Written guide topics are included because it became apparent whilst carrying out this review, that although many published case studies will indicate where things went wrong for them, they often assume a level of specialist knowledge that potential readers may not have. The literature review has tried to redress this by attempting to seek out key works from outside the academic area of information studies that may contain relevant information within them. The reasoning behind this was that publications in these areas may express relevant ideas in language that is more appropriate to CPR than that conveyed using the terminology assigned by information practice. In particular this has been attempted with research into photography, television and theatre.

The third, shorter branch, looks at background information. It explores the general background of digitisation by following the outline of the history of digitisation and at two projects in particular.

The systems map of the complete literature review, Figure 5, shows two separate boundaries within the literature review that enclose the literature relating to practical topics and general topic literature.

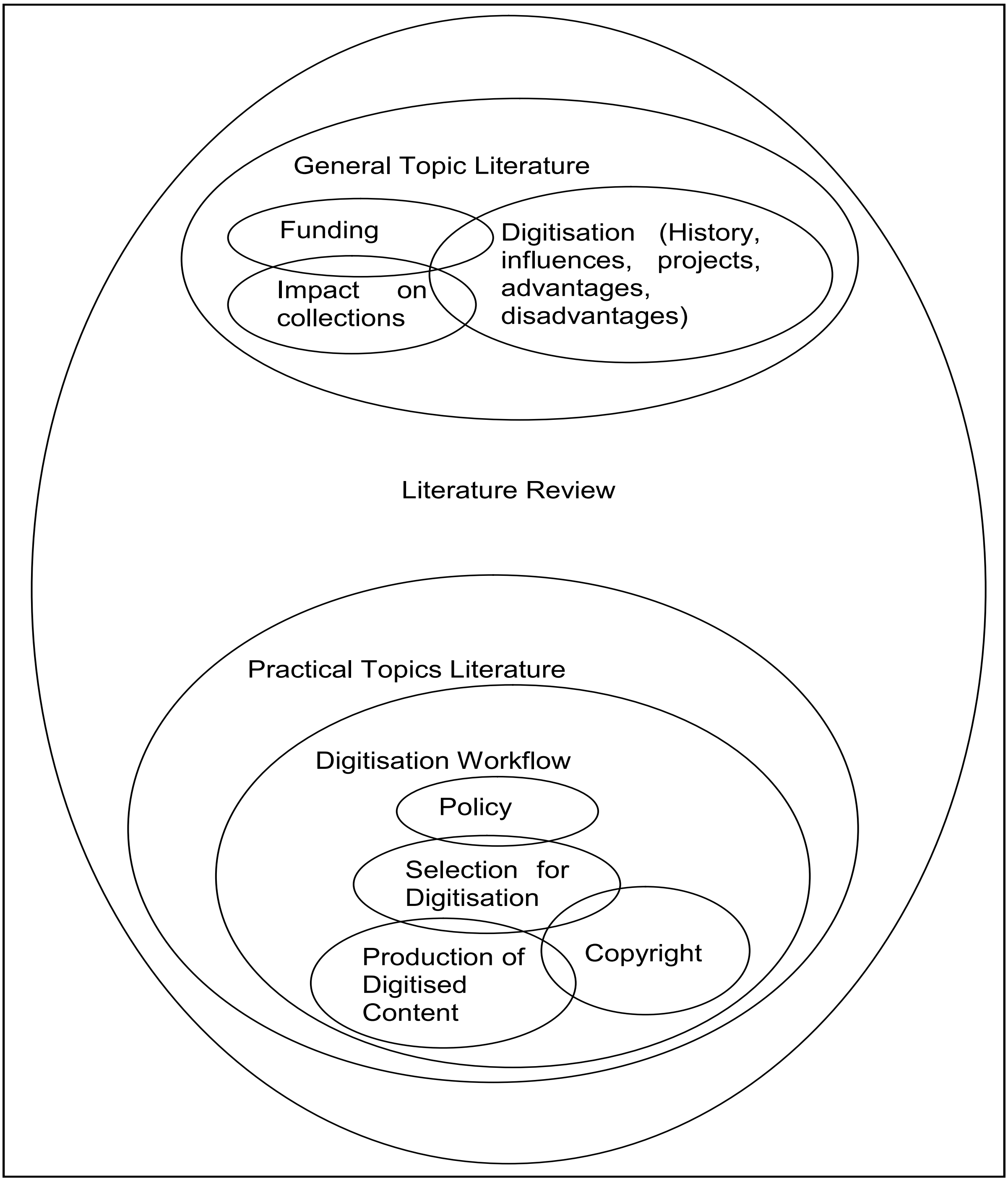


Figure 5: Systems Map of Literature Review

3.2.1 Key Works

Key works were selected based upon their relationship of the content to the topics and aims of the research. Having been identified, key works were arranged in Table 5 in order of the topic area, then by the authors' names.

Hedstrom (1997), Ross (2000) and Rothenberg (1999) are generally accepted to have written key works in the area of digital preservation, but the principles of data management they discuss apply across all digital files. Bulow and Ahmon (2011) aim their book at collection managers. While they cover in depth issues such as the different types of fastener and how ideally they should be dealt with, the thrust of their writing is the project management of the digitisation workflow (p. xv). Together these works provide background information on the types of issues of general digital file management and an outline workflow for digitisation.

Author and Date	Area of Key Work	Section In Figure 4
Bulow, A. & Ahmon, J. (2011)	Collection preparation	Background Information
Hedstrom, M. (1997)	General digital preservation awareness	Background Information
Ross, S. (2000)	General digital preservation awareness	Background Information
Rothenberg, J. (1999)	General digital preservation awareness	Background Information
JISC Digital Media (various dates)	Guides to digitisation practice	Written guides
Langford, M. & Bilissi, E (2011)	Images and their adjustment	Written guides
Padfield, T. (2010)	Copyright	Written guides
Zeng, M.& Quin, J. (2008)	Metadata – general technical advice	Written guides

Table 5: Table of Key Works

JISC Digital Media have published many guides that are aimed at those carrying out projects in many areas. The key facts in the guides are clearly and concisely expressed and therefore present useful examples for the research. Langford and Bilissi (2011) have written a generally accepted key work for photographers. The topics of interest to the research were the perception of images, how scanners work and digital file formats. Padfield's (2010) book on copyright is generally accepted as a key work as it comprehensively describes the restrictions and exceptions to copyright in the UK. There are many books written about metadata, but I felt that Zeng and Quin (2008) describe it in terms that allow the reader to make the links between the outputs of technical web sites and a basic description.

3.3 Digitisation

In their works Hedstrom (1997) and Ross (2000) primarily addressed the intricacies of the problems surrounding the preservation of digital material; they also made points relevant to the treatment of all digital materials. When commenting upon the frailty of the media, they predicted many of the problems that have become apparent in their preservation. Both described why there is a need to consider carefully the planning and management of the preservation of digital objects. While this may seem irrelevant to a digitisation project, in order to maintain accessibility, it is generally agreed that the preservation of digital objects should be a consideration from the point the decision to digitise is made (Bulow & Ahmon, 2011, p. 12).

Writing at the same time as Hedstrom and Ross, Reid writes from the perspective of a practising archivist in Powys, Wales (Reid, 2000). He outlined how the profession needed to adapt to changing circumstances to ensure it remained at the centre of information delivery and avoided becoming relegated to specific, highly specialised areas. He indicated this should be done by viewing the challenges posed by ICT advances as an opportunity rather than a threat (2000, pp. 143 - 146). He characterised digitisation projects into three types: digitisation of finding aids, digitisation for remote access and digitisation to aid preservation of the original. He further divides digitisation for access into the project outputs of exhibitions and virtual repositories. Using his definitions he described how for digital exhibitions, users are targeted (p.148 - 149) through interaction during the compilation process (p.153) as well as when they are actually online (p.150). He also described his view of the democratisation of the archive through the influence of the Web and the associated changes in user profile. He saw this as a natural extension of the current practices of museums, libraries and archives sector, not a threat (2000, p. 155).

This contrasts with the later review, received by Flinn (2010, p. 1), on a research proposal concerned with user generated content. He felt the review neatly encapsulated a powerful and genuine strand of thinking within the archive profession and academia more generally.

He argued that this review points to the possibility of instigating practices that exclude sections of society through limiting points of access (p. 2). Feather (2008, p. xviii) argues about exclusion from a slightly different position, preferring to express his view from the more positive stance that the availability of information is the driver of change. And that if they have access, people will find a way to use information.

The views of professionals working in archives and libraries are important to this research because they did much of the original work in this area. As the pioneers carrying out large investigative projects such as eLib and NOF-digitise, they have shaped much of the current thinking and processes associated with digitisation and allied areas.

3.3.1 History of Digitisation

Using the figures available on the JISC and the Arts and Humanities Data Service (AHDS) web sites, and the contents listings in each edition of Ariadne to August 2011, a graph of the number of completed projects per year in the UK and Ireland, was drawn (Figure **6** 6). This graph showed a large spike in 2009 with 14 projects being completed for JISC. The numbers of projects have been edited to include digitisation projects and exclude projects that only examined aspects related to digitisation such as metadata gathering.

Two of the formative digitisation projects carried out in the UK were the Electronic Libraries Programme (eLib) and the New Opportunities Fund – Digitise (NOF-Digitise).

ELib took place between 1996 and 1999. The prompt for the project was the 1993 Follett Report. The report reviewed the position of libraries for the Higher Education Funding Council (HEFC) with a view to saving physical space in Libraries (Field, n.d.). It had funds of £15 million from JISC, for all three phases and included two digitisation projects in its remit. The aim for the complete eLib project was to "*engage the HE community in developing and shaping the implementation of the electronic library*" (Kirriemuir, 1998).

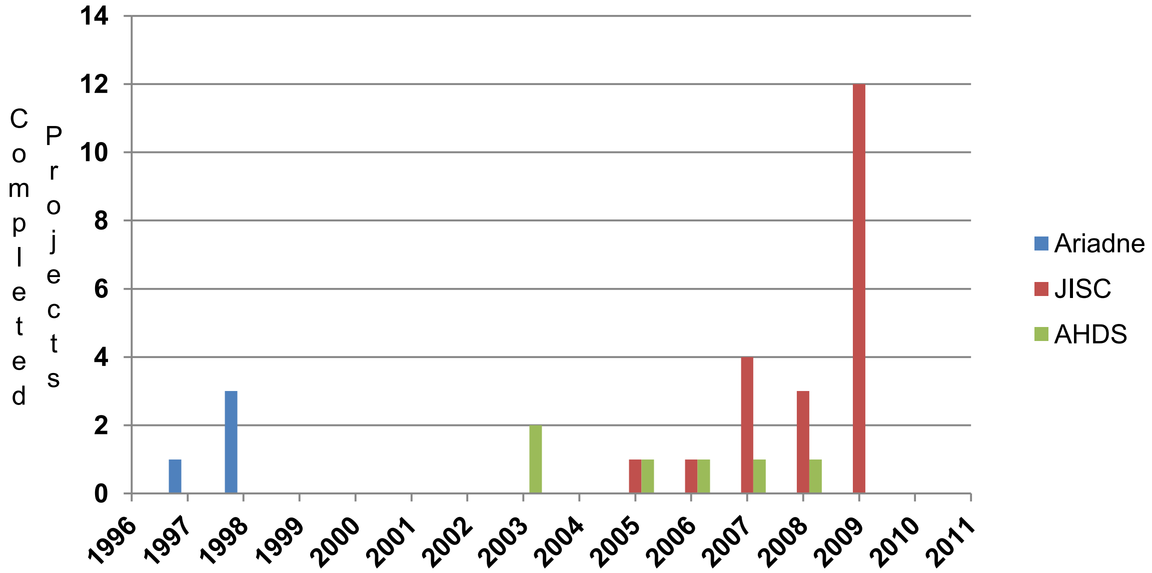


Figure 6: Number of Digitisation Projects Completed Per year

One project was called the Internet Library of Early Journals. Its aim was to digitise 20 years each of six journals from the 18th and 19th century and display them on the internet. While doing this they were to investigate digital object creation, indexing and the access platform for its users.

For digital objects they investigated the effect of the archival and digital format, scanner resolution and file compression. For users they investigated how the indexing and access platform affected their experiences (UKOLN, 1996b).

The second project was the Digitisation in Art and Design (DIAD). Its aim was to create a record of selected core journals via databases of digitised content. The expected delivery method was expected to be on CD-ROM or possibly the internet. Over the two years of the project it was expected that the titles would be selected, copyright holders identified, the prototypes for delivery built and evaluated and the final product made and launched (UKOLN, 1996a).

The New Opportunities Fund - Digitise (NOF-Digitise) was launched in 1999 with a budget of £50 million. To put this in perspective over the same time span as phases one and two of eLib, JISC spent around £50 million on networks (Rusbridge, 1998).

The money was to be divided between 150 national and local bodies with the aim of developing lifelong learning resources in three areas. These areas were:

- Citizenship in a modern state
- Cultural enrichment
- Re-skilling the nation.

Among the first schemes to be given money were in the subject areas of archaeology and science and nature. The formats being digitised were varied. Another group of projects were the Sense of Place projects that aimed to give resources with a local emphasis. Projects that gave advice and developed skills in digitisation techniques were also funded (NOF-Digitise, 2009).

3.3.2 Influences on Digitisation

In Figure 7 7 the publications from Table 5Table 5 have been re-arranged in order of copyright date and show how there have been two distinct groups of activity. These periods of action are grouped around the introduction of major legislation that changed the practice of file management. The acts responsible for this change were the Data Protection Act 1998 implemented in 2000, and the complete implementation of the Freedom of Information Act 2000 in 2005 (Information Commissioners Office, n.d.). These foci also echo periods of increased momentum in preservation practices, probably because the two are closely interlinked (Wood-Fisher, 2010, p. 18).

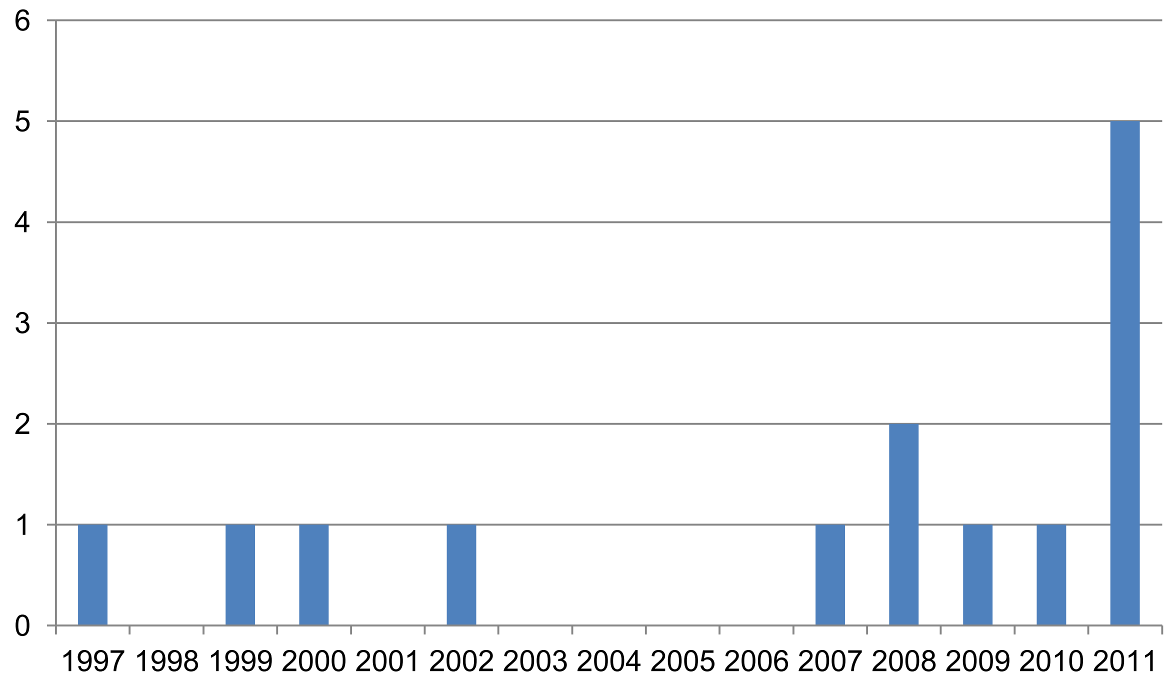


Figure 7: Graph of the Number of Digitisation Projects per Year

Using these legislative levers, bodies such as JISC have had an influence upon digitisation, especially in the area of education and other public services. Because of this influence it is uniquely placed to exert influence upon the practices within those areas through the allocation of funds to projects.

3.3.3 JISC Projects

Of the 135 projects that JISC Digital Media list as digitisation strand projects they specifically list three as being directly of relevance to the digitisation of still images (JISC Digital Media, 2011b). Not all of these projects were funded by JISC, but were useful case studies. These are:

- University of Bristol Theatre Collection
- Staffordshire Past Track
- History of Art, University College Cork.

3.3.3.1 University of Bristol Theatre Collection

This case study was set in the University of Bristol, which is a large institution with its own digitisation unit. The scale of the three projects explored is also relatively large, given the timescale in which they were accomplished (JISC Digital Media, 2007a, p. 6). This required an agreed and tested workflow (JISC Digital Media, 2007a, p. 2 & 11).

The technical details of the project equipment are described in depth. Examples of the technical difficulties encountered and the solutions adopted to solve them are also given. Each of the collections were selectively digitised, with different criteria being applied according to the desired outcomes. Access to the digital objects was provided via the Library catalogue (JISC Digital Media, 2007a).

3.3.3.2 Staffordshire Past Track

Unlike the University of Bristol, the Staffordshire Past Track case study was a collaborative digitisation project funded by the NOF-digitise funding scheme. Its aim was to produce a publicly available resource for a range of different users. It took place over the relatively long time scale of approximately two years (JISC Digital Media, 2004).

The project captured approximately 97% of its digital objects using scanners bought for the project, and two percent with a digital camera. Some of these digital objects were reprocessed from previous projects to meet the current project standards (JISC Digital Media, 2004, p. 6). The first half of the paper describes the digitisation process in some detail to show how the different parts of the project fed into one another. It also gives a comprehensive table of standards applied to the project (JISC Digital Media, 2004, p. 7). This table would have been useful for the separate partners in the project to calculate the storage requirements of their digital objects.

Metadata was collected from the holders of the archival material and recorded in the bespoke data management system by hand. The metadata schema used was based upon DC and uses the same information to describe both the archival materials and digital objects.

3.3.3.3 History of Art, University College Cork

This project was not funded by JISC, but it is one of three case studies used on their web pages to highlight the use of digitised still images in education.

This case study examined how a newly established university department created pertinent, image based teaching resources. Compared to the University of Bristol, the department was small, with part-time staff that accommodated the production of resources alongside other duties (JISC Digital Media, 2005a, p. 1). The paper explored the compilation of images from a variety of sources. Some were produced by digitising slides and some by linking to external resources. The workflow to achieve this was not detailed.

The results obtained were driven by the aim to produce materials for viewing on a screen or digital projection (JISC Digital Media, 2005a, pp. 2-3). The aim was to produce images that complement a learning text. These digital objects are made available using the course Virtual Learning Environment (VLE).

Each of the three case studies used a different size or type of institution with different user groups. It also indicated that digitisation meant different things to different suppliers. This is shown by the different solutions employed by each of the case studies to create their collections. It also demonstrated how an experienced institution can efficiently extend its collections by completing digitisation projects as funds allow. In this case study the University of Bristol achieved this because they had previously invested in staff skills and had a tried and tested workflow that is used, with suitable adaptations, time and again.

3.3.4 Advantages of Digitisation

The generally accepted advantages of digitisation are an increased ease of access for users and support for preservation of archival materials.

3.3.4.1 Access

The advantages of digitising archival materials will vary in detail depending upon the collection and the users uses of the resources. In general, it is agreed that digitisation for dissemination via either an intranet, or the internet, can bring new audiences to view a collection just because of the utilisation of a new method of access (JISC Digital Media, 2007a, p. 1). The benefit this new audience brings is a wider knowledge and greater dissemination of the holdings of the collections in general. These benefits can also produce a diversification in the audience base for collections (Skarstein, 2010a, p. 1).

The NLW has pursued a strategy of in-house digitisation of its resources, with staff trained and funded from the core budget (Jones, 2008, p. 104). Similarly the Norwegian National Library chose pay to maintain open access to current digitised content in copyright. Here the National Library negotiated with publishers to have permanent access to digitised copy distributed through the public library network (Skarstein, 2010a). It is also carrying out the digitisation of nationally significant literature and hopes to complete this "*as fast as possible"* once copyright has expired (Skarstein, 2010b, p. 52).

Both the National Libraries of Wales and Norway have placed digitisation at the centre of their policies and provided nationally and culturally significant digitised content through the Web. Both have achieved access for their users to a selection of their content. The similarities between Wales and Norway do not end there as both countries have a small, but geographically distributed population. This remoteness of users from physical access to materials is being overcome by the national libraries making their content available on the Web. In pursuing this strategy both national libraries have overcome one of the main barriers this portion of the population has in accessing information distributed using physical media.

The 2008 policy of the British Library (BL) sees them embark upon mass digitisation of some of their current holdings with external partners after pursuing a policy of selective internal digitisation until then. Their view is that this digitisation will reflect their collections, enable researchers in remote locations using the internet and

provide a compelling user experience that facilitates innovative methods of research and meets 21st century requirements for interacting with content (British Library, 2008a, p. 1).

This is again different to the NLW and Norwegian policies in that the BL seek to represent their entire holdings rather than provide access to content that is culturally significant to their users.

3.3.4.2 Benefit to Collections

The second usually agreed advantage of digitising archival materials is to aid the preservation of the original item by directing users to an alternative means of access (Hughes, 2011a). This is not the contradiction it first appears as increased access, and an increased audience size, doesn't always lead to more requests to access the original. For most users, access to the digital surrogate is generally enough to satisfy curiosity (Bulow & Ahmon, 2011, p. 8). Accepting this premise means that far from destroying the scholarly reputation of archives as some argue (Flinn, 2010), the original is available for those that need it, and possibly in a better condition than if the merely curious were not directed to other means of access (Bulow & Ahmon, 2011, p. 171).

For an individual item of archival material, reducing the number of times it is accessed reduces the risks associated with handling for that item. Reducing handling also reduces general accumulations with respect to time and therefore, the number of preservation treatments required. For the item, fewer treatments will mean less stress from being subjected to processes that over time could still lead to degradation. By extrapolation there are accumulating benefits to the collection, and institution, through reducing these treatments. Aside from the obvious cost implications, fewer treatments could mean that in the future items may be passed on in better condition than might have been expected had they not been digitised (Bulow & Ahmon, 2011, p. 171).

3.3.4.3 Raising Institutional Profile

The most widely acknowledged benefit of digitisation projects to an institution is the increased awareness of a collection such a project brings to its audience. With much content digitised to be displayed on the Web, the existence of collections can also be disseminated to a more geographically distributed audience (Australian Government Information Management Office, 2004).

The measurement of the institutional benefits can be difficult to separate from other related activity e.g. the pay back from the costs of staff training may not be apparent for many years (Hughes, 2003, p. 14). Another institutional benefit that is difficult to quantify, is the future use of resources by audiences targeted by specific exhibitions (Reid, 2000, p. 153).

3.3.4.4 Academic Benefits

An often overlooked benefit of digitisation is that people with visual impairment can access items when they choose, rather than when a guide is available, for the first time. Through online access they can use the audio description facilities associated with the Web to ‘see’ items they would normally be unable to touch. (Sitts, 2000, p. 16).

Writing in 2003, Hughes (p. 12), gives examples of projects to re-unify distributed materials through the digitisation of each separated part. This was an advantage foreseen by Reid where he cites the example of an addition to their holdings which contained correspondence with famous literary figures. He describes how the followings of famous individuals could be interested in archival material within the collection he curates, but may not travel to Powys unless they could be assured of a positive outcome. By digitising, Powys County Archive has made the content and condition of the archival materials clear as well as advertising that they hold them (2000, p. 148).

In an educational setting there are two popular uses for digitised content; in an institutional web site and the Virtual Learning Environment (VLE) (JISC Digital Media, 2005a).

3.3.5 Disadvantages of Digitisation

The disadvantage of digitisation most often acknowledged is that it represents an on-going cost which is not quantifiable because of the continually progressing nature of the technical infrastructure needed to keep access to digital objects (Hughes, 2003, p. 7). The main contributors to this on-going cost are the overheads of preservation of the stored files and the need to refresh Web pages. Calculating the monetary costs of storage is difficult to separate out from other on-going costs. The observation that little is written on the subject was made during a lecture at Aberystwyth University by Professor Lorna Hughes, University of Wales Chair in Digital Collections (Hughes, 2011a). Part of the problem for institutions is arriving at a definition for 'value'. Reasons for digitising collections vary, and the value of digital objects varies from one user to the next (Hughes, 2011b, pp. 6-7). Additionally, the situation is often complicated by the need to manage files for preservation (Davis-Perkins, Butterworth, Curzon, & Fields, 2005, p. 7; JISC Digital Media, 2008c, p. 3; Spence, 2005, pp. 373 - 374).

For institutions starting out in digitisation practice additional costs may also occur where training is required for staff to be able to manage this new resource, and again for preservation of the resource (JISC Digital Media, 2008a). A worry of some of the archives community is that digitisation may be carried out at the expense of other preservation actions (Spence, 2005, p. 366). Digitisation is not the same as curation or preservation, and preservation of the original item is still as important (Hughes, 2003, p. 8).

A second cost to institutions starting out is the up-front cost of equipment and, once acquired, the need to maintain and upgrade it periodically (Sitts, 2000, p. 17). It is generally accepted as good practice that when creating digital files long-term storage conditions should be a consideration (JISC Digital Media, 2008c). In this context long-term is generally taken to mean "*that digital information lasts forever—or five years, whichever comes first*" (Rothenberg, 1999, p. 2). Another, more technical way, of expressing the longevity of digital media is shown in Table 6 (Triantaphillidou & Jacobson, 2001, p. 5 - Paraphrased):

Hardware: new generation	occurs approximately every two years
This new generation is:	compatible with two previous generations
Device drivers:	new generation within device lifetime
Devices can be:	operated by four generations of drivers
Operating systems:	new generation every one to three years
New generation:	can run same drivers as last two software generations

Table 6: Table Summarising Relationship between Obsolescence and Re-generation

An often unacknowledged disadvantage is if the digitisation process damages an item. To avoid this Bulow and Ahmon (2011, pp. 18-20) advise consulting an expert, just as archivists would usually do with archival materials they were not experienced in handling.

3.4 Funding

The sources of funding for digitisation projects have been varied and sometimes sector specific (JISC Digital Media, 2009c). Estimates of funding costs vary greatly with no ‘one-size-fits-all’ solution available to make calculation of the costs easier for novices. JISC (2009a) lists some principles to be considered, and the Western States Digital Standards Group based in Michigan, USA, estimate the cost of the digitisation process to be about one third of the total with the other two thirds coming from costs associated with compiling metadata (2003, p. 9). Both also note other considerations such as storage costs over time and the cost of obtaining reproduction rights (JISC Digital Media, 2009a; Western States Digital Standards Group: Digital Imaging Working Group, 2003, p. 9).

Davis-Perkins, et al. (2005, p. 3) describe how the projects they researched, were only allocated money to complete the digitisation of the archival mateiral, not any other aspect of a digitisation project. However they felt that the greater problem was the emphasis on access as an output as this skewed the whole process towards an outcome, rather than towards the application of traditional sector skills on archival materials. This stems, they believed, from a belief that the value in a primary source is greater than that in its digital surrogate (p. 2). Spence also believes that many projects carried out lack a long term vision (2005, pp. 371 - 372).

Whilst it is undoubtedly true that funders want to be able to see value for money (Davis-Perkins, et.al., 2005, p. 7) they also make this clear when institutions apply for funding through the wording of their respective application forms (Digital Curation Centre, 2011b).

Knowing how much a project is going to cost to be done in house is a pre-requisite to establishing whether out-sourcing a project may be a more cost effective operation. When these costs are fully appreciated it can be easier to assess whether the aims of a project can be satisfied more fully through out-sourcing. The monetary advantages of out-sourcing projects can be that the facility doesn't have the expense of maintaining and updating equipment and the actual costs may be easier to ascertain. On the other hand a project manager needs to know exactly what they are asking for before approaching the supplier so that they are able to discuss and describe requirements precisely (JISC Digital Media, 2008e, pp. 1 -2).

3.5 Impact of Digitisation on Collections

3.5.1 Small Collections

In describing themselves Birmingham Institute of Art and Design (BIAD) used words such as "small" and "specialist". Hence their decision to actively pursue collaboratively funded initiatives to achieve digitisation of their archival materials (Everitt, 2005, pp. 312-313). BIAD made the point about their size relative to other partners in projects and from that drew conclusions about the comparative loudness of their voice (p. 322). They also questioned the representation of their resources on the web resulting from adopting the strategy of joining cooperative projects. They concluded that each online resource "*impose their own criteria and offer vastly different levels of information*" (p.321) but, having accepted that any online presence would be impossible without external assistance, felt their collections were not misrepresented. BIAD felt the relative invisibility of their data within collaborative search engines was a problem to be overcome and discovered collection level descriptions (p.316) were usually more compatible with the host site requirements and made for consistent search terms across a number of hosts. Latterly, of more concern to them was the upkeep of the sites they had been part of, and whether, in using these, they had sufficiently addressed the scholarly needs of all their users.

Of the papers written about small collections a proportion generally seem to be written from the perspective of a large institution digitising a small, specialised collection. Sometimes this is within an establishment with experience and equipment (JISC Digital Media, 2007a), others document setting up a new, permanent facility with a small selection from their holdings (Hampson, 2001; Rayner, 2009). Conversely, writing in 2005, Spence draws attention to the wealth of information being generated and lost in small, culturally significant organisations without statutory responsibility to deposit records (2005, p. 367). They felt this was due to a lack of cohesive, strategic, sustainability planning (Spence, 2005, pp. 371 - 374).

3.5.2 Large Collections

Many of the comments made by BIAD about the difficulties of being a small partner in collaborative projects could apply equally to the large collections garnered by collaborative projects such as those funded by NOF-digitise. Yeates and Guy (2006, pp. 149-158), indicated that the impact on individual collections was not assessed at time of their paper, but did give two learning points for future collaborative projects. Both related to the time a large scale project takes and the quality of the digital object produced. The first was solved by a resolution to disseminate training more effectively and to keep the staff committed throughout the project. The second, to ensure that agreed standards were kept at all levels of the project so that local adjustments did not threaten the reliability of the whole.

The University of Bristol case study previously referred to makes no mention of the impact on its collections of the digitisation projects it describes. This was, however, one of the qualifying factors in the selection of archival materials prior to digitisation (JISC Digital Media, 2007a, p. 3). They appear to have none of the representation problems BIAD reported (Everitt, 2005, p. 321); perhaps because they have their own resources for web display (JISC Digital Media, 2007a, p. 2).

The Sphere project set out to assess the impact of the Stormont Papers, Northern Ireland, by using the Toolkit for the Impact of Digitised Scholarly Resources (TIDSR). TIDSR is an impact evaluation method developed by Oxford University for JISC (Meyer & Eccles, 2010; Meyer, 2011). The results of this evaluation showed that the resource was well known, had high usage, but that use by academic researchers was not as high as expected. It also concluded that the web pages needed redesigning to include links to and from other related resources (Hughes, Knight, Ell, Yeates, & Dobreva, 2011, pp. 21-22).

In their digitisation strategy the BL state that so far they have digitised less than one percent of their collection (British Library, 2008a), and the NLW declare a commitment to creating a "*critical mass*" of digitised content (National Library of Wales, 2009, p. 6). Neither give an indication of the impact that digitisation has had on their collections at the time of writing.

3.6 Digitisation Workflow

In discussing the models they used, case studies written about digitisation projects inevitably focus on an area specific to their project outcomes (Leslie, 2004). A number of writers include some level of workflow descriptions in their pieces and include references to the areas indicated in Table 7Table 7.

Papers Scanned for References to:	Creating a Digital File	Some Form of Metadata Creation and Use	Quality Assurance of Created Image	Storage Considerations
Davis-Perkins, Butterworth, Curzon, & Fields, 2005	√	√	√	√
Everitt, 2005	√	√	X	√
Hampson, 2001	√	√	√	√
JISC, 2007	√	√	√	√
Jones, 2008	√	√	X	√
Keys, 2011	√	√	√	√
Rayner, 2009	√	√	√	√
Triantaphillidou & Jacobson, 2001	√	X	X	√

Table 7: Comparison of Papers Read and Their References to Elements in a Digitisation Workflow

In the article written by Jones (2008) he pointed to an early expression of what has become the core principle of digitisation policy at the NLW. In an introduction to a web site that was compiled from October 1999 to May 2000 (p.103) he described the intent to develop relevant skills in-house. When this was allied to the decision to adopt Metadata Encoding Transmission Standard (METS) in 2002, it was as he describes it "*the most significant decision*" (p.106). This was because having decided to adopt the strategy of digitising items in-house, NLW had committed resources of every type to carrying it out. Having taken these steps, a means to manage the results was needed. The digital objects were accessed via a database and the digital descriptions via another (2008, p. 108). As both of these used XML based schema, the two systems 'spoke' to each other and collected all the data needed to publish a digital object on the Library's web site.

Overall the article concluded, as does Hampson's (2001, p. 273) article on the BUILDER project, part of eLib, that digitisation is only part of a much larger workflow. Davis-Perkins, et al. refer to this idea in terms of the funds supplied for digitisation projects being only for the process of producing the digital object and not for cataloguing and placing them in an accessible environment (Davis-Perkins, et al., 2005, p. 3).

From the point of view of a project manager trying to design a workflow, there is far more unity in the form of advice in publications from the major national bodies. JISC Digital Media and the Western States Digital Imaging Group give comprehensive lists of what should be considered for inclusion in a workflow and why.

3.7 Policy

3.7.1 Digitisation

In line with the UK governmental policy of publishing policy documents online, all the National Libraries have the relevant documents available for public scrutiny.

The NLW policy acknowledges straight away the place digitisation has within the whole institution and goes on to explain in some detail how it plans to execute its aims and objectives for the period of the document (National Library of Wales, 2009). The NLS approaches the subject from the point of view of how digital materials fit into the acquisitions strategy (Newton, 2008, p. 1). In its more recent strategy document, covering the period 2011 – 2014 (National Library of Scotland, 2011, p. 5), listed first in its 'Content' section is the long-term preservation of digital, film and sound materials, but nothing about the place of digitisation specifically. In contrast to the densely detailed document of the NLW and the NLS, the BL digitisation strategy document (British Library, 2008b) covers the same points, but is written using a visually more simplistic layout. A comparison of the areas written into these policies is in Table 8.

	BL	NLW	NLS
Position of policy in institution	√	√	√
How policy relates to other institutional policies	X	√	1 other
Digitisation objectives	√	√	√
Implementation of digitisation objectives	√	√	√
Scope of digitisation activities	√	√	X
Accountability for digitisation	√	√	X
Glossary	X	X	X

Table 8: Comparison of the Digitisation Strategies of Three UK National Libraries

3.7.2 Preservation

As the guides written for this research were tested on a higher educational institution, the JISC Web site was consulted for an ideal list of items to include in a preservation policy. In a review written for JISC, and published in October 2008, Beagrie et al. stated that the model developed should be used selectively when applying it to an institution (2008, p. 16). A table of items to be considered for inclusion was followed by an explanation and examples. This was supplemented by an explanation of possible preparatory work in the preceding section (2008, pp. 12 - 15). This report appeared to have been superseded by a document published as part of the JISC Digital Media pages in January 2009. The later, much shorter and more easily digested document lists eight points for inclusion, after a brief general positioning statement for the policy. It ended with a recommendation that using DRAMBORA (Digital Curation Centre, 2008a) could be a good means of assessing the institution's current position.

The DCC preservation policy template drew together statements from four bodies on what they expected to see in a policy statement for repositories. These were culled from documents intended as an aid to funding applications. Unlike the JISC 2008 paper, this listed seven sections to be included and what should be in each. The paper (Digital Curation Centre, 2011b) started with a headline list of things to be included. This is then expanded using selected statements presented in coloured text to make following one source easier. Within these statements the different approaches and philosophy of each of the different bodies to the task of preservation is revealed. For example: data share is aimed at repositories taking in research data and is expressed in plain, scientific language with example statements that can be used and expanded upon. The use of plain language is important because the users of data repositories are from a range of disciplines, not necessarily familiar with the processes which underpin a repository service. It is therefore important that the policy is understandable by a wide range of audiences.

Comparing the digitisation and preservation policy tables (Table 8 and Table 9) shows that in both cases the BL policy does not clearly place itself in relation to other institutional policies phrasing its position in brief, highly interpretable phrases (British Library, 2007; British Library, 2008b).

Comparison of Elements in a Digital Preservation Policy						
	JISC 1	JISC 2	DCC	BL	NLW	NLS
Position of policy in institution setting	√	√	X	√	√	√
How policy relates to other institutional policies	√	√	X	X	√	X
Preservation objectives	√	√	√	√	√	√
Implementation of preservation objectives	√	√	√	√	√	X
Scope of preservation activities	√	√	√	√	√	√
Standards to be used	√	X	√	X	√	X
Accountability for preservation	√	√	√	X	√	X
Glossary	√	√	X	X	X	X
Policy version control	√	√	√	X	X	X

Table 9: Table Showing a Comparison of Different Preservation Policy Components
(JISC 1 = Beagrie, et.al., 2008: JISC 2 = JISC Digital Media, 2009)

3.8 Selection for Digitisation

When archival materials are being assessed for digitisation it can help to use a formula to provide a consistent basis for comparison. In their digitisation project handbook, the North East Document Conservation Centre, Massachusetts, USA, aligns the skills used to assess items prior to acquisition by archives and libraries with the kinds of qualities being assessed for digitisation selection (Sitts, 2000, pp. 46-47). Worked examples of how selection may be documented were followed by a list of what should be assessed. This list was supplemented by definitions of the assessment criteria so that all are clear what the terms mean (Sitts, 2000, pp. 53-58). An alternative approach was adopted by Bulow and Ahmon where they covered the topic by concentrating on the preparation and preservation of archival materials (2011).

How to choose archival materials for preservation was the subject of the PLATO project, the outcome of which can be found on the PLANETS web site. Preservation and Long Term Access Through Networked Storage (PLANETS) was a four year project that sought to construct tools to aid the preservation of digital assets (PLANETS, 2007). After the funding stream ended, maintenance of the site and tools was under taken by the Open Planets Foundation (OPF) (PLANETS, 2007). PLATO asks detailed questions that are designed to help assess which archival materials should be prioritised for digitisation and to plan how the digital object will be preserved (PLANETS, 2011).

3.9 Copyright

Copyright is the legal protection conferred upon the external expression of original work that originates in the mind (Padfield, 2010, p. 4) by the creator of that work. The duration of copyright depends upon a number of factors including the type of work created. More than one person or media type can hold the copyright for a work.

The clearance of copyright is noted by workflows as an obstacle to be cleared at the earliest opportunity, and that much effort can be wasted in digitising items where copyright is not cleared and images then have to be destroyed (Bulow & Ahmon, 2011, p. 49; Padfield, 2010, p. 13). Padfield also states that an inability to discover who holds copyright is not a reason to assume it does not exist (2010, p. 187).

3.9.1 Photographs

In his book *Copyright for Archivists and Records Managers*, Padfield (2010) indicated that the type of copyright protection visited on a photograph depends upon the reason for the photograph being taken, in the same way as the words and music of a song have their own individual copyright. He also stated that in the UK all photographs are treated as documentary images (p.20) so all receive the same level of protection. Even this clear and unambiguous rule had codicils however. Historically, photos have relied on authorial definitions as they were not always seen as artistic works. This is important because the standard duration of copyright for artistic works is the life of the author plus 70 years or if published, 70 years after publication. For a documentary photograph however, it depends upon when that photograph was taken (p.46). The same principles are generally taken as applying to the electronic world, but the web adds another layer of complexity to this.

Langford and Bilissi (2011, pp. 427 - 428), in their advice to those wishing to become professional photographers, intimate a much simpler interpretation of the law. This is possibly because of the number of countries the book sells in. Their outline is written from the perspective of a photographer needing to guard against others copying or claiming their intellectual output. In it they state that in most countries, copyright belongs to an author unless they surrender it. For guarding against theft from the web they advocates the use of watermarking software, and a copyright statement strictly limiting the use of an image, or when infringement would be deemed to have occurred.

A JISC funded working party studied copyright issues faced by HEI's when trying to negotiate access to digitised journals (Bide, Oppenheim, & Ramsden, 1997). They showed the universality of difficulties such as identifying the true owner of copyright and from that obtaining clearance. They concluded that a unified copyright clearing service as a desirable outcome in combating the confused situation which they found in 1997 for journals. Secker (2010, p. 20) describes a similar situation where the complexity of deciding which of the various licenses terms applied in a situation took too long.

3.10 Production of Digitised Content

3.10.1 Scanners and Cameras

A scanner is a piece of hardware that uses software to interpret and transmit data to a linked computer. The data is generated by the transformation of an analogue object to the 1's and 0's of digital code.

The types of scanner most used in digitisation projects are reflectance or transmission scanners.

3.10.1.1 Reflectance Scanner

Reflectance or flatbed scanners use sensors arranged in rows and activated by reflected light. The head containing the sensors move in a predetermined sequence over the item being scanned. The sensors either collect the information for all three colours in one pass or make one pass for each of the colours in the array. The software in the scanner then collects and processes this data to produce the information on a computer screen (JISC Digital Media, 2010c, pp. 175 -176).

3.10.1.2 Transmission Scanners

The scanners used for scanning film are called transmission scanners. Film uses a different scanner type because it needs to be lit and transmitted to capture the positive that is the digital object. This is because the image on film is a negative, and the positive is the one needed for display. Some of these scanners use similar techniques to those outlined for flatbed scanners, others use a block onto which the complete image is projected and captured, then reversed by the accompanying software (JISC Digital Media, 2010c, p. 176).

3.10.1.3 Cameras in Digitisation

Another method of capturing a digital object is to photograph it using a digital camera with a fast back. A fast back is a back that exposes the complete sensor at once. This is different to a film Single Lens Reflex (SLR) camera which exposed a strip at a time. Using this method, means that additional equipment in the form of a tripod and camera studio flash lights are required (JISC Digital Media, 2008f; JISC Digital Media, 2010e).

Studio flash lights are usually used in addition to, or instead of, the flash placed on top of a camera. This is because lighting from two angles will provide more modelling for the archival material (JISC Digital Media, 2008f). Using only the one on top of a camera can only provide light from the front and can have the effect of flattening the appearance of the archival material in the digital object (Millerson, 1972, pp. 66-80).

This indicates that cameras are a viable alternative to a scanner for creating digital objects from two dimensional archival materials and make more visible the skills required to produce a digital object of quality (JISC Digital Media, 2008f, p. 7; JISC Digital Media, 2010e, pp. 1-8).

3.10.2 Software

All hardware needs software to allow it to ‘talk’ to a linked computer. For scanners this software has two purposes. One is to transmit the code sequence that the computer can then use to display on its screen an image of the scanned item. The other is to allow adjustments to be made to the scanner to change properties of the scan being produced. For film scanners these are often expressed as choices between differing film formats e.g. Kodachrome, to compensate for the differing manufacturing techniques (JISC Digital Media, 2010c, p. 181).

From the point of view of digitising archival materials, automatic features on scanners should be examined carefully so as to ensure that they can be turned off if required. This is because some software automatically compensates for features such as scratches. From an archival point of view, when digitising an item these imperfections should be retained because they are part of the defining features that make the archival material what it is. Changing the appearance could be seen as constituting the misrepresentation of an item, which is one of the objections of some archivists to the digitisation of items (Davis-Perkins, et al., 2005, p. 5).

3.10.2.1 Scanner Adjustment and the Effect on Digital Objects

Adjustments made to scanners via their software play a critical part in determining the appearance of the resulting digital object of archival materials. In general, scanners for digitisation should be equipped with a preview facility so that adjustments can be made to capture more data from the archival material prior to the scan being completed (JISC Digital Media, 2006e, p. 3). By making adjustments before scanning and therefore only capturing data that is already present, the argument that digitisation has somehow augmented the archival material is nullified (Davis-Perkins, et al., 2005, p. 6).

The resolution of a scanner is usually measured in pixels per inch (ppi) and is usually quoted in the literature along with the optical resolution or interpolated resolution. Optical resolution is the number of pixels the sensors can record and is normally the first number quoted in manufacturers' details e.g. 1200 x 1200 ppi. The interpolated resolution is the optical resolution combined with that the scanner software can add to it. This results in an increased file size, but not more data from the scanned item in the file (Langford & Bilissi, 2011, pp. 178-179).

3.10.3 What Constitutes a Good Digital Object

Assessing a picture is something that viewers often do unconsciously. They have to be made aware of the relationships between the contents and the technical elements that make up a picture. Even without this awareness people automatically make adjustments for a variety of everyday situations (JISC Digital Media, 2010g, p. 87; Millerson, 1972, p. 63; Nurnberg, 1969, p. 10). When digitising archival materials awareness of these adjustments mean that the controls on a scanner can be used to gather as much of the data present as possible to create the best digital object possible (Western States Digital Standards Group; Digital Imaging Working Group, 2003, p. 24).

3.10.3.1 Colour Theory

The primary colours used to reproduce all other colours depend upon the method being used to produce them. There are two methods of producing a range of colours; additive mixing and subtractive mixing (Langford & Bilissi, 2011, p. 445 & 468).

Langford and Bilissi (2010c, p. 17) and Millerson (1972, pp. 46 - 52) both indicate that the human eye is good at identifying colour when seen in comparison to others, even in differing lighting conditions, but less so when seeing them in isolation. This is because humans build a set of references which enable them to compare colours. Because of this ability, neutral surroundings are recommended for digitisation projects (JISC Digital Media, 2005b). To avoid mistaking the colours in archival materials a comparison strip should be placed, in view, at the edge of each scan where appropriate.

Comparison strips are a small strip of graduated tones that are either white to black or from violet to red in standardised colours. These graduations are at regular intervals on the electromagnetic spectrum and provide an accurate comparison with the defined colours (Stouffer Industries Inc., 2006). By using a strip like this it can be seen what specific colours look like on the viewing monitor and, therefore, those in the archival material.

Langford and Bilissi and Millerson both offer straight forward descriptions of this topic for their very different audiences. In comparison Nurnberg (1969, p. 93) cites examples of composers who saw particular musical keys in specific colours and claims the differing colour to key matches mean that there are no rules for assessing colour.

3.10.4 Storage Media

It has long been appreciated by archivists that storage conditions affect the rate of decay of items. When first developed it was predicted that media such as the compact disc would have a lifetime of around thirty years, add on the decreasing costs of capacity compared to size of digital storage media and problems with digital storage media should be decreasing not increasing. However, as recorded in a report to the US Government in 2000, environmental conditions such as handling, humidity and temperature have as great an effect on this type of storage media as any other (Sadashige, 2000).

3.10.4.1 File Formats for Storage

In exactly the same way as different analogue media make a better job of recording different types of output, the same is true for differing types of digital file format. The most appropriate for each media type however is not always clear. Most of the usually selected formats are open source and widely available. Being open source it means that their encoding is freely available. In practical terms, it means that the format is likely to be supported and therefore have more than the standard two generations of backward compatibility (Bulow & Ahmon, 2011, pp. 42 - 43; JISC Digital Media, 2006b).

Some file formats for digital files are widely accepted as being suitable for storage e.g. for uncompressed image files .TIFF is often used. More recently JPEG2000 has gained ISO recognition and promises both lossless and lossy compression in the same file format. The advantage of using this lossless format over .TIFF files is the resulting file size; it is much smaller thus leading to equivalent reductions in the space required for storage (JISC Digital Media, 2006b; Western States Digital Standards Group: Digital Imaging Working Group, 2003, p. 28).

The International Organisation for Standards (ISO) is a network made up of the standards institutes from 163 countries. Its technical committees develop and publish standards whose contents have been agreed by it members (International Organization for Standardization, 2011a).

The Wellcome Institute makes decisions upon accepting donations to all their collections based upon a set of criteria, similar to a standard risk assessment, relating to the future management of digital materials (Thompson, 2010, p. 2). The file format of the digital object is only one of the criteria used. Part of the rationale for this is that traditionally archivists have had to manage their holdings no matter what the media (2010, p. 1) and on this basis they choose to not restrict the file types they will accept.

By accepting a wide range of file format types the Institute expects that the problems associated with the management of digital files will increase proportionately; particularly those associated with obsolescence (Thompson, 2010, p. 5). The threat of obsolescence is a major concern of records managers and archivists (Pearson & Webb, 2008, p. 90). Tools to combat this are focussed upon detecting when files in specific formats become unusable and in particular upon using a risk assessment approach. In their conclusions to the Automating Obsolescence Notification System (AONS) project, Pearson and Webb suggest that the way forward is perhaps through a community based watch system (2008, p. 102). By accepting many differing file formats it could be argued that the Wellcome Institute is spreading its risk with respect to format obsolescence. Conversely by restricting the formats in which digitised files are stored it could also be argued that the risk of losing access through obsolescence is reduced because there are fewer file types to be watched.

Table 10 is a table of file formats that are generally accepted for use in digitisation projects.

File Format Stem (Name)	ISO number	Media Used For	Why	Suitable for Master File (Y/N)	Suitable for Access Copies (Y/N)
.TIFF (Tagged Image File Format)		Still Images	Flexible, widely adopted. Lossless storage files therefore can recreate all of saved picture. Specification publicly available.	Y	N
.JPEG2000 [jp2] (Joint Photographic Experts Group)	ISO 15444-2	Still images	Combination of both lossless and lossy files. Platform independent. Specification publicly available.	?	Y
.MPEG4 (Moving Picture Experts Group)	ISO 14496	Audio Visual	Convergence across technical production paradigms.	Y	Y
.MP3	ISO11172-3	Audio	Lossy, compressed audio format. Open standard. Widely used.	N	Y
.WAV		Audio	Uncompressed audio files.	Y	N
.PNG	ISO 15948	Still Images	Uses lossless compression. Application independent. Not Widely used	Y	N
. PDF/A-1 (Portable Document Format)	ISO 19005-1	Documents	Open Source, Platform independent.	Y	Y
.ODF (Open Document Format)	ISO 26300	Documents including spreadsheets and databases/	Specification widely available.	Y	Y

Table 10: Table of File Formats for Digitisation

3.10.4.2 Digital Record Object Identification (DROID)

DROID is a software tool that analyses and indentifies digital files and their extensions. It works in conjunction with the PRONOM database. PRONOM is a database that contains details of file formats and their supporting software. Both were developed by The National Archives in the UK, to provide technical information about software and its versions and the file formats that run on them. It does this by keeping a registry of regularly updated information that is accessible via The National Archives (TNA) web site (The National Archives, 2011b; The National Archives, n.d.b).

3.10.4.3 File Compression

When digitising a picture the course usually recommended is to adopt a scan once procedure (Jones, 2008, p. 103; Western States Digital Standards Group; Digital Imaging Working Group, 2003, p. 5) to acquire an archival quality master and then create files for all other requirements from this. This is because using a large file size for other applications is not necessarily the best approach. For instance the time to download a large .TIFF file over a slow web connection may cause a user to give up on the site altogether. So creating user accessible files often means using a format type which involves compression. In file creation terms compression involves losing particular parts of the data stream which the algorithms of the software decide are not important. Two of the most popular for still images are JPEG and PNG.

3.10.5 Systems for Managing Digital Objects

Managing digital objects requires a combination of information, software and hardware (Higgins, 2008). The information comes from collecting good metadata (Bulow & Ahmon, 2011, p. 43; Haynes, 2004, p. 149). The hardware and software can be chosen from many on the market, but it needs to be suitable for the needs of the collection it is being chosen for (JISC Digital Media, 2007b; JISC Digital Media, 2009f; JISC Digital Media, 2010c). One specific type of system is a repository. This is defined by the repositories support project as "a *mechanism for managing and storing digital content*" (Repositories Support Project, n.d.b).

The Society of American Archivists (Pearce-Moses, 2005a) indicate that the origins of institutional repositories were in academia. Their function was envisaged as enabling the sharing of data in any format; an extension of the function of a conventional repository holding archival materials in the format they were deposited. By storing the intellectual output, repositories enable the re-use and dissemination of the resources of an organisation. In order to be open to discovery, their contents need to be able to be seen by their users. In general, this means adopting agreed standards so that resources can be retrieved. As well as using the supplied metadata to retrieve and access their contents, repositories use metadata to manage their deposits. This management can include long-term preservation as well as tools for reporting (Repositories Support Project, n.d.b).

How this capture and distribution is achieved varies from one organisation to the next. The Welsh Repository Network (WRN) is a group of 12 repositories; one from each Higher Education institution Wales. Even within this grouping there were variations in the configurations chosen by each of the repositories. These variations reflected local circumstances. The arrangements were:

- Nine were locally hosted using open source software,
- Two were hosted by a neighbouring institution and
- One commercially hosted (Welsh Repository Network, 2009, p. 3).

3.10.6 Catalogue

In order to retrieve content successfully it first needs to be organised in such a manner as to allow that to happen (Rowley & Hartley, 2008, p. 4). For digitised content this often involves the use of metadata. This metadata is generally organised into some type of catalogue. Catalogues come into their own when researching new areas as users don't know what the best sources are (Reid, 2000, p. 144). This is because the simple keyword search facility that exists in 'office' type documents can find instances of word use, but in them you know what you are searching for.

A catalogue is the representation of the facility (Davis-Perkins, et al., 2005, p. 6). The catalogue also expresses the basis upon which trust in the facility is founded because it is the intellectual representation of an archive. The catalogue should contain all the information that, in records management and archival practice, endows individual items with the attributes considered necessary to be able to prove that the item is what it claims to be. These attributes, and representations, combine to endow an object with authenticity. Without these attributes the items within an archive are worthless (Davis-Perkins, et al., 2005, p. 6).

Taking a purely mechanical view, Rowley and Hartley refer to a computer based catalogue as being

> *used to list the print and electronic documents within a library collection for purpose of inventory or access (2008, p. 10)*

within a framework of three principles for organising knowledge (p. 12). These principles acknowledge the need for catalogues to be user centric through the use of language that reflects the users' values and will therefore generate correct results from their point of view.

3.10.7 Metadata

Metadata means quite literally data about data (Deegan & Tanner, 2002, p. 112; Haynes, 2004, p. 6). In the case of digital files it is the data that is attached to, but not part of it that describes the file. It also has an important role to play in identification and management, both short and long term, of digital files (Bulow & Ahmon, 2011, p. 43; Haynes, 2004, p. 126 & 149; JISC Digital Media, 2008c).

In a collection of writings on electronic records management issues Cumming (2005, p. 36) states that:

> *the application of appropriate metadata can help to maintain records and preserve their structure, content, accessibility and meaning but, at a higher level, implementation of metadata can also help testify to the authenticity of records....Because of their inherent vulnerabilities and the threats that they face, it is critical that records are able to be demonstrably valid and accountable chronicles of the business they document. Applying metadata can help testify to the authenticity of records by documenting their effective maintenance and management.*

If you replace the word “record” with digital file, this statement neatly encapsulates the issues faced in digitisation projects too. While authenticity is not a regulatory requirement in this particular instance, demonstrable, effective documentation processes can only help with finding solutions when problems do occur. This includes documenting the decisions made during the digitisation process (Haynes, 2004, p. 128).

The issues surrounding representation, and in particular, the compilation of metadata were raised by Jones in his report on the lessons learnt at the NLW where one of his conclusions is that metadata creation needs to keep pace with digitisation (Jones, 2008, p. 112). On page 105 he indicates how in relation to staffing in the 2005 NLW strategy the fact that metadata is usually individually hand crafted, because it often cannot be mechanically created, would mean that any increase in digitisation would require a reciprocal increase in metadata output.

This manual creation brings about a conflict with quality assurance. Irregularities in file metadata can be an indication of a damaged file, or it can be due to human error. Irregularities need to be checked and repaired by appropriately skilled people as there is currently no technological method to do this (JISC Digital Media, 2008b). In 2011 there was a project to attempt to provide some answers to this conundrum. As part of the Open Planets Foundation (OPF) it is hoped that the outputs generated from the AQuA project will lead to the development of sustainable tools available to the whole digital preservation community (Wheatley, 2011). This project was important because, providing they are packaged in an accessible manner, these outcomes will have uses for all those who create digital files.

3.10.7.1 Types of metadata

Metadata is usually accepted as being further divisible into 3 parts: administrative, descriptive and structural metadata (Deegan & Tanner, 2002, pp. 116-120). This is summarised in Table 11.

3.10.7.2 Administrative metadata

Administrative metadata is the data that is needed to manage the digital object and includes preservation metadata (Haynes, 2004, p. 120). The types of data include scanner settings, the date of capture of the digital object and the file format. This is the section that is updated whenever a process is carried out on the digital object. In a lending situation this is also included here.

	What it Does	**Example**	**Function**
Administrative metadata	Describes the features of an object(s) used to manage the object(s)	Who printed the book and when; where it is kept and if it is in good condition. If the book has had conservation treatments or been loaned.	Used to manage an object throughout its life. Includes all data needed for managing preservation (Deegan & Tanner, 2002, p. 120)
Structural metadata	Describes the internal structure/arrangement of the object(s)	Describes the arrangement of the item or the set of items e.g. the order of the pages of a book or which page an illustration is in that book	Structure is part of the meaning of an object. Especially important for digital objects as structure doesn't have an enduring physical form. Structure is part of meaning (Deegan & Tanner, 2002, p. 117)
Descriptive metadata	Describes the appearance of the object(s)	Describes what the book is about, what it looks like, the paper it is printed upon and when etc.	Object descriptions help humans identify an object. Also aids attributing meaning to an object (Deegan & Tanner, 2002, p. 116)

Table 11: Table of Metadata Types and their Function

3.10.7.3 Structural metadata

The structure of archival materials is often obvious to sighted users. For those with visual impairment it is less so. For digital objects physical structure needs to be copied.

The structure of, for instance, a book is a very important part of the archival material and this should be repeated in the digital object so as to recreate this. In recreating the page order of a book as well as recreating the physical appearance the relationship between the pages is also recreated. This relationship also recreates meaning inherent in the archival material in the digital object e.g. if the archival material was in large print or on a postcard rather than plain A4 paper, recreating the meaning recreates the subliminal meanings carried by an archival object (Bulow & Ahmon, 2011, pp. 93-94).

Structural metadata is also used as an aid to navigation. If the pages in a book are randomly arranged then reading from one page to the next is at best extremely difficult (Haynes, 2004, p. 118).

3.10.7.4 Descriptive metadata

Descriptive metadata is the kind of information found in a traditional catalogue. It fulfils the same function for a digital object and is often used to link the archival material to the digital object (Bulow & Ahmon, 2011, p. 43).

3.10.7.5 Metadata Schema

Having been collected, metadata needs to be arranged for humans to be able to use it effectively. This can be done using individual methods, but it is easier for others to understand if the way data is arranged is consistent. The advent of computer systems made this an arrangement issue as they cannot interpret and re-arrange data in an order to suit the individual, as humans can (Rowley & Hartley, 2008, pp. 42-43). To facilitate the exchange of data, a series of rules governing the presentation and contents of metadata were drawn. Some of these are illustrated in Table 12.

	Has Part(s) Called	**Does what?**	**Example**
Metadata Structure standards	Element sets/schemes	Rules for the way the contents are arranged	Dublin Core Metadata Element Set (DCMES)
Metadata Content standards		Rules for writing the contents down e.g. the way a date is written	Anglo American Cataloguing Rules (AACR2)
Metadata Exchange standards	Formats	Rules for the way the contents are sent or received	MARC21
Metadata Values standards	Thesauri/Controlled Vocabulary	Rules for the words used to describe contents e.g. World War 1914 – 1918 rather than WWI.	Library of Congress (LoC)

Table 12: Table of Metadata Arrangement Rules and Applied Standards

These standards show how the different functions of metadata interact to produce information that can be accessed in different forms and formats depending upon the situation (Rowley & Hartley, 2008, p. 44). The strength of a standard depends upon the intended end-use of schema e.g. the strength of Encoded Archival Description (EAD) is in structural description (Deegan & Tanner, 2002, p. 118; Zeng & Qin, 2008, p. 15). The schema listed were chosen either because they were used in NLW, or because they seemed to warrant closer scrutiny for possible use in CPR.

3.10.7.6 ISAD(G)

The General International Standard for Archive Description (ISAD(G)) is a standard devised to achieve consistent description of content and context through the use of rules for organisation (International Council on Archives Ad Hoc Commission on Descriptive Standards, 2010, p. 7). There are paths for crossing between other standards such as EAD as interoperability was also a design feature (Haynes, 2004, p. 57).

3.10.7.7 EAD

The EAD schema was devised to turn into machine readable code, archive descriptions and catalogues. Based on ISAD(G) it uses XML to communicate. Its strength is in describing the structure of archives (Deegan & Tanner, 2002, pp. 127 - 128).

3.10.7.8 Dublin Core Metadata Element Set (DCMES) or DC

DC was devised to stand between internet search engines and professionally drawn catalogues to enable resource discovery. It has 15 parts, called elements, which provide some of all the metadata types needed for electronic information exchange and can be arranged in any order. The elements can be refined by extending the basic element or repeated if required (Zeng & Qin, 2008, p. 18).

DC has been used and adapted by many users. In the UK the TNA has used it as the basis of the minimum acceptable requirements for its Electronic Document and Records Management systems (EDRMs), a specialist type of computer system for managing records (Haynes, 2004, p. 52 & 105) and Europeana collects a subset of DC along with its own additional details (Day, 2010, p. 8).

Europeana collects metadata from cultural and scientific collections around Europe and links back to the original source. It was launched in November 2008 and was fully functional in April 2009. It gives access to more than 1000 collections throughout the European Union. The digital objects that can be accessed came from a wide variety of institutions. The information was aggregated at a national level before being uploaded to the Europeana system (Europeana, n.d.). The use of DC by Europeana indicates that it is adaptable to many situations and collections.

3.10.7.9 AACR2

AACR2 stands for Anglo-American Cataloguing Rules 2 and is a standard for writing catalogues. It regulates details such as whether to place a full stop after the word 'Mr' or how names should be written in a catalogue entry e.g. it prescribes that the author's name should appear as Shakespeare, W. rather than W. Shakespeare. These rules give entries consistency and make it easier for a user to scan the detail in a catalogue. It also enables accurate automated searching for specified entries as machines cannot 'see' variation (Rowley & Hartley, 2008, p. 72).

3.10.7.10 MARC

Machine Readable Cataloguing Standard (MARC) was established to allow the mechanisation of many of the processes associated with cataloguing. After evolving through several different generations, MARC21 was designed in 2000 and subsequently adopted widely. Created originally for traditional bibliographic records it has maintained its industry place by adapting to new requirements through the adaptation and introduction of new fields for description (Rowley & Hartley, 2008, pp. 85-92).

3.10.7.11 METS, MADS and MODS

METS, Metadata Authenticity Description Schema (MADS) and Metadata Object Description Schema (MODS) are schema which when used together create a complete suite for transferring information about a digital object. METS transmits the structure, MADS the authority information and MODS describes the digital object. They are expressed in XML and maintained by the Library of Congress. The strength of METS is that it is a modular packaging standard. This modularity means that it can carry data from other schema making it interoperable with any schema or schema profile that can be expressed in XML. METS is also able to construct cross walks to other schema to allow metadata exchange (Day, 2010; Digital Library Federation, 2001; Haynes, 2004, p. 158; Library of Congress, 2011c; Library of Congress, 2010; Zeng & Qin, 2008, p. 308).

3.10.8 Preservation

The preservation of photographs and negatives is relatively simple if they are not in need of any treatment (Cartwright, 2011; The National Archives, n.d.a). The general recommendation is to wrap prints and negatives in acid free paper or cellophane and stand on their edge in a box also made from acid free card (Library of Congress, 2011b).

Digitisation is not preservation. The effect of digitising an item can affect the preservation workflow of the original, but this could also be seen as an opportunity to develop a workflow to incorporate creating a digital object (Bulow & Ahmon, 2011, p. xvi). To aid preservation of digital files, tools have been developed to aid different parts of the workflow. The development of PLANETS, which is a complete suite of tools for preservation of files, has incorporated and developed some of these (PLANETS, 2007).

The Open Archival Information System (OAIS) Reference Model is an ideal model that describes the processes needed to ensure the preservation of digital objects (Consultative Committee on Space Data Systems, 2002). ISO 14721 is applicable to any institution that keeps archival materials or digital objects for the long term (International Organization for Standardization, 2011b).

3.11 What is an Image?

The definition of an image is an example of an everyday word which has been imbued with a set of specific, context dependent values when applied to that particular instance.

In everyday use, an image is a static representation of the perspective of the author of an event or the visualisation of an artistic creation. This is consistent with dictionary definitions (Cambridge University Press, 2011; Merriam-Webster Inc,

2011) which indicate the imaginary, artistic aspect of an image. This coincides with the philosophical position of archives seeing an image as a record of something.

Definitions from the computing discipline, however, call an image an array of pixels arranged in two dimensions (Howe, 2010; Merriam-Webster Inc, 2011). This definition avoids other complications that occur when digital files are stored. For instance, if you wrote a letter on a computer, what you would see as you composed it, according to the computing discipline definitions, is an image. If you saved the letter as an image e.g. by using the print screen function, when you re-opened it you would need to use imaging software to be able to see what was on the screen when you saved it. If you saved the letter as a text file, then to open it you need to use text enabled software. This has implications for digitisation and preservation because the future use of files created through these processes need the appropriate software to render them accessible. In addition to this, if files had an additional function to fulfil, such as being a record, this also needs to be taken into account when processing the archival material from its analogue state to its digital one.

So there is no agreement about what images are and how users interact with them. What about photographs? In his book tracing what he called his "*catalogue of displeasure*" (2011, p. 219) on aspects of photography that he found puzzling, Elkins recognised his own reactions to a copy of a photograph of a selenite window (2011, p. 28). Throughout, he compared these to Barthes descriptions of his reactions to the same window in *"Camera Lucida"*. He described his reactions to touching this copy and how these compare to those for the digital object (Elkins, 2011, p. 24).

> *....but in another sense photographs are all about touching. When I hand someone a photograph, I am touching its surface….if the photo is onscreen, I may touch the glass to point out something, smearing it a little with the grease in my fingertip. I can't agree with the notion that photography has become "information in the pure state" just because it is digital. There is always the surface, and now there's light from the screen.*

And on page 26:

> *My eyes can touch the surface of a photograph. If it is a print made in a darkroom, I can see its surface as a griffonage (an illegible handwriting) of*

marks and scratches. If it's onscreen, I can just barely make out the fuzzy mosaic of RGB sub-pixels or, if it's an older monitor, the woolly RGB phosphor dots….The handling of photographs is a social act, and the optical feel of a photographs surface is something that almost everyone who writes on photography ignores.

Berger on the other hand was only interested in the relationship between the image and the viewer, although he does go on to indicate that care needs to be taken to keep the context of an image in mind when viewing it (1972, p. 4).

Berger based much of his commentary on the depictions of art over time where as Mitchell and Giles seemed to feel that images carry much stronger messages today (Giles, 2010, p. 3; Mitchell, 1984, p. 503). It is generally accepted that advertising works on the theory that the viewer places themselves into the image created. This contrasts with photography and art which work on the premise that the viewer sees themselves in relation to the subject of the image (Chandler, 2000).

These descriptions of images and the impression they make, appear to support the idea that everyone in digitisation process should be trying to convey the same things, including the context, that are present in the archival materials. But that the view of what this is, is individual. And that this applies to the photograph, whether it is viewed on paper or a screen.

In archives and libraries, Conway indicated that evaluations done to date have concentrated upon retrieval, and not whether the digital image had a different effect upon the user to the archival material. In other words whether being a digital object affected its context significantly for the user (Conway, 2010, p. 426).

3.11.1 Viewing Digital Objects

Accepting that there is a paucity of information recorded on images and their influences on archives it is unsurprising that there is also little that relates to the interaction between archives and digital objects.

An understanding of the way colour is made and viewed is a complex physical concept, but it is necessary in order to understand the processes carried out by the software and hardware used in digitisation. Becoming aware of the use of colour is a good way of becoming aware of the way a particular image is composed. This in turn can act to make the user aware of the way technology can affect how a digital object is viewed and how tools such as scanners can alter that perception during digitisation if the user is unaware (JISC Digital Media, 2010c, pp. 4 - 5; Millerson, 1972, p. 41). It can also serve to show the limitations of the equipment being used. Humans have an array of methods for mentally comparing the colours of objects they encounter. In comparison the reproductive ability of man-made devices is limited to three colours, red, green and blue, and any technological input achieved by its manufacturers. On this topic Langford and Bilissi conclude with the observation that while this may make the achievement of accurate colour rendering seem very difficult it is worth doing to make human sensitivity the variable in a situation (Langford & Bilissi, 2011, p. 21). Developing this skill is central to achieving a consistent standard when creating digital objects.

In theatre, the fourth wall is an acknowledged and accepted idiom which is used, or ignored, depending upon the style of performance being acted out. Less openly acknowledged, until the 1960's, was that associated with television, when Mary Whitehouse was one of the first to question the effect of the movement of the wall from public spaces, such as theatres or cinemas, to the home. More recently, that movement has become more prolific with the advent of screens for computers and telephones.

When used by a production that wall is being openly acknowledged; but in a computing environment where the involvement of the individual can be officially sanctioned through the workplace it is less obviously accredited. This lack of recognition of a sequence of actions can lead to the normalising of a situation, which in turn can lead to an implicit level of trust being placed upon it.

Broadening this enquiry into the area of media studies, it should be noted that 'media' generally means radio, television and newspapers, or a specifically mentioned sector.

McCullagh, a sociologist, and Giles, a psychologist, have both pointed out that concern over the influence of media on lives has varied according to which was source prevalent at the time (Giles, 2010, p. 4; McCullagh, 2002, p. 4). Both also record that there is a lot of media influence on individuals. So much so that:

we sometimes exist in a state of 'media blindness', where we fail to acknowledge the presence, much less the influence, of media in our decision-making, beliefs and attitudes (Giles, 2010, p. 3).

As Giles indicated media are not simply communication devices, and that sorting the information wheat from the chaff takes a level of media literacy users may not have. To achieve the required literacy, users need to use the range of adaptive skills and knowledge of conventions they brought from other media to exploit the possibilities of the internet (Giles, 2010, p. 8 & 12; McCullagh, 2002, p. 174).

In general two theoretical views of the media are accepted. One is that held by the media, which sees itself as transparent. The other adopted by sociologists, sees the media as presenting its own selective view of the world (McCullagh, 2002, pp. 14-15). Whichever view is dominant, it is generally accepted that users often just go along with the views espoused by their chosen broadcaster (McCullagh, 2002, pp. 167-168).

This view stems from user research which generated the theory, screen theory, that an audience is made up of individuals whose language skills place them in society; and that this society is limited by the same language. In screen theory this model is thought to echo the way that media operate (McCullagh, 2002, p. 156). Other research models have moved from seeing the audience as a mass of receivers to a mass of needs seekers. The latter of these is research on the Uses and Gratification Theory which expands on Maslow's Hierarchy of Needs. The relevance of this theory is that it describes how users formulate favourite information sources by seeking sources that fulfil their personal hierarchy of needs (Giles, 2010, p. 21). Both of these models share the view that the transmission and reception sides of media are divorced from each other (McCullagh, 2002, p. 152). McLuhan on the other hand saw each media device as having a direct cultural effect and therefore, that the generating device and the effect of receiving a transmission were linked (McLuhan & Fiore, 1967).

In his introduction, Giles indicates that prior to the advent of the internet, psychology had, in general, avoided studying various media and that previously only the fact that media deliver stimuli was acknowledged by experiments (2010). The image subject used by most audience studies has been violence. These experiments were often conducted by measuring the response to an isolated stimulus and the response to it measured. This is called effects research (Giles, 2010, pp. 16-19). However, as McCullagh (2002, p. 168) indicates, attempts by research to separate the stimulus from its cultural surroundings have also produced questionable results. This is because audience studies have been divorced from reality by placing them in a laboratory setting. It is felt that this leads participants to construct different meanings to those they would build in their own environment. Experiments that have viewed users in their own home had a similar problem as participants must be conscious of being recorded, if only for a proportion of the research time.

Commentary specifically on the internet in media studies sources is generally focussed on its use by large corporations as an add on to an already effective selling and distribution network (Burton, 2010, pp. 192-193; McCullagh, 2002, p. 167).

As a means of accessing information the internet is promoted as being democratic because it may lessen the influence of traditional media outlets. It is felt that this will happen because it will increase the pool of information providers. However what site owners choose to place on the internet will still be selected (McCullagh, 2002, p. 109). But as he goes on to point out

> *while one of its [the internet] strengths is that there is no formal filter or censor to restrict the kind of information that is available, equally there is no mechanism to guarantee the accuracy and completeness of that information (McCullagh, 2002, p. 119).*

Early research focussed on what users accessed on the internet now, from a psychological point of view, it focuses on user preferences between online and offline social interaction (Giles, 2010, p. 170).

These studies have implications for archives and libraries because they host collections that transmit information to their users. Much of the work on photography has concentrated upon its artistic properties, whereas that on advertising has assessed the message being relayed. Because research has concentrated on specific traits of that media in set circumstances only general conclusions on their applicability to archives could be drawn. This is because viewing digital objects and images in archives and libraries is very context dependent. Studies that exclude this therefore remove a vital part of what the metadata and the items are trying to convey.

3.12 Chapter Summary

This review of literature covered different media types and a range of issues.

The most often quoted advantage of digitisation is increased ease of access to items, usually achieved through a wider broadcast of the file than the original can obtain (Hughes, 2003, p. 9). The main disadvantages of digitisation are the associated long term cost commitments to manage and maintain the digital file, and the possibility of damage to the original whilst being processed. (Davis-Perkins, et al., 2005, p. 7; JISC Digital Media, 2008c, p. 3; Spence, 2005, pp. 373 - 374).

The impact on small collections has been difficult to assess as many of the available case-studies relate to large facilities carrying out work on small collections they hold as opposed to small facilities trying to digitise their collections (Everitt, 2005; Hampson, 2001; Rayner, 2009).

The tools required to implement a digitisation workflow include those that facilitate management of the digital files produced and hardware such as scanners. It is generally accepted that the decision to preserve a digital file should be made as soon as possible after creation so as to mitigate the effects of obsolescence.

There exists a wealth of information to allow practitioners the opportunity to improve their practices but there is little written that is aimed specifically at those in small facilities which lie outside the remit of traditional memory institution provision. In order to comply with the indicated outcomes of the project, research needs to be carried out into the adaptation and scaling of the ideas which underpin digitisation and preservation of digital files.

Chapter 4: Research Design

4.1 Research Outline

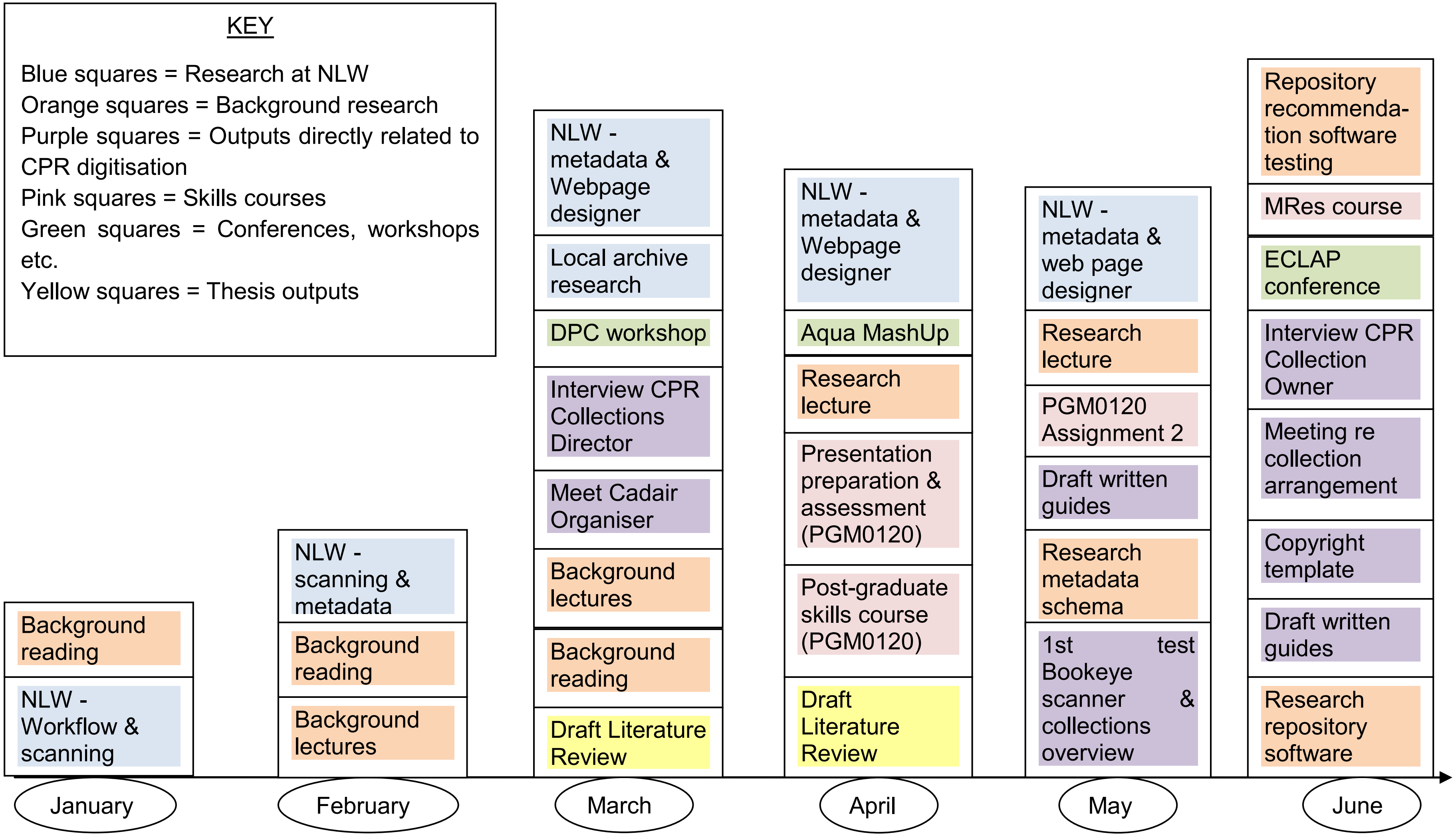
The research timeline in Figure 8 shows how all the elements of the research fitted together during the year. The key colours match those used in Figure 9 which shows how the separate strands of the workflow combined to produce the system for CPR.

In the timeline there were three additional colours used: green, yellow, and pink compared to the workflow. The yellow squares represent drafts of the written outputs for the thesis. The green squares represent external events attended to add to the research background and the pink, study skills courses.

The background reading and lectures, coloured orange, provided the theoretical background needed to re-enforce the practical learning at the NLW. The skills acquired at the NLW, coloured blue, were then adapted and applied at CPR, the purple.

The processes in the general outline of the research method were listed and expressed in a shortened form in the flow diagram, Figure 9. These were:

- Acquire the practical skills needed to design the written guides from the NLW
- Conduct a search of the available literature
- Interview selected information professionals regarding their views on the past, present and possible future of digitisation.
- Extract from the learning any special considerations to be applied to the digitisation of independent small collections
- Apply skills learned to:
  - Design the written guides, and
  - Test the guides by digitising a collection in CPR
- Note any areas of difficulty or compromise
- Take action to resolve difficulties
- Discuss issues raised by the guides and process testing
- Evaluate the written guides using volunteers from CPR
- Note and take action on any recommendations arising from the evaluation


KEY
Blue squares = Research at NLW
Orange squares = Background research
Purple squares = Outputs directly related to CPR digitisation
Pink squares = Skills courses
Green squares = Conferences, workshops etc.
Yellow squares = Thesis outputs
Background reading
NLW - Workflow & scanning
January
NLW - scanning & metadata
Background reading
Background lectures
February
NLW - metadata & Webpage designer
Local archive research
DPC workshop
Interview CPR Collections Director
Meet Cadair Organiser
Background lectures
Background reading
Draft Literature Review
March
NLW - metadata & Webpage designer
Aqua MashUp
Research lecture
Presentation preparation & assessment (PGM0120)
Post-graduate skills course (PGM0120)
Draft Literature Review
April
NLW - metadata & web page designer
Research lecture
PGM0120 Assignment 2
Draft written guides
Research metadata schema
1st test Bookeye scanner & collections overview
May
Repository recommenda-tion software testing
MRes course
ECLAP conference
Interview CPR Collection Owner
Meeting re collection arrangement
Copyright template
Draft written guides
Research repository software
June

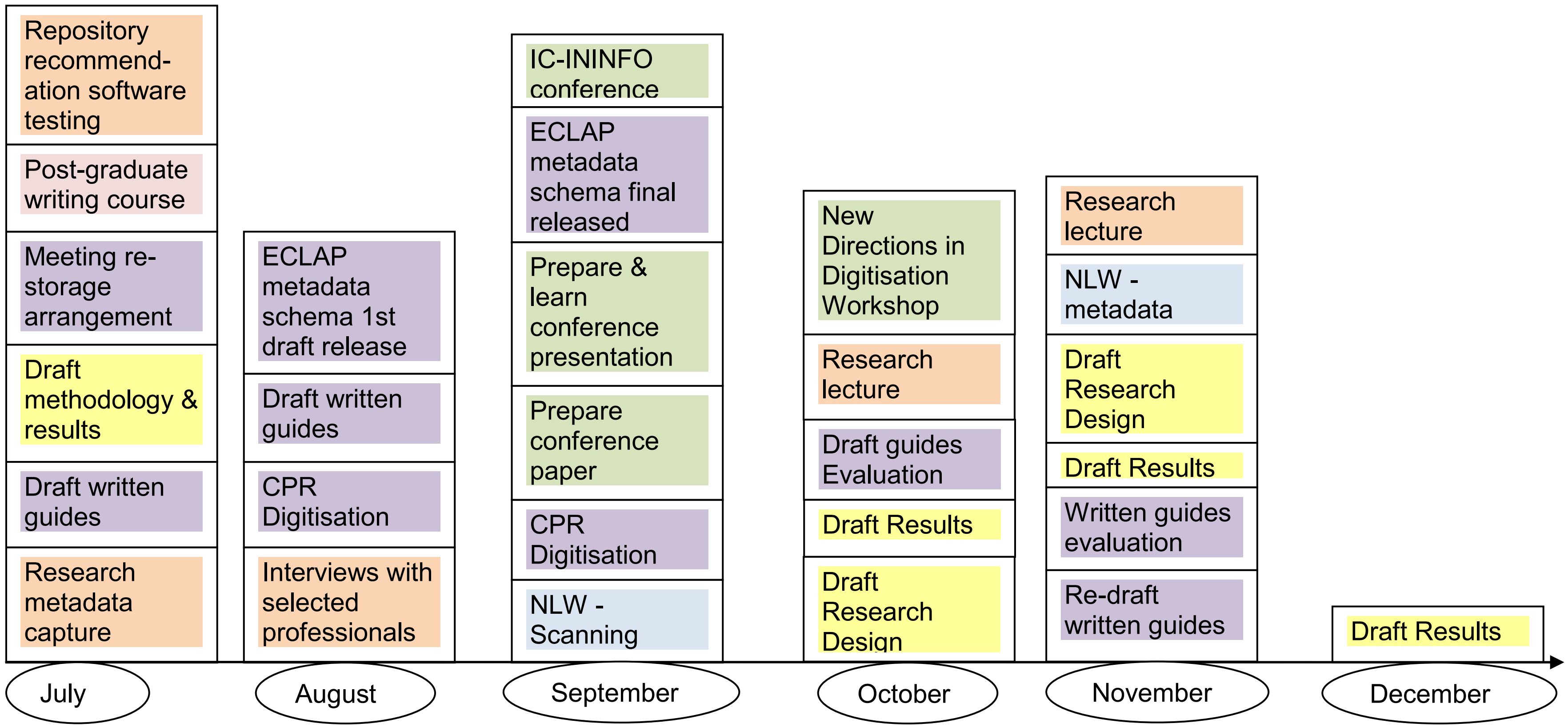


Figure 8: Research Timeline (2011)

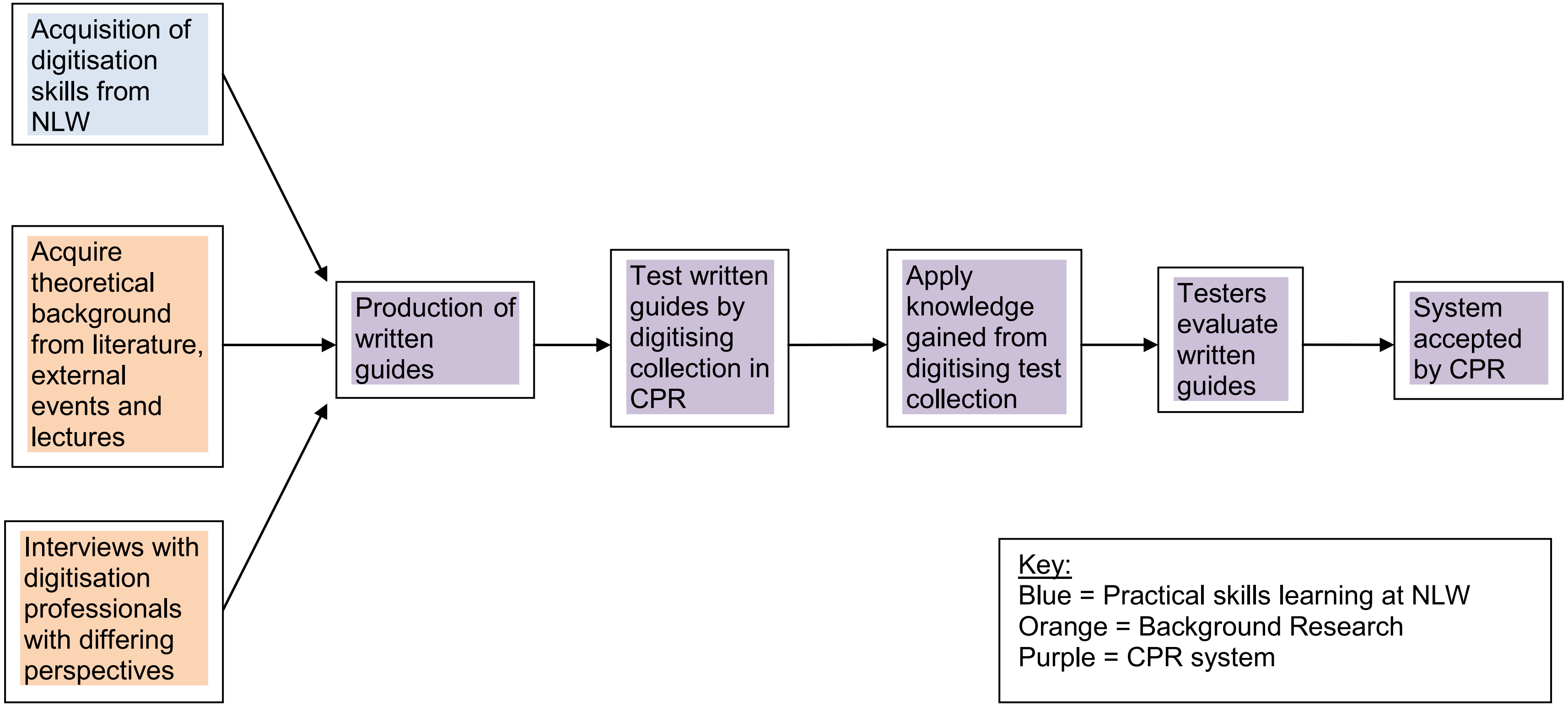


Figure 9: Workflow to Transfer Knowledge to CPR

From this list it could be suggested that the research design followed a scientific enquiry. This design is generally accepted as one where a theory is advanced (learning skills at NLW), tested (design of written guides and test collection digitisation) and then, either proved or disproved (evaluation, results and conclusions).

In many scientific experiments the materials being tested are consistent in their nature; for example, if you heat water at a consistent pressure to a given temperature, it will produce steam. This consistency means that repeating the same procedure gives the same result. But scientific research design does not take into account that in this research, it is the people that influence the outcomes.

In this research the people involved were being tested as they had to understand and apply the techniques learned. The learning and understanding of the techniques took place at the NLW. This was then applied: first, when the guides were written; and second, when the users of the guides digitised archival materials. Because of this transfer to a second group, when the written guides were used, the research does not have reliability as it cannot be repeated. This was because the same population could not be used a second time as they will have learned from the first experience (Bryman, 2001, p. 29). If the research was repeated with a different population, the result would not necessarily be the same because either the culture of the population, or the situation of a similar culture, would be different and this could affect the outcome.

In a diagram on page 22 of his book, Bryman (2001) lists five factors that influence social research. These factors are: theoretical considerations, practical considerations, values, epistemology and ontology. Each of these has affected the research with varying weight as the priorities during the study fluctuated. For instance during the acquisition of skills, theoretical and practical considerations took precedence over the ontology of the research as this was the objective during this stage. Similarly whilst applying the learned skills, the priority was communicating the values held within the cultural sector to a small department outside it. Theoretical considerations and their associated values were conveyed during the training at the NLW. This was most noticeable in relation to metadata gathering and the arrangement of the archival materials.

The internal validity of the research depended upon what was taken as the independent variable. Internal validity comes from the research being repeatable. In this investigation if the people were the independent variable, then the research loses internal validity. The research does, however, have external validity (Bryman, 2001, p. 30) as it can be applied to other situations with differing circumstances through extending the learning gained from the written guides.

Considering all of the above indicates that rather than being a scientific enquiry, the research is a case study that uses knowledge transfer to "*reveal[ing] important features about its nature*" (Bryman, 2001, p. 55).

4.2 Ethics

For this research ethical considerations centred on two areas. One was selecting, accessing and interviewing information professionals, the other digitisation process issues.

For the interviews information professionals in the Aberystwyth area, were selected to represent both academia and digitisation practice. Aberystwyth was chosen as it is the location of both the NLW and CPR. The selected professionals were contacted initially by email. A copy of this email is in Appendix D: Example Letters. If they consented to take part, they nominated a place and time for the interview which was expected to take no more than an hour.

Whilst at the NLW I sought advice from my supervisor on acceptable ways to communicate my research so that I used terms that were acceptable to them.

The evaluators of the written guides were told the ethics policy with particular emphasis given to their being free to stop and leave at any time during the process. This was done verbally as, until they came into the room, I did not know who had been asked, and, therefore, the best way to communicate with them.

The recorded answers from the interviews and evaluations were transferred to a computer, to which I am the only person with access.

The issue of the privacy of the subjects of the photographs digitised in CPR was addressed as all were informed of the intended purpose of the photographs, and signed a form to allow their use, at the time the photographs were taken.

No vulnerable groups were contacted during the research and none of the data accessed would be considered personal data under the Data Protection Act 1998.

4.3 Acquisition of Digitisation Skills

The main channel for acquiring the relevant practical skills for transfer was the NLW. This was the starting point for the whole research project. The aim for this part of the research was to track the digitisation process, from beginning to end, by following the digitisation of two collections. This involved all aspects of the process from attending workflow meetings, to visiting the relevant local archives office to research the subjects in the archival materials.

Prior to my arrival, the collections to be followed through the workflow had been selected using the digitisation policy of the NLW. As well as the policy, there were additional restrictions on the number of archival materials in the chosen collections. The number of archival materials was restricted, and the NLW had to have copyright so that the collections would be digitised within the time available for the research.

The digitisation process started with a meeting to establish the best workflow for the collection. Conservation treatments were then carried out by the conservation unit, followed by my starting the workflow on the NLW database. Turning the archival materials into digital objects was carried out using the allocated scanner in the digitisation unit. These were then adjusted and re-sized to make the digital objects for displaying on the web.

I then learnt how to write the metadata by spending time sitting at the shoulder of one of the metadata writers observing and discussing how the form fields were completed. After this templates, appropriate to the collection I was to write about, were drawn. There was a lot of detail to absorb and it took a while, but having completed the basic metadata and attached the digital object, I began to search local newspapers for some of the events depicted.

When the metadata was complete there was further training in the use of the software Typo3. This was used to write web pages to make an exhibition to highlight the collection to users. To get user feedback, a form that would email a comment to an allocated address was included in this exhibition.

As time allowed, a second, smaller collection was digitised later on in the project timeline. For this second collection, the initial metadata was written prior to digitising the collection, and the digital object attached to the record so that it can be accessed via the Library catalogue.

4.4 Acquisition of Theoretical Background

A search for relevant literature revealed that there are many papers written about digitisation and its practice, but that it can be hard, without some prior knowledge of a similar area, to divine what is meant. Readings on related topics, such as photography were informative about specific common areas, such as colour composition, but they did not deal with digitisation specifically.

It was recognised when the scholarship was granted that there would be gaps in the background knowledge in one or more project areas of the chosen candidate. In my case, my knowledge of archives and their practices was thin. This was addressed by attending relevant lectures and reading accompanying course materials on the Archives Administration course and the Archival Description and Digitisation module at the Department of Information Studies, Aberystwyth University.

There were also opportunities to attend lectures as part of the Aberystwyth Information Research Seminars (AIRS) series. Of these, lectures by Professor Lorna Hughes, University of Wales, Chair in Digital Collections, and Andrew Green, Head Librarian of the NLW were particularly relevant. I also took the opportunity to attend an open lecture given by Dame Lynne Brindley, Chief Executive of the British Library on the possible future role of libraries.

4.5 Research Training

The research training was delivered in two courses, PGM0120 and MRes, that were compulsory elements of the KESS sponsorship, and a voluntary writing course.

PGM0120 Research Training took place over four days in April 2011 and gave an overview, through the modules presented, of some of the issues to be considered while completing, and writing for research. There were two compulsory assignments which had to be passed to complete the course. The first took the form of a five minute presentation. The topic given was how the course I had just completed would influence how I wrote about my research. The second assignment was an analytical piece on different styles of academic writing and how these style issues appeared in our research area. Eight people took part in the course, most of whom were Doctoral students in their first year.

The MRes course was held in June 2011 at Nantlle, North Wales. This took the form of two days of tasks designed to develop skills that could be used to gain, or during employment. They concentrated for the most part on team working and the roles people occupy within them. A total of nine KESS sponsored students from all over Wales attended.

The Summer Writing Course took place at Aberystwyth University during July. The three day course covered a range of writing topics. In particular it expanded upon some of the topics in PGM0120. The structures of writing were discussed in depth. There were two individual tutorials and two department specific group discussions. These discussions were led by group members on topics that had been covered during the course. Our group consisted of two in the room, and one taking part using a video link.

4.6 External Events Attended

The external events to attend were selected because it was felt that to go to them would have a valuable impact upon the research. The criteria adopted considered whether the events appeared to offer valuable insights into digitisation practice that were highlighted by the literature review as possible areas of difficulty.

The events chosen, after consultation with my Supervisors, were:

- A DPC Workshop Getting Started in Digital Preservation (21st March 2011): held at Glamorgan Archives, Cardiff, Wales.
- Aqua MashUp (11-13th April 2011) held at Leeds University, Leeds, Yorkshire. An event that placed collection owners with IT experts to solve common problems associated with digitisation.
- ECLAP Conference (9-10th June 2011) held at La Bellone (Theatre), Brussels, Belgium. A two day conference to highlight particular difficulties and discuss solutions to them, for the repository.

- AEIOU focus group (June to July 2011) held at Aberystwyth University, Aberystwyth, Wales: testing and evaluation of recommendation software.
- IC-ININFO 2011 (29th September – 3rd October 2011) held in the International Convention Centre, Kos, Greece: an international conference to explore the theme of communication, and innovation, in the areas of information practice represented
- New Directions in Digitisation Workshop (28th October 2011) held at NLW, Aberystwyth, Wales: a one day event that presented findings into recent research into digitisation practice.

The events that had the largest influence on the research were the DPC Workshop and the ECLAP conference.

The DPC Workshop introduced digital preservation and gave tips on the first steps to take. This was important to the research because preservation of the digital object should be considered when digitisation takes place to ensure that the relevant metadata is collected (Bulow & Ahmon, 2011, p. 43; Haynes, 2004, p. 126). A simplified approach similar to that used in PLATO in the PLANETS project was recommended and exercises to illustrate this were carried out (PLANETS, 2011). These exercises indicated that adapting, or extending, the kind of assessment often carried out on archival materials when they are being accepted into a collection, was a suitable method for the research to investigate.

ECLAP held a conference aimed at all involved in the collecting of performing arts materials in digital form. I discovered the organisation and the conference from an email sent to the JISC repositories mailing list and, having learnt more about the organisation, I asked to attend the conference.

The eventual aims of the project are similar to those of CPR; to have a repository of performing arts materials that are acknowledged to be searchable and of consistent, quality. The problems they had anticipated had similarities to those that could be expected for this project. Two large differences were, the scale of the proposal and the fact that the costs of digitisation would be borne by contributors rather than the repository.

Prior to the conference I had discovered that some of the companies that had been mentioned in relation to some of the holdings at CPR were already members of the organisation. At this point I felt that by uploading selected resources, CPR could use ECLAP as a pointer to the rest of their collections when digitised. After discussing the conference outcomes and the common interests of other member groups with the Director, CPR took the decision to become formal members of the network. This decision had a large influence on the metadata (ECLAP, 2011).

The Aqua MashUp was a JISC sponsored event held in Leeds and London. I attended the one held in Leeds. It aimed to put experts and collection owners together to attempt to solve some of the known problems associated with digitisation and preservation. The technical outcomes from this event were of limited use to the research as they would have needed substantial adaptation. This adaptation required advanced IT skills which were beyond the scope of the project.

The AEIOU focus group was another JISC sponsored project that met on two afternoons, approximately a month apart. AEIOU is recommendation software that searches titles and authors for a specified term. The search conducted is nominated as successful if an item is chosen. Having done this, any further searches within a specified time are deemed to be related to the initial search. When another user enters the same term, once they choose an item it lists the top five of the items appearing on your search as 'possibly useful'. The focus group was introduced to the software and asked to start creating links using selected terms in specific repositories within the Welsh Repository Network (WRN). We were then asked for feedback on our experiences with the software. The second meeting focused on the time limit given to making associations' between items. Again feedback was collected and collated. This event reminded me of the importance of reliable results being retrieved during searches for resources.

At the IC-ININFO 2011 conference meeting professionals, in similar areas, from different parts of the World was rewarding. A paper giving an outline of this research was submitted and chosen for presentation. The process of writing and presenting a paper for the conference was useful since it emphasised summarising and presenting the research in a manner that could be understood by a varied audience. The theme of the conference was the future of the field of information. This was addressed in a variety of ways, but the common theme was how information practice communicates with an outside audience.

The New Directions in Digitisation workshop focussed on the outcomes from recent digitisation projects. This was held at the NLW. It concentrated on projects that showed the value or impact of digitised collections and how additional technology could be used to provide greater user interaction with a resource. An example was the use by Putting Pictures in Their Place, a Royal Commission for Ancient and Historical Monuments, Wales (RCAHMW) project, which uses Google Maps to provide additional information on sites for mobile phone users. This kind of application could be utilised by CPR in the future to pinpoint places relevant to their holdings.

4.7 Interviews with Professionals

A list of potential candidates to be interviewed was compiled after consultation with my Academic Supervisors. Interviewees who had many years experience of digitisation from a variety of standpoints were approached, by email, and asked to take part. The same questions were asked of each participant and their answers recorded using a Dictaphone. These answers were transferred to the computer without any editing. It was estimated that answering the questions would take up to one hour of the interviewee's time.

Those who kindly consented to take part were:

- Prof. Lorna Hughes (University of Wales, Chair in Digital Collections)
- Dr Susan J. Davies (Aberystwyth University, Former Director Archives Administration Programme - retired)
- Lyn Lewis Dafis (Head of Digital Development, National Library of Wales), and
- Vince Jones (Learning Development Officer, Aberystwyth University and Visiting Fellow for JISC Digital Media)

The questions asked during interview were:

- What do you think have been the main developments in digitisation practices over the last 10 years?

- How do you think these developments impact on the use of digital resources for scholarly practice?
- How do you see these developments carrying forward into the future?
- What effect do you think these developments will have on small projects?
- Thinking of small collections, what do you feel are the main issues with regard to digitisation projects? How do you think they can be solved?
- Which of these issues do you think should be prioritised?
- How will these issues affect users?
- What would you consider a good method of evaluating the value of digitisation?

The questions were designed to take the interviewees through from a time around the turn of the millennium to date, focussing on the difficulties of small collections. The answers given were analysed by listening to them for commonalities in the first instance, and then any alternative views.

4.8 Production of Written Guides

Having learnt the principles of the digitisation process from the NLW, the system used needed to be adapted before it could be implemented in CPR (Beagrie & Jones, 2008; JISC Digital Media, 2008d). In a similar manner to the process at NLW, time was spent learning about CPR by talking to various members of staff and familiarising myself with the equipment already in the department. I initially approached this by using a series of open questions as an aide memoire during discussions with the Collections Director. My aim in drawing these questions was to prompt a conversation from which I could build my background knowledge of CPR rather than to provide a definitive set of answers to queries. These exchanges also provided the platform for building relationships within CPR.

From these conversations it was decided to produce written guides to transfer digitisation knowledge. These were to be written so that they could be used by a range of individuals with different learning styles. The guides would also be used to direct the level to which data would be collected (Deegan & Tanner, 2002, p. 115). Another consideration in choosing to write the guides was the amount of time that could pass between digitisation projects in CPR.

Producing the written the guides also showed good project management practice as by following the instructions in them, they document why and how decisions were made (Graham, 2010, p. 79).

Copies of the guides are in Appendix B: Generic and Organiser's Guide and Appendix C: CPR Quick Start Guide.

There are two separate guides because one, the CPR Quick Start Guide, is a guide that describes how to operate the equipment in CPR to produce digital objects to the required standards. The other, the Generic and Organiser's Guide, describes the workflow before digitisation, plus it contains some explanation of why the data is collected as described in the CPR Quick Start Guide. The Generic and Organiser's Guide is also written to a more general audience, because the actions described are not prescribed by the equipment available and its operation. Both of these guides are designed so that they can be used without reference to the thesis.

To compare the outputs of the equipment described in the guides, I tested the output of the scanners available in CPR during the research by using selected photographs and comparing the digital objects created on the same computer screen.

I also composed a template copyright letter for CPR to contact all their copyright holders.

Storage options available were investigated and the learning summarised in Appendix G: Storage Tables. Each of the options discussed was summarised by listing the positive and negative influences they could have on CPR; whether the storage type was suitable for long term storage and its usual use, the possible cost, and how much of the storage type could be required.

4.9 Test Written Guides on a Collection and Apply Learning

I tested the written guides by working through them producing digital objects from the Blaenau Ffestiniog collection. These digital objects were stored and had metadata recorded for them. This was a thorough test of the CPR Quick Start Guide as this is the guide aimed at those carrying out the conversion of archival materials and metadata gathering.

For the Generic and Organiser's Guide it was not possible to carry out all the processes because the archival materials had already been selected. Instead these sections were tested by thinking which questions would have been asked to arrive at a decision to digitise. Other processes, such as allocating individual numbers to archival materials, were carried out according to the guides.

4.10 Evaluation of Written Guides

Evaluation means to assess the quality of something. In this situation the reasons to conduct an evaluation of the guides are to verify that:

- They are useable by those other than the writer;
- That the potential users are able to use them as intended, and
- That they convey the meanings they are meant to, to potential users.

It was decided to evaluate the guides using a method similar to the way CPR expects to use the guides. This was a small sample, but as the staff numbers in the department are very small, had an iterative method been used, then there wouldn't have been the numbers left, who had not already learnt about the guides, to carry out an evaluation. Three staff members took part in the evaluation. The method used was:

- A group of five items, three slides and two photographs, group one, were digitised and metadata gathered in advance of the evaluation by myself
- The testers, with myself acting as coach, used the guide to digitise and write metadata on a second group of five items
- The testers then digitised group one while being observed. Feedback on the experience was provided on a form devised for the purpose (APPENDIX F)
- This feedback was then used to amend the guides.

The reasoning behind this mode of testing was that if a person unconnected with the writing of the guides can complete a cycle, without intervention, it will show that they are well written, and can be followed without help. If this happens, it will also indicate that the initial 'jargon' barrier identified by CPR has been overcome.

This type of evaluation utilised student-centred learning because it recognised that giving guidance on the process, via the instructor in the first instance, and then the written guides, is the knowledge that is required rather than a series of direct commands (Canadian Ski Instructors Alliance, 2000, p. 4.2).

The difficulty with this method of evaluation is that it was only the opinion of three testers which made the results obtained subjective and singular in view point.

The testers were chosen by CPR. Who had been asked to carry out the test was not known by me, until they came into the room for the evaluation. The results from their digitisation were compared to the digital objects and metadata written by myself using the same archival materials. If these were the same it would have showed that the written guides had passed on the practices I was trying to convey.

The questions for the evaluation of both guides are in Appendix F: Evaluation Questions. The questions written for the evaluation were given to the testers before they carried out the trial so that they were aware of the questions before they began. The contents of the questions were compiled from the objectives to be achieved by writing the guides.

4.11 System Accepted by CPR

After the evaluation of the guides had been completed the Director of CPR evaluated, and provided feedback on the research outcomes. This was done by recording on a Dictaphone his thoughts on four questions. These questions were:

- What did you expect the research to achieve before it started?
- Have these expectations been met?
- If not met, where do you think it went wrong?
- Where do you think it was successful?

4.12 Chapter Summary

The structure of the research design showed that the acquisition of practical skills to digitise archival materials were to be learnt at the NLW. This would be complemented by learning about the theoretical aspects of the process by attending University lectures on the aspects of archival theory that I had previously not encountered. In addition to the University tutoring there was the chance to attend a range of external events to add to the total information. The final piece of information was to be gathered by talking to selected information professionals.

It was decided that the transfer of knowledge was to be carried out by writing guides to digitisation practices. The written guides were to be formatted so that they allowed for different learning styles. This also followed good project management and metadata gathering principles by writing down decisions being made (Haynes, 2004, p. 11).

The guides were tested by my digitising a collection held at CPR. The written guides were then evaluated by volunteers chosen by CPR.

Chapter 5: Results

5.1 General Results

The necessary practical digitisation skills were acquired at the NLW and the remaining theory from literature, lectures and conferences. Why the detailed processes were important to digitisation was assessed and the results in each section from both the NLW and CPR were written. The outcomes of this comparison were summarised for each section. The evaluation results were assessed and the results incorporated into the written guides. The interviews with the professionals were compared for similarities, and then for significant comments. These were summarised for each question asked.

Table 13 shows a summary of some of the vital statistics of the NLW and CPR at the beginning of the research.

	National Library of Wales	**Centre for Performance Research**
Length archive shelving	15km (Jones 2008)	Approximately 0.07km (i.e. 70m)
Type of holdings in archive	Photographs, maps, wills and probate records, significant books in Welsh and relating to Welsh culture, examples of all the creative arts including audio visual materials	Photographs, posters, programmes, documentary materials and audio visual materials relating to performing arts
Other holdings	5,000,000 + printed items and copyright library	Produces journal *Performance Research* and small library of printed texts
Staff	41 in digital developments unit 2007 (Jones 2008)	Total = 4
Equipment at beginning of project	7 Scanners and Macs (in room where I was based) + additional computers for creating OCR files etc.	2 scanners + 1 PC

Table 13: Summary of Situation in NLW and CPR at Start of Research

5.1.1 Outline of Workflow at NLW

The workflow followed for the HW Lloyd collection, at the NLW is outlined below and illustrated in Figure 10.

- Obtain copyright clearance

- Select archival materials to be digitised
- Hold a benchmark meeting.
- Start the workflow on the database
- Conservation processes
- Digitise archival materials
- Digital objects quality check
- Metadata writing
- Complete the metadata and attach the digital object to the record
- Publish the catalogue record with the digital object attached
- If required make an exhibition to introduce the collection using the Web design software.

This workflow was followed so that each part of the workflow was learnt completely and separately. This was so that I was clear about where each part of the workflow began and ended. For the Dwynwen Belsey collection I followed the accepted workflow where, for me, the process began by writing the metadata, followed by the digitisation of the archival materials rather than the other way round as I did for the HW Lloyd collection. The metadata collection was completed and the digital object attached before publishing.

For the HW Lloyd collection, each batch, or group, of archival material followed the workflow, and one batch followed on from another. For a large Library this allows them to keep the work flowing, and to use their equipment and staff time to its fullest. In turn this means that the NLW receive value from their investment in staff and equipment.

The investment in staff is re-paid through the expansion of the digital collections and the implicit understanding of the value of each item of archival material to the NLW. This is the most valuable outcome of this investment and is unquantifiable. Because of this culture, each member of staff did their best for the archival materials being processed by their area.

The systems map of the workflow, Figure 11, shows two clear points of contact with the Library for NLW users, the catalogue and web-based displays. The relevant polices loop connects with all of the other loops, signifying that it had an influence on all of the areas of the workflow.

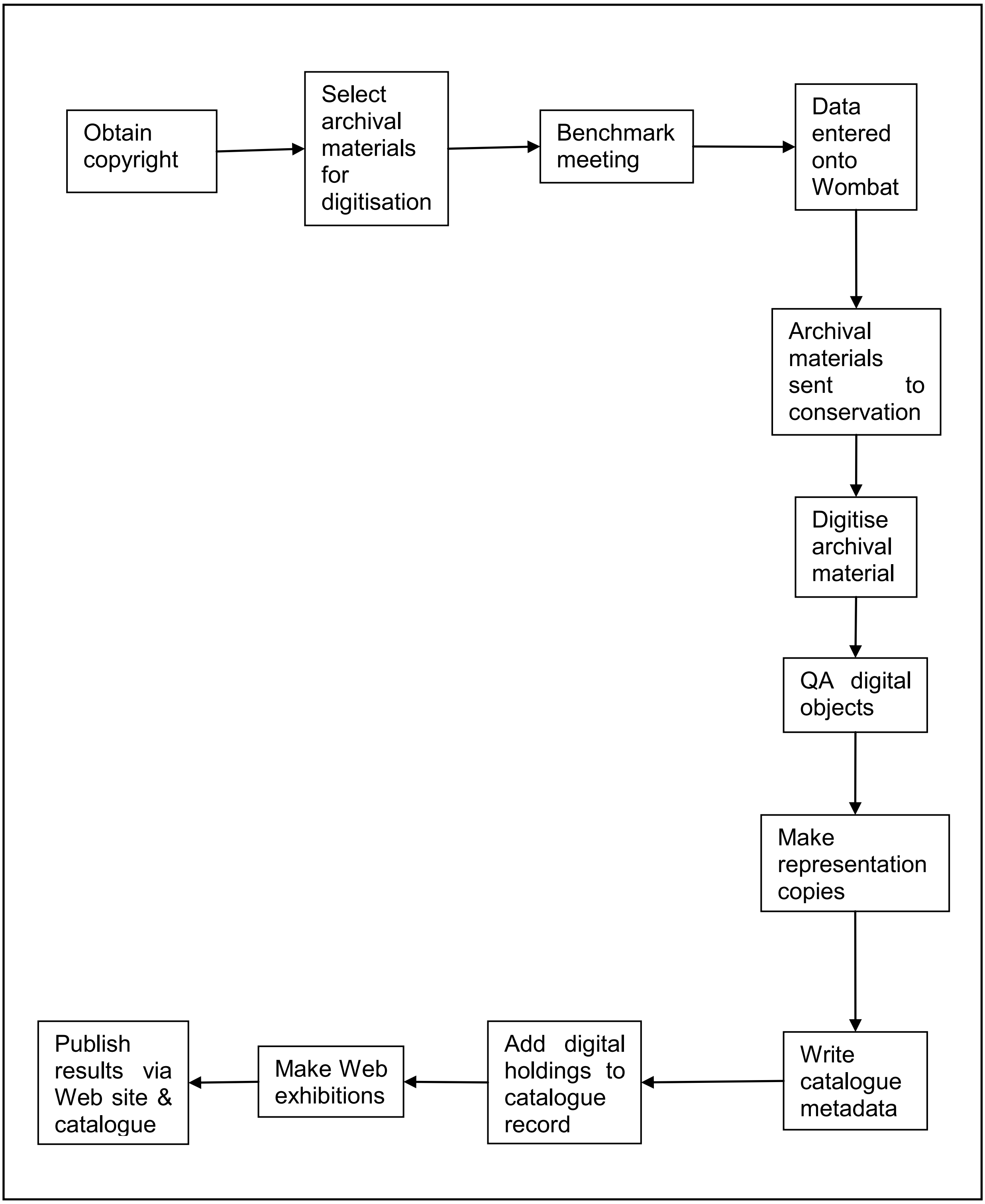


Figure 10: NLW Digitisation Workflow for the First Collection Digitised

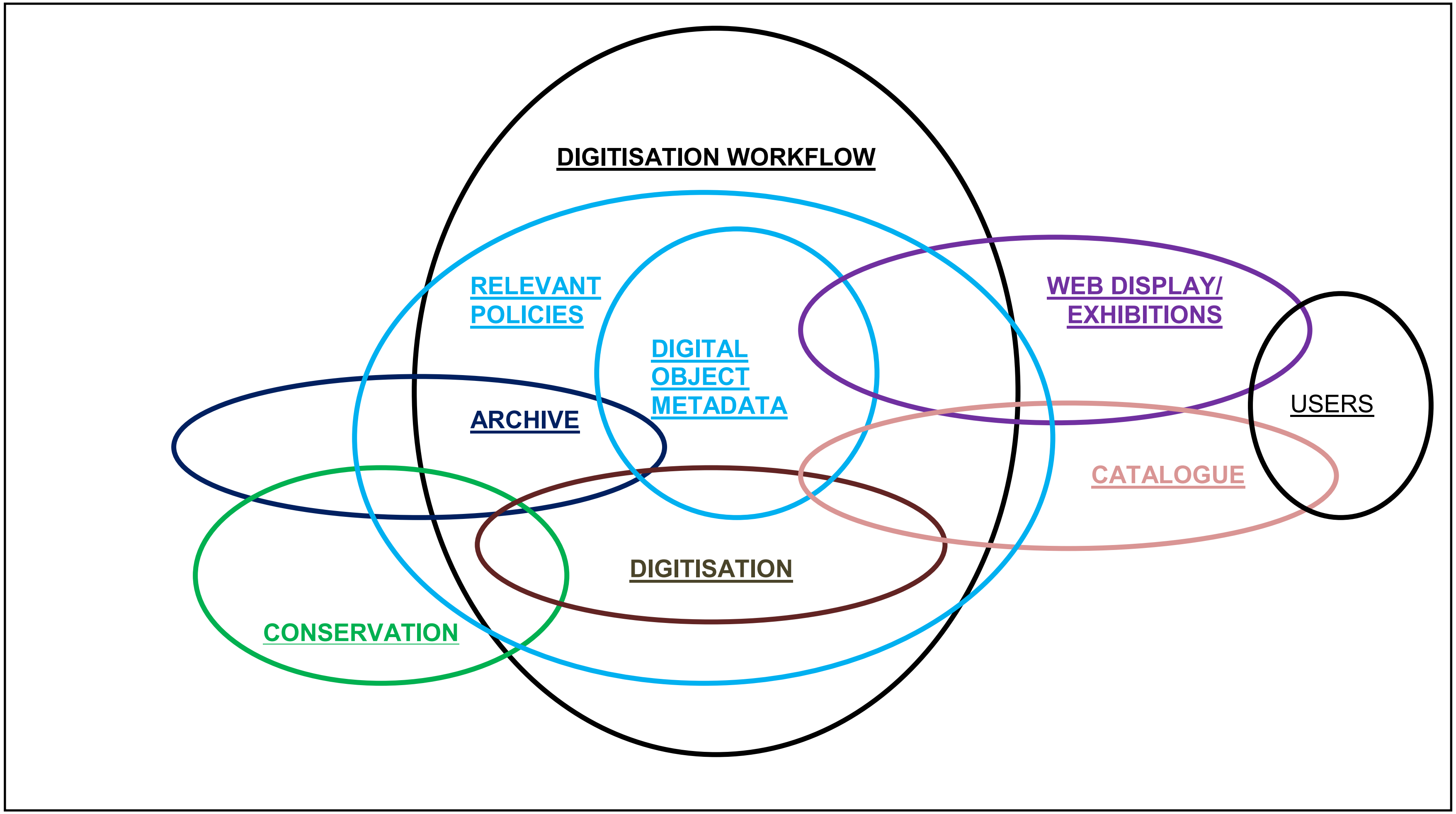


Figure 11: Systems Map of NLW Digitisation Workflow

5.1.2 Outline of Workflow at CPR

As shown in Figure 3, digitisation is not a primary function of CPR. However, as Bulow and Ahmon (2011, p. 17) indicate digitisation projects impact upon an institution in a variety of ways. For CPR the cost of acquiring the required knowledge by staff is relatively large when compared to the principal functions of the centre within TFTS. Therefore identifying areas where there may be comparable, or transferable, skills that could be developed was important. The isolation of these skills was prepared from observations made during the period spent at the NLW and were recorded in tables at the start of the Generic and Organiser's Guide.

Considering both the place of digitisation in CPR, and the use by staff of their transferable skills, it became apparent that it was important to transfer the knowledge they required in a form that could be readily absorbed. One way of doing this was to separate the management of the digital objects from the technical requirements of a system. This was done by setting up a technically simple system, and writing two guides to digitisation for CPR. The first, the Generic and Organiser's Guide, describes processes undertaken prior to digitising archival materials. The second, the CPR Quick Start Guide, describes how to operate the equipment in CPR to create the digital objects and record the metadata. It was important that both of these guides were accessibly written as CPR staff needed to be able to assimilate, and use the information quickly. Together these guides describe the steps for CPR to successfully digitise archival materials and collect appropriate metadata.

Having learnt the principles of the process from the NLW, the methods used needed to be adapted before they could be implemented in CPR (Beagrie & Jones, 2008; JISC Digital Media, 2008d). Adapting the outline workflow at the NLW to the starting conditions at CPR gave this list of objectives:

- Understand the eventual aim for digital objects
- Obtain copyright permission to digitise archival materials
- Formulate and obtain storage options
- Scan archival materials
- Digital objects quality check

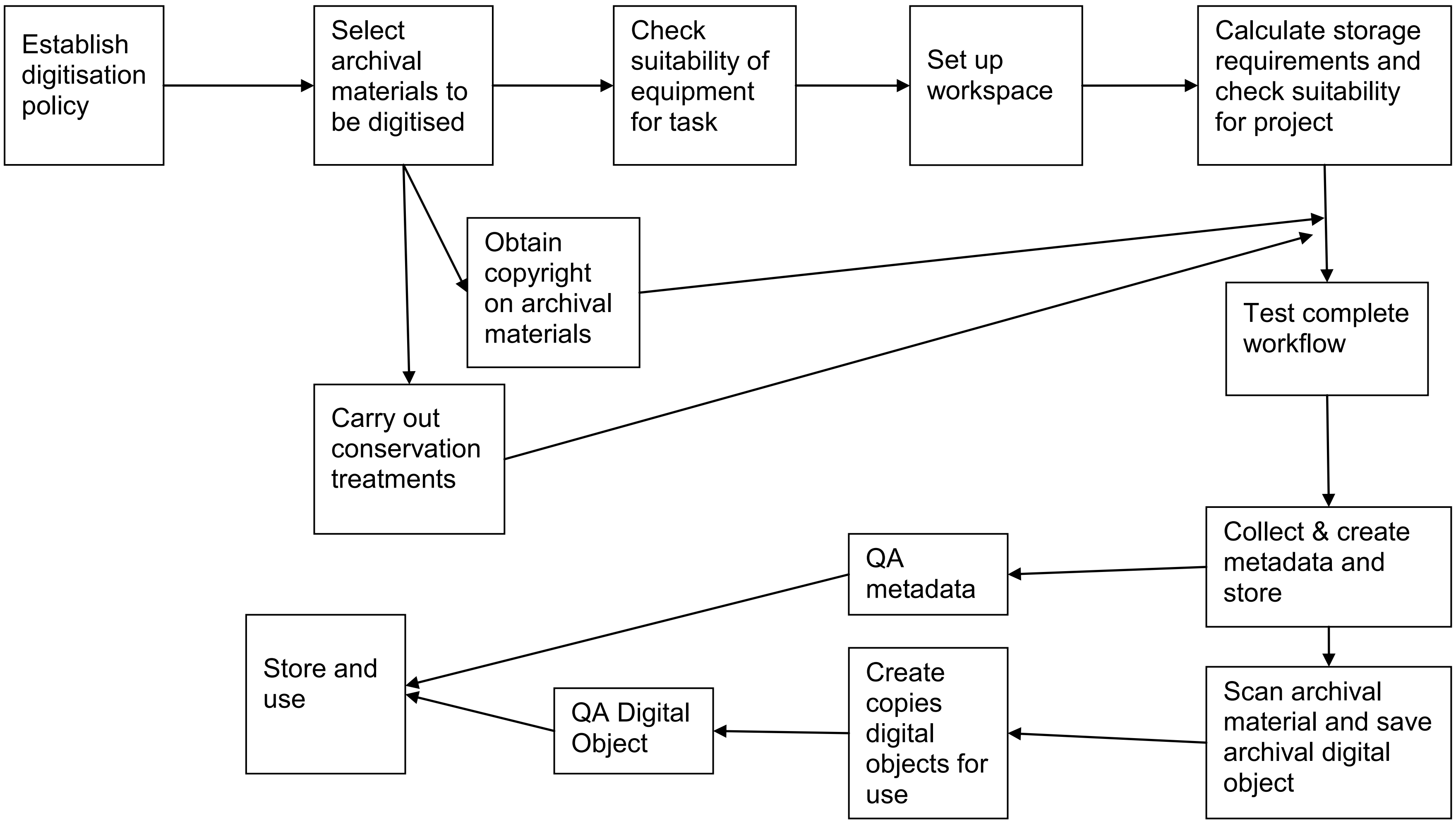


Figure 12: CPR Digitisation Workflow

- Collect and write metadata
- Upload digital objects for display to the institutional repository, CADAIR.

This list is expanded and illustrated in Figure 12. As the outline shows the system in CPR began at an earlier stage than at the NLW. This is because archival materials without copyright permissions are not generally considered for digitisation at the NLW and therefore do not enter into the digitisation workflow (Padfield, 2010, p. 15 & 186). Figure 12 shows the complete suggested process outline for CPR. It was then used to indicate the tasks to be completed for writing the guides. After being tested, the guides were evaluated by CPR staff.

As can be seen from Figure 13, the technical system utilised in CPR for the test digitisation used a PC linked to two scanners.

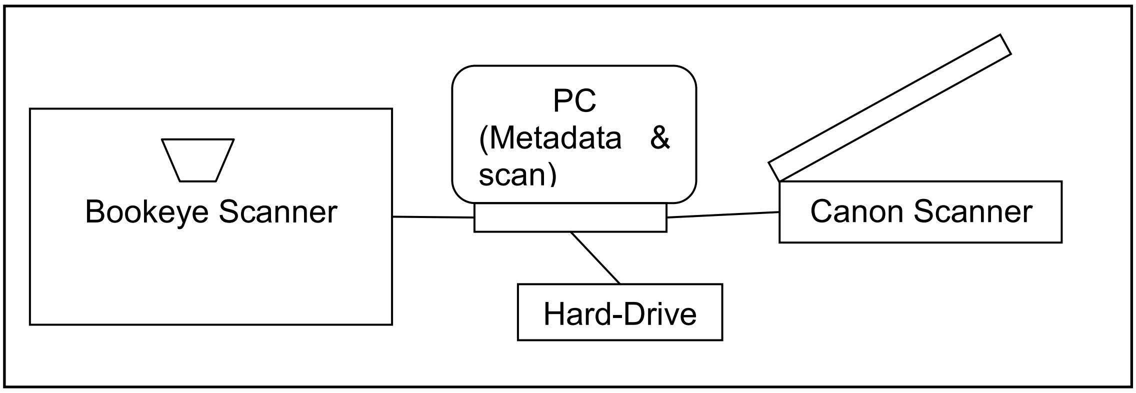


Figure 13: Equipment Layout for Test Digitisation

The two scanners were the Bookeye and the Canon domestic scanner. Storage for the archival digital objects and all the metadata was on the attached hard-drive. The decision to purchase new equipment was taken in consultation with the Director of CPR. Having decided what was required, and what we wished to acquire, it was anticipated that the system would change to that in Figure 14.

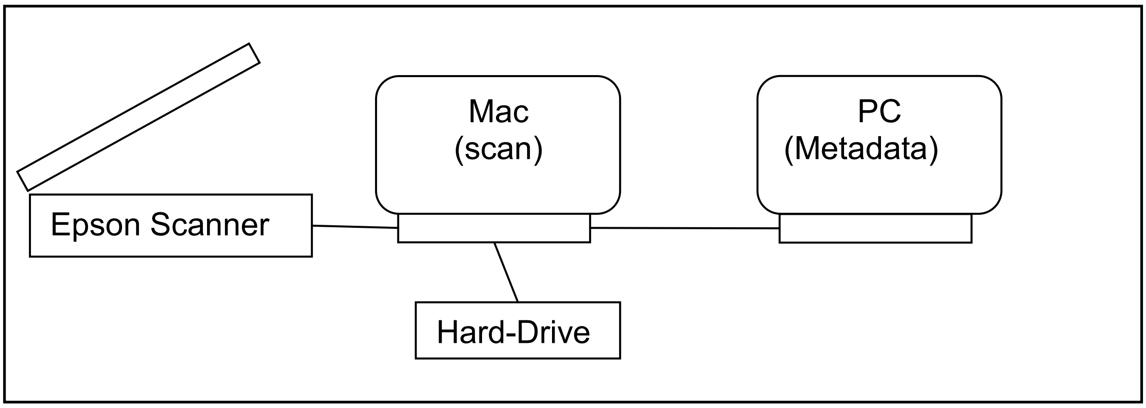


Figure 14: Expected Layout with New Equipment

During the installation process it became apparent that system in Figure 15Figure 15 would be the best compromise available.

5.1.2.1 Evaluation of the Written Guides

The evaluation of the guides progressed smoothly with willing volunteers using the process outlined in section 4.10. Despite being the views of individuals the volunteers were a realistic representation of the expected users and therefore of the system being proposed.

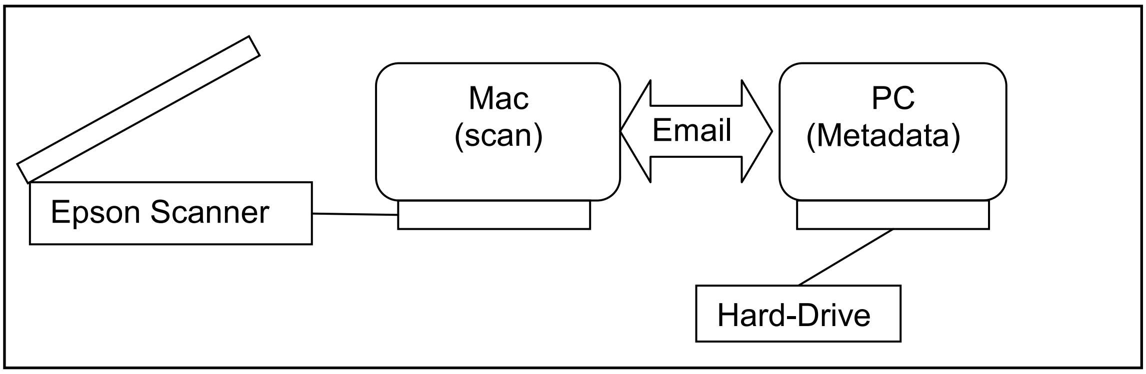


Figure 15: Compromise Equipment Layout

In this regard the use of a small number of testers is a realistic depiction of the expected circumstances of use for the written guides. The volunteers completed a questionnaire on their experience when using the written guides after they had used them.

5.1.3 Comparison of NLW and CPR

The diagrams of the NLW workflow showed it is a linear process with digitisation interacting with many of the other departments. Users make contact with the NLW collections via two points. The workflow drawn for CPR was based on the NLW workflow and contains many of the same elements, but sometimes in different parts of the workflow. The differences between the workflows were that CPR had boxes for:

- Establish digitisation policy
- Check equipment suitability
- Set up workspace
- Calculate storage requirements and

- Test workflow.

Whereas the NLW had boxes that CPR didn't have for:

- Benchmark meeting
- Data entered onto Wombat, and
- Add digital holdings to metadata.

The differences occurred because the workflow in CPR began with processes that are well established in NLW, e.g. the digitisation policy, and because the scale of the operation at NLW is very different to that in CPR.

The evaluation showed that the testers felt they could use the guides successfully to carry out a digitisation project.

The differences between the NLW and CPR workflow were:

- The physical space used for digitisation in the NLW was dedicated to digitising archival materials. In contrast the equipment and space around it in CPR was checked each time it was used because it was also used for other functions
- In the NLW as each project ends, the staff moves on to the next. In CPR a workflow test each time a digitisation project is begun will help the staff to remember the steps to take if there is a large time gap in between
- The workflow is controlled in the NLW by specially designed software. The workflow in CPR is controlled through the exchange of emails between the staff carrying out the digitisation and metadata writing. This was because the bespoke software had a high cost attached to it.

5.2 Pre-Digitisation Actions

Pre-digitisation actions are those which are carried out before digitisation takes place. Sometimes these can be hidden as they can be included as part of a wider strategy for the care of items in an archive or repository.

5.2.1 Policy

A digitisation policy is needed to guide how archival materials are selected for digitisation. Formalising this process means that selection is applied consistently across the holdings of a facility and that each person involved understands their part in the process.

5.2.1.1 NLW

The NLW digitisation policy for 2008 - 2011 is freely available on the library web site (National Library of Wales, 2009). In its opening lines it indicates that it sees digitisation as a key activity (2009, p. 2). After placing the digitisation strategy within the context of the whole Library, it goes on to explain its objectives and their implications over the next three years.

5.2.1.2 CPR

As this was the first project of its type there was no formal written digitisation policy. However, the situation was helped by the fact that the archive is Archon registered, as described in section 2.2. The deposited policy states that "*only items which in the judgement of the Collections' Director are of sufficient quality for permanent preservation will be accepted*". This general collections policy indicates there is an on-going commitment to maintaining access to archival materials through appropriate methods; digitisation could be considered one of these. It also states that CPR is committed to providing access to its resources. Digitisation can be considered as a method of providing access (Centre for Performance Research, 2009).

5.2.1.3 Comparison of NLW and CPR

The NLW have had a digitisation policy for more than 10 years. By contrast CPR had no formal statement of digitisation policy. For this research the statement of a commitment to preserving access in the general policy was taken to include a commitment to digitisation for preservation.

5.2.2 Selection of Archival Materials

Which archival materials should be selected for digitisation is often covered by a digitisation policy. Some policies are deliberately written in general terms so that an organisation can adapt to changes in circumstances e.g. when funders agree to meet the cost of the digitisation of specific archival materials.

In large collections, if the priorities are not clear, archival materials could become too fragile to be digitised if they are requested a lot by users. Small discrete collections can be completed to target specific events, e.g. an anniversary, and use this to target publicity at the newly digitised collection. Eventually these could link up and become one large collection. Complete, but small collections can also be linked to each other virtually by researchers.

5.2.2.1 NLW

Archival materials for digitisation are selected according to the Library's criteria as set out in their preservation and digitisation policies. In selecting the archival materials that were suitable for the research, extra criteria, such as ensuring that the collection size was manageable in the time allowed, were considered in addition to these. Other possible considerations were the fact that in 2014 it will be 100 years since the beginning of the 1914 to 1918 war and the digital objects of Frongoch Prisoner of War camp could be of interest to users researching projects associated with the anniversary. Similarly 2015 will see the 150th anniversary of the establishment of Y Wladfa, the Welsh settlement in Patagonia, and the subject of the Dwynwen Belsey collection.

5.2.2.2 CPR

For this study the materials to be digitised were pre-selected, but represented a very significant work for the department. In the future, for CPR, the selection process is addressed in the Generic and Organiser's Guide. It suggests it is approached by creating a table of questions similar to those completed when filling in a risk assessment form. Questions are chosen according to the priorities of the collection and the archive these are often indicated in a digitisation policy. For CPR, all their archival materials are distinct, so using their uniqueness as a measure would not be meaningful to them. This is because each piece of archival material would return the same value. One aim of selecting archival materials for digitisation therefore, is to show this uniqueness through digital objects. This is achieved in the way the questions asked of the collection in the risk assessment are framed (Bulow & Ahmon, 2011, p. 49).

The format of this table is similar to the type of risk assessment often used in a business setting. This is in Section B3.4 of Appendix B: Generic and Organiser's Guide.

5.2.2.3 Comparison of NLW and CPR

In the NLW, selection for digitisation is governed by the digitisation and preservation policies.

In CPR the Blaenau Ffestiniog collection was chosen because it is the first CLT production that was recorded in photographs.

The value of an archival resource is based upon its significance as a record of what it contains (Pearce-Moses, 2005b). To the CPR archive the record of each CLT performance has the same high value. This is because no two performances were the same and there are few records to show what they did.

In future the selection of archival materials in CPR can be carried out using the assessment framework in the Generic and Organiser's Guide. Unlike NLW all CPR's collections would return the same value if uniqueness was used as a criteria for selection. Therefore basing the choice of archival materials for digitisation upon the value assigned by this quality to CPR would be inappropriate. Instead using criteria such as the time available to a project, the condition of the archival materials, and whether the materials could be used during tutorials and lectures in TFTS, would be more appropriate for them.

5.2.3 Obtain Copyright Clearance

Digitisation is copying. Therefore to remain on the right side of the law it is necessary to obtain permission from the copyright holder to scan and hold a copy of archival materials. For some collections this is a larger problem than others as tracing the owner of the copyright on archival materials can be problematic (Padfield, 2010).

5.2.3.1 NLW

In the NLW the archival materials had been selected and copyright permissions obtained prior to the commencement of the research. The NLW strategy explains that generally, archival materials that are out of copyright are prioritised for digitisation or these permissions are sought after appropriate research (National Library of Wales, 2009, p. 15).

5.2.3.2 CPR

At CPR it was decided to write to all holders of copyright on photographs, even if they did not have photographs in the BF1981 collection, as they were relatively few. This was achieved using a template letter a copy of which is in Appendix D2: Copyright Example Letter. The decision to write to all copyright holders was taken so that future projects to digitise photographs could proceed without delay from this phase. A draft of the letter template was written, and its phrasing and contents discussed with the University Data Protection and Copyright Manager. It asked for permission to digitise all the photographs they had taken for CPR and detailed how the digital copies were expected to be used. This was because the number of photographers was few, because each took all the photographs over several years. As there was a direct line of communication from CPR to the copyright holder for the Blaenau Ffestiniog collection, obtaining permission was achieved quickly. CPR has now obtained permission to digitise all their copies of photographs taken by this photographer.

5.2.3.3 Comparison of NLW and CPR

In this research obtaining permission from the copyright holders to digitise archival materials did not present any problems. This is because the NLW had secured the rights before my arrival, and because CPR knew who owned the copyright on their archival materials. In NLW, if not already out of copyright, permissions are sought for individual collections as they are put forward for consideration for digitisation. The decision by CPR to obtain copyright for all the archival materials deposited by one contributor will prove useful in the future because they can create digital objects at their own pace, once these holdings have been given permission. At present they have permission to digitise most of their photographic holdings. This is because the copyright to most of CPR's photographic holdings are held by just two photographers. Had CPR contacted them each time they wanted to digitise a group of photographs they could soon have become irritated and ceased to co-operate.

5.2.4 Storage Considerations

The conditions digital objects are kept in after creation affect their longevity and, similar to archival materials, should be kept in controlled conditions if their exact needs cannot be met (Sadashige, 2000, p. 8 & 80). The condition of the storage media should be reviewed, by using appropriate metadata, within the manufacturers recommended schedule (Beagrie & Jones, 2008, p. 154).

5.2.4.1 NLW

At the NLW digital objects are stored on allocated servers when they are created. The archival digital object is stored separately to the representational digital objects. The representational digital objects are the viewing and thumbnail copies and are attached to the catalogue record for users to access.

5.2.4.2 CPR

As described in section 4.8, I was asked to prepare the information in Appendix G: CPR Storage Options in July 2011. The options put forward for consideration were assessed against the criteria listed in section 4.8. In the tables presented to CPR DSpace was used as an example of repository software in the tables, because this is the software used by CADAIR. Using this example meant that the external appearance of the software was familiar to CPR. It also means that there is experience in managing digital objects in a repository that CPR can access if they wish to.

The information on storage arrangements were presented to CPR as a set of tables which are summarised in Table 14.

Storage Type	Why it Should be Considered
External hard drive + University server space	If a combination of external hard-drives and, if it can be afforded, purchased University server space were used it would give CPR a technically simple, but flexible system that can be refined in the future. It would also allow them to keep more than one copy of each digital object. If one hard drive is kept away from University premises, it will mean that all the copies are not in one place and therefore not subject to the same storage conditions as those at the University.
Repository software	Repository software varies in complexity, but all require knowledge of digital preservation workflows to successfully manage the digital objects they contain. If CPR does decide to install their own repository software then they should consider systems that are relatively simple to install and run, such as ICA Atom. Some repository software is free and open source, some is not. All repository systems have the same basic technical requirements, but the scale of the needs vary.

Table 14: Table Showing a Summary of Research into Storage Options for CPR

The possible future demands of the management of CPR's collections were given considerable thought while researching possible storage options and their implications. The system I recommended for CPR was illustrated to them using Figure 16 and showed how, by using a hard-drive to store the digital objects, the current digitisation project could be stored, accessed and preserved.

The main considerations for long-term storage were (Bulow & Ahmon, 2011, p. 40):

- Cost, including the cost of metadata storage (JISC Digital Media, 2007b)
- Whether CPR could maintain it independently.

The decision was made to use an external hard-drive attached to the computers available in the centre for this digitisation project. This storage method had the advantages to CPR of:

- Being very similar to the arrangement staff use for storing the files they use everyday
- Being portable, if required
- Allowing digital objects to be collected until a critical mass is reached without on-going cost implications
- The digital objects can become safety copies in the future when repository software is installed

- Allowing the data arrangement to reflect the structure of the archive. This means that digital objects are clearly associated with the archival materials both through arrangement and name.

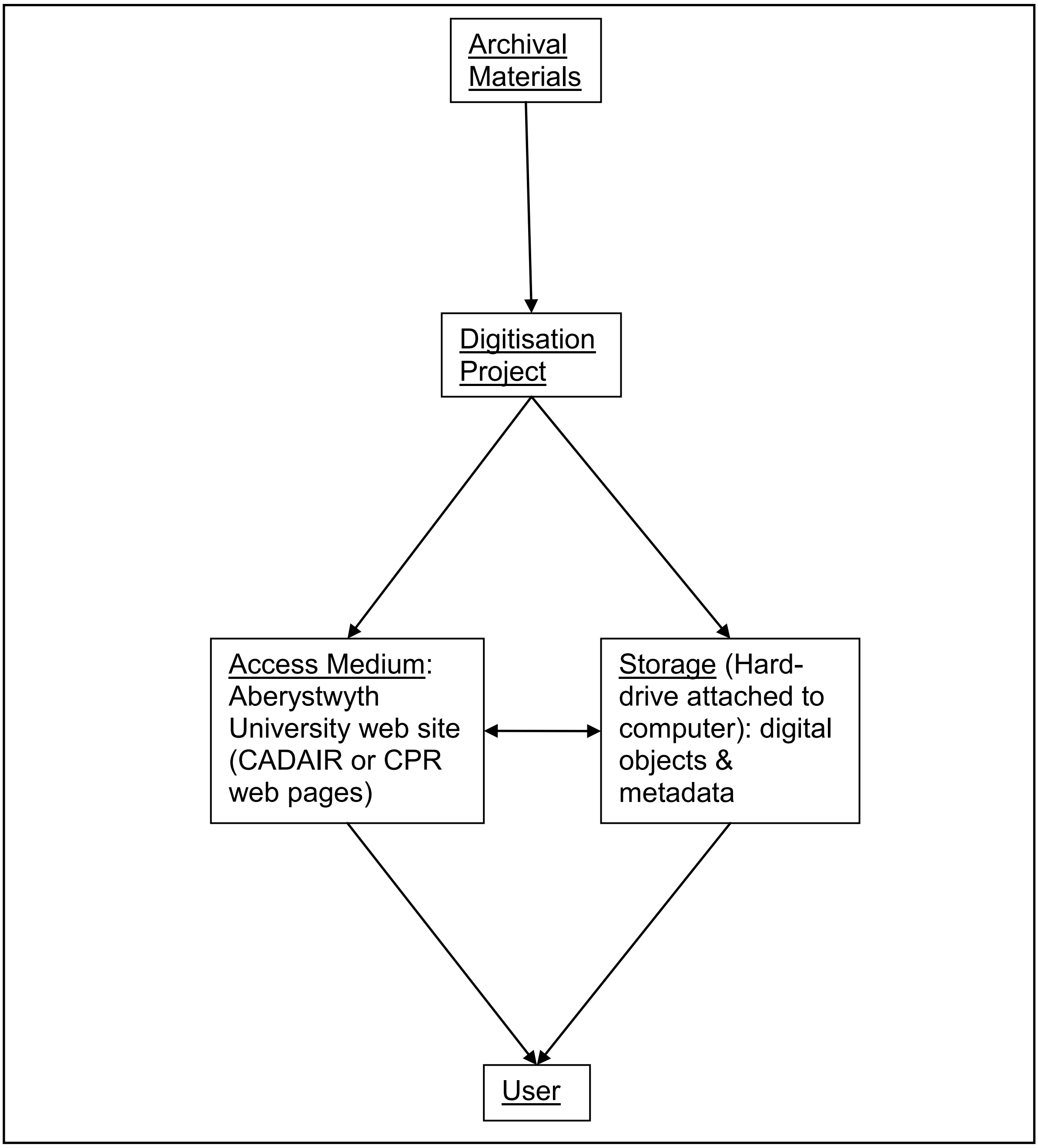


Figure 16: How Digitisation Research Could Feed Into Future Plans of CPR

5.2.4.3 Comparison of NLW and CPR

At the NLW digital objects are stored on an allocated server. In CPR, a similar situation could have been achieved by buying space on the University servers. However, for CPR this was inappropriate because the size of the allocated space would have meant buying more each time similar sized digitisation projects had been completed. This represented an on-going expenditure for the department that would be hard to justify. It was therefore decided, in consultation with CPR, to invest in an external hard-drive that is attached to the PC in the centre. In reaching this decision, consideration was given to the intended future of CPR.

5.2.5 Physical Workspace

The physical locations of, and the equipment used by, those carrying out the roles of digitisation are discussed in general terms in the literature. This is clearly because there are so many variables that if they attempt to describe the features needed in any detail they risk missing some out. The locations and working conditions of those carrying out the processes of digitisation are important as uncomfortable conditions makes concentrating on fine adjustments difficult. It follows from this that inappropriate conditions can lead to mistakes being made during the creation of digital objects, which leads to extra cost when they have to be re-made (Bulow & Ahmon, 2011, pp. 160-163; Health and Safety Executive, 2012; JISC Digital Media, 2005b).

5.2.5.1 NLW

In NLW the digitisation workspace is painted in a neutral colour. The lighting is placed along the centre of the room which directs as little light as possible onto monitors on the computers and there are blinds on every window. Each computer has a space next to it that allows archival materials to be arranged prior to being placed on the scanner (JISC Digital Media, 2005b).

5.2.5.2 CPR

In CPR the digitisation took place in a room away from the usual hustle and bustle of the centre. The room has blinds at each window and overhead lighting. The walls are painted in a neutral colour. A large central table is next to the computers and can be used to arrange archival materials ready for digitising (Bulow & Ahmon, 2011, p. 160; JISC Digital Media, 2005; JISC Digital Media, 2011).

5.2.5.3 Comparison of NLW and CPR

Having access to amenable surroundings in which to create the digital objects results in higher quality output as staff are able to concentrate better, for longer. Both NLW and CPR have rooms with neutrally coloured walls and lighting which can be arranged to minimise the reflections on monitor screens. Both also have space for handling and arranging archival materials adjacent to the scanning station (JISC Digital Media, 2005b).

5.3 Workflow

An agreed workflow should be worked out and documented before a digitisation project begins. This workflow can be used to:

- Track where archival materials are within a project
- To improve a system
- Keep a project running to time, or
- To track where a project has made mistakes.

This workflow becomes part of the project documentation.

5.3.1 NLW

The workflow at NLW is regulated by the specially designed database, Wombat. Wombat was initially designed by CASIS (Centre for Advanced Software and Intelligent Systems) at Aberystwyth University for NLW with further development undertaken by the NLW. Wombat is completed with the decisions made at the project benchmark meeting.

Prior to the benchmark meeting, archival materials are removed from their permanent storage and into the intermediate storage area. This storage area acts as a halfway house between the controlled conditions of the permanent store and the warmth of the digitisation studio.

This workflow, Figure 10, shows the agreed stages from the benchmark meeting and lists the processes a project undergoes. It also notes any distinguishing marks or aspects of the archival material that could influence digitisation. An example would be a postcard with writing on the back: this postcard would be scanned on both sides and two separate digital objects made. These would be separate digital objects but would still be linked together, and appear on the same catalogue record as the front and reverse of the postcard. The workflow notes that there is a need to scan the front and back of an item.

When entering items into the workflow it is important to leave a gap between groups of file numbers as they are allocated. The reason for this became apparent during digitisation. These gaps allow for the insertion of extra scans if, during the digitisation process, it is discovered that particular features would benefit from being scanned separately. Additionally by running a check to see where there are gaps in the numbering highlights where scans could have been missed or, where the beginning and end of sections are.

Because the workflow is separated into logical sections e.g. a book chapter or box of photographs, this makes the re-assembly of a complete book or photo album as a series of digital objects easier (Haynes, 2004, p. 118). A perhaps unintended consequence of this separation is that when producing additional digital objects, for example for web pages, it is easier to see where you have got to in this process if breaks happen at section ends.

5.3.1.1 Benchmark Meeting.

The purpose of the benchmark meeting is to discuss any project specific processes, such as preservation treatments or storage and handling of archival materials, as well as to agree a timetable of actions to the completion of the project. The agreed actions are recorded and circulated so that all involved are aware of the commitment and processes required for completion. Recording and distributing the meeting outcomes complies with the guidelines set out in the Prince2 methodology which expand the acronym SMARTER (Set, Measured, Agreed, Recorded, Timed, Evaluated and Repeatable) to achieve successful project implementation (Graham, 2010, pp. 10 - 11). Prince2 is a structured method that describes how to run a project (Graham, 2010, p. 1).

The benchmark meeting is attended by representatives of the departments with a role in a specific project. For the digitisation projects I carried out there were attendees from conservation, digitisation, workflow writers, IT and photography experts. The meetings only lasted around 15 minutes, but everyone had an input.

5.3.2 CPR Workflow

At the beginning of the research period CPR had no recorded experience of digitisation and therefore no documentation that had been designed with this in mind. The general collections policy that had been deposited with TNA provided a starting point for working out how to design the written guides to digitisation.

The workflow seen at the NLW was used as the starting point for determining a workflow for CPR.

The initial outline of the workflow in CPR was:

- allocate individual identifying numbers to archival materials
- create metadata using these individual numbers to gain control of the archival materials
- use this metadata to manage the digitisation process by using it to:
    - indicate what was to be scanned
    - indicate how the archival materials were arranged
    - identify the archival materials, and
    - indicate when digitisation had taken place
- create required digital objects
- store digital objects in allocated places
- complete the metadata

Many of these processes are in the same order as at the NLW, but there were also differences. Some of these differences were due to the fact that this was a first project for CPR. Some were due to the difference in scale between the NLW and CPR. A typical example is that the NLW has a suite of scanners and operators, who occupy a large room, dedicated to the digitisation of its holdings whereas CPR now has two scanners linked to two computers and no permanent staff.

The workflow outline in Figure 12 remained unchanged throughout the iterations to accommodate the changes in equipment. What did change between the test digitisation and the evaluation was the physical location of the tasks.

The test digitisation was the digitisation of the Blaenau Ffestiniog collection (BF1981). During the test, the digital objects were created on a PC and adjusted using imaging software on a Mac in a separate office. This was not ideal, but was the best arrangement available at this point in the research.

Having made the decision that the equipment in situ was not suitable for the task, it was expected that the replacements would allow the complete workflow to be accommodated on one computer. However this became two. One, an iMac with the scanner and image processing software available on it, and the second, a PC that is used for metadata gathering. This second computer also has the external storage for the digital objects created attached. Technical metadata was collected by using DROID to analyse the contents of the folders where the digital objects were stored. The decision to use a second computer was made when it was discovered that DROID command lines needed modification for it to run on an iMac. It was decided to transfer this software and the external storage to the PC. This meant that DROID reports would be compiled on the same PC at the same time as metadata is written. This arrangement has several advantages over using one computer:

- two people can work on one, or more, projects in parallel rather than sequentially, and
- the skills required to write the metadata are similar to those to operate DROID software
- using a PC to run DROID means that if it becomes necessary to re-install it (perhaps to update to a new version), this can be accomplished by CPR staff without having to adjust command lines.

The disadvantage is that:

- the transfer of the digital objects between the computers could lead to corruption of the digital object, although this would hopefully be spotted by regularly running DROID reports.

Using DROID to generate the required metadata is important to the research. Transferring the data and analysing it as a group, avoids the increased risk of making a mistake that the alternative of individually transferring each item of data into the spreadsheet carries.

In the NLW control of the digitisation process is achieved by marrying the item accession number to a unique file number for the digital object in a database. This means that each item of archival material can be tracked through the process. This was not an option in CPR as this kind of software is not available without significant investment in both time and money.

It was initially envisaged that the workflow in CPR could be controlled using a simple slip of paper that was signed on completion of a stage and passed on to the next along with the archival materials. However the physical layout of the equipment meant this was refined to sending an email containing the relevant data, and the archival materials kept in one place in one room. This meant that workflow and item control was achieved by completing and forwarding a copy of the metadata spreadsheet to the scanner operators. They used it to see where batches of archival materials began and ended and what the individual identifying numbers were. When items had been digitised, the scanner operators added the date, the scanner and software used to the spreadsheet, and returned it to the metadata writer. The NLW database is used in a similar manner as it records when a batch has completed the digitisation process for the metadata writers to see (Bulow & Ahmon, 2011, p. 9 & 165; JISC Digital Media, 2005; JISC Digital Media, 2008b)..

Most of the collections in CPR are un-catalogued. The quickest way to achieve control of the complete holdings would be to compile a catalogue at collection level because the number of records to be completed is less than if the collection were catalogued at the level of archival materials. However control at this level can present difficulties for digitisation projects because precise identification of individual archival materials is required. Hence a decision was made to allocate unique identification numbers to individual archival materials.

The test collection, BF1981, came from the CLT section of CPR's holdings and was not catalogued before this research. Individual identifying numbers were written, using a soft pencil, onto the reverse of photographs and, during the test, slide numbers were allocated via the metadata spreadsheet as each box was scanned. It was recognised that for slides, this method could easily lead to confusion so it was amended to placing identifying numbers for slides onto stickers on the outside of the slide storage boxes. This process has been placed immediately after the assessment of archival materials phase of the workflow so that these two processes can be done together. Allocating numbers here means that number and archival material are associated.

The identifying numbers are entered onto the metadata spreadsheet ready for the start of the metadata writing and digitisation part of a project. For those involved in a digitisation project this tells them that the archival materials have:

- been assessed for digitisation, and
- a rudimentary catalogue entry which allows a level of control over the collection to be established.

Identifying numbers for individual archival materials were devised using the three initial letters chosen for the digitisation project plus five digits. Because the slides in the collections have plastic mounts the identifying numbers could not be written on to them. Placing stickers on the outside of the boxes with the accession number written on will identify that a particular slide should be in a particular box. However, it will not stop slides being moved, or falling out, and becoming disassociated with its number unless linked on the metadata spreadsheet.

This numbering system also allows for several possible variations in application in the future when cataloguing different parts of the holdings. For instance, it could be decided to distinguish between the four collection areas in CPR by placing a signifying letter in a nominated position or similarly, to distinguish between formats.

For the scanner operators receiving a copy of the spreadsheet with its completed data is the trigger for the digitisation of a collection to start. Control of the digitisation phase comes from each digital object requiring an entry for the creator, and the equipment and software used, before the spreadsheet is returned to the metadata writers.

After the digitisation process specific technical digital object metadata is collected by utilising DROID (The National Archives, n.d.b). The report function on this tool provides data in the form of a spreadsheet. The results in this spreadsheet are copied into the metadata spreadsheet and the writing process completed.

5.3.3 Comparison of NLW and CPR

The NLW uses bespoke software to control a workflow that interacts with several departments. This ensures that a project is controlled, or can be regulated to work in parallel to others.

Individual archival materials are traced through the workflow by a number allocated by Wombat at the NLW. Similarly the individual identifying number is used to trail the archival material in CPR. The arrangement of archival materials was used to help to break the collection into easy to manage sections for digitising. This also helped with their reconstruction for display.

Investing in workflow software was not an appropriate solution for CPR as the financial investment would not be recouped. Instead the workflow is controlled by the sending and receiving of data by email between the CPR staff. Like the NLW, the system breaks the archival materials into sections to make digitising and the assembly of the digital objects easier.

Allocating individual identifying numbers to archival materials is an important part of the system for CPR. This is because for many of the archival materials this is their primary identification. This number associates the archival materials and digital object.

5.4 Conservation

For their general conservation, archival materials need to be placed in the best environment that can be achieved for their needs. This includes the physical wrappings around archival materials, e.g. photographs should be kept in wrappings that have passed the Photographic Activity Test (PAT) test, as well as those of the general environment (Library of Congress, 2011b).

One motivation for digitising archival materials can be to reduce wear on popularly accessed archival materials. To achieve the best digital object possible the archival materials should be in the best condition possible. Therefore placing conservation before, but in juxtaposition with, digitisation achieves this. Wear and tear on archival materials is reduced by encouraging users to access the digital object before requesting access to the archival materials. This is because for many users access to the digital object is sufficient to satisfy their needs. Thus reducing the number of times users access the archival materials, but not limiting users' views of the archival material (Bulow & Ahmon, 2011, p. 7).

5.4.1 NLW

For both collections, after the archival materials had been assessed for conservation needs, they were passed on to have their details entered onto the workflow database. In the Conservation Unit traditional skills are used to make sure the archival materials are in the best condition possible for digitisation. Conservation is placed here so that any delays due to treatments do not impact on the next part of the workflow (Bulow & Ahmon, 2011, p. 149).

In the case of the HW Lloyd collection the age of the glass negatives meant that they were individually inspected by a conservator before digitisation commenced. Any glass plates considered too delicate to be subjected to the digitisation process were substituted with a negative created by the Library. These negatives were made when the collection was accessioned to the Library and used to make a photograph album.

Handling the archival materials during the digitisation process also carries risks. In addition to the possibility of damage being caused by scanning archival materials, there are oils that occur naturally on the skin which can also cause harm. To counteract this gloves were worn while handling all the archival materials (Library of Congress, 2011b).

5.4.2 CPR

During the research no conservation processes were undertaken, but a general discussion on the types of actions that could be taken to improve collection care did take place.

In future digitisation projects, conservation actions can be slotted into the workflow parallel to the process of gaining copyright permissions. The archival materials were not treated during the research process because the majority of the items were in good condition.

5.4.3 Comparison of NLW and CPR

In the NLW the glass negatives were assessed, and treatments carried out, immediately prior to digitisation. For the modern collection no extra treatments were considered necessary. In CPR no conservation treatments were carried out during the project, but plans have been put in place for steps towards the general conservation of the holdings.

5.5 Digitisation

The outcomes of digitising archival materials are digital objects. The initial digital object is used to make copies, in different file formats and sizes, for use in different situations.

It is generally acknowledged that roles, such as those found in conventional conservation situations, are highly skilled. That they require dedication and perseverance to acquire is not in doubt, but what a skill is requires some clarification. In sports, skill has been defined as the ability to select and adapt from learned techniques an appropriate method to solve a problem (Sheddon, 1993, p. 119). Many people have computers in the home and a proportion of those also have a scanner attached. However being able to use these pieces of equipment does not automatically confer the ability to create and see digital objects at the highest levels, just as the possession of a football does not make everyone, a Premier League player. Scanning archival materials to create digital objects involves using a series of skills that can easily be overlooked. There is a need to be able to see and understand the composition of a picture of whatever is being digitised; a need to understand how the digital object is processed by the scanner and a need to understand how the software works on the digital object. These skills affect the decisions made during the creation and transformation into copies of digital objects (Langford & Bilissi, 2011, p. 198; Western States Digital Standards Group: Digital Imaging Working Group, 2003, p. 24).

5.5.1 NLW

In the NLW digitisation workspace there are a series of scanners of differing types which are used to accommodate different archival materials. There is also the option of using a camera if all the scanners prove unsuitable.

Each type of scanner was designed to do a specific job. For the research a flat bed Kodak scanner with a solid lid was used for scanning the glass plates from the HW Lloyd collection and the slides from the Dwynwen Belsey collection. For the photographs and postcards, a different flat bed scanner with a tracking overhead Charge Coupled Detector or Device (CCD) unit was used.

The lidded flat bed scanner had templates that meant that several plates or slides could be placed onto the bed of the scanner at once and the digital object made as a separate file. Each space on the template was shown as a preview. When this had been checked it was re-scanned by the scanner and saved directly to the allocated folder.

The CCD scanner worked in a different way. Each piece of archival material had to be set up individually against a neutral background. In this case the background was a uniform grey. To show that the colours achieved in the scan were consistent, a comparison strip was included in each scan. The strip consisted of four shades ranging from white to black (Stouffer Industries Inc., 2006). These shades were used to measure the values of the colours being produced by the scanner as they appear on the monitor, before the scan is accepted. The head of the scanner passed over the area of the scan, then showed it on the monitor. When it was accepted the scan passed straight to the allocated folder space without any further processing.

Training to use the scanners to the expected standard of output began with direct supervision of me adjusting the digital objects achieved from a set of selected photographs. Mistakes were pointed out and suggestions on how to make suitable corrections were made then and there. This was so that the difference the adjustments made could be seen immediately. This method of training utilised the fact that people working in this field have strong visual skills and that what are small differences to most, appear large to them. It is also an example of student-centred learning as it recognises that making mistakes is part of the learning process. It also encourages the use of an individual’s visual skills to assess the digital objects being produced (Canadian Ski Instructors Alliance, 2000, p. 4.4).

The steps to create each digital object were:

- Create a preview scan
- Check the preview is suitable
- Create the scan
- Save to the designated folder as .TIFF files
- Complete the workflow database
- Repeat for each piece of archival material
- Quality assessment of the batch of digital objects
- Create copies for viewing as .jpg files
- Save to the allocated folder

The steps in the process for the Dwynwen Belsey collection, the more modern collection of photographs, postcards and slides; were exactly the same as for the glass negatives except for the emphasis on the data collected by the scans. For the glass negatives the emphasis was on adjusting the scanner controls to collect the largest amount of data possible. For the modern collection, adjustments were made to collect data that gives the closest possible match to the original source followed by the amount of data. The reason for this is because of the relative dates and condition of the original source. The glass negatives dated from around 1900 - 1920 and are, purely because of their age, in a more fragile condition. Because of this fragility, it was decided to collect as much data as possible so that a record was made before further deterioration meant it was lost forever. The second collection has dates of around 1955 to 1965 and has the benefit of being made with more contemporary materials and newer processing techniques. The collection of as much data as possible was still a priority for these archival materials but, because much was still apparent, it was known that the majority would be gathered by creating a copy that matched the original.

Quality assurance of the digital objects was carried out by another person as they would not be as familiar as I with the archival materials. I was told if any corrections were needed. These were made by adjusting the digital object until they were satisfactory.

Representational copies of the digital object were created by using Photoshop software to make digital objects for viewing and thumbnail use. In the case of the glass negatives some further adjustment was required to some digital objects to make them presentable; so that the emphasis was on the same parts of the picture as the plates showed originally.

The mixture of formats of the archival materials added to the variety of techniques and skills to be learned. For instance, the glass plates required slower, more considered handling compared to the photographs, slides and postcards.

5.5.2 CPR

Unlike the NLW, digitisation in CPR takes place in an area that generally functions as an archive and occasionally as a classroom. There are blinds to each window and the walls are painted in a light neutral colour. Through experiment it was found that most of the reflections on the monitor screen could be removed if one set of lights was used on its own. Removing as many reflections as possible made viewing the digital objects easier.

At the beginning of the project CPR was already equipped with a flat-bed scanner, a Bookeye, whose operating software was installed on an attached PC. The PC monitor was adjusted to the maximum resolution that it allowed. It was noted that the maximum refresh rate of 65Hz was lower than the rate recommended by the prompts by Microsoft software when it was investigated. The technical details of the monitor were not available, but it was observed that the print on web pages with a lot of text had visible square shapes making up the letters. It was also noted that there was no means, such as calibration software, to verify the colour reproduction of the monitor. While none of these things were covered directly as part of the learning at the NLW the opportunity had been taken to discuss the need to setup equipment correctly.

The description of the Bookeye scanner in its accompanying manual says that it is suitable for producing digital copies of reflective sources such as newspapers, periodicals, catalogues; books and delicate materials; bound and stapled documents and describes itself as being best suited to ‘scan to print’ operations (Imageaccess, nd). From this it was reasonable to assume that it scanner should produce a reasonable copy of a photograph as journals are on often produced on glossy paper.

The manual also indicated that some metadata is transferred when the digital objects are transferred to storage using a particular type of 'order', as it describes it. Which metadata is transferred is not made clear in the written description.

The technical details, summarised in Table 15, of most importance to the research given in the scanner manual were (Imageaccess, nd):

Parameter	**Value**
Colour depth	30 bit internal
Colour depth	24 bit external
Sampling pattern	Not known
Bed size	Max A2 paper size (16.5" x 23.4")
Bed Size in pixels	4800 x 6775
Other resolution values	Optical resolution (calculated by the software) 150dpi, 200dpi, 300dpi & 400dpi
Native resolution (in instructions)	600dpi
Actual resolution	(4800/16.5) x (6775/23.4) = 290ppi (to 3 sig fig)

Table 15: Bookeye Scanner Technical Detail

During the pre-digitisation phase for the research, tests were carried out to determine the suitability of this scanner for the project. This was done by examining digital objects made from pre-selected photographs. The outcomes from these scans of most concern to the project were:

- That the scans produced were very noisy. This means that there was information missing from the scan.
- The resolution actually worked out at approximately 290ppi. This means that in order to achieve a native resolution of 600dpi the software is interpolating the contents of approximately 1 in 2 pixels
- Examination of scans made at settings other than 300dpi on the scanner showed interpolation errors
- The scanner has no apparent preview available on the monitor, just post scan adjustments. There is control via the settings buttons over the digital objects produced, but the adjustments are made blind and a scan had to be re-made to discover the result. This is described as a printer output by the instruction manual that came with the scanner
- Unresolved was whether the image on the monitor had been converted from RGB to CMYK

Further research showed that better quality digital objects could be obtained from a domestic scanner located elsewhere in the centre. This was used instead of the large flat bed scanner, the Bookeye, for most of the test digitisation of BF1981 as this eliminated most of the concerns raised during testing of the large scanner. The only archival materials the large flat-bed scanner was used for were the few photographic prints that were too large for the domestic scanner. It was still necessary to acquire a new scanner for CPR because the domestic scanner was normally attached to a PC, and used, in an office elsewhere in the department.

The one concern that was not completely removed by using the domestic scanner related to the scan preview. The scan preview was of concern because of its size on the monitor and the inability to show and compare adjustments before re-scanning. The scans resulting from the use of this scanner were judged to be of sufficient quality for producing archival digital objects.

The steps followed to create a digital object in CPR were the same as those at the NLW, except that the batch assessment was carried out after viewing and thumbnail copies had been created.

The test digitisation, BF1981, had three general implications for the research:

- Testing the workflow with an open mind can lead to results from unexpected quarters
- Answers to process problems could be simpler, and cheaper than expected
- The general advice indicates that there should be ample time allowed to gain copyright clearance (JISC Digital Media, 2011b). Adopting a similar tactic for testing a workflow allows for greater experimentation and therefore the exploration of possible methods to make the required digital objects
- I decided that the Bookeye scanner was not suitable for the test digitisation and demonstrated this by comparing the results of test scans made by the Bookeye with my own domestic scanner. I then researched a replacement that is appropriate for similar digitisation projects in the future,
- It was decided to use a Canon scanner already in CPR that gave acceptable results for the test digitisation, and

- It was also decided to acquire a new computer to run the new scanner on so that the issues with the resolution of the monitor on the previous one were also resolved.

Replacing the Bookeye scanner was important for the research in a number of ways as can be seen from the list of requirements I drew up when assessing which scanner to buy. The list included items such as the replacement:

- Must be easy to install
- Must be quick to set up for each set of scans
- Must be easy to use. One button operation if possible.
- Should walk the user through each stage, each time it makes a new scan, or set of scans, from the same place each time.
- Must produce scans of higher quality than those produced by the Bookeye
- Must be able to scan both reflective and non-reflective archival materials

After researching the models of scanner available, it was decided to purchase an Epson Perfection Pro V750 scanner. The technical details of interest to the research are in Table 16 (Seiko Epson Corporation, 2012):

Parameter	**Value**	
Color Depth	48-bit or 24-bit	
Sampling Pattern	Alternate 6 line color CCD line sensor	
Bed size	Maximum A4 paper size (8.25 x 11.7 inches)	
Bed size (in pixels)	Lens1: 4800	Lens 2: 6400
Actual Resolution (ppi)	Lens1: (4800/8.25) x (4800/11.7) = 238695. Or ≈ 240,000 ppi (to 3 sig fig)	Lens 2: (6400/8.25) x (6400/11.7) = 424346. Or ≈ 424,000 ppi (to 3 sig fig)

Table 16: Epson Scanner Details

The Epson scanner satisfied all of the requirements and came with a version of the software, Photoshop. This software version was different to that I had used in the NLW. It was written so that some functions, such as re-sizing digital objects, became easier to set up. In this software version, re-sizing was accessed from one clearly labelled menu item and one tick box of options completed. On the software used in the NLW re-sizing digital objects had involved setting up four different actions in total: one for portrait and one for landscape; plus one for each set of viewing and thumbnail digital objects. This meant that re-sizing had to be done by selecting each digital object in turn, so that the automatic numbering remained in sync. The version of the software in CPR worked out whether a digital object was portrait of landscape and adjusted itself to re-size according to the dimensions entered in the option boxes. This meant that re-sizing could be done by pointing the software to a prepared folder of digital objects, and, after setting it up, clicking on 'Go'. The software would then re-size, re-name and place the new digital objects where directed. This still needed to be done separately for viewing and thumbnail digital objects, but the ability to adjust for portrait or landscape made it a quicker task.

5.5.3 Scanners

Particular types of scanner are designed with a particular type of material in mind (JISC Digital Media, 2006e). Sometimes this can be limited to archival materials of particular size e.g. A4 or less. Or it can be a particular type of material e.g. paper. Choosing a scanner that is not designed for digitising archival materials affects the digital objects produced. Therefore, selecting a suitable scanner affects the project outcomes as it affects the digital object produced.

Most scanners work by using software to create a digital copy of archival materials (JISC Digital Media, 2006e). The digital object produced is determined by the settings used to capture each digital object. Some scanners allow the size of the digital object to be set and adjust the settings to achieve this, and some do not. Ideally this should be done using a preview of the digital object (JISC Digital Media, 2006e; JISC Digital Media, 2008a; Western States Digital Standards Group: Digital Imaging Working Group, 2003, pp. 32-34). These digital objects are then transformed using image processing software to produce additional copies that are used for access (Bulow & Ahmon, 2011, pp. 38-39).

5.5.3.1 NLW

The selection of a suitable scanner was discussed at the benchmark meeting. This evaluated a balance of whether the archival materials were of suitable size and condition to accept the rigours of being placed onto a scanner against the required output.

A lidded scanner was chosen for the glass negatives and slides because it was suitable for non-reflective materials. This means that it was designed for archival materials that let light through them. To be able to read a negative it also has to be able to reverse the data so that the digital object appears as if it were a photograph.

The CCD scanner with a tracking, overhead arm was chosen for the reflective items because it was able to cope with different sizes of archival material, when the basic parameters had been set up, without needing constant re-alignment and adjustment.

Each scanner had its own initial calibration process. Taking time to complete this procedure meant that adjustments for individual digital objects were kept to a minimum as well as producing the most faithful copy of the archival materials.

5.5.3.2 CPR

As described previously, CPR already had a large flat-bed scanner, a Bookeye, which uses a fixed overhead camera to acquire a digital object. It is designed primarily to scan and print text-based materials rather than make digital objects for archiving.

The domestic scanner, used for the digitisation of BF1981, was a Canon that was designed to be able to scan reflective and non-reflective materials. It achieved this through the use of templates and by inserting or removing a section in the lid. The Canon was used to complete the test digitisation.

Both of these were replaced for archival materials of A4 size or less by the Epson that was delivered immediately before the evaluation of the written guides. In a similar manner to the Canon it adapted to reflective and non-reflective materials by inserting templates and a lid section. Its greatest advantages were the size and quality of the scan preview, the ease of use of its software plus the quality of the final digital object.

5.5.4 Comparison of Scanners

Having discovered that the digital objects produced by the Bookeye did not meet the required standard it was decided to evaluate the scanners CPR had access to during this research. To compare the results it was decided to make a set of scans using pre-selected photographs. The photographs were chosen for specific features. These were:

- the range of contrast i.e. the level of black to white (the photograph of the white and black cat) (Millerson, 1972, p. 21)
- the colour range (the photograph with tulips in the foreground), and
- the subtly of shades of one colour (the photograph of bluebells in a wood) (Langford & Bilissi, 2011, p. 227), and
- the long whisker on the cat's head was used to assess the clarity of the digital object produced

Scans were made using the setting labelled 600dpi for each scanner and at a setting that would normally be used to make archival digital objects. The resulting large file sizes from scanning at 600dpi would be unacceptable for use but, this was the only setting common to all the scanners tested. The digital objects were saved to the same USB stick and compared on the same monitor without any additional processing. The table below shows the results observed from both sets of digital objects produced for each scanner. The pictures have been reproduced below to illustrate the comments made. Apart from being cropped to a suitable size and compressed by the same amount to produce .jpg files, there has been no additional processing.

Only the Bookeye came equipped with targets for checking black and white and colour reproduction. These were used before the test was carried out and each scanner was turned on until the lamps were judged to have warmed up by the equipment.

The results in Table 17 show that, in this test, the Bookeye scanner produced digital objects that contained the most variation compared to the original photographs. The performance of the domestic scanners was very similar to the photograph except that it, the Canon, put slightly more red into the detail compared to the original photographs. The Epson scanner produced the best copies of the photographs with the sharpest detail and most faithful colour reproduction.

Parameter	Bookeye	Canon CanoScan	Epson Perfection Pro
Clarity (cat)	Whisker not easily seen; very red in colour; edges where cat & grass/soil cross are not distinct, they seem to merge into black	Whisker easily seen; coloured white; is a clear line on the digital object; edges of cat & soil/grass are distinct. Colour changes clearly between soil/grass and white of cat	Whisker easily seen; coloured white; is a clear line on the digital object; edges of cat & soil/grass are distinct. Colour changes into soil/grass and white of cat. Detail present is sharpest of all the equipment tested.
Colour (tulips)	Overall impression is a good match. Red is very slightly too light on the highlights and too dark on the low lights, i.e. picture looks too pink (look at the stones) Green is too dark on the low lights	Overall impression is a good match. Red is not quite dark enough at lower end Green is even across range and a very good match Blue is even across range but slightly too much as stones appear slightly purple and tubs too dark	Overall impression is a very good match. Red is a good match Green is even across range and a very good match Blue is even across range and a very good match as stones appear same as in photograph
Colour (bluebells)	Range of colours present in original source not reflected in the digital object. Very little of the variation in the colour of the bluebells as the eye goes from the fore ground backwards captured.	Range and colours in the original source accurately re-created with the flowers appearing blue in the foreground and purple in the background. Soil in foreground is accurately recreated	Range and colours in the original source accurately re-created with the flowers appearing blue in the foreground and purple in the background. Soil in foreground is accurately recreated
Contrast (cat)	Too much black as the sides of the cat are not distinct from the soil/grass. Little detail present.	Contrast good as cat is clearly sitting on the soil/grass as in the original source. Detail present in both high and low lights.	Contrast very good as cat is clearly sitting on the soil/grass as in the original source. Detail present in both high and low lights

Table 17: Table of Observed Results from Scanners on the Same Photographs

Bookeye scanner

Canon CanoScan scanner

Epson scanner

Bookeye scanner

Canon CanoScan scanner

Epson scanner

Bookeye scanner | Canon CanoScan scanner | Epson scanner

Figure 17: Prints of the Digital Objects Created by the Three Scanners Compared

5.5.5 Comparison of NLW and CPR

In the NLW the project choices are made and recorded at the benchmark meeting. This allowed work to flow through the unit with the machines being used to their fullest capacity. The NLW has an array of scanners for its dedicated, highly skilled staff to use. These abilities are used, on the selected equipment, to produce the best digital object possible. The variety of equipment also allows a wide range of archival materials to be digitised. The quality of the digital objects created is verified by a second person, before representation copies are produced. They are then signed off on the workflow software.

In CPR there are now two scanners. The decision to obtain an Apple iMac and a second scanner was key. This new scanner, the Epson, is expected to be used for most archival materials in most projects. This is because it can accommodate the majority of archival materials in the collections. These archival materials are photographs, slides and paper based materials that are up to A4 size. Another advantage is that the staff appear to find the Epson easy to use. Finding the scanner easier to use gave the staff confidence that they would produce a good outcome. Creating digital objects is not a full time occupation for CPR staff so they need to be positive that they can adjust and operate the equipment.

The quality of the digital objects will be checked in batches after the representational digital objects are produced. This is because a second person may not be on hand to carry out the check when the digital objects are produced. Rather than hold up the digitisation process, it makes sense for the representational digital objects to be made and all the digital objects checked when the person is available. This is possible because, unlike the NLW, the digitisation can be split up into relatively small sections. For example the test collection could have been sub-divided into four smaller projects. The largest would have created 88 digital objects, the smallest 25.

The different scanners in CPR were compared using a set of photographs selected for the purpose. Each picture had different properties that were used to compare the resulting digital objects to the original. The Epson scanner showed the best reproduction of the original for all the digital objects produced.

5.6 Metadata Writing

During a digitisation project metadata is used to control the workflow. The metadata collected is used to fulfil several purposes. Some of these are to:

- Complete catalogue data
- Locate the digital object
- Control the digital object
- Control the collection, and
- Preserve the digital objects

5.6.1 NLW

Throughout the digitisation process as much metadata as possible is gathered automatically and married up with the catalogue data via the object management system (JISC Digital Media, 2008d). However there still exists a need for some description to be written individually and the correct digital object attached to the correct record.

Metadata writing was learned by examining and discussing examples already completed for other collections. These examples were then used to make a series of templates for each type of holding for the digitised collection. The templates were then used to create records for each item of the archival materials and digital objects in the collection. These were stored on a test database until they had been checked and passed. Feedback was given on the mistakes I made and how to avoid them in future. In the NLW the catalogue records are written using AACR2 rules in MARC records with the controlled language sourced from Library of Congress Subject Headings lists (Rowley & Hartley, 2008, pp. 88 - 89).

Having completed the fields which represented a data input, e.g. the size of the original photograph, the next step was to research the subjects of the images. This was done to try and ascertain more details to add to catalogue information. First those who may possess information within the Library were asked if they could add to the information. After this further details were sought on the HW Lloyd collection by using the resources of the Library and associated archives. A visit was also made to the Gwynedd local authority archive to try and identify the places and events in the digital objects through other images kept there.

Looking at the image and deciding what to include in the descriptive metadata for the library catalogue took quite a time at first. For example, a number were black and white images of German prisoners of war in uniform. Through research it was discovered that the differences between the uniforms of different sections of the German army at this time were small. For instance one section wore a red hat band another, dark green. Because of the age and possible deterioration of the glass negatives, trying to distinguish between these two colours in black and white images would have been arbitrary and probably not accurate. For this reason it was decided to write a generic description of the German forces uniforms of that period and to include it on each photograph containing uniformed prisoners of war. Working through this decision making process took time, but became easier with practice because seeing what was important in the image became quicker. An example of the digital objects from the NLW and its accompanying metadata is shown in Figure 18. The digital object attached to the metadata, in the example, has been reduced to 47% of the transmitted size to fit to the page.

The view I saw while completing the MARC forms would have been similar to the blue coloured fields in Figure 19. The fields to the left of the data are taken from the MARC web site and those to the right are from the notes that I made while being taught (Library of Congress, 2011d).

An advantage of completing catalogue descriptions after the digitisation process was that because the digitisation process often involves closely examining the screen image it was possible to note fine detail that may not show up in the archival materials. For instance the lace edging on some period women's dresses or the writing on a sign could have been missed. Having noted these during digitisation they could be added to the descriptive metadata. There is space for notes in Wombat next to each item that could be used for this.

Metadata writing requires highly developed skills. Many of the required fields have repetitive data requirements and can be set to be filled automatically. Others require subject knowledge and, more importantly, knowledge of the user, to arrive at terms that will enable successful searching.

	10 German POWs, Frongoch Camp, 1 a sailor
Authors:	Lloyd, H. W., of Bala.
Material type:	Picture
Subject(s):	Prisoners of war, Fron-goch Concentration Camp, 1910-1920, Wales, Frongoch, Frongoch (Wales), German, Glass negatives
Language:	English
Publisher:	[s.n.]
Year:	1916
Annotation:	During WW1 the German Army wore a Feldgrau (field grey) uniform. A soldiers' rank was denoted by the decoration on the collar, cuff and epaulettes. Officers wore a peaked cap while other ranks had a Feldmütze, a cap without a peak, with the unit they belonged to denoted, until the issue of a grey camouflage band, by the colour of the band around the cap. The contingent a soldier belonged to was denoted by the upper of 2 buttons worn on the front of the cap. Puttees (gamaschen) were often worn by storm troops (Sturmtruppen). During WW1, men who served afloat in the German Navy wore blue, and those who served on land, Feldgrau (grey) uniform. A peaked cap and insignia on the left sleeve denoted rank. The German army air force during WW1 was called the Deutsche Luftstreitkräfte (or before 1916 – Die Fliegertruppen des deutschen Kaiserreiches). They wore the same uniform as regular army troop except for the insignia on the shoulder boards.
Description:	1 negative : glass, b&w 119 x 163 mm.
Summary:	Image shows group of 10 German POWs, with 4 seated on a bench and 6 standing behind. Person seated right appears to be Naval personnel and the others from the army.
In:	H. W. Lloyd photographic collection. (WlAbNL)006082796

Online:	

Figure 18: Copy of Metadata and Digital Object from "10 German POW's, Frongoch Camp, 1 Sailor" (Accessed 02/02/2012 from: http://discover.llgc.org.uk/default.ashx?q=h+w+lloyd)

MARC Description						My Description
Group Name	Description of Field	Tag	In 1	In 2	Data	Which Information to Record in Field
Control fields	Control number	001			vtls006084950	Record identification number
	Control number identifier	003			WlAbNL	Organisation identification number
	Date & time last transaction	005			20110324103700.0	Time is to the second so it acts as version control
	Physical description field	007			cr bn ||aap	Set tables of descriptions for physical description
		007			gt bd||s|	
	Fixed-length data elements	008			110222s1916 wlk k|eng d	Set tables of cataloguing data
Numbers and code fields		019			$a c $c g	
		039		9	$a 201103241037 $b cwf $c 201103240841 $d cwf $c 201103170918 $d cwf $c 201102241424 $d cwf $y 201102221504 $z cwf	Record of each edit made (date, time & editor)

MARC Description						My Description
Group Name	Description of Field	Tag	In 1	In 2	Data	Which Information to Record in Field
	Cataloguing source	040			$a WIAbNL $b eng $c WIAbNL	Code for organisation writing data & data language
Main entry	Heading - personal name	100	1		$a Lloyd, H. W., $c of Bala.	Name of Author or creator
Title & title related fields	Title Statement	245	1	0	$a 10 German POWs, Frongoch Camp, 1 a sailor $h [graphic] / $c H. W. Lloyd.	Title & format of holdings
Edition & imprint	Publication, distribution, imprint etc.	260			$a Bala : $b [s.n.], $c ca. 1916.	Information on the publication or edition
Physical description fields	Physical description	300			$a 1 negative : $b glass, b&w ; $c 119 x 163 mm.	Extent & dimensions of item(s) e.g. 1 photograph b&w
Note fields	General note	500			$a During WW1 the German Army wore a Feldgrau (field grey) uniform.…They wore the same uniform as regular army troop except for the insignia on the shoulder boards.	
	Restrictions on access	506			$a Restricted, please refer to digital copy or photo album.	
	Summary, etc.	520			$a Image shows group of 10 German POWs, with 4 seated on a bench and 6 standing behind. Person seated right appears to be Naval personnel and the others from the army.	Description of contents
	Preferred citation of described materials	524			$a H. W. Lloyd Collection, National Library of Wales.	Citation format
	Additional physical form available	530			$a Available in Llyfr Ffoto 1561 and as digital image.	Different formats of item(s) available

MARC Description						My Description
Group Name	Description of Field	Tag	In 1	In 2	Data	Which Information to Record in Field
	Immediate source of acquisition note	541			$e PZ8254/97	
Subject access fields	Subject added entry – personal name	610	2	0	$a Fron-goch Concentration Camp $z Wales	
	Subject added entry – topical term	650		7	$a Prisoners of war $x German $z Wales $z Frongoch $y 1910-1920. $2 lctgm	
	Subject added entry – geographic name	651		0	$a Frongoch (Wales).	
	Index term-genre/form	655		7	$a Glass negatives. $2 gmgpc	Genre or format of item(s).
Linking entry fields	Host item entry	773	0	8	$7 p1kc $i Part of: $a Lloyd, H. W., of Bala. $t H. W. Lloyd photographic collection. $w (WIAbNL)006082796	Collection details + organisation code

Figure 19: Mock MARC Data Entry Record Form Completed with Data from Fig.17 (Accessed 15/03/2012 from: http://cat.llgc.org.uk/cgi-bin/gw/chameleon)

On reflection I observed that there are two actions that are key to successfully carrying out this process. One, is being methodical to ensure that all the fields on the MARC forms are completed accurately; the second, adopting several perspectives at once. Adopting several perspectives at once is useful because when subject entry points are selected the choices made need to reflect the expectations of users when researching items via the catalogue (Rowley & Hartley, 2008, p. 44).

5.6.1.1 AACR2

In the context of the NLW, AACR2 cataloguing rules are used in MARC records to create catalogue entries for the holdings. In relation to digitisation the main entry is the item of archival material while digital objects are added entries.

5.6.1.2 MARC Records

MARC records provide a standardised layout of information. Main entry forms for each format of archival material are drawn up with as much information as possible automatically completed. Digital objects are added to the relevant main entry via an added entry form. These provide the data that is displayed as a catalogue entry on the web site.

5.6.1.3 Library of Congress Subject Headings (LCSH)

Library of Congress Subject Headings tables provide lists of terms and the standardised way they should be written. This is so that searches carried out using these terms give the largest possible number of successful results. This is because computerised systems cannot read, unless programmed to do so, that 'WWII' probably means the same as 'World War 1939 - 1945'.

5.6.2 CPR

At the beginning of the research a collection of posters had been catalogued by previous students using a paper based system, and some books had been catalogued using the University Library system. The Library system was investigated with the view to using it to catalogue the metadata for the digitisation project. It uses Marc records and LCSH in a similar manner to the NLW, but I was advised by Information Services staff that it was not suitable for what I wished to do.

I decided to write the metadata into a spreadsheet, designed specifically for CPR, using Open Office software after examining several alternatives. These alternatives were rejected on the grounds that they either incurred significant cost, or would require significant upkeep by CPR. A spreadsheet was chosen over a bespoke database due to the generally acknowledged difficulties with the future maintenance of legacy databases. Open Office software is known to be compatible with systems that utilise XML, allowing for the possibility of data exchange in the future. One effect of these sheets is that they allow the archive to gain control of its holdings by recording the identification of individual archival materials. This identification will add to the authenticity and reliability of the archive.

Unfortunately with this arrangement, the facility for searching and retrieving data and digital objects was limited to finding a word in the document. So the accurate retrieval of digital objects from the system, by linking it directly to its metadata was not achieved. However, the arrangement of files and folders that the digital objects are stored in does help with finding resources. This is because the folders have been nested so that they mimic the arrangement of the archival materials. Therefore, although the digital object and its metadata are not physically linked, the metadata does link both the archival material and digital object and the arrangement of the virtual storage copies the archival arrangement (JISC Digital Media, 2009f, p. 2; Western States Digital Standards Group: Digital Imaging Working Group, 2003, p. 18).

The metadata spreadsheet column headings were written using refined DC as it provides the possibility of crosswalks to other schema if required for data transfer. It also provides some of all categories of metadata generally accepted as the minimum needed for effective exchange between facilities. So using DC allows for the expansion or transfer of data if needed in the future. This also allows for the possibility of future exchanges of data (Haynes, 2004, p. 52; Zeng & Qin, 2008, p. 17). A second spreadsheet formatted with the metadata designed by ECLAP for submission was also drawn and filled. This schema is also based on DC and is aimed specifically at the performing arts community. This second sheet has been drawn up because CPR are members of ECLAP and are, therefore, precisely the type of group the metadata format is aimed at. It was written into a second separate document because at present the ECLAP metadata schema goes beyond the usual refinements of DC and has not been formally ratified (DCMI, 2011).

Using a spreadsheet works well for individual archival materials as each row would represent an item of archival material and its related digital objects. This would also work with complex archival materials that require several scans to complete the item of archival material. For example, when books are digitised the item accession number would remain unique. What would change is that instead of an item of archival material with one accession number occupying one row with just one digital object identifier associated with it, there would be several rows of digital objects associated with one accession number.

Keeping metadata up to date is important to ensure its accuracy as intimated by Deegan and Tanner (2002, p. 3) where they state that as much care should be taken in its making as in the making of the digital object itself. This is because a catalogue becomes a metadata level copy of the archive.

At the start of the project, descriptive details for the collections existed only in the memory of the Director of the productions performed. After discussion it was felt that the quickest and easiest way to gather the metadata from the Director was to interview him. This was recorded on a Dictaphone to make playback for extracting the relevant data for metadata easier. Having completed the interviews it was felt that these recordings provided such a context rich resource that they were transferred to a disc and added to the collection. These discs have been made available to users so that they can view and listen to the description of the digital objects together.

Metadata is collected to enable the long-term management of the created digital objects. For CPR it also has the function of creating a catalogue of the archival materials, at item level, in the collections. There are two actions that are key when collecting this data:

- Keeping it up to date, and
- Keep on checking its accuracy

For this study most metadata was written after the digital object was created. This balanced the demands of the metadata writer, who need to be able to see archival materials to create descriptions, and the archival material, which is protected from accidents during handling, as the digital object can be used instead. It also meant that all the remaining metadata could be written for one digital object at one time.

5.6.2.1 Metadata Format

Various metadata schema were investigated and rejected in the course of the research for the project. The advantages and disadvantages of each were assessed carefully. For instance MPEG-7 was developed to provide metadata for audio visual content and is not particularly suited to describing books (Zeng & Qin, 2008, p. 15 & 309). The 15 element Dublin Core Metadata Element Set (DCMES) was devised to provide a middle ground between professionally drawn catalogues and search engine indexes to provide a description method for the majority of materials (Cumming, 2005, p. 35; Zeng & Qin, 2008, p. 17). These findings are summarised in Table 18. DC was chosen for this research because it:

- has been widely adopted,
- is easily understood by non-specialists, and
- it contains some of all the data required to successfully manage the digital objects in the future (DCMI, 2011).

Schema	Why Rejected/Pros and Cons for Research
Dublin Core	Widely used for exchange; OAI-PMH exposure; crosswalks to XML have been ratified by DC board; easy to use and understand by non-specialist staff; during project ECLAP have adopted it as basis for own submission metadata (DCMI, 2011).
Anglo American Cataloguing Rules	Designed for use in libraries and constructing catalogues; rules need to be learnt but access can be bought online; once learnt easy to use, but more complex than Dublin Core; need to be used in MARC records – more cost; crosswalks are constructed using the MARC format not the AACR rules (Zeng & Qin, 2008).
International Standard Archive Description (General)	Strength is in describing content and context of archive materials; needs to be mapped to XML to crosswalk to other metadata schema; can be used just at fonds level; only 6 description fields are considered essential; access based upon description; can be more difficult to extract data due to field arrangement; not complex to learn (Haynes, 2004, p. 57; Zeng & Qin, 2008, p. 15).

Table 18: Table of Metadata Schema and Features Relevant to the Project

5.6.2.2 Writing Rules

The use of a spreadsheet imposed some rules on writing as relevant cells were formatted to ensure this e.g. the cells that have dates inserted were formatted so that the date will appear dd/mmm/yyyy. Other writing rules were written into the relevant guide. An example of these rules is how the content for the field ‘dc.identifier’ is worked out. Figure 20 shows how this is done.

The dc.identifier structure also echoes that of the holdings in the archive and that of the folders on the hard-drive. The names for the derivatives made from the archive digital object have the letters ‘v’ and ‘th’ added after the file name to distinguish between the viewing and thumbnail copies. In this way all the digital objects are related to the archival material and to each other (JISC Digital Media, 2009f, p. 2;

Western States Digital Standards Group: Digital Imaging Working Group, 2003, p. 18).

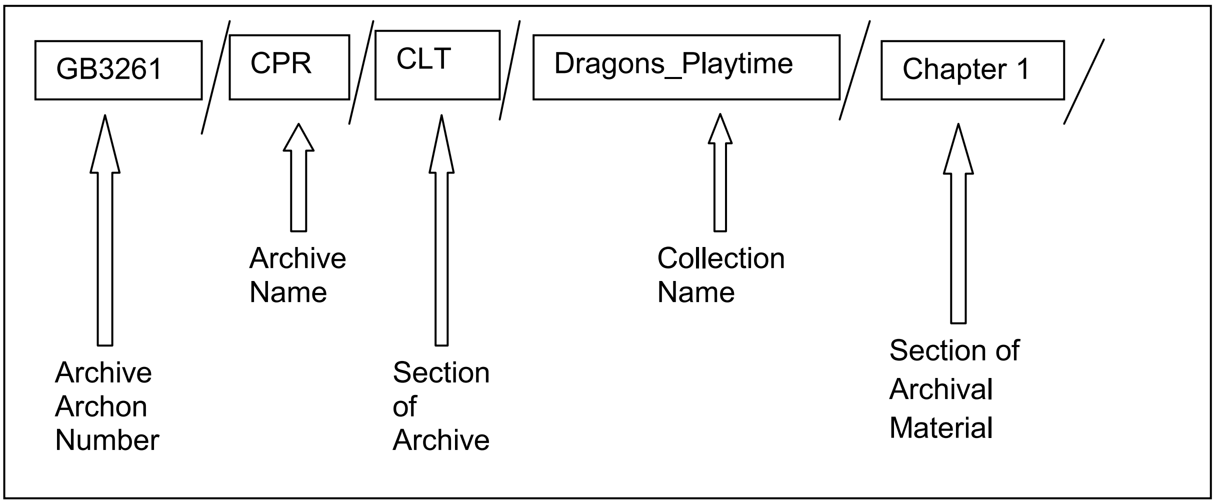


Figure 20: How to Write the dc.identifier in CPR

5.6.2.2.1 File Name

The file name of the digital objects follows standard guidelines with regard to using lower case letters and numbers combined. Figure 21 shows a screen shot of the front page for The Widow's Dream section in CADAIR. It can be seen that there is an abstract that briefly describes the work, the page URI and the date it was created on the site. Below this "Files in this Item" lists the file name of each digital object, the file size and format, a button to click to view and a description.

Figure 21: Screen Shot Showing Metadata as it appears in CADAIR

A copy of the data on this page had been reproduced in Table 19. The row highlighted in blue is the description for the reproduced digital object Figure 22.

The Widow's Dream	
Abstract:	This production took place in the Llechwedd Caverns. The performance journeyed down through the caverns stopping to tell the tale at each location. The lighting from the attraction was supplemented to highlight features specific to the performance. The story tells the tale of a woman widowed by a mining accident. She journeys into the mine and strikes a deal to free the body of her husband so that she can ensure he receives a fitting funeral.
URI:	http://hdl.handle.net/2160/7716
Date:	2011-12-12

Files in this Item				
Files	Size	Format	View	Description
twd00002_v.jpg	62.66Kb	JPEG image	View/Open	Lights inside the mine give the effect of a fire lighting the woman's face.
twd00003_v.jpg	74.65Kb	JPEG image	View/Open	2 female characters sit at a table and drink together while discussing their situation
twd00004_v.jpg	42.93Kb	JPEG image	View/Open	Very large hand pointing the way underground
twd00005_v.jpg	74.01Kb	JPEG image	View/Open	2 people in hats and cloaks lit up as though ghosts underground
twd00006_v.jpg	95.66Kb	JPEG image	View/Open	Person carrying a large light passes the audience candles for light as they wait to leave the caverns on the train
twd00007_v.jpg	104.1Kb	JPEG image	View/Open	Some of the old workings and rock face lit up at the edge of the lake - the entrance to Hades. The husband approaches the gate keeper to the afterlife behind the desk
twd00009_v.jpg	42.02Kb	JPEG image	View/Open	Possibly en-route to the funeral processions
twd00010_v.jpg	74.08Kb	JPEG image	View/Open	Another viewing angle of 2 female characters sitting together at a table, drinking and discussing the situation
twd00011_v.jpg	121.3Kb	JPEG image	View/Open	Funeral' boat lit up and floating across the water
twd00012_v.jpg	94.30Kb	JPEG image	View/Open	Audience leaving, the mine workings after the show, wearing hard hats. Musicians play and other cast members hold lit torches
twd00020_v.jpg	84.35Kb	JPEG image	View/Open	A woman wrapped in a black shawl and carrying a black umbrella is carried by a man wearing a top hat
twd00021_v.jpg	108.6Kb	JPEG image	View/Open	Group blowing whistles and banging drums process down some steps taking the audience to the mine
twd00023_v.jpg	58.76Kb	JPEG image	View/Open	The woman and her husband sit and spend time together before 'his funeral' underground

twd00025_v.jpg	42.54Kb	JPEG image	View/Open	Man carrying a candelabra and wearing evening dress and sunglasses - the character is imagined as the servant of Pluto and Persephone
twd00030_v.jpg	69.00Kb	JPEG image	View/Open	Group of instrumentalists call the audience to the mine
twd00031_v.jpg	91.04Kb	JPEG image	View/Open	The widow meets with Pluto & his servant
twd00032_v.jpg	87.21Kb	JPEG image	View/Open	View of the 'funeral boat' floating on the water in the cavern.
twd00033_v.jpg	98.19Kb	JPEG image	View/Open	The widow dances as she leaves the mine, passing an audience member wearing a hard hat as she does so
twd00034_v.jpg	82.03Kb	JPEG image	View/Open	2 men lit by candlelight in the cavern
twd00035_v.jpg	68.44Kb	JPEG image	View/Open	2 men lit by candlelight in the cavern

Table 19: Table Showing the Metadata for Screen Shot Displayed in CADAIR

Clicking on the View/Open button opens a digital object like that in Figure 22. Under Figure 22 it can be seen that the last part of the URL is the equivalent of the dc.relation field in the metadata spreadsheet.

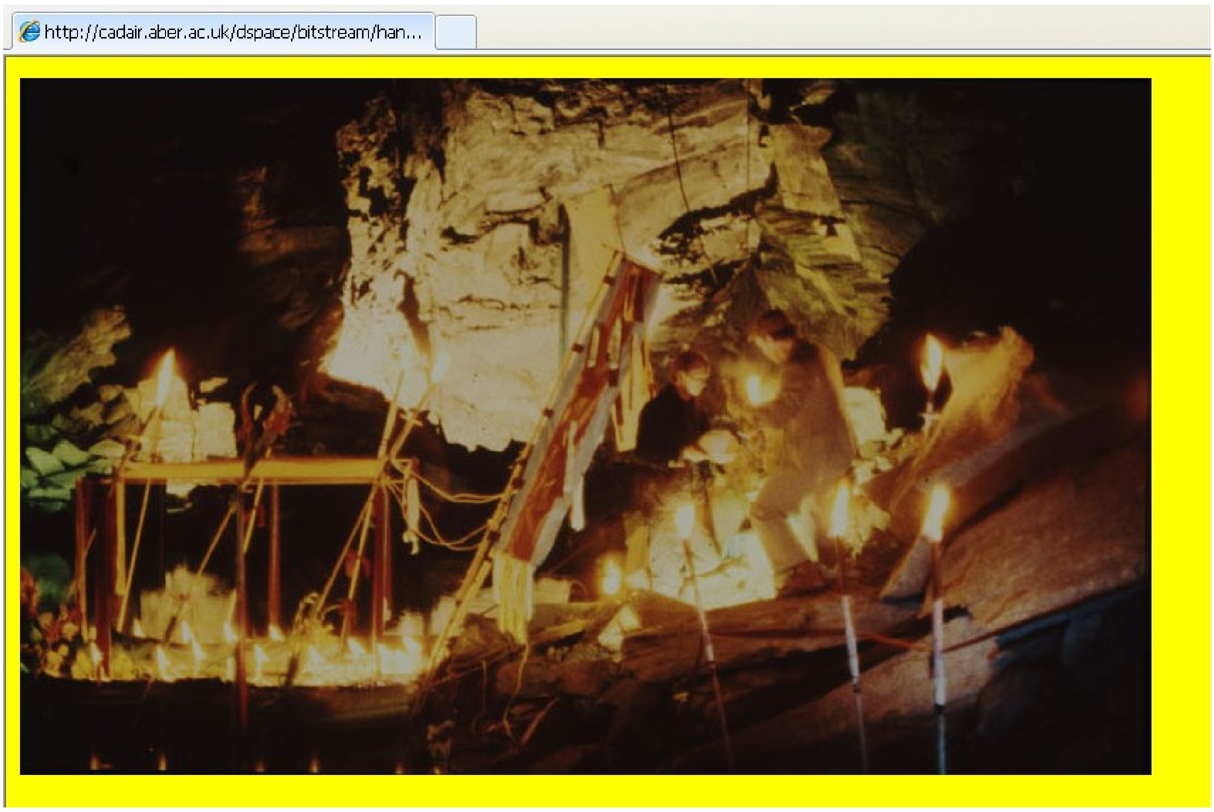


Figure 22: Cropped Screen Shot Showing a Retrieved Digital Object (Accessed 20/12/2011 from: http://cadair.aber.ac.uk/dspace/bitstream/handle/2160/7716/twd00007_v.jpg?sequence=15)

It can be seen that the URL comprises the three letter section code, in this case 'twd' standing for The Widow's Dream, and five digits, '00007'. The letter '_v', just before the format extension, indicates that this is the viewing copy of the digital object (Western States Digital Standards Group: Digital Imaging Working Group, 2003, p. 18). The last part of the URI, the '?sequence=15', indicates that this was the 15th digital object in the sequence uploaded.

5.6.2.3 Controlled Language

Language control was imposed by using the Dictionary of the Theatre: Terms, Concepts Analysis (Pavis, 1998). This dictionary was used as it was the reference resource recommended by CPR. A copy of this book is kept in CPR Resource Centre.

5.6.2.4 DROID

DROID is a piece of software designed and distributed by TNA. It is an automatic file identification tool (The National Archives, 2011b). It works by comparing the file format of a digital object to those in its related database, PRONOM.

PRONOM is a database that contains information on file formats and the software versions that can open them (The National Archives, n.d.b).

In this research it was installed on the PC already in CPR, and used to generate reports on the technical properties of the digital objects created. These reports contain several columns whose contents could easily be taken down incorrectly from individual digital objects e.g. the file size or the URI. It is then copied into the metadata spreadsheet to complete these fields.

The software says it is "*a platform-independent Java tool*" (The National Archives, n.d.b) so it should run in any Java enabled environment. However when downloaded to the iMac in CPR it failed to run. The advice pages enclosed in the software zip file suggest that in order to rectify this, command lines in the software may need to be rewritten. Command lines are lines of programming code that tell the computer what to do. It was decided that to re-write these for CPR this went against the principle of developing a system that CPR could manage without external help, as they could not re-install or update the software, if needed, without re-writing these lines.

5.6.3 Metadata Quality Assurance

Metadata, like digital objects, should be checked for consistency. This is so that an archive can demonstrate the authenticity of its holdings. The method used for this depends upon the system being employed to store the metadata. It can be as simple as checking for spelling errors or as complex as running software to interrogate the consistency of the file. The proof of error checking should be kept as these can show when an error was first detected or if it was the result of receiving a faulty digital object from the digitising process (Haynes, 2004, p. 154; JISC Digital Media, 2008b; Western States Digital Standards Group: Digital Imaging Working Group, 2003, p. 38).

The description of DROID reveals that it was designed as an interrogator of the PRONOM database to help organisations check whether file formats or software versions were in danger of becoming obsolete (The National Archives, 2011b). In the research it was used to check that the metadata of the digital objects in the folders matched that in the metadata spreadsheet. Another way of doing this would be to check the sheet consistency by using a function to check the data is the type of data it should be e.g. that a data is written in a specific format (Haynes, 2004, p. 162).

5.6.4 Comparison of NLW and CPR

At the NLW as much metadata as possible is harvested automatically. That which cannot be automated is written into MARC records using AACR2 rules and LCSH. These measures ensure that the metadata is as accurate and complete as possible. This accuracy is therefore reflected in the Library catalogue.

Collecting and collating metadata using automated processes requires an investment in equipment and software, plus the expertise in how to use it. An understanding of what metadata to collect was one the areas highlighted by CPR as problematic at the start of the research. There was also a need for CPR staff to be able to manage and maintain any system installed. It was eventually decided to use a spreadsheet formatted with extended DC and written using Open Source software, to store the metadata. In CPR, the staff were already familiar with spreadsheet software and therefore have only had to learn what is meant by each of the headings to know what to collect. Because it is written using Open Source software that uses a recognised XML dialect, it is expected that it will be compatible, having been converted, with a system that imports using XML. Because DC contains some of all the metadata required for successful data management, in the future the collected data can be used to manage the digital object.

There were rules applied to the metadata writing to ensure that the information is accurate and reliable across the collection digitised for the research. Some detailed metadata was collected using reports generated by using DROID software (The National Archives, 2011b).

5.7 Creating Access

Using the internet to display digital objects increases the opportunities for access by a wider audience of users which is one of aims of many digitisation projects.

As indicated in the definitions, digitisation is the process of turning archival materials into digital objects. For the NLW and CPR one reason to digitise materials is to increase access to the archival holdings by creating digital objects and the opportunity to view them.

Creating opportunities for viewing the digital objects is the final part of the process at the NLW.

5.7.1 NLW

In general there are two methods used to display digital objects on the NLW Web site. One is by attaching the digital object to the catalogue record and the second is by developing an exhibition in the Digital Mirror area of the web site.

Within the Digital Mirror section there are two main types of presentation:

- A static display where digital objects are placed on a page and a short piece written about them or the collection they are part of, and

- A rotating display of selected images from a collection with a caption underneath.

Each section is repeated with the captions written in both English and Welsh for the two versions of the web site.

For the HW Lloyd collection a rotating display was created showing the two main themes of the collection. I created this display using Typo3 software. It is generally accepted that most computers have, at least, basic software that will read text out loud. The NLW delivery model is predicated upon users using this, or their own specialist software, if they wish to access resources this way. This delivery model fulfils a requirement to provide access to content for users with a disability. This is why the contents of the Digital Mirror have written descriptions attached to them.

For both the HW Lloyd collection and the Dwynwen Belsey collection the digital object was attached to the catalogue record and can be accessed by users via the Library web site.

Direct access to the individual digital objects digitised for the research is achieved by typing the collection name into the catalogue search box on the opening page of the NLW web site. Filtering the results from the choices on the right hand side of the page will lead to the chosen collection. Alternatively access can be gained by navigating to the catalogue pages.

5.7.2 CPR

During the research project it was decided to display the digital objects produced in CADAIR, the University Repository. This would make them accessible via the Web to researchers outside the University. The uploading procedure was straight forward and the digital objects are displayed in the order of their file numbers not their upload order. This display feature has implications for the arrangement of the archival materials prior to their digitisation as the numbers allocated here are carried forward through the system.

In future CPR expect to make digital exhibitions of digital objects on their Web pages showcasing specific collections or specific aspects of the performing arts. Making the digital objects accessible in this manner satisfies one of CPR's aims of making their collections more accessible to users, while giving some protection to archival materials.

5.7.3 Comparison of NLW and CPR

Access to the digital objects was created at the NLW by attaching the digital object to the catalogue description. In this research an exhibition was created using Typo3 software for the HW Lloyd collection in the Digital Mirror part of the NLW web site.

Access to the digital objects created in CPR was made by displaying them in CADAIR. The importance of arranging archival materials before commencing digitisation was shown when the software re-arranged the digital objects into numerical order after they had been uploaded.

5.8 Written Guides Results

At the beginning of the study, CPR had no recorded experience of digitisation projects. Meetings were held with CPR staff with the objective of ascertaining their expected requirements for their digitised outputs. They had identified a barrier to acquiring the information they required as the use of jargon. They explained that as they read advice on digitisation projects, the word metadata appeared to have many different meanings. In initial meetings with CPR it became apparent that this jargon barrier was, in part, due to the failure by publications to convey exactly what they meant with their descriptions, and in part due to the knowledge that there was a wealth of hidden meaning in the use of the terms they encountered.

Having considered the points made during the meeting I used the conversational channels that had been opened to informally discuss ideas for transferring the learning process. Using this method, the design of the written guides was arrived at.

The written guides naturally split themselves into two sections; they are called Generic and Organiser's Guide and CPR Quick Start Guide. The guides are not written as academic pieces; therefore, the language used is less formal than that required in academic texts. There are also no citations, but the references for these are included in the text of the thesis.

Among other topics the Generic and Organiser's Guide contains details on how to work out which items of archival materials to prioritise for digitisation, how to work out how much storage space is needed as well as why and what metadata is collected. It concludes with a section detailing the next steps to take to ensure the sustainability of the digital objects created. This guide is intended for those who are carrying out a function similar to that of a project manager.

The CPR Quick Start Guide contains detailed instructions on what metadata to collect and where to store it, and directions on how to use the equipment to scan archival materials.

Throughout the guides an imaginary collection was used to illustrate the principles being discussed.

As part of the process various other strategies for learning transfer were considered and rejected. Table 20 shows some of the possible methods and an outline of why they were rejected.

Strategy	**Reason for Rejection**
Shoot a video	• Actions take place in real time which can make them difficult for viewers to follow. • Video would require re-shooting and editing when updates are needed. Re-shooting would probably require more than one person: re-editing would probably require another. • Viewing adjustments to digital objects would rely upon the sensitivity of the camera to capture them
Record an audio description	• Pictures can make it easier for learners to identify that they are carrying out the same action by comparing their screen to the one they are viewing. • Recording would require re-recording and editing when updates are required.
Social networks/Q & A forum	• Interactive forums can be useful to build up a database of FAQ's to research answers, but until populated the value of the forum to a questioner could be small. • A second, suitably qualified person is needed to answer the question. There is also a time delay in forming an answer which when there is limited time available can lead to longer delays in completing the project.

Table 20: Table of Rejected Communication Strategies

5.8.1 Evaluation of Guides Results

The three testers were chosen by CPR. From CPR's point of view it is realistic that one person at a time will be working on the process. Each tester was assigned to test either the Generic and Organiser's Guide or the CPR Quick Start Guide depending upon their particular skill set. This was done when they were selected by the Director of CPR. He compared the testers' skills to the tables at the beginning of the Generic and Organiser's Guide. All the testers were at post-graduate educational level.

Two testers completed tasks from the CPR Quick Start Guide and one the sections labelled 'For Organisers' in the Generic and Organiser's Guide. All were engaged and interested in the research, and how the department could make use of it in the future.

The tester of the Generic and Organiser's Guide had previous experience as a book editor and chose to approach understanding the guide by taking it to a room to be on her own to read. A questionnaire was completed and has been included in Appendix F: Evaluation Questions. After she had read the guide and made her comments, she chose to give me verbal feedback and showed me where she thought improvements could be made. Overall she felt the guide was well laid out and useable in that form. However, she felt that sometimes the sentence structure was too complex and that this hindered understanding of the process being described. She also felt that one table could be improved with a minor change in the way it was formatted. These issues have been addressed by incorporating the suggested changes in the current version. This tester gave three hours to reading and marking up the guide.

There were two testers for the CPR Quick Start Guide. One completed both the metadata and digitisation sections and the second, just the digitisation section. The second tester was a last minute replacement for someone who was ill and had limited time to give to the evaluation. They therefore completed the digitisation task, but not the feedback form or the metadata section of the guide. The feedback form from the first tester is included in the appendices. Both testers use both PCs and Macs for work related tasks.

The testers were told that the task involved digitising three slides and two photographs. They were told that first time through they would be helped, where necessary, by myself, and the second they were to use the written guide. They were then asked if they wished to read the guide before commencing the assisted run through. Both elected to read after completing the task the first time. I then told them that I would be observing them to see if the process could be improved and if they thought of anything as we went through I would like to hear it.

Both testers completed the assisted run through of the task. The type of problems they encountered were those that could be anticipated for people using computers and equipment they are new to. For instance, forgetting the short cut to a path to a folder, or missing the need to change the prefix on the file name of the digital object on the scanner. During the second run through it was interesting to note that both testers relied upon their memory to complete the task, rather than refer to the guide. At the end of the session both testers were asked if they would now feel confident to carry out the task on their own. With the first tester this discussion moved onto changes they would make. They suggested making a strip to sit along the top of the computer key board. This would have a print out of the column headings and examples of their content so that filling in the metadata spreadsheet would be easier. Further discussion also indicated that a 'cheat sheet' with things such as the relative sizes of 1megabyte (Mb) to 1 gigabyte (Gb) etc. could prove useful.

Both testers took around two hours to complete the assisted cycle, and around one hour for the un-assisted run through. One tester asked why 24-bit colour was still selected even though they were scanning black and white photographs. The reason for this is that black and white settings on scanners generally reproduce only in black and white. A black and white photograph is made up of shades of grey so the setting chosen would need to produce a digital object in shades of grey.

The test was limited in scope because of the number of testers, but for CPR it showed them that their staff understood what the guides were asking them to do. It also showed that the way they intend to use the guides will work for them.

The new equipment was installed immediately prior to the evaluation being carried out. The installation progressed smoothly except that DROID had not been installed on the iMac as anticipated. It was however still installed on the PC. The tester was shown the software on the PC, how it worked, and how the results wanted were selected and transferred to the metadata spreadsheet. They felt that they had understood the principle, but this was not assessed.

Having watched and talked with the testers while they were using the equipment, I decided to keep DROID on the PC and to adopt the equipment layout in Figure 15 15.

5.9 CPR Feedback

Feedback on the system installed in CPR was sought after the evaluation of the written guides had been completed. This was collected by interviewing the Director of CPR as he had been involved with the research from the beginning.

<u>Question 1</u>: What were your expectations at the beginning of the research?

At the outset CPR wanted to discover a well-organised mode of digitising its archive. The outcome the Director most wanted was a system that could be used by those with little or no experience to allow them to digitise the collections successfully to acceptable standards.

The thing that made most impression on him was the amount of metadata that was recorded for each digital object. The Director was aware, through his own professional contacts, that ‘having the eye’ to be able digitise and assess the digital object produced was something that some people had, and was often developed over many years. They felt that the investment in equipment would allow CPR to digitise many archival materials to a high standard without too much trouble or worry about quality.

<u>Question 2</u>: Do you feel that your expectations have been met by the system?

Yes, he felt that CPR’s expectations had been met. Beyond his initial requirements he felt that having discovered the deficiencies in the operation of the [Bookeye] scanner, the replacement [Epson] scanner extended the facilities they [CPR] now have to enable them to digitise many formats of archival materials well.

The success of the project has also made them realise how much is involved in digitising archival materials to acceptable standards.

Question 3: What do you feel was not successful about the research?

He felt that there was nothing that was not successful, as problems were situations from which to learn. In that respect the largest opportunity to learn from was the discovery that the Bookeye scanner was not able to do what I wanted.

One of the personal learning points the Director took from the research was, how much of the context and knowledge of the archival materials was dependent upon him.

Question 4: What do you feel was successful about the research?

The most successful part of the research he felt were the recordings describing the digital objects and by association, the archival materials. He had found it an interesting exercise to try and build a running order for some archival materials. This was because many of CLT's productions had no obvious narrative as they were performed in the streets of a town and progressed using what was happening around them at the time.

The Director described how he had seen the backup hard-drive in storage and had decided to look at the digital objects. Prior to digitising them, to view the slides the digital objects were made from, would have taken some time to set up the required projector and screen correctly. He described how to look at the digital objects now there was no delay in accessing them, unlike viewing the archival material. To him this epitomised the value of the access they wanted to obtain for their users.

5.10 Interview Results

Interviews were conducted to supplement the background information available on digitisation because it was felt that a personal account might indicate where changes were led by utilising the available technology or adapting what was available to fill a gap.

The interviewees who agreed to take part in the study had experienced digitisation from different perspectives. Some had been involved with projects to produce digital objects and some had also spent time educating resource users. All had been involved professionally with digitisation for more than ten years. This is a summary of their remarks.

Question 1: What do you think have been the main developments in digitisation practices over the last 10 years?

The main developments that were highlighted by participants' were the growth in project size, and the way preservation of the digital objects created has moved from being an afterthought to being part of the planning for digitisation projects.

Participants A, B and D described how digitisation projects being carried out now have a scale that could not have been coped with ten years ago. The development of Google Books was cited as being a driver for change. Google Books is a database of digitised books (Google Books, 2011).

It was also felt that changes to the technology available led to developments in practices. They felt this was seen particularly in developments in workflows and data management (Participant A, B, and D). These changes in techniques led, in turn, to reduced production costs. These costs occurred because: the speed at which projects could be completed meant a lower cost in man hours; the way storage costs reduced over the time and technological developments that have led to better management of the digital object.

Participant C described how initially, some funded projects had concentrated on the immediate outcomes of digitisation with little awareness of the long-term needs of the resource created. They also highlighted the impact on the user of different user models. In particular how places that charge for access had developed their resource contents based upon a model that assumed that those who wished to access their services could and would pay.

Question 2: How do you think these [developments] impact on the use of digital resources for scholarly practice?

The views on how the last ten years have impacted upon scholars were varied. Participant A highlighted how the developments in metadata collection meant that finding a resource was no longer as arbitrary as it had been. However Participant C indicated that the impact had been significant, and that a skills gap had been highlighted in the skills required to analyse the suitability of resources by students.

Participant D felt that in order to allow new ways of working to emerge, scholars need digital materials.

Participant C felt that there was a mis-match between what was available, and what was needed to help inexperienced users to use digital resources to their capacity.

Participant B felt that the usability of finding aids was still an afterthought in terms of project management and that the use of digitised content should be further embedded into the practice of scholars. In common with Participant D, they felt that there are two communities involved in digitisation; developers and users. And that not many changes in practice had yet been observed as resources were being used in traditional ways once accessed.

<u>Question 3</u>: Thinking of small collections, what do you feel are the main issues with regard to digitisation projects? How do you think they can be solved?

For small stand alone collections all of the Participants' agreed that making the collection long-lived and useable was the major issue facing small or specialist collections at the present time.

The Participants had differing views on how the possible effects the change in emphasis by funding agencies may manifest themselves. Participants B and A both felt that funders should lead projects towards ensuring the sustainability of the created resource. Participant C felt that small collections were in a good position to make good quality resources. They felt that the developments in off-the-shelf technology were making smaller more easily managed projects more likely.

Participants C and D highlighted a variety of practical issues. They also indicated that they felt that copyright issues can be particularly difficult for isolated collections.

Participant D felt that small collections had a distinct advantage with their flexibility.

<u>Question 4</u>: Which of these issues do you think should be prioritised?

In no particular order, the priorities given by the Participants were:

- Sustainability of the resource
- Making the resource available through the web and exposure to web crawlers
- Greater use and recognition of the resources use

How the greater use and recognition of resources could be achieved was described in a variety of ways by the Participants. They all felt that greater use should be made of the opportunity to publicise resources and collections so that they became known to a wider audience. Ways given to achieve this were through exposure to web crawlers and training in the use of the digital resources and by things such as ensuring that resources are correctly cited. It was felt that this was away to make resources more valuable and therefore more likely to attract funding for sustainability.

Participant C felt that a priority for small collections could be to make digitisation a whole concern for the whole organisation, and that this would in turn encourage sustainability. They explained that they felt that sometimes projects failed due to a lack of knowledge on a variety of fronts. This lack of knowledge included things such as what the applicable standards were.

<u>Question 5</u>: How will these issues affect users?

Adopting the issues outlined will affect users by:

- Giving the ability to do more sophisticated searching via the semantic web (Participant D)
- Giving users what they want (Participant D)
- Its sustainability only affects users when it is under an external threat (Participant D)
- Funding issues could be tackled by more collaboration taking place. One way of tackling this could be to map nationally where these federations could occur for particular types of project bids (Participant B).
- The effect depends upon the initial concept of the project (Participant C)

<u>Question 6:</u> What would you consider a good method of evaluating the value of digitisation?

Participant D and C drew attention to the resources such as Google metrics and (Participant D) indicated adopting readily available software to gather quantitative data is relatively simple and does not cost a lot of money. But Participant D also felt that this cheapens the value of the collections. That using qualitative methods would produce results that reflected the value of the collections to the user more. The downside is that they cost a lot more to implement. Participant B highlighted the work that has already been done on this [TIDSR], but that it currently lacked peer review to give it credence.

Participant B indicated that a step in this general direction would be to demonstrate the value of collections already held. They indicated that one way of doing this would be to put on each web page how the resource should be cited and to collecting data related to its use.

Participant C suggested using the initial parameters of a [funded] project and comparing them to data retrieved from the individual projects after completion to see whether the two matched.

Participant A also suggested adopting a statistical analysis of users. They then went on to outline how folksonomies could be utilised to engage the user with the resource.

<u>Question 7</u>: How do you see these developments [those from the last 10 years] carrying forward into the future?

In the future it was felt that technically digitisation is likely to become more automated. This could allow an automatic marrying of the digital object and metadata so that publication is much quicker (Participant D). Scholarly practice changes could be here quite soon, and this means the approach of digitisers must change.

Work has already been done that shows a virtuous circle of digitisation is created by adopting best practice. And that this circle leads to a more sustainable resource that users want to use (Participant B).

<u>Question 8</u>: What effect do you think these [possible developments] will have on small projects?

All felt that for small collections it is their very size is the key to their future. Participant D indicated how small collections are more flexible than large ones. That they do not have the time lag created by having to change working practices on a large scale when, for instance, new equipment is adopted. So using this flexibility would be their key to their survival.

Participant A however felt that the most likely outcome in the near future was that small projects would merge to ensure sustainability, but that long-term, because so much had happened in the last 10 years that beyond the immediate future it was impossible to predict.

5.11 Chapter Summary

The knowledge transfer was achieved by the production of writing guides for CPR that describe both the higher level processes and the nuts and bolts of how to create digital objects from archival materials. Information on how to carry out a digitisation project was gained from being embedded at NLW, attending relevant external events and lectures, reviewing available literature and soliciting the views of leading professionals.

All the information gathered was developed into a system for digitisation by CPR. This system consists of a straight forward technical system and two written guides. The technical side of the system uses two computers and a scanner. One guide describes the pre-digitisation processes the other how to operate the equipment to create digital objects and write relevant metadata. The guides were independently evaluated by volunteers from CPR. This was successful as they were able to use them without intervention after receiving suitable training.

There are a number of related processes to be carried out before digitisation and metadata writing takes place. Assessing which materials to digitise is governed in the NLW by its policy and any additional restrictions for specific projects. CPR's general collections policy was used as the basis for the care of the collections, and selection uses a process similar to risk assessment tables. Copyright permission to digitise all the holdings for the photographer was obtained using a letter template. At NLW this permission had been obtained before I arrived. Working out how much storage is required for a project is not usually done in great detail at NLW as the digital objects are stored on specific servers and moved to long-term storage when appropriate. The storage option chosen in CPR was to place the digital objects on an external hard-drive attached to the PC being used for writing metadata. This hardware can continue to be used to fulfil different functions as their future plans are realised.

Bespoke software is used to control the workflow at NLW. Such a solution was not an option for CPR. Instead, workflow is controlled by a flow of emails between contributors. Control of individual archival materials is achieved via a number allocated to them in both CPR and NLW. In CPR this is the individual indentifying number and in NLW a file number, not the accession number, allocated when the workflow is started on Wombat.

At NLW there is a whole department that digitises archival materials. This is equipped with several different types of scanner that are used by highly skilled staff to create the digital objects. CPR has two scanners attached to two different computers and staff with other primary functions. The workflow to create a digital object was the same at both NLW and CPR, but assessment of the quality was carried out at a different point.

An assessment of the differences between the scanners available in CPR was carried out. The scanner that produced the best digital objects was the Epson.

At the NLW, MARC21 records are completed using AACR2 writing rules, and LCSH as a controlled language, with the data that cannot be automatically collated for a digital object. The chosen metadata scheme for CPR was DC which was written into a spreadsheet and stored with the digital objects.

The evaluation showed that the guides can be followed successfully by the intended users, and used to produce digital objects and metadata of a suitable quality.

The professionals contacted and asked to take part in the research were asked questions pertaining to the history of digitisation, its associated problems, and possible future direction. All the interviewees highlighted gathering appropriate metadata as a past and continuing problem. In the future it was felt that more would need to be done to prove the usefulness of the digital object by extending its scholarly worth and, that part of this may be achieved through extending metadata applications.

Chapter 6: Discussion

The system adopted by CPR to create digital objects was reproduced by the actions written into the guides. During the design and testing phase of the system there were a number of issues that became apparent. Two of these were:

- The communication of values to users outside the information sphere
- Adding relevant context to metadata

These areas were addressed by collecting additional metadata relevant to CPR users in alternative formats. Collecting this extra metadata added to the authority of the archive.

6.1 Communication with Centre for Performance Research

The needs and wants expressed by CPR were enquired about and noted throughout the period of the project. As indicated in section 4.8, this was initially approached by asking questions on a range of topics that I wished to learn more about and by exploring the physical archive. Time was spent observing how users interacted with the archive and the learning centre. Having discovered the weaknesses from CPR's point of view, it was possible to arrive at a workable solution.

Having listened to the comments made, it was determined that there was great enthusiasm within CPR for the project and the future they hoped for. As a department that utilises strong visual skills, they were aware of where these skills were used in a digitisation project. However they were also wary of them. It also became apparent that there was nervousness with regard to the fact that there were professional skills and standards, of which they had no knowledge, being used by those who were carrying out digitisation professionally.

One of the first ways this issue was clearly articulated by CPR was when they described how they found the usage of some terms a barrier to understanding what was meant by some papers. From their point of view they found the different ways the word was used confusing as it seemed to be used to describe very different things. The confusion occurred because CPR always understood that the terms used were loaded with values specific to the archives world. Finding explanations of these values they could relate to was where they felt the difficulty, or the gap, lay.

From my perspective the communication of ideas to CPR was key to successfully carrying out the project. I well remember being told that I wouldn't understand hierarchies when I started working in libraries. As a mathematician, I understood that hierarchies are a series of numbers, letters, or both that follow a previously worked out pattern. Dewey Decimal Classification (DDC) is a hierarchy generally applied to arranging books. What the person was trying to tell me was that giving a book a number from the DDC was based upon their knowledge of the students search patterns, the subject syllabus and the contents of the book: that it was not just giving it the number that would probably be assigned by someone who did not know the library. Allocating a book a number and some descriptive detail some information professionals would call accessioning, some cataloguing, some classification and some applying metadata. When being taught how to do this, the principles, not the rules, being taught are what need to be transferred. The principle in this example is how a book is described using an allocated code for that library, not the strict allocation of numbers. It was being able to isolate these skills from the techniques I was taught at NLW that allowed me to see what needed to be transferred to CPR, and the existing skills to align them to.

One of CPR's driving forces was the wish to set up a system that functioned correctly from the start. Because of this they were open to ideas and suggestions of how a complete system to preserve and use the digital objects they created could be implemented. This complete system view was an important consideration throughout the project and strongly influenced the methods chosen for capturing the metadata and digital objects. Because I was taking a whole system view I researched the creation and storage of digital objects as though I were tasked with implementing a repository project. This work then fed back towards to the methods used to store the digital object and the chosen metadata scheme to ensure that the projected plans could be achieved.

6.1.1 Strategy Adopted to Transfer Learning

The strategy for transferring the learning process was expressed explicitly during the project when adapting the workflow learnt at the NLW for transfer to CPR. This learning was implicitly transferred during the evaluation of the written guides when the evaluators created, stored and wrote the metadata for digital objects. Values were transferred by the written guides and during their evaluation.

As shown in the systems map, Figure 3, unlike the NLW, the digitisation workflow is isolated from the rest of the functions of CPR and TFTS as it is completely within the archive and has little interaction with other parts of the department.

The final digitisation workflow in CPR is illustrated in systems map Figure 23. This systems map uses the same colour scheme as the NLW digitisation workflow systems map, Figure 11. Boundaries that are drawn in the same colour carry out the same function in CPR as they do in NLW. Functions not present in the NLW workflow are drawn in red. From these maps it can be seen that the two workflows share the functions:

- Digitisation
- Catalogue
- Conservation
- Relevant policies/general preservation policy/ written guides, and
- Metadata

In addition the CPR workflow has functions for:

- Using the Generic and Organiser's Guide
- Using the CPR Quick Start Guide
- Seeking copyright permission
- Selecting archival material
- Calculating the storage requirements
- Allocating an individual identifying number
- Using DROID, and
- Managing digital object storage

The key to the CPR workflow was the allocation of an individual identifying number to archival materials. This was shown as the bridge between the Generic and Organiser's Guide and the CPR Quick Start Guide in the systems map.

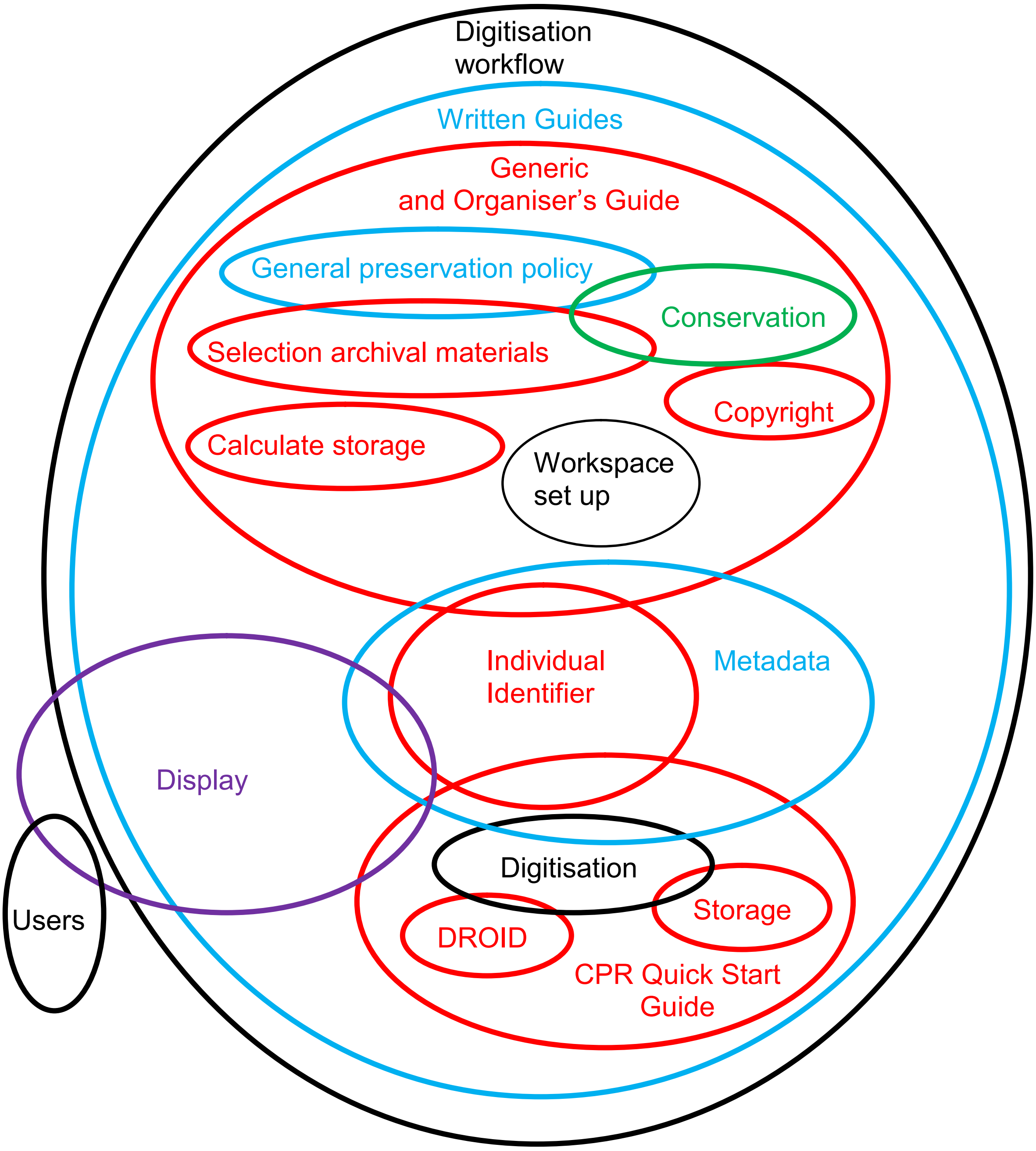


Figure 23: Systems Map of Final CPR Workflow

An important point of understanding was that the transfer of skills information could be achieved through the identification of common values. Having recognised these, the learning sequence, with related values encapsulated, could be aligned to them.

As a department, TFTS had different traditions to those experienced at the NLW. This was unsurprising as the industry their students ultimately wish to join, traditionally projects a very different public face to that of libraries and archives. It would have been unreasonable therefore, to expect that an archive that is wholly within TFTS had ingrained the culture of an archive within the information sector. CPR wished to actively maintain and use their collections to the best of their ability. Similarly professional archives also undertake to maintain collections in their care to the best of their ability. Another common value was using specifically phrased, written details to repeat an action. Many performances actively seek to repeat the same thing, in exactly the same order each time. This is achieved by relating the business of a performance back to, for example, a script. This business is recorded in a specific order, using an agreed language. This is similar to a catalogue record being completed using a controlled language in regulated fields. The catalogue record controls which item is accessed from the holdings, and retrieves the same one each time it is recalled. Similarly, in writing down how an action is performed, a person can go back to the text and repeat the action to recreate the business or to digitise archival materials (JISC Digital Media, 2008c). Both of these examples are the same principle expressed using different words (Reid, 2000, p. 144).

Extrapolating from this I isolated the skills outlined in Table 21 from the tasks I was learning at NLW. I then arranged them having first identified three distinct phases in a digitisation project. I called these the pre-scanning phase; the scanning phase and the metadata phase. These names were chosen to reflect the main focus of the activity of the phase and not the place within the project where they are carried out.

Skill	**Digitisation Project Area**
Project organisation	Pre-scanning actions
Team building/Leadership skills	Pre-scanning actions
Quick assimilation of information	Pre-scanning actions
Understanding of technical information	Pre-scanning actions
Communication of ideas	Pre-scanning actions
Picture balancing/ differentiate fine detail in a digital object	Scanning phase
Colour balancing a digital object	Scanning phase
How technical adjustments affect digital object	Scanning phase
Proof reading skills	Metadata phase
Adopting differing perspectives	Metadata phase
Broad general knowledge	Metadata phase
Written communication skills	Metadata phase
Ability to repeat a task while self monitoring performance	All phases
Methodological approach to tasks	All phases

Table 21: Skills Identified in Digitisation Projects

Having identified the skills I was using I considered where else I had used them previously. This began the process of identifying other roles that use the same or very similar skills. Table 21 shows that there are many skills that are utilised in a digitisation project. It also shows how few of them are common between the different phases. It is this range of skills that makes digitisation a complex operation.

Unfortunately, whereas some of the skills in an area such as conservation can be visible, those used by scanner operators and metadata writers remain hidden except to those also in possession of them. This is because unless a person has, for instance, highly developed visual skills they do not see the difference that small changes make to the digital object produced. The same applies to metadata writing; unless a user knows the subject matter very well they are unlikely to appreciate that the resources they are being directed to are all pertinent to their search. It is this challenge that CPR were confronting when they identified the language used about metadata as confusing. This has implications for the digitisation process as without the skill to see, for example, the faults in digital objects, quality assurance will be unsuccessful. Analysing the technical qualities of a digital object only goes so far, as the trained human eye is still by far the most sensitive instrument available (Langford & Bilissi, 2011, p. 3). It is often not acknowledged that the most important part of any digitisation system is the human eye as it transfers the data from the monitor to the brain of the viewer.

Written guides became the strategy adopted for transferring the skills learning for two reasons:

- CPR asked for this kind of guidance and
- Following written instructions is a familiar method to many people.

From the point of view of their users, written guides allow a learner to:

- progress at their own pace,
- have them read out loud while using them as visual aid.

Writing a guide means that it is a description of the system at the time they were written, so they need to be updated at least as frequently as the equipment is changed. The current validity of the actions described in the guides should be checked periodically. One possible way of triggering this could be when files formats are checked for possible obsolescence. Implementing a system to periodically check all aspects of the digitisation process would ensure that CPR could be sure that their actions continued to comply with current recommendations for producing digital objects and that their files remained in good health.

When writing the guides it was also decided not to adopt a style that included a series of explanatory facts linked, but placed elsewhere in the guides. This was because a user could be in the position of flicking from one place to another if using a hardcopy or from screen to screen if on a computer. This could make the process appear more complicated than needed. This decision was reinforced in discussion with CPR where they indicated a preference for a simple descriptive document containing direct instructions and another with more detailed explanations of the process.

6.1.2 Sustainability of Guides

The final versions of the guides have been supplied as a digital file to CPR as well as in a hard copy. The digital file can be sustained by preserving it in the same manner as the other digital objects. The relevance of the text can be maintained by making alterations to the copy as appropriate.

The directions on the operation of the equipment and its software in the Quick Start Guide will remain valid until they are replaced or upgraded. At that point the instructions for this part of the equipment may need to be rewritten to reflect this. Writing a new set of instructions can be carried out by CPR. This is because the descriptions of the adopted standards are in the relevant section of the Generic and Organiser's Guide and then repeated alongside the operating instructions. Therefore to update instructions CPR would need to write down what they do to operate the equipment and software to achieve the required digital object while ensuring they keep to the standards that the present instructions include.

One of the limits to the guides is the validity of the information they contain. The standards used to write the guides remain applicable until they are changed by the information community. Changes to these can be monitored by accessing relevant email discussion lists or social media services following contributors such as TNA or

BL. An alternative could be to establish a parallel process alongside the periodic monitoring for preservation where periodicals and selected websites are reviewed for relevant changes to the digitisation process. Establishing a once a period sweep of all the relevant actions for changes to recommendations alongside that for digital objects and metadata, will ensure the sustainability of the digital collection and its adherence to standards.

6.1.3 Genericity of Guides

In writing the guides it was envisaged at first that there would be one guide. It became clear during the writing process that while the instructions to operate the equipment were case dependent, much of the process prior to the scanning of archival materials was less case dependent. This realisation prompted the separation of the guides into two; one generic and one case specific.

In this research, skills transfer information was mainly presented via the instructions in the written guides. But these guides cannot make up for the human factor. The kind of skills possessed by those carrying out digitisation at NLW are either present or it is not. To an extent, successfully using the written guides relies upon these invisible skills being brought forward and used consciously, at least at first. The problem here is that to people that have the skill it is very visible, to those that don't, it is not.

The issue of how the field of information communicates with all audiences was the theme of the IC-ININFO conference I attended in September. There several speakers called for a unified thesaurus or at least convergence between terms so that the same frame of reference was used by the whole information profession. Convergence would help organisations similar to CPR to digitise, or at least to understand, what the information area is trying to say.

6.2 Adding Relevant Context to Metadata

As previously indicated, the adoption by ECLAP of extended DC confirmed the choice of schema, and meant that only one was being used in CPR. The great advantage of this schema is how the contents of fields are expressed, can be locally defined. This allowed CPR to put across information that was relevant to their users in appropriate fields (Zeng & Qin, 2008, pp. 18-22). Expressing data relevant to their users was a key point to be engaged with when gathering relevant metadata for CPR. The textual copy of the metadata was enriched by taking details from an audio description of the productions and digital objects by the Director.

The main avenue through which the metadata for CPR was collected, and given added context, was through the recorded commentary. At the time, this was chosen as the most expedient way of collecting the required data from the Director; the only person who possessed it. As the audio recording was being made, I became aware that the nuances of vocal delivery gave it even deeper context. This was because I could hear there was much more being transmitted than could ever be caught by any written metadata scheme. There was also a deal of facial movement not being captured by the recording. So much, that I suggested that in future, if the equipment was available, CPR might want to consider videoing the metadata collection interview. Videoing, instead of just collecting an audio file, would capture the movement and facial expressions being transmitted by the person talking as well as the audio description. This would be particularly useful because much performing arts material is heavily context dependent.

The context conveyed by this audio recording meant that the level of metadata available is much richer than that in a simple written scheme. It is generally accepted that the stronger the metadata the better the user experience. The metadata CPR collected was much stronger because the recording contained more data of relevance to the performing arts sector. This is important because from an archival point of view metadata should be expressed neutrally but, in this case, this doesn't serve its intended community of users as well as the recollections of one of the people involved in the productions.

6.3 The Role of the Catalogue

Connecting the written descriptions, the audio file, the digital objects and the archival material together, establish the connections that create a catalogue for CPR. By extension the authenticity and the authority of the archival materials could be established and maintained. For an archive, achieving this state has important ramifications.

6.3.1 Authenticity

Archival authority comes from trust because, like mathematical formulae, it is backed up by logical processes that can be applied successfully to many situations and provide reliable outcomes. Some archivists feel creating digital objects from archival materials violates some of these basic principles and therefore the bond of trust between an archive and its users. The principle most often felt to be violated is that relating to the principle of ‘not changing the item’ (Davis-Perkins, et al., 2005, p. 5). Others see new forms of carrying media as being able to be absorbed through the adaptation of current practice to reflect the emerging new paradigms (Reid, 2000, p. 1). In this way authenticity is maintained by adopting the principles of current practices rather than by applying them as dogma. The InterPARES glossary describes authenticity as:

> *authenticity*
>
> *n., The trustworthiness of a record as a record; i.e., the quality of a record that is what it purports to be and that is free from tampering or corruption. [Archives] (InterPARES, 2011).*

This means that to be considered authentic a record must satisfy these criteria. Therefore by definition an item that cannot fulfil these is not authentic. In the archival profession this causes some anxiety when it comes to digital objects. This is because for some, the proofs associated with these definitions have become intractably allied to physical manifestations.

In the NLW an archival digital object is the copy from which all others are taken. Other copies are made according to known presentational parameters such as formatting all the digital objects as pdfs’ so that the text of an item can be searched or as a .jpg for display on a web site. The transformations are documented and form part of the metadata for the digital object thus proving they are what they say they are.

Digitised items, because of decisions made during the process, are inevitably a reflection of the view of the person carrying out the scanning of an item. This is because the way a scanner is set up will vary from one person to the next. Even if different operators use the same settings, differences in the way people perceive the digital object make personal variations in perception inevitable. Users in turn also see their version of the scanner operator's view. This means that their view of the digital object may be different to the view they form of the archival material. These differences in perception may be unimportant to users, but never-the-less it is still there. The danger in this situation has most effect on the future in facilities where the digital copy becomes the replacement for the original in all its forms (Conway, 2010).

Images placed on web sites are often subject to further change beyond those made when processing a digital object for archival purposes. These changes can range from being compressed to ease transmission, to changing the colour through the receiving monitor set up. Most reputable institutions do not change the appearance of the digital object beyond replicating the appearance of the archival material, but research on the level at which a users' perception is altered is scarce (Conway, 2010, p. 455). The available evidence is also, inevitably, subjective, if only due to the manner in which such a topic can be assessed (JISC Digital Media, 2010c). In the *American Archivist*, Conway (2010) outlines research he conducted using digital objects from the Library of Congress collection. He assessed their use and comments made by individuals of their personal image choice. He concluded that the personal information system used by his subjects overcame any shortcomings in the digital objects they were using, according to their personal view point. These findings contrast with Davis-Perkins, et al. (2005) who felt that the adjustments made during the digitisation process were an intolerable breach of the trust placed in archives by users. They felt this trust was eroded because, they felt, users did not realise that digital objects had been manipulated to produce them. I felt that there were two things to be considered about the way this view is articulated:

- Davis-Perkins, et al. do not detail how the bond of trust is broken by these actions i.e. they do not provide instances where this has happened and the consequences for the user, and
- They do not present evidence that users are not aware of the processes involved in creating digital objects.

From the point of view of an archive; a disclaimer or introduction could be used to explain what processes had taken place and thereby allow the user the chance to weigh up the evidence for themselves.

6.3.2 Representation

In his paper Conway said that the process of "*building collections of photographs through digitization is fundamentally a process of representation*" and then after recognising the possibility of miscommunication asserts that it is a "*means of communication*" (Conway, 2010, p. 427)

Representation is not the same as authenticity, although they are inextricably linked through the catalogue of the archives holdings. The representation of a collection can affect the authenticity of an archive, but the authenticity of an archive cannot affect its representation. This is because authenticity reflects the integrity of the whole archive whereas representation is about only one aspect of an archive.

To an archive its integrity is the foundation of trust in its materials. Trust comes from being able to show it has certain procedures in place, but also from reputation. The catalogue, and now any web presence, is the basis of that reputation. The catalogue is the basis of trust because it is the intellectual representation of an archive. The catalogue contains all the information that, in records management and archival practice, endows individual items with the attributes considered necessary to be able to prove that the item is what it claims to be. These attributes, representations, combine to endow an object with authenticity (International Council on Archives Ad Hoc Commission on Descriptive Standards, 2010, p. 7).

If the catalogue was persistently wrong, then trust would be eroded. If a web presence mis-represents the holdings of an institution, this would have the same effect as an incorrect catalogue. When Davis-Perkins, et al. (2005, p. 6) pointed out the deficiencies in the web presence generated by the images they surveyed, they are highlighting the deficiencies in the cataloguing practices of those institutions. When so much content is available on the web, the catalogue is not the only expression of representation. Their web sites are too.

Another point in the digitisation process where the representation of a collection was an issue was where an institution was a partner in a collaborative project. Birmingham Institute of Art and Design (BIAD) questioned whether the output from collaborative projects accurately reflected their holdings (Everitt, 2005). They felt this was a particular problem when project catalogues, because they only contained some of their material, were not drawn in the same way across all the projects they were involved with. They solved this by only providing high level descriptions for all projects. This ensured all the projects they took part in had the same level of information on the collections and therefore represented them fairly.

Although the primary aim of digitisation for CPR is to provide a means of access, it will also enable them to record the significant properties of their archival materials. For CPR, carrying out the assessment for the digitisation process gave the opportunity for archival materials to be catalogued. This was because they could be assigned an individual identifier, even if the decision is not to digitise. The allocation of this individual identifier means that generally accepted ways of demonstrating the legitimacy of the holdings can be shown. This demonstration of legitimacy means that the value of the holdings can be demonstrated. This demonstration of legitimacy and value becomes more important in an era where increasing levels of funding are being sourced from outside the University. Demonstrating the value of what that funding would achieve, as well as show casing their record of achieving results from previous funding streams, could be of importance to CPR in the future.

6.4 Reflections on My Role

Throughout this process my role as an intermediary has been influential on its completion. The role I undertook could be described in different terms depending upon the project phase. While at the NLW, I was a pupil or trainee; at CPR I became a translator or communicator of the digitisation process. As an evaluator I took the role of reflector or coach for those taking part.

To carry out the role of pupil I needed to use listening and questioning skills to separate the elements of what I was being taught. While learning about the workflow software I used skills I had acquired previously in understanding spreadsheet software and married them to my skills in understanding systems and workflows to visualise how the software might control the process. While scanning I used my strong visual skills to assess the colour and composition of the archival materials to compare them to the digital objects produced. These were married to my software using skills in assimilating what each piece of software was doing and how it affected the digital object created. Metadata writing primarily used my skill at searching a database, reading and understanding its contents and applying it to each piece of archival material. While writing I used my knowledge of the subject area to investigate areas of further research to acquire additional details for the metadata. I also used my questioning skills to ask experts questions I couldn't resolve myself. In parallel I used my information assimilation skills to understand and apply the theory I was learning from the University and through reading relevant items.

For each of these skills I tried to answer four questions:

- Do I really understand this
- Can I explain it to someone else in manner they can understand
- Can I write this explanation down clearly, and
- If I cannot; what do I need to change to achieve be able to do this?

I then expressed the learned skills on the equipment in CPR by using the equipment and carrying out a series of scans of archival material. I did this because I learn best by doing, rather than reading about something. Sitting at the equipment and placing myself in the position of the people in CPR confirmed to me that in order to receive consistent guidance over a period of time, I needed to structure a set of guidelines. The guidelines would lend consistency by telling each user the same thing in the same order, assuming they were accessed in the same order. CPR had indicated this would be a suitable method for them at the beginning of the project; but I had difficulty initially visualising how it would be most effective. I felt strongly that making it simple to understand would be key to its adoption. I subsequently realised that by simple I really meant clear.

During the evaluation phase of the project I resisted becoming an instructor and became a reflective questioner, or coach. This was because during the evaluation I realised that the testers were asking questions to be reassured that they had understood the instruction correctly. I felt that this was partly because I was there with them as they often indicated they already knew the answer by, for instance, pointing to the correct place on the screen.

6.5 Chapter Summary

In summarising his research into what he calls "*modes of seeing*" Conway indicated that the time had come for digitisation practices to look more deeply into the needs of users of digitally based collections (2010, p. 460). This research found a way to do this for CPR users by making an audio recording of relevant metadata. This added context gains in importance the further we look into the future of metadata use through the possibilities afforded by the extending of the capabilities of metadata enhancing software. However these functions rely upon the skills of the human operators to capture the required details.

Chapter 7: Conclusion

The aim of this research was to investigate how the knowledge of digitisation practices and connected standards may be transferred from a large organisation to a small one through the digitisation of special collections.

The objective of acquiring the knowledge of practical skills to achieve the digitisation of two collections to high standards, were learnt at the NLW. The supporting theories to complement these skills were learnt from reading, lectures and by attending relevant events. Interviews were also carried out where selected professionals discussed their experiences of the past, the present and the future of digitisation. The knowledge gained was transferred to CPR by writing guides to digitisation. These guides were tested when I digitised the Blaenau Ffestiniog collection in CPR. They were also tested by CPR staff. The guides described the creation of digital objects and metadata using the equipment in CPR. Having discovered the added context given to visual materials by audio description, ways of continuing to record this are actively being investigated by CPR.

- In addition to achieving these objectives a paper describing the project, to that date, was given at the IC-ININFO 2011 Conference on the Island of Kos, Greece, in September 2011.
- A template of a copyright permission letter was written. It asks for permission to digitise and explains why CPR wish to do this to the archival materials in their collections.
- The written guides may be further tested by students taking part in the MSc Econ Archives Administration course at Aberystwyth University. This would demonstrate if there are additional needs for other groups not yet included in the guides.

The area of investigation was whether you can successfully transfer a knowledge system for digitisation from the NLW to CPR. Together, the written guides and the investment in suitable equipment make a complementary system. The system can be used by CPR staff and students to carry out digitisation and metadata collection un-aided. The guides were written so that they do not compromise accepted standards or values by being imprecise.

The smooth running of the workflow and therefore its output relies upon the traditional skill of arranging the archival materials. This is a skill that relies upon subject knowledge and visual skills. If this does not happen as expected then the allocation of identifying numbers could become confused. If this is carried through the complete digitisation process, there is the possibility that the digital object and the archival materials could become disassociated. This is the main issue that could arise through adopting the strategy in the written guidelines.

The limitations, as described in section 1.5, imposed by the research boundaries on the transfer of digitisation practices were generally minimised. This is because:

- the funds available were sufficient to allow CPR to acquire key pieces of equipment that were more suited to their needs
- digitisation practices and future preservation requirements were considered throughout by collecting appropriate metadata
- the collections pre-selected by the NLW were similar in nature to those I expected to encounter at CPR and
- the collection chosen by CPR was representative of their holdings.

7.1 Recommendations for Future Use: CPR

For CPR it is suggested that they examine their current preservation policy, Appendix E: CPR Policy Statement, with the additional management needs of digital objects in mind (Deegan & Tanner, 2002, p. 36). In particular they should consider drawing a timetable for inspecting the digital objects created and their storage conditions (Deegan & Tanner, 2002, p. 102; Haynes, 2004, p. 102). Guidance on some of the elements to include can be obtained from web sites such as the DCC and JISC (Digital Curation Centre, 2011a; JISC Digital Media, 2008c; JISC Digital Media, 2009b; Jones, 2010).

The preservation policy can then be used to write a digitisation policy and guidelines for the priorities to be applied when selecting archival materials for digitisation. This will help ensure that the archival materials selected for digitisation, form a coherent resource.

An additional benefit for CPR of completing a series of digitisation projects is that more of their collections will be catalogued. However, the current system will not draw together the separate catalogues that exist. Installing repository software could bring together all the different catalogues into one resource, along with born digital resources. This means that users only need to access one finding aid to discover all CPR's resources. As the primary aim for CPR is to build a curated resource for its users, it is recommended that CPR install repository software as soon as the finances are in place to enable this.

Which software should be adopted will depend upon the assessment of CPR's circumstances at the time, however a possible configuration is illustrated in Figure 24. CADAIR the University repository uses D-Space software, but research carried out for this project suggested that ICA Atom and other similar software, are other options that could be considered (ICA Atom, 2011). The adoption of software such as ICA Atom would require support from University Information Services, but because it is free and open source software (FOSS), there would be no purchasing costs. There is more on some of the options available in Appendix G: CPR Storage Options.

At present the audio descriptions made of the digital objects are only available for use in CPR, but there are several different plans for user events where this kind of metadata may be generated and added to the collection. If suitable software is chosen these could also be stored in the repository. These events could be utilised to provide more materials for researchers by recording them too.

7.1.1 Semantic Linking

The possibility of adding contextual data to metadata was highlighted particularly by Participant D during their interview. They referred to the fact that now MARC records have been described by a Resource Description Framework (RDF) there is a possibility of constructing a cross walk to schema, such as METS and MODS, that already utilise XML. This answer prompted me to think about how further into the future, there is the possibility of using semantic linking for users to access just the metadata they want to see, or to extend it to see more detail.

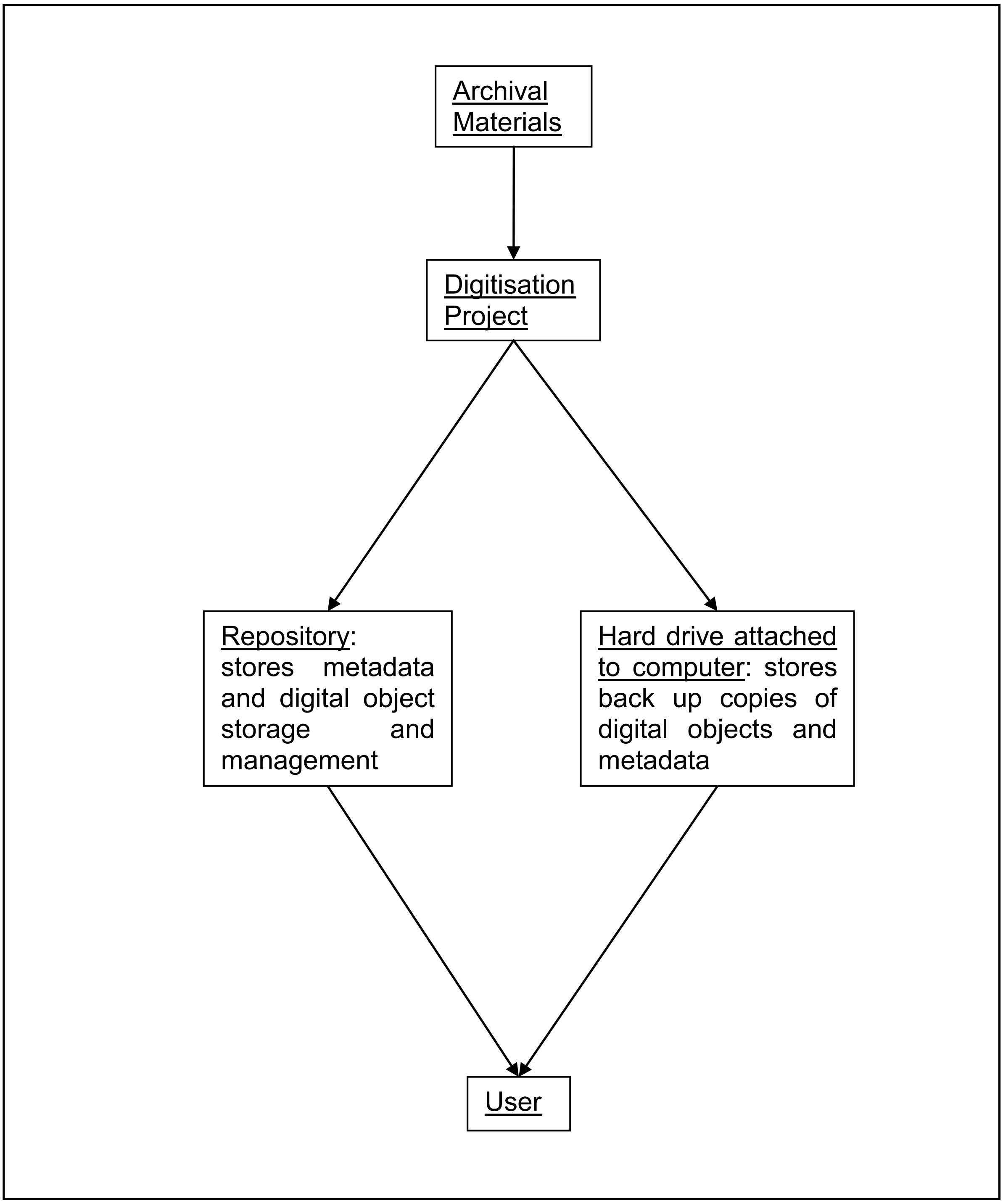


Figure 24: A Possible Future CPR System

RDF is both human and machine readable. It uses XML to describe structure of whole digital object, not just one aspect. One digital object is allocated one Uniform Resource Identifier (URI). A URI is a sequence of numbers and, or, letters that uniquely identifies a resource (Berners-Lee, Fielding, & Masinter, 2005, p. 1). A URI is also known as a triple because it contains three pieces of information that uniquely identify the digital object (W3C, 2012; W3C Schools, 2012a).

The importance of these crosswalks is that, where XML based systems are already in use, it offers a direct route to the possibility of matching data in a MARC record with contextual data that is not catered for in MARC records. This is illustrated in Figure 25Figure 25. This means that data, currently not visible, but contained in the METS of the digital object, can be seen as part of that digital object. Having created the link between the data and the object, the possibility of searching using the semantic web also becomes possible as shown in the rest of Figure 25.

In the future the possibilities of the semantic web are also open to CPR with some modification of the metadata. If a system similar to that in Figure 25 were adopted the data on the metadata spreadsheet would need configuring in XML and then wrapping with METS to produce an information package that could be stored in a suitable database. The resource would then need to be described using RDF and a link back to the information package. For a user to access this information, it then needs to be transformed into a web language, which software can interpret, to deliver it to the desktop (W3C Schools, 2012c). Links between resources are made by the user recognising common qualities such as the collection an item is from. If repository software that is already configured with XML has been adopted by CPR, then the process could be much simpler.

If all possible relevant staff were trained in the use of the complete digitisation process, then the problems associated with staff changes could be minimised. Following this practice would ensure that firsthand knowledge of the process would remain alive in the department. Additionally, if a strategy where a person carried the complete process through for a collection were adopted, then this could increase the sum of knowledge of the archive holdings as the person who carried out the process would have in depth knowledge of the collections they had processed.

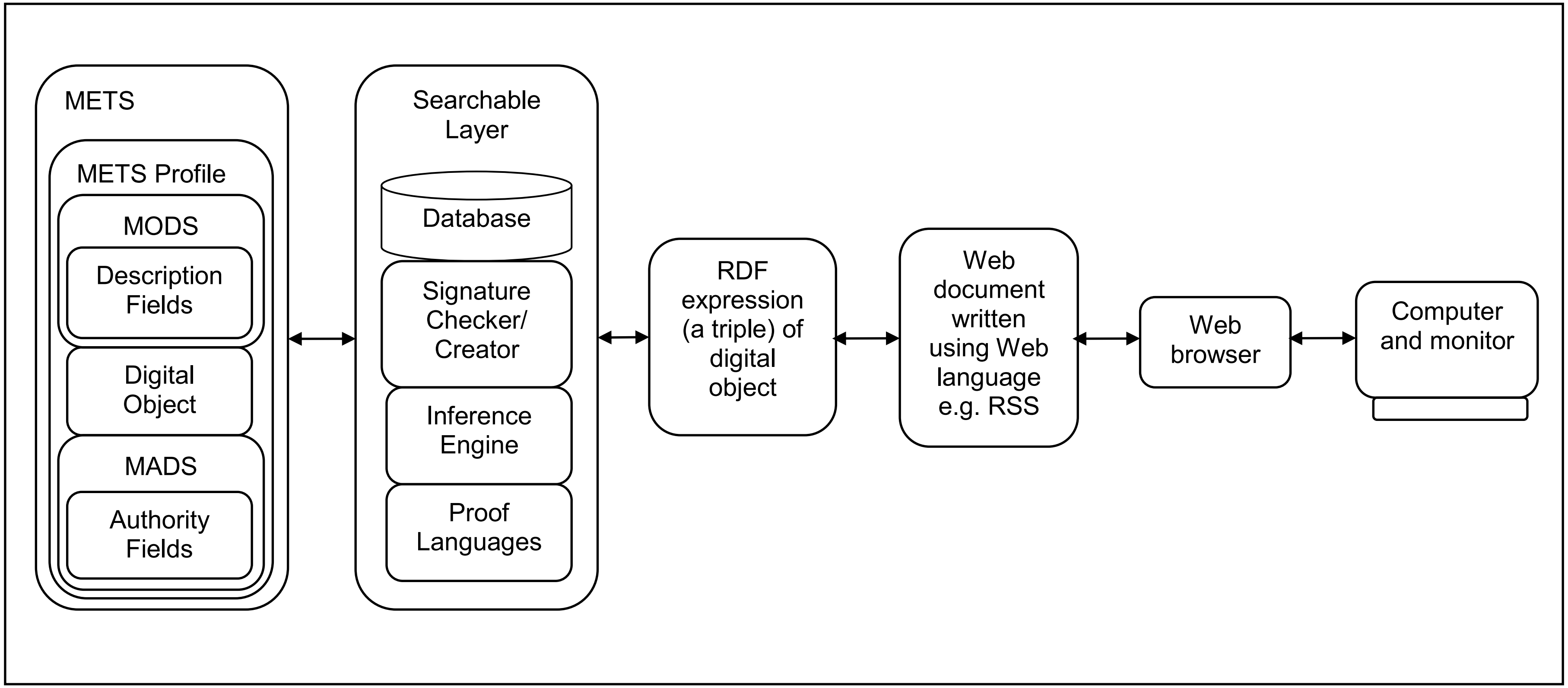


Figure 25: Possible Semantic Web Configuration

.

7.2 Further Work

There are two areas that suggest themselves for further exploration.

- There could be more investigation into which archival values have been transferred and why. Having observed that both NLW and CPR staff value their archival materials, it could be of benefit to other groups outside information practice to discover which of their values align with those of an archive.
- More investigation into the relationship between digital objects and the user of archival materials. Work that led to the TIDSR toolkit has already been carried out on the impact of resources in educational settings. As yet I am unaware of any work that has been carried out on how digital objects influence the emotions of the user. Discovering if digital objects have the same impact as the archival material could be of benefit to archives

Chapter 8: Bibliography

Bibliographic referencing style used: American Psychological Association Style

American Psychological Association. (n.d.). *APA Style.* Retrieved 12/03/2012, from APA: http://www.apastyle.org/

AQuA. (2011). *AQua MashUp*. Retrieved 2011, from AQuA: http://wiki.opf-labs.org/display/AQuA/AQuA+-+The+Leeds+Mashup

Arts and Humanities Data Service. (2005). *About the AHDS*. Retrieved 07/03/2012, from AHDS: http://www.ahds.ac.uk/about/index.htm

Arts and Humanities Data Service. (2008). *AHDS Projects*. Retrieved 07/03/2012, from AHDS: http://www.ahds.ac.uk/about/projects/index.htm

Australian Government Information Management Office. (2004). *Digitisation of Records: Better Practice Checklist.* Retrieved 12/03/2011, from Department of Finance and Deregulation: http://www.agimo.gov.au/archive/better-practice-checklists/docs/BPC18.pdf

Baca, M. (2008). *Introduction to Metadata: Online edition.* Los Angeles: Getty Publishers.

BBC. (2007). *BBC - Mid Wales National Library - History*. Retrieved 24/01/2012, from BBC: http://www.bbc.co.uk/wales/mid/sites/national_library/pages/history.shtml

BBC. (2012). *Data, Information and Knowledge*. Retrieved 16/02/2012, from BBC GCSE Bitesize: http://www.bbc.co.uk/schools/gcsebitesize/ict/databases/0datainforev_print.shtml

Beagrie, N., & Jones, M. (2008). *Preservation Management of Digital Materials: The Handbook.* Retrieved 02/05/2011, from DPC: http://www.dpconline.org/component/finder/search?q=media+and+formats

Beagrie, N., Beagrie, R., & Rowlands, I. (2009). Research Data Preservation and Access: The Views of Researchers. Retrieved 20/05/2010, from *Ariadne* (60): http://www.ariadne.ac.uk/issue60/beagrie-et-al

Beagrie, N., Semple, N., Williams, P., & Wright, R. (2008). *Digital Preservation Policies Study.* Retrieved 22/04/2011, from JISC: http://www.jisc.ac.uk/publications/reports/2008/jiscpolicyfinalreport.aspx

Berger, J. (1972). *Ways of Seeing.* London: BBC Books: Penguin.

Bermingham, A. (1989). *Colour Temperature, Colour Temperature Correction and Neutral Density Filters in Television Lighting.* London: The Society of Television Lighting Directors.

Berners-Lee, T., Fielding, R., & Masinter, L. (2005). *Uniform Resource Identifier (URI): Generic Syntax.* Retrieved 03/03/2012, from Internet Society: http://www.ietf.org/rfc/rfc3986.txt

Bide, M., Oppenheim, C., & Ramsden, A. (1997). *Copyright Clearance and Digitisation in UK Higher Education: Supporting Study for the JISC/PA Clearance Mechanisms Working Party.* Retrieved 06/03/2011, from UKOLN: http://www.ukoln.ac.uk/services/elib/papers/pa/clearance/

Blue Ribbon Task Force on Sustainable Digital Preservation and Access. (2010). *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information.* Retrieved 18/07/2011 from Blue Ribbon Task Force Website http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

British Library. (2007). *British Library Digital Preservation Strategy.* Retrieved 20/04/2011, from British Library: http://www.bl.uk/aboutus/stratpolprog/ccare/introduction/digital/digpresstrat.pdf

British Library. (2008a). *Digitisation Strategy 2008-2011.* Retrieved 16/04/2011, from British Library: http://www.bl.uk/aboutus/stratpolprog/digi/digitisation/digistrategy/

British Library. (2008b). *Digitisation Strategy 2009 - 2011*. Retrieved 16/04/2011, from British Library: http:www.bl.uk/aboutus/stratpolprog/digi/digitisation/digistrategy/

British Standards Institute. (2003). *Space Data and Information Transfer Systems - Open Archival Information Systems - Reference Model.* British Standards Institute.

Brown, C. (2010). Let's Not be Luddites. *Society of Archivists* (January 2010), pp. 10-11.

Bryman, A. (2001). *Social Research Methods.* Oxford: Oxford University Press.

Buckley, R. (2008). *A Technology Watch report: JPEG 2000 a Practical Digital Preservation? A Tech Watch report.* Retrieved 28/04/2011, from DPC: http://www.dpconline.org/docs/reports/dpctw08-01.pdf

Bulow, A., & Ahmon, J. (2011). *Preparing Collections for Digitzation.* London: Facet.

Burton, G. (2010). *Media and Society: Critical Perspectives* (2nd ed.). Maidenhead: Open University Press.

Cambridge University Press. (2011). *Cambridge Dictionaries Online*. Retrieved 20/04/2011, from Cambridge University Press: http://dictionary.cambridge.org/dictionary/british/image_1

Canadian Ski Instructors Alliance. (2000). *Skiing and Teaching Methods.* Montreal: CSIA.

Caplan, P. (2001). *International Metadata Initiatives: Lessons in Bibliographic Control.* Retrieved 02/2012/16, from Library of Congress: http://www.loc.gov/catdir/bibcontrol/caplan_paper.html

Caplan, P. (2009). *Understanding Premis.* Retrieved 16/05/2011, from Library of Congress: http://www.loc.gov/standards/premis/understanding-premis.pdf

Cardiff University. (n.d.). *Knowledge Economy Skills Scholarships (KESS) at Cardiff University*. Retrieved 02/02/2012, from Cardiff University: http://www.cardiff.ac.uk/racdv/expertise/kess/index.html

Cartwright, J. (2011). *Information for Community Archives: Photographs.* Retrieved 16/03/2011, from Society of American Archivists: http://www2.archivists.org/groups/lesbian-and-gay-archives-roundtable-lagar/information-for-community-archives-photographs

Centre for Performance Research. (2009). CPR Collections' Collection and Appraisal Policy. Retrieved 20/04/2011 from: http://www.aber.ac.uk/en/media/CPR-Collections-Policy.pdf

Centre for Performance Research. (2011). *Centre for Performance Research*. Retrieved 12/12/2011, from Centre for Performance Research: http://www.thecpr.org.uk/index2.php

Chandler, D. (2000). *Notes on 'The Gaze': The Eye of the Camera*. Retrieved 16/02/2012, from Aberystwyth University: http://www.aber.ac.uk/media/Documents/gaze/gaze06.html

Choudhury, S., DiLauro, T., Szalay, A., Vishniac, E., Hanish, R., Steffen, J., et al. (2007, 11). Digital Data Preservation for Scholarly Publications in Astronomy. *The International Journal of Digital Curation , 2*(2). Bath, UK: UKOLN. Retrieved 26/05/2010, from The International Journal of Digital Curation.

Consultative Committee on Space Data Systems. (2002). *Reference Model for an Open Archival Information System.* Washington, DC, USA: CCSDS Secretariat.

Conway, P. (2010). Modes of Seeing: Digitized Photographic Archives and the Experienced User. *The American Archivist , 73* (Fall/Winter), 425 - 461.

Cumming, K. (2005). Metadata Matters. In J. Mcleod, & C. Hare (Eds.), *Managing Electronic Records* (p. 36). London: Facet.

Dappert, A., & & Farquhar, A. (2011). Implementing Metadata that Guide Digital Preservation. Retrieved 01 07, 2011, from *The International Journal of Digital Curation* 6:1: http://www.ijdc.net/index.php/ijdc/article/view/176

Davis-Perkins, V., Butterworth, ,. R., Curzon, ,. P., & Fields, ,. B. (2005, 3652/2005 p.278 - 289). *A Study into the Effect of Digitisation Projects on the Managment and Stability of Historic Photographic Collections.* Retrieved 01/15/2010, from Lecture Notes in Computer Science,: http://www.springerlink.com/content/5y2drmh99fvt3p3q/

Day, M. (2010). *IMPACT Best Practice Guide: Metadata for Text Digitisation & OCR.* Retrieved 19/05/2011, from UKOLN: http://opus.bath.ac.uk/23311/1/IMPACT-metadata-bpg-pilot-1.pdf

DCMI. (2011). *User Guide*. Retrieved 23/02/2012, from DCMI Wiki: http://wiki.dublincore.org/index.php/User_Guide

Deegan, M., & Tanner, S. (2002). *Digital Futures: Stratgies for the Information Age.* London: Facet.

Digicult. (2004). *Resource Discovery Technologies for the Heritage Sector.* Retrieved 08/07/2011, from Digicult: http://www.digicult.info/downloads/digicult_thematic_issue_6_lores.pdf

Digital Curation Centre. (2008a). *DRAMBORA Interactive: Digital Repository Audit Method Based on Risk Assessment.* Retrieved 22/07/2010, from DCC: http://www.repositoryaudit.eu/about/

Digital Curation Centre. (2010a). *History of the DCC*. Retrieved 07/03/2012, from DCC: http://www.dcc.ac.uk/about-us/history-dcc

Digital Curation Centre. (2010b). *Trustworthy Repositories.* Retrieved 17/08/2010, from Digital Curation Centre: http://www.dcc.ac.uk/resources/tools-and-applications/trustworthy-repositories

Digital Curation Centre. (2010c). *What is Digital Curation?* Retrieved 07/03/2012, from DCC: http://www.dcc.ac.uk/digital-curation/what-digital-curation

Digital Curation Centre. (2010d). *What is the Digital Curation Centre?* Retrieved 21/07/2010, from DCC: http://www.dcc.ac.uk/

Digital Curation Centre. (2011a). *All Standards for Any Lifecycle Action*. Retrieved 28/4/2011, from DCC: http://www.dcc.ac.uk/resources/standards/diffuse/standards?framework_id=0&lifecycle_id=0&sort=type#File Formats

Digital Curation Centre. (2011b). *Checklist for a Data Management Plan (v3.0, 17 March 2011).* Retrieved 21/03/2011, from DCC: http://dmponline.dcc.ac.uk/system/attachments/8/original/DCC_Checklist_DMP_v3_md_sj.pdf?1300724157

Digital Curation Centre. (2011c). *How can the Digital Curation Centre help You?* Retrieved 20/04/2011, from DCC: http://www.dcc.ac.uk/

Digital Curation Centre. (2011d). *OpenDocument (ODF)*. Retrieved 29/04/2011, from DCC: http://www.dcc.ac.uk/resources/standards/diffuse/show?standard_id=111&sort=type

Digital Library Federation. (2001). *<METS> Metadata Encoding and Transmission Standard: Primer and Reference Manual.* Retrieved 22/07/2010, from Library of Congress: http://www.loc.gov/standards/mets/METSPrimerRevised.pdf

Digital Preservation Coalition. (2009a). *About the DPC*. Retrieved 07/03/2012, from Digital Preservation Coalition: http://www.dpconline.org/about

Digital Preservation Coalition. (2009b). *Home*. Retrieved 20/04/2011, from Digital Preservation Coalition: http://www.dpconline.org/

Digital Preservation Coalition. (2009c). *Introduction - Definitions and Concepts.* Retrieved 16/04/2011, from Digital Preservation Coalition:

http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts

Dunning, A. (2006). *Copyright and other Rights Issues in Digitisation*. Retrieved 06/03/2011, from Arts and Humanities Data Service: http://www.ahds.ac.uk/creating/information-papers/copyright-introduction/

ECLAP. (2011). *Eclap - E-library for Performing Arts*. Retrieved 05/05/2011, from ECLAP Connecting Stages: Best Practice Network: http://bpnet.eclap.eu/drupal/?q=en-US/home

Elkins, J. (2011). *What Photography is.* New York: Oxford: Routledge.

Europeana. (n.d.). *About Us*. Retrieved 10/03/2012, from Europeana: http://www.europeana.eu/portal/aboutus.html

Everitt, S. (2005). Opportunistic collaboration: Unlocking the archives of the Birmingham Institute of Art and Design. *Program: Electronic Library and Information Systems , 39* (4), 312 - 323.

Farquhar, A., & Hockx-Yu, H. (2007). Planets: Integrated Services for Digital Preservation. Retrieved 6/10/2010, from *The International Journal of Digital Curation* 2:2 : http://www.ijdc.net/index.php/ijdc/article/viewFile/45/31

Feather, J. (2008). *The Information Society: a Study of Continuity and Change* (5th ed.). London: Facet.

Field, C. D. (n.d.). *The Internet Library of Early Journals: Preserving and Disseminating the British Journal Heritage.* Retrieved 16/03/2012, from University of Oxford: Bodleian Library: http://www.bodley.ox.ac.uk/ilej/papers/paper02.htm

Flinn, A. (2010, 62). *An Attack on Professionalism and Scholarship?: Democratising Archives and the Production of Knowledge.* Retrieved 20/01/2011, from *Ariadne* (62): http://www.ariadne.ac.uk/issue62/flinn/

Giles, D. (2010). *Psychology of the Media.* Basingstoke: New York: Palgrave Macmillan.

Glynn, R. (1996). *HELIX (Higher Education Library for Image eXchange).* Retrieved 03/08/2011, from Ariadne (5): http:www.ariadne.ac.uk/issue5/helix/

Google Books. (2011). *About Google Books*. Retrieved 12/03/2012, from Google Books: http://books.google.com/intl/en/googlebooks/history.html

Gough, R. (2009). *Out of the Box and Dusted Down: Foraging and Findings.* Retrieved 29/02/2012, from CADAIR: http://cadair.aber.ac.uk/dspace/bitstream/handle/2160/3288/Gough,%20R.pdf?sequence=1

Graham, N. (2010). *Prince2 For Dummies* (2nd ed.). Chichester: John Wiley & Sons, Ltd.

Greenstone. (2007). *General Information*. Retrieved 05/05/2011, from Greenstone: http://wiki.greenstone.org/wiki/index.php/General_information

Guenther, R. S. (2008). *Battle of the Buzzwords: Flexibility vs. Interoperability When Implementing PREMIS in METS.* Retrieved 16/05/2011, from D-Lib Magazine 14:7/8: http://www.dlib.org/dlib/july08/guenther/07guenther.html

Hampson, A. (2001). Practical experiences of digitisation in the BUILDER Hybrid Library Project. *Program: Electronic Library and Information Systems ,35* (3) pp.263-275.

Haynes, D. (2004). *Metadata for Information Management and Retrieval.* London: Facet.

Health and Safety Executive. (2012). *Display Screen Equipment (DSE)*. Retrieved 25/02/2012, from Health and Safety Executive: http://www.hse.gov.uk/msd/dse/

Hedstrom, M. (1997). Digital Preservation: a Time Bomb for Digital Libraries. *Computers and Humanities , 31* (3), pp.189 - 202. Retrieved 07/07/2010, from JSTOR: http://www.jstor.org/stable/30200423

Higgins, S. (2008). The DCC Curation Lifecycle Model. Retrieved 26/01/2011, from *International Journal of Digital Curation* 3:1: http://www.ijdc.net/index.php/ijdc/article/view/69/48

Higgins, S. (2011). Digital Curation: the Emergence of a New Discipline. Retrieved 03/02/2012, from *The International Journal of Digital Curation* 6:2: http://www.ijdc.net/index.php/ijdc/article/view/184

Howe, D. (Ed.). (2010). *FOLDOC: Free Online Dictionary of Computing*. Retrieved 20/04/2011, from http://foldoc.org/

Hughes, L. (2003). *Digitizing Collections: Strategic Issues for the Information Manager.* London: Facet.

Hughes, L. (2011a). Digital Collections: Use, Value and Impact. *MS Power Point Presentation* . Presented at as part of module programme at Aberystwyth University.

Hughes, L. (2011b). *Evaluating and Measuring the Value, Use and Impact of Digital Collections.* (L. Hughes, Ed.) London: Facet.

Hughes, L., Knight, G., Ell, P., Yeates, E., & Dobreva, M. (2011). *Stormont Parliamentary Hansards: Embedded in Research and Education*. Retrieved 28/10/2011, from JISC: http://www.jisc.ac.uk/media/documents/programmes/digitisation/impactembedding/SPHERE%20Impact%20analysis%20summary-final.pdf

Ibrahim, K. (2007). *Newnes Guide to Television and Video Technology* (4th ed.). Oxford:Burlington: Newnes.

ICA Atom. (2011). *ICA Atom: Open Source Archival Software*. Retrieved 12/08/2011, from ICA Atom: http://ica-atom.org/

Imageaccess. (nd). *Bookeye3: version R1:set up and assembly manual.* Retrieved 28/03/2011, from Image Access: http://www.imageaccess.com/PDFs/BE3-R1_SetupAndAssembly.pdf

Information Commissioners Office. (n.d.). *History of the ICO*. Retrieved 12/03/2012, from ICO.: http://www.ico.gov.uk/about_us/our_organisation/history.aspx

International Council on Archives. (2009). *About ICA*. Retrieved 12/03/2012, from ICA.org: http://www.ica.org/102/about-ica/an-introduction-to-our-organization.html

International Council on Archives Ad Hoc Commission on Descriptive Standards. (2010). *ISAD(G): General International Standard Archival Description - Second Edition.* Retrieved 26/07/2011, from International Council on Archives: http://www.ica.org/7102/public-resources/isadg-general-international-standard-archival-description-second-edition.html

International Organization for Standardization. (2011a). *About ISO*. Retrieved 10/03/2012, from International Organization for Standardization: http://www.iso.org/iso/about.htm

International Organization for Standardization. (2011b). *ISO14721:2003*. Retrieved 10/03/2011, from ISO: http://www.iso.org/iso/catalogue_detail.htm?csnumber=24683

InterPARES. (2010). *InterPARES Project Overview.* Retrieved 21/07/2010, from InterPARES Project: International Research on Permanent Authentic Records in Electronic Systems: http://www.interpares.org/welcome.cfm

InterPARES. (2011). *The InterPARES 2 Project Glossary.* Retrieved 25/0/ 2011, from InterPARES 2: http://www.interpares.org/ip2/display_file.cfm?doc=ip2_glossary.pdf&CFID=1772518&CFTOKEN=61016539

JISC. (2007). *Case Study: University of Bristol Theatre Collection.* Retrieved 13/03/2011, from JISC Digital Media: http://www.jiscdigitalmedia.ac.uk/stillimages/advice/case-study-university-of-bristol-theatre-collection

JISC. (2011c). *Who We Are*. Retrieved 04/08/2011, from JISC: http://www.jisc.ac.uk/aboutus/whoweare.aspx

JISC. (2012a). *About JISC*. Retrieved 03/07/2012, from JISC: http://www.jisc.ac.uk/aboutus.aspx

JISC. (2012b). *Digitisation and Content*. Retrieved 07/03/2012, from JISC: http://www.jisc.ac.uk/whatwedo/programmes/digitisation.aspx

JISC. (n.d.). *JISC Beginners Guide to Digital Preservation*. Retrieved 16/02/2012, from UKOLN: http://blogs.ukoln.ac.uk/jisc-beg-dig-pres/content/what-is-digital-preservation/definition-of-digital-object/

JISC. (n.d.). *The JISC Digitisation Programme*. Retrieved from JISC.

JISC Digital Media. (2004). *Case Study: Staffordshire Past Track.* Retrieved 04/08/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/stillimages/advice/case-study-staffordshire-past-track

JISC Digital Media. (2005a). *Case Study: History of Art, University College Cork.* Retrieved 04/08/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/stillimages/advice/case-study-history-of-art-university-college-cork

JISC Digital Media. (2005b). *Setting Up a Workspace For Digitising Images.* Retrieved 22/04/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/stillimages/advice/setting-up-a-workspace-for-digitising-images#ws8

JISC Digital Media. (2006a). *Basic Guidelines for Image Capture and Optimisation*. Retrieved 08/08/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/stillimages/advice/basic-guidelines-for-image-capture-and-optimisation

JISC Digital Media. (2006b). *Choosing a File Format for Digital Still Images.* Retrieved 04/08/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/stillimages/advice/choosing-a-file-format-for-digital-still-images

JISC Digital Media. (2006c). *Converting your Slides into Digital Images*. Retrieved 18/05/2011 from JISC: http://www.jiscdigitalmedia.ac.uk/stillimages/advice/converting-your-slides-into-digital-images

JISC Digital Media. (2006e). *Still Images: Converting your Slides into Digital Images.* Retrieved 18/05/2011, from JISC: http:www.jiscdigitalmedia.ac.uk/stillimages/advice/converting-your-slides-into-digital-images

JISC Digital Media. (2007a). *Case Study: University of Bristol Theatre Collection.* Retrieved 13/03/2011, from JISC Digital Media: http://www.jiscdigitalmedia.ac.uk/stillimages/advice/case-study-university-of-bristol-theatre-collection

JISC Digital Media. (2007b). *Choosing a System for Managing Your Image Collection.* Retrieved 29/06/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/stillimages/advice/choosing-a-system-for-managing-your-image-collection

JISC Digital Media. (2008a). *Digitisation Staff Training.* Retrieved 19/03/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/digitisation-staff-training

JISC Digital Media. (2008b). *Quality Assurance and Digitisation Projects.* Retrieved 19/03/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/qulaity-assurance-and-digitisation-projects

JISC Digital Media. (2008c). *Sustainability of Digital Collections.* Retrieved 19/03/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/sustainability-of-digital-collections

JISC Digital Media. (2008d). *To Outsource or to Digitise In-house?* Retrieved 19/03/2011, from JISC: http://jiscdigitalmedia.ac.uk/crossmedia/advice/to-outsource-or-to-digitise-in-house

JISC Digital Media. (2008e). *The Digital SLR Camera for Teaching, Learning and Digitisation.* Retrieved 03/08/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/stillimages/advice/the-digital-slr-camera-for-teaching-learning-and-digitisation

JISC Digital Media. (2009a). *Budgeting for a Digitisation Project*. Retrieved 29/04/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/budgeting-for-a-digitisation-project

JISC Digital Media. (2009b). *Establishing a Digital Preservation Policy.* Retrieved 23/04/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/establishing-a-digital-preservation-policy

JISC Digital Media. (2009c). *Sources of Funding for Digitisation Projects*. Retrieved 29/04/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/potential-sources-of-funding-for-digitisation-projects/

JISC Digital Media. (2009d). *Systems for Managing Digital Media Collections.* Retrieved 19/03/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/systems-for-managing-digital-media-collections

JISC Digital Media. (2010a). *Image Management Software.* Retrieved 29/06/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/stillimages/advice/image-management-software

JISC Digital Media. (2010b). *Tethered Camera Operation.* Retrieved 03/08/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/stillimages/advice/tethered-camera-operation

JISC Digital Media. (2010c). *Uncompressed Audio File Formats*. Retrieved 22/04/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/audio/advice/uncompressed-audio-file-formats

JISC Digital Media. (2010d). *User Guide to MP3*. Retrieved 29/04/2011, from JISC: http://www.jiscdigitalmedia.ac.uk/audio/advice/user-guide-to-mp3

JISC Digital Media. (2011). *Still Images: Managing a Project.* Retrieved 06/03/2011, from JISC: http://ww.jiscdigitalmedia.ac.uk/stillimages/docs/category/managing-a-project/

JISC Digital Media. (2012). *Still Images, Moving Images and Sound Advice*. Retrieved 12/03/2012, from JISC Digital Media: http://www.jiscdigitalmedia.ac.uk/

Jones, R. A. (2008). A Marathon Not a Sprint: Lessons Learnt from the First Decade of Digitisation at the National Library of Wales. *Program: Electronic Libraries and Information Systems , 42* (2), pp.97 - 114.

Jones, S. (2010). *Preservation Policy Template for Repositories.* Retrieved 01/25/2011, from DCC: http://www.dcc.ac.uk/sites/default/files/Preservation%20policy%20template.pdf

Kirriemuir, J. (1998). *A Brief History to, and Background of, the eLib Programme Background.* Retrieved 12/03/2012, from UKOLN: http://www.ukoln.ac.uk/services/elib/background/history.html

Klein, Y. (2009). *East London Theatre Archive.* Retrieved 13/03/2011, from JISC: http://www.jisc.ac.uk/media/documents/programmes/digitisation/eltafinal.pdf

Knowles, J. (2010). *Collaboration Nation: the Building of the Welsh Repository Network.* Retrieved 05/07/2011, from CADAIR: http://cadair.aber.ac.uk/dspace/handle/2160/4636?show=full

Koenen, R. (Ed.). (2002). *Overview of the MPEG-4 Standard*. Retrieved 04/29/2011, from International Organisation for Standardisation: http://mpeg.chiariglione.org/standards/mpeg-4/mpeg-4.htm

Lagoze, C. (2001). Keeping Dublin Core Simple: Cross-Domain Discovery or Resource Description*?* Retrieved 18/07/2011, from *D-Lib Magazine* 7:1: http://www.dlib.org/dlib/january01/lagoze/01lagoze.html

Langford, M., & Bilissi, E. (2011). *Langfords Advanced Photography: The Guide for Aspiring Photographers* (8th ed.). Oxford: Burlington: Focal Press.

Leslie, F. (2004). Bringing collections to Life, Digitising Local Studies and Special Collections: the COLLAGE Project. *22:* (3), pp.261-263. Retrieved 03/04/2011, from *The Electronic Library*: http://www.emeraldinsight.com/journals.htm?articleid=862060&show=abstract

Lewis, S. (2008). *The DSpace Course – Look & Feel Customization.* Retrieved 11/07/2011, from DSpace: http://cadair.aber.ac.uk/dspace/handle/2160/640?show=full

Library of Congress. (2009). *What is a MARC Record and Why is it Important?* Retrieved 29/11/2011, from Library of Congress: http://www.loc.gov/marc/umb/um01to06.html

Library of Congress. (2010). *Metadata Encoding and Transmission Standard: Primer and Reference Manual.* Retrieved 16/06/2010, from Library of Congress: http://www.loc.gov/standards/mets/METSPrimerRevised.pdf

Library of Congress. (2011a). *Collections Care: Caring for your Photographic Collection.* Retrieved 20/04/2011, from Library of Congress: http://www.loc.gov/preservation/care/photo.html

Library of Congress. (2011a). *Authorities and Vocabularies*. Retrieved 24/07/2011, from Library of Congress: http://id.loc.gov/search/?q=world+war+one&cs=All

Library of Congress. (2011b). *MADS: Metadata Authority Description Schema*. Retrieved 02/02/2012, from Library of Congress: http://www.loc.gov/standards/mads/

Library of Congress. (2011c). *MARC21 Format for Bibliographic Data*. Retrieved 12/03/2012, from Library of Congress: http://www.loc.gov/marc/bibliographic/

Library of Congress. (n.d.). *PREMIS:Preservation Metadata Maintennance Activity*. Retrieved 16/06/2010, from Library of Congress: http://www.loc.gov/standards/premis/

LOCKSS. (2008). *Home*. Retrieved 02/07/2011, from LOCKSS: http://www.lockss.org/lockss/Home

McCullagh, C. (2002). *Media Power: A Sociological Introduction.* Basingstoke: New York: Palgrave.

McLuhan, M., & Fiore, Q. (1967). *The Medium is the Massage.* London: New York: Toronto: Dublin: Victoria: New Delhi: North Shore: Johannesburg: Penguin.

Merriam-Webster Inc. (2011). *Merriam-Webster.com*. Retrieved 20/04/2011, from Merriam-Webster Online: http://www.merriam-webster.com/dictionary/image

Meyer, E. T. (2011). *Splashes and Ripples: Synthesizing the Evidence on the Impacts of Digital Resources*. Retrieved 28/10/2011, from JISC: http://www.jisc.ac.uk/media/documents/programmes/digitisation/Impact_Synthesis%20report_FINAL.pdf

Meyer, E. T., & Eccles, K. (2010). *Toolkit for the Impact of Digitised Scholarly Resources:Usage and Impact of JISC Phase I Digitisation projects*. Retrieved 28/10/2011, from JISC: http://www.jisc.ac.uk/media/documents/events/2009/06/eric_meyer_2d.pdf

Michelson, A., & Rothenberg, J. (1992). Scholarly Communication and Information Technology: Exploring the Impact of Changes in the Research Process on Archives. *The American Archivist , 55:2*, pp.236 - 315. Society of American Archivists.

Millerson, G. (1972). *The Technique of Lighting for Television and Motion Pictures.* London: New York: Focal Press.

Mitchell, W. (1984). What Is an Image? Retrieved 04/20/2011, from *New Literary History* via JSTOR: http://links.jstor.org/sici?sici=0028-6087%2189 8421%2915%3A3%3C503%3AWIAI%3E2.O.C0%3B2-S

Moles, S. (2010). Education: a Bridge Not Far Enough. Retrieved 04/02/2011, from *Lighting and Sound International*: http://edition.pagesuite-professional.co.uk/pdfspool/Vw0OtdZRLoGP49vIpTV2PS1cBI=.pdf

Morrisey, S. (2010). The Economy of Free and Open Source Software in the Preservation of Digital Artefacts. *Library Hi Tech , 28* (2), pp.211-223.

National Information Standards Organization. (2004). *Understanding Metadata.* Retrieved 02/02/2011, from NISO: http://www.niso.org/publications/press/UnderstandingMetadata.pdf

National Library of Scotland. (2011). *National Library of Scotland Strategy 2011 - 2014.* Retrieved 23/06/2011, from National Library of Scotland: http://www.nls.uk/about-us/corporate-documents/2011-2014-strategy

National Library of Wales. (2009). *Digitisation Strategy 2008 - 2011.* Retrieved 16/03/2011, from National Library of Wales: http://www.llgc.org.uk/fileadmin/documents/pdf/2009_Digi_Strat.pdf

National Library of Wales. (2011a). *About NLW*. Retrieved 24/01/2012, from National Library of Wales: http://www.llgc.org.uk/index.php?id=aboutnlw

National Library of Wales. (2011b). *Digitisation Projects*. Retrieved 24/01/2012, from National Library of Wales: http://www.llgc.org.uk/index.php?id=4723

National Library of Wales. (2011c). *The Agile Library: The Library's Strategy 2011 - 2012 to 2013 - 2014.* Retrieved 21/06/2011, from National Library of Wales: http://www.llgc.org.uk/fileadmin/documents/pdf/Strategy2011-12_2013-14.pdf

National Screen and Sound Archive Wales. (2011). *About*. Retrieved 24/01/2012, from NSSAW: http://www.archif.com/index.php?id=4180

NDIIPP. (2011). *PDF/A-1, PDF for Long-term Preservation, Use of PDF 1.4*. Retrieved 29/04/2001, from Sustainability of Digital Formats Planning for Library of Congress Collections: http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml

Newton, C. (2008). *Integrated Collecting Strategy 2008.* Retrieved 09/08/2011, from National Library of Scotland: http://www.nls.uk/media/22389/2008-collecting-strategy.pdf

NOF-Digitise. (2009). *Home: Digitise Launch.* Retrieved 12/03/2012, from NOF-Digitise: http://www.nof-digitise.org/launch.htm

Nurnberg, W. (1969). *LIghting for Portraiture* (8th ed.). London, Focal Press.

OCLC. (2012). *Dewey Decimal Classification Summaries.* Retrieved 12/02/2012, from OCLC: http://www.oclc.org/dewey/resources/summaries/default.htm

Open Source Initiative. (n.d.). *Mission*. Retrieved 07/03/2012, from Opensource.org: http://www.opensource.org/

Padfield, T. (2010). *Copyright for Archivists and Records Managers* (4th ed.). London: Facet.

Pavis, P. (1998). *Dictionary of the Theatre: Terms, Concepts, Analysis.* Toronto: University Press Toronto.

Pearce-Moses, R. (2005a). *A Glossary of Archival and Records Terminology.* Retrieved 12/03/2012, from Society of American Archivists: http://www.archivists.org/glossary/term_details.asp?DefinitionKey=2713

Pearce-Moses, R. (2005b). *A Glossary of Archival and Records Terminology.* Retrieved 12/03/2012, from Society of American Archivists: http://www.archivists.org/glossary/term_details.asp?DefinitionKey=6

Pearson, D., & Webb, C. (2008). Defining File Format Obsolescence: A Risky Journey. Retrieved 22/01/2011, from *The international Journal of Digital Curation* 3:1: http://www.ijdc.net/index.php/ijdc/issue/view/4

PLANETS. (2007). *Home*. Retrieved 16/06/2010, from Planets: http://www.planets-project.eu/

PLANETS. (2011). *PLATO*. Retrieved 27/07/2011, from Planets: http://www.ifs.tuwien.ac.at/dp/plato/intro.html

PREMIS Editorial Committee. (2008). *PREMIS Data Dictionary for Preservation Metadata: version 2.* Retrieved 22/07/2010, from Library of Congress: http://www.loc.gov/standards/premis/v2/premis-2-0.pdf

Rayner, S. (2009). *Furer-Haimendorf Archive Digitisation.* Retrieved 15/02/2011, from JISC: http://www.jisc.ac.uk/media/documents/programmes/digitisation/edrsoasfinalreport.pdf

Reid, G. (2000). The Digitisation Of Heritage Material: Arguing for an Interpretive Approach Based on the Experience of the Powys Digital History Project. *Program: Electronic Library and Information Systems , 34*:2.

Repositories Support Project. (2007a). *Key Stakeholders & Benefits.* Retrieved 05/07/2011, from Repositories Support Project: http://www.rsp.ac.uk/start/setting-up-a-repository/

Repositories Support Project. (2007b). *Preservation & Storage Formats for Repositories.* Retrieved 05/07/2011, from Repositories Support Project: http://www.rsp.ac.uk/start/setting-up-a-repository/

Repositories Suppport Project. (2008a). *Metadata.* Retrieved 05/07/2011, from Repositories Suppport Project: http://www.rsp.ac.uk/start/setting-up-a-repository/

Repositories Support Project. (2008b). *Pilot to Production.* Retrieved 05/07/2011, from Repositories Support Project: http://www.rsp.ac.uk/start/setting-up-a-repository/

Repositories Support Project. (2008c). *Specifiying Repository Requirements.* Retrieved 05/07/2011, from Repositories Support Project: http://www.rsp.ac.uk/start/setting-up-a-repository/

Repositories Support Project. (2008d). *Workflows.* Retrieved 05/07/2011, from Repositories Support Project: http://www.rsp.ac.uk/start/setting-up-a-repository/

Repositories Support Project. (n.d.a). *Technical Maintenance.* Retrieved 12/03/2012, from Repositories Support Project: http://www.rsp.ac.uk/start/resourcing-for-sustainability/technical-maintenance/

Repositories Support Project. (n.d.b). *What is a Repository?* Retrieved 15/06/2011, from Repositories Support Project: http://www.rsp.ac.uk/start/before-you-start/what-is-a-repository/

Ross, S. (2000). *Changing Trains at Wigan: Digital Preservation and the Future of Scholarship.* London: UK: National Preservation Office.

Rothenberg, J. (1999). *Ensuring the Longevity of Digital Information.* Retrieved 22/07/2010, from Jeff Rothenberg: http://www.clir.org/pubs/archives/ensuring.pdf

Rowley, J., & Hartley, R. (2008). *Organising Knowledge: An Introduction to Managing Access to Information. 4th Ed.* Surrey: Burlington, VT: Ashgate Publishing Limited.

Rusbridge, C. (1998). *Towards the Hybrid Library.* Retrieved 12/03/2012, from *D-Lib Magazine* July/August 1998: http://dlib.ukoln.ac.uk/dlib/july98/rusbridge/07rusbridge.html

Sadashige, K. (2000). *Data Storage Technology Assessment: Storage Media Environmental Durability and Stability.* Retrieved 22/02/2010, from DPC: http://www.dpconline.org/component/finder/search?q=storage

Saunders, C. (2005). Digital imaging South Africa (DISA): a case study. *Program: Electronic and Library Information Systems ,39*(4), 345 - 352.

Schonefeld, R. (2005, p. 337 - 344 39:4). *JSTOR: a case study in the recent history of scholarly communications.* Retrieved 13/06,/011, from Emerald: http://www.emeraldinsight.com/journals.htm?articleid=1528717&show=abstract

Schrier, R.A.(201*1).* Digital Librarianship & Social Media: the Digital Library as Conversation Facilitator. (n.d.). Retrieved from *D-Lib*. *Magazine 17:*7/8 from http://www.dlib.org/dlib/july11/schrier/07schrier.html

Secker, J. (2010). *Copyright and E-learning: a Guide for Practitioners.* London: Facet.

Seiko Epson Corporation. (2012). *Epson Perfection V750 Pro*. Retrieved 25/09/2011, from Seiko Epson Corporation: http://www.epson.co.uk/Scanners/Epson-Perfection-V750-Pro/Tech-Specs

Sheddon, J. (1993). *Skiing: Devleoping Your Skill.* Marlborough: Crowood Press.

Sims, J. (2008). Librarianship in the 21st Century: a British Library Perspective. *Legal Information Management* (8), pp.84-91.

Sitts, M. (Ed.). (2000). *Handbook for Digital Projects: a Management Tool for Preservation and Access.* Retrieved 16/03/2011, from Northeast Document Conservation Center, Andover, Massachusetts,: http://www.nedcc.org/resources/digitalhandbook/dman.pdf

Skarstein, V. (2010a). A National Strategy for the Digital Preservation and Dissemination of our Cultural Heritage. Retrieved 01/25/2011, from

Skandinavian Public Library Quarterly: http://www.splq.info/issues/vol43_2/05.htm

Skarstein, V. (2010b). The Bookshelf: Digitisation and Access to Copyright Items in Norway*.* Retrieved 02/02/2011, from *Program:Electronic Library and Information Systems* 44:1 pp.48-58

Spence, J. (2005). Small Organisations and Cultural Insititutions - a Digital Future? Retrieved 18/03/2011, from *Program:Electronic Library and Information Systems* 39:4 pp.366-380

Staniforth, D. A. (2010). *International Theatre Collection Seminar—Introduction.* Retrieved 29/02/2012, from CADAIR: http://cadair.aber.ac.uk/dspace/bitstream/handle/2160/4160/AS_intro.pdf?sequence=1

Stouffer Industries Inc. (2006). *Reflection Step Tablets*. Retrieved 23/02/2012, from Stouffer Industries Inc.: http://www.stouffer.net/Reflection.htm

The Archives Hub. (2011). *Cataloguing Standards*. Retrieved 01/08/2011, from The Archives Hub.

The National Archives. (2011a). *Archives for the 21st Century in Action.* Retrieved 22/01/2011, from TNA: http://www.nationalarchives.gov.uk/documents/information-management/archives-for-the-21st-century-in-action.pdf

The National Archives. (2011b). *Droid*. Retrieved 07/12/2011, from Source Forge: http://sourceforge.net/projects/droid/

The National Archives. (n.d.a). *Caring for your Photographic Collection.* Retrieved 16/03/2011, from TNA: http://www.nationalarchives.gov.uk/documents/archivesconservation_photo.pdf

The National Archives. (n.d.b). *Pronom.* Retrieved 22/07/2010, from The National Archives: http://www.nationalarchives.gov.uk/PRONOM/Default.aspx#

The National Archives. (n.d.c). *Register of National Archives*. Retrieved 11/03/2011, from TNA: http://www.nationalarchives.gov.uk/nra/help/nra/nrafaq.htm#5

Thompson, D. (2010). A Pragmatic Approach to Preferred File Formats for Acquisition. Retrieved 22/01/2011, from *Ariadne* 63: http://www.ariadne.ac.uk/issue63/thompson/

Treolar, A., & Wilkinson, R. (2008). *Access to Data for eResearch: Designing the Australian National Data Service Discovery Services.* Retrieved 26/05/2010, from International Journal of Digital Curation: http://www.ijdc.net/index.php/ijdc/article/viewFile/95/66

Triantaphillidou, S., & Jacobson, R. E. (2001). *A Simple Approach to Digitising a Photographic Collection.* Retrieved 03/04/2011, from Eprints; University of Westminster: http://eprints.wmin.ac.uk/8951/1/pics38.pdf

UKOLN. (1996a). *DIAD: Digitsation in Art and Design.* Retrieved 12/03/2012, from UKOLN: http://www.ukoln.ac.uk/services/elib/projects/diad/

UKOLN. (1996b). *Internet Library of Early Journals.* Retrieved 12/03/2012, from UKOLN: http://www.ukoln.ac.uk/services/elib/projects/early/

UKOLN. (2010). *Mission Statement*. Retrieved 07/03/2012, from UKOLN: http://www.ukoln.ac.uk/ukoln/mission/

UKOLN. (2011). *About UKOLN*. Retrieved 07/03/2012, from UKOLN: http://www.ukoln.ac.uk/about/

W3C. (2010). *XML Essentials.* Retrieved 19/09/2010, from W3C: http://www.w3.org/standards/xml/core

W3C. (2012). *RDF.* Retrieved 15/02/2012, from W3C: http://www.w3.org/RDF/

W3C Schools.com. (2011). *XML Tutorial.* Retrieved 06/06/2011, from W3C Schools.com: http://www.w3schools.com/xml/default.asp

W3C Schools. (2012a). *RDF Tutorial.* Retrieved 15/02/2012, from W3C Schools: http://www.w3schools.com/rdf/rdf_intro.asp

W3C Schools. (2012c). *The Semantic Web*. Retrieved 29/02/2012, from W3C Schools: http://www.w3schools.com/web/web_semantic.asp

Welsh Repository Network. (2009). *WRN: Repository Hardware Case Studies*. Retrieved 05/07/2011, from CADAIR: http://cadair.aber.ac.uk/dspace/handle/2160/1881

Western States Digital Standards Group: Digital Imaging Working Group. (2003). *Western States Digital Imaging Best Practices: Version 1.0.* Retrieved 12/03/2011, from Western States Digital Standards Group: http://www.mndigital.org/digitizing/standards/imaging.pdf

Wheatley, P. (2011). *AQuA The Leeds Mashup*. Retrieved 29/04/2011, from AQuA: http://wiki.opf-labs.org/display/AQuA/AQuA+-+The+Leeds+Mashup

White, A. (2005). The Development of Digital Resources by Library and Information Professionals and Historians: Two Case Studies from Northern Ireland. *Program: Electronic Library and Information Systems , 19:* (4), pp. 324 - 336.

Wood-Fisher, C. (2010). *What are the Issues Surrounding the Preservation of Dynamic Databases?* Aberystwyth University, Aberystwyth, Wales : Unpublished Masters Dissertation.

Wood-Fisher, C., Gough, R., Higgins, S., Morgan, M., Staniforth, A. and Tedd, L., (2011). *The Transfer of Knowledge from a Large Organizations' to Small: Experiences from a Research Project on Digitization in Wales.* In the Proceedings of the International Conference on Integrated Information 2011, Advances on Processing and Management, Vol 1, pp. 289-292.

Woodyard-Robinson, D. (2007). *Implementing the PREMIS Data Dictionary: a Survey of Approaches.* Retrieved 16/05/2011, from Library of Congress: http://www.loc.gov/standards/premis/implementation-report-woodyard.pdf

WRN Staff. (2009). *WRN Repository Hardware Case Studies.* Retrieved 05/07/2011, from Cadair: http://cadair.aber.ac.uk/dspace/handle/2160/1881

Yeates, R., & Guy, D. (2006). Collaborative Working for Large Digitisation Projects. *Program: Electronic Library and Information Systems , 40* (2), 137-156.

Zeng, M., & Qin, J. (2008). *Metadata.* London: Facet.

APPENDICES

These appendices are arranged in the order in which the topics they address are first mentioned in the thesis.

Appendix A: **Technical Background**

Appendix A: Technical Background explains some of the technical terms and theory that come together in digitisation and is an extension of the technical topics covered by the literature review. It also explains why knowledge of this term is important to be able to successfully complete a digitisation project. The terms are grouped so that the terms that describe the appearance of a digital object are together. This table has also been reproduced in the Generic and Organiser's Guide.

Term	**What it Means/What it is**	**Why it is Important for Digitisation**
ppi	Pixels per inch	Imagine a pixel is a small square. The number of pixels per inch tells you how many of these squares will fit into one inch on your monitor or scanner resolution scale. Each pixel can be one colour (or shade of grey). Therefore the more pixels you have, the more detail you can have (Langford & Bilissi, 2011, p. 463).
dpi	Dots per inch	Imagine a dot on a page. The quoted number of dots per inch tells you how many of these will appear per inch if you *print* the digital object (Sitts, 2000, p. 95; Western States Digital Standards Group: Digital Imaging Working Group, 2003, p. 4)
CMYK	Cyan, magenta, yellow, key (black) – the primary colours used for printing (Sitts, 2000, p. 101)	When viewing a digital object on a monitor, if the monitor has been set up to view the digital object as though it has been printed, it will appear different when it is printed because monitors cannot display the colours in quite the same way as they will be the printer has put them onto paper. If the monitor is set up to view just red, green and blue then this is fine for viewing a digital object on screens, but, again, it will appear different when printed (JISC Digital Media, 2010c, pp. 4-5; Millerson, 1972, p. 41).
RGB	Red, green, blue – the primary colours used by monitors	

Term	What it Means/What it is	Why it is Important for Digitisation
Subtractive colour mixing	Uses CMYK, generally the palette used in printing.	This method subtracts unwanted colour from white by adding cyan, magenta or yellow and is used in printing (Langford & Bilissi, 2011, p. 468)
Additive mixing	Mixing colour to make another by adding to that already present	This method of mixing red, green and blue colours is used by monitors (Langford & Bilissi, 2011, p. 445)
Tone	The tone of a digital object is where on a range between pure black and pure white the contents of a digital object lies. This range of values is often displayed using a type of graph called a histogram (Western States Digital Standards Group: Digital Imaging Working Group, 2003, p. 35). Ideally it should have a smooth appearance, like a histogram, without any spikes.	In digitisation projects spikes on a histogram usually occur when the black and white points in a scan are not set to true black or white for the digital object being created (Langford & Bilissi, 2011, pp. 116-118). The expected smooth shape is an ideal, and it may be that a series of spikes on the graph are the best that can be achieved.
Curves	Curves are a graphical representation of the density of the digital object against the range from black to white in the digital object (Millerson, 1972, p. 331).	They are a drawing of the range of brightness and contrast values over the whole digital object. Their distinctive shape can be used in a digitisation project to achieve a more balanced digital object (Langford & Bilissi, 2011, p. 260).
.TIFF	Tagged Image File Format is a lossless, uncompressed file format (Langford & Bilissi, 2011, p. 223).	Lossless means that all the data saved in a digital object is retrieved when it is re-opened. Uncompressed means that the original file data is not altered when it is saved. This can result in a large file if there is a lot of data in a digital object (Western States Digital Standards Group: Digital Imaging Working Group, 2003, p. 28).

Term	What it Means/What it is	Why it is Important for Digitisation
Scanner resolution	The number of pixels per inch the scanner sensors can copy	The higher the number of pixels a scanner can copy, the finer the quality of the digital object (Langford & Bilissi, 2011, p. 178). This can never be improved upon, whereas you can print it with fewer or more dots per inch. The higher the resolution the larger the file size of the digital object. The accepted compromise resolution is 300ppi (JISC Digital Media, 2006a).
Monitor resolution	The number of pixels the monitor can display (Ibrahim, 2007, p. 145)	As in scanners, the higher the number of pixels the better the quality of the digital object on the screen.
Colour temperature	The temperature at which a black body radiator emits the same wavelength of light as the element in the lamp (Bermingham, 1989, p. 11; Langford & Bilissi, 2011, p. 451).	This is important because it explains why 'white' is different for different objects and lamp elements. Lamps on scanners take time to heat up to their steady temperature and therefore to maintain a steady output of light energy at a given temperature.
Comparison strip	A strip consisting of a series of (often eight) colours. These are either from white to black through a series of greys, or white to black through a series of colours (Stouffer Industries Inc., 2006)	These strips show how the primary colours in the digital object appear on your screen. This means that the colours in the digital object can be compared to these ideals so that you can see how the archival material would appear.
Slide making process	Very temperature sensitive. Variation due to imperfect control of the process results in slides which are not the same as the original negative would have been (Langford & Bilissi, 2011, p. 276).	The slide becomes the original source as the negative is used to make them.
Contrast	The contrast in an image represents either the difference between the blacks and whites or the extremes in the amount of colours present (Langford & Bilissi, 2011, p. 93).	Imagine the shadows created on a very sunny day. On a sunny day it is much harder to stand in the sunlight and see the detail in the shadows compared to a cloudy day (Millerson, 1972, p. 21). This is why if there is little detail in the dark areas of a digital object, reducing the contrast and re-scanning the archival material might allow more detail to be seen.

Term	What it Means/What it is	Why it is Important for Digitisation
.jpg	Joint Photographic Experts Group (JPEG) file format compresses digital objects before saving them (Langford & Bilissi, 2011, p. 224).	A file format that compresses a digital object reduces the amount of data stored when it is saved. For instance if the file is compressed by 50% only half the data in the original will be kept. When it is opened the information cannot be retrieved. However these files require less storage space.
.jp2	JPEG2000 is a lossless and lossy compressed file format	This is a file format that compresses the data when it is saved, but re-constitutes it when the file is re-opened. The result is a file that looks the same as the original, but requires less storage space. Web display requires a suitable plug-in (Langford & Bilissi, 2011, p. 224)

Appendix B: **Generic and Organiser's Guide**

N.B Both Appendix B: Generic and Organiser's Guide and Appendix C: CPR Quick Start Guide have been written using a plain and simplified style, rather than following an academic tradition that uses references and citations. The spellings adopted copy those in manuals, or as they appear on a computer screen as appropriate. The guides are intended to stand alone and be accessible to a wide group of people.

B1. Introduction

These two booklets, the Generic and Organiser's Guide and CPR Quick Start Guide, are aimed at anyone who wants to digitise archival materials to accepted standards. This first booklet uses the workflow adopted by CPR to discuss and to work through each stage. Each point is discussed from a general point of view and then how it was used in the research project. Therefore while reading you should always have in the back of your mind how your situation differs from that of the case study used i.e. CPR. So that you can easily identify the different parts to each stage, tips and pointers to their successful completion are in round edged boxes. To illustrate the use of the principle being discussed specific examples are in bevelled boxes.

The second booklet, the CPR Quick Start Guide, is a set of instructions developed using the equipment in CPR describing how they can create digital objects from their collections to recognised standards. You may want to create a similar set of instructions for your own equipment incorporating appropriate guidance for your circumstances. To use these guides to create your own, you may find it helpful to start by taking the headings from each section as a guide to the steps to be worked through. This list is the same as the list of stages at the beginning of the relevant chapter in the Generic and Organiser's Guide. By comparing your situation to that of CPR you will be able to work out where adaptations need to be made.

So, before you start:

This section, the Generic and Organiser's Guide, is aimed at giving those in charge a description of the many parts that comprise a successful digitisation project. It has four parts. The B3 addressed the actions needed before a scan is made and B6 considerations for ensuring the preservation of the digital objects created. Sections

B4 and B5 give an outline of the digitisation and the metadata gathering process followed by detailed information on specific aspects of these tasks. These sections should be read along with the shorter, CPR Quick Start manual which gives instructions, using the equipment available at the time of writing, on the actions needed to be successful in digitisation project.

It is suggested that project participants read the outline processes at the beginning of each section to familiarise themselves with the complete process before concentrating upon any which are aimed specifically at them.

It is generally agreed that digitisation is only a small part of a workflow that begins with creating a policy and ends with a digital object and its metadata stored in the nominated environment. This is because digitisation literally means turning something into a series of 0's and 1's.

Digitisation can appear to have a lot of jargon associated with it so, to try and avoid any confusion there is a list of key terms used throughout this written guide in the glossary in B2. But listed here are the three key ones to remember:

- *Archival material* is what you start with
- *Digital objects* are produced from the archival material
- *Metadata* is the data that is written about the archival material and digital objects

This guide is divided into five sections listed in Table B1:

Section Heading	Contents
Pre-scanning phase;	For the person in overall control. This section describes the factors to consider and actions needed to set up the system before any scanning is carried out
In More Detail … Scanning Phase	For the person(s) who carry out the scanning of archival materials, and
In More Detail … Metadata Phase	For those who write and collate all the information about a digital object into the database. Sections two and three contain extra information beyond the instructions given in the second booklet that may be of interest, hence the title
What Happens Next?	This section describes some of the actions needed to ensure the long-term survival of the digital objects created
General Pages	Contain tables listing general sources of help, some factors that may need to be considered when dealing with archive materials and a short bibliography

Table B1: Contents and Page Numbers

For consistency an imaginary collection is used for all examples given to aid clarity. It comprises:

- a copy of a script for an imaginary play called “Dragons Playtime” annotated with directors notes
- a publicity poster
- 100 production photographs, and
- Three digital objects for each photograph and the annotated play script

B2.1 Glossary

B2.1 Glossary contains terms used in the guides that may need some additional explanation.

Term/Acronym	Explanation
Archival Materials	What you start with e.g. Photographs or a script
Archive Digital Object	The digital object that has the largest file size and is used to make other copies from
BL	British Library
Digital Object(s)	Are what you get when you have scanned the archival material.
Dublin Core Metadata Element Set (DCMES)	The metadata scheme used for this collection. It has 15 parts to be filled in. Each of which can be repeated or left out, unless it is made mandatory locally.
ECLAP	European Collected Library of Artistic Performance
Format	The physical manifestation of a resource. This could be paper, a DVD, a digital file etc.
Metadata	The data, or information, collected about a digital object. This can be just an 'add on' to information already collected, perhaps in a catalogue, and is used to manage a digital object throughout its life.
NLS	National Library of Scotland
NLW	National Library of Wales
Open Source (OS)	Applied to software resources whose source code is freely available. Sometimes this is at no cost, sometimes not.
Optical Character Recognition (OCR)	Software applied to a digital object to make text contained within it readable. This is usually displayed as a separate file.
Original source	Same as archival material
Quality Assurance (QA)	The process of checking that a digital object and its metadata meet minimum standards
Resource	An item or group of items that are a unit in the collection
Record	An item or group of items that can show that they are what they claim to be e.g. a letter can be shown to be from a particular person or margin notes in a script made by a particular person
Thumbnail digital object	A low quality, small file size digital object created to give an impression of the viewing digital object
TNA	The National Archives
Viewing digital object	The digital object created for viewing the resource

B2.2 Technical Background

B2.2 Technical Background explains some of the technical terms and theory that come together in digitisation. It also explains why this term is important for digitisation. The terms are grouped so that the terms that describe the appearance of a digital object are together.

<table>
<tr><th>Term</th><th>What it Means/What it is</th><th>Why it is Important for Digitisation</th></tr>
<tr><td>ppi</td><td>Pixels per inch</td><td>Imagine a pixel is a small square. The number of pixels per inch tells you how many of these squares will fit into one inch on your monitor or scanner resolution scale. Each pixel can be one colour (or shade of grey). Therefore the more pixels you have, the more detail you can have (Langford & Bilissi, 2011, p. 463).</td></tr>
<tr><td>dpi</td><td>Dots per inch</td><td>Imagine a dot on a page. The quoted number of dots per inch tells you how many of these will appear per inch if you print the digital object (Sitts, 2000, p. 95; Western States Digital Standards Group: Digital Imaging Working Group, 2003, p. 4)</td></tr>
<tr><td>CMYK</td><td>Cyan, magenta, yellow, key (black) – the primary colours used for printing (Sitts, 2000, p. 101)</td><td rowspan="2">When viewing a digital object on a monitor, if the monitor has been set up to view the digital object as though it has been printed, it will appear different when it is printed because monitors cannot display the colours in quite the same way as they will be the printer has put them onto paper. If the monitor is set up to view just red, green and blue then this is fine for viewing a digital object on screens, but, again, it will appear different when printed (JISC Digital Media, 2010c, pp. 4-5; Millerson, 1972, p. 41).</td></tr>
<tr><td>RGB</td><td>Red, green, blue – the primary colours used by monitors</td></tr>
<tr><td>Subtractive colour mixing</td><td>Uses CMYK, generally the palette used in printing.</td><td>This method subtracts unwanted colour from white by adding cyan, magenta or yellow and is used in printing (Langford & Bilissi, 2011, p. 468)</td></tr>
</table>

Term	What it Means/What it is	Why it is Important for Digitisation
Additive mixing	Mixing colour to make another by adding to that already present	This method of mixing red, green and blue colours is used by monitors (Langford & Bilissi, 2011, p. 445)
Tone	The tone of a digital object is where on a range between pure black and pure white the contents of a digital object lies. This range of values is often displayed using a type of graph called a histogram (Western States Digital Standards Group: Digital Imaging Working Group, 2003, p. 35). Ideally it should have a smooth appearance, like a histogram, without any spikes.	In digitisation projects spikes on a histogram usually occur when the black and white points in a scan are not set to true black or white for the digital object being created (Langford & Bilissi, 2011, pp. 116-118). The expected smooth shape is an ideal, and it may be that a series of spikes on the graph are the best that can be achieved.
Curves	Curves are a graphical representation of the density of the digital object against the range from black to white in the digital object (Millerson, 1972, p. 331).	They are a drawing of the range of brightness and contrast values over the whole digital object. Their distinctive shape can be used in a digitisation project to achieve a more balanced digital object (Langford & Bilissi, 2011, p. 260).
.TIFF	Tagged Image File Format is a lossless, uncompressed file format (Langford & Bilissi, 2011, p. 223).	Lossless means that all the data saved in a digital object is retrieved when it is re-opened. Uncompressed means that the original file data is not altered when it is saved. This can result in a large file if there is a lot of data in a digital object (Western States Digital Standards Group: Digital Imaging Working Group, 2003, p. 28).
Scanner resolution	The number of pixels per inch the scanner sensors can copy	The higher the number of pixels a scanner can copy, the finer the quality of the digital object (Langford & Bilissi, 2011, p. 178). This can never be improved upon, whereas you can print it with fewer or more dots per inch. The higher the resolution the larger the file size of the digital object. The accepted compromise resolution is 300ppi (JISC Digital Media, 2006a).
Monitor resolution	The number of pixels the monitor can display (Ibrahim, 2007, p. 145)	As in scanners, the higher the number of pixels the better the quality of the digital object on the screen.

Term	What it Means/What it is	Why it is Important for Digitisation
Colour temperature	The temperature at which a black body radiator emits the same wavelength of light as the element in the lamp (Bermingham, 1989, p. 11; Langford & Bilissi, 2011, p. 451).	This is important because it explains why 'white' is different for different objects and lamp elements. Lamps on scanners take time to heat up to their steady temperature and therefore to maintain a steady output of light energy at a given temperature.
Comparison strip	A strip consisting of a series of (often eight) colours. These are either from white to black through a series of greys, or white to black through a series of colours (Stouffer Industries Inc., 2006)	These strips show how the primary colours in the digital object appear on your screen. This means that the colours in the digital object can be compared to these ideals so that you can see how the archival material would appear.
Slide making process	Very temperature sensitive. Variation due to imperfect control of the process results in slides which are not the same as the original negative would have been (Langford & Bilissi, 2011, p. 276).	The slide becomes the original source as the negative is used to make them.
Contrast	The contrast in an image represents either the difference between the blacks and whites or the extremes in the amount of colours present (Langford & Bilissi, 2011, p. 93).	Imagine the shadows created on a very sunny day. On a sunny day it is much harder to stand in the sunlight and see the detail in the shadows compared to a cloudy day (Millerson, 1972, p. 21). This is why if there is little detail in the dark areas of a digital object, reducing the contrast and re-scanning the archival material might allow more detail to be seen.
.jpg	Joint Photographic Experts Group (JPEG) file format compresses digital objects before saving them (Langford & Bilissi, 2011, p. 224).	A file format that compresses a digital object reduces the amount of data stored when it is saved. For instance if the file is compressed by 50% only half the data in the original will be kept. When it is opened the information cannot be retrieved. However these files require less storage space.

Term	What it Means/What it is	Why it is Important for Digitisation
.jp2	JPEG2000 is a lossless and lossy compressed file format	This is a file format that compresses the data when it is saved, but re-constitutes it when the file is re-opened. The result is a file that looks the same as the original, but requires less storage space. Web display requires a suitable plug-in (Langford & Bilissi, 2011, p. 224)

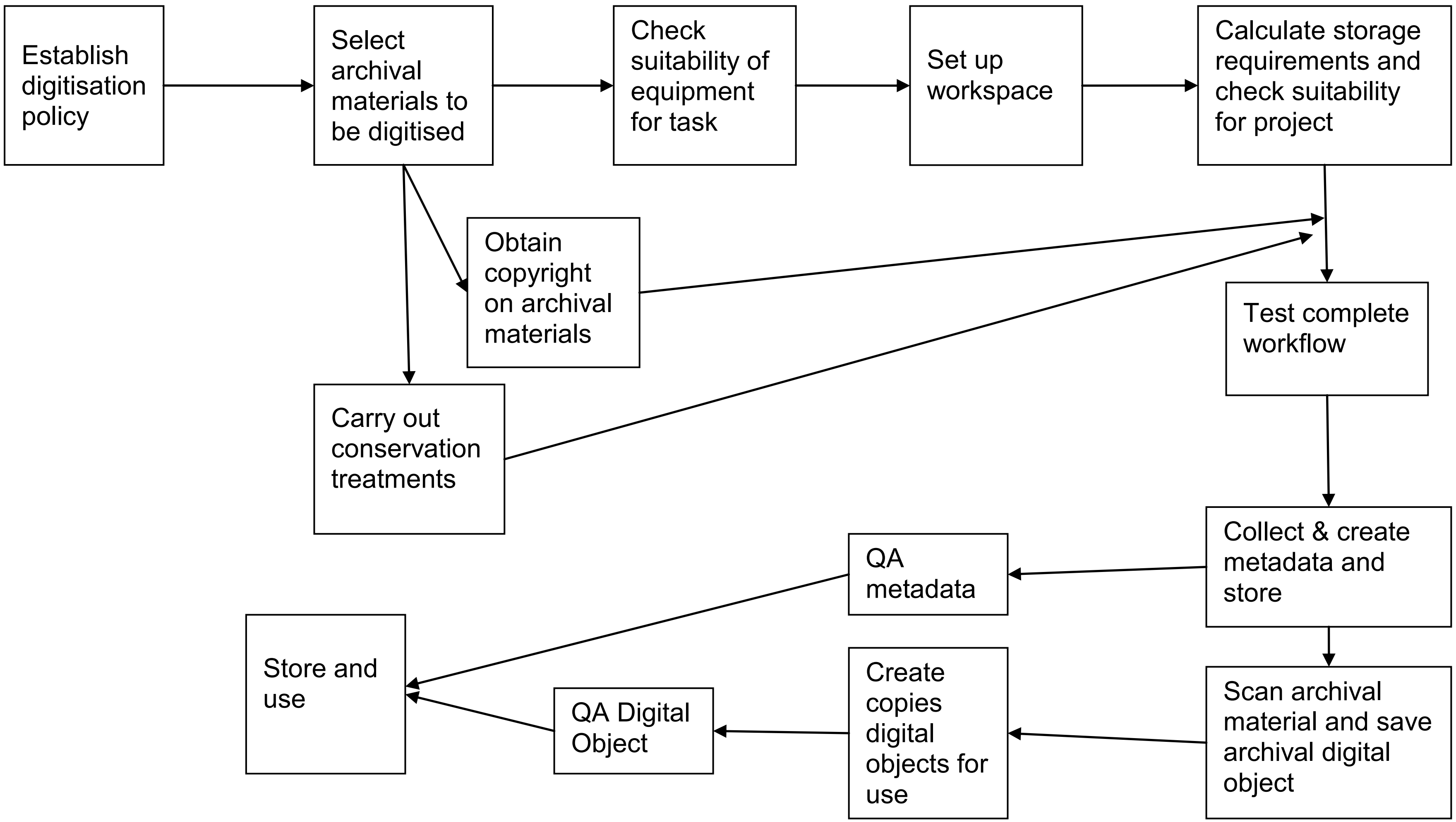


Figure B1: Sketch of the Digitisation Workflow Used

B3. Pre-Scanning Phase

B3.1 Process Steps

At first glance there appear to be a lot of stages to go through to achieve the digitisation of archival materials. However, some of these can be achieved in parallel with other activities so reducing the complete time scale.

So what did I mean when I wrote ‘accepted standards’ in the introduction? Well, there is currently no list of ISO standards drawn by the digitisation community; but those that are applicable from intertwining fields, such as archival practice are often used instead. For instance, ISO1472:2003 often referred to as the OAIS is a standard that describes the composition and functions of an ideal archival system, whether storing analogue, digital or mixed formats. For digitisation projects this standard can be used to design a system with a structure for long term management of the digital objects and their metadata.

The overall aim of any digitisation project is to convert archival materials to digital objects for pre-defined uses. Because the reasons for conversion will vary, the digital objects created will reflect this. Figure B1 shows the workflow that this guide adopts.

As well as the workflow above (Figure B1) there is a form that can be adapted and used to chart the steps achieved in the process (Table B17). Documenting the process is important because:

* It can help predict where difficulties could arise and therefore aid the organisation of the project under way.

* If there is some time between projects it can help to refresh the memory on the procedures used previously.

* It can help new staff understand the procedure.

* The documentation becomes part of the collection metadata i.e. it becomes part of the collection documentation because it records an action carried out on archival materials. If this is not recorded then the value of the archival material can be diminished because an archive cannot prove that the archival material is exactly what it purports to be

* It helps with tracking archival materials as they progress from their storage, through the imaging process and back again.

The pre-defined uses, or the reasons for conversion, affect the digital objects created because the output from the scanning process can be different. For example if you are scanning a page of text, the resulting digital object is a picture of the text. If the end aim is to make the text on the page available to users then the scan needs to be completed so that it captures all the characters in the text so that complete words and their associated grammar are encoded properly. The resulting digital object is processed again using software that extracts the code for the text so that it can be reproduced using a computer with the right software. You may have come across Web pages that claim they are 'OCR enabled' (Optical Character Recognition); well this is an example of this type of processing. If however, the aim is to show what the page of text looks like then there is no need to go through the additional process and the scanner can be set to capture an accurate impression of the look of the text on the page.

B3.2 People

The people that carry out a digitisation project have to use a variety of skills to produce digital objects and metadata of the required standard. These skills are wide and varied. Some are obvious, some not. For instance scanner operators need good visual skills; only one of these visual skills is the ability to assess whether what they see on the screen is exactly the same as the archival material they started off with. Theoretically one person could do all of the work involved in a digitisation project, but it is more likely that more than one person will be needed. Tables B2, B3 and B4 show some of the skills identified for the three principal roles and could be found in other roles. These lists are not exhaustive or in any order of priority.

In addition to these skills I have assumed a level of basic computer skills that include the ability to:

- Use a variety of software appropriate to the situation having received basic instruction on their use
- Locate and save files on differing drives (directories)
- Locate and create new folders in a specified place, and
- Isolate and then to copy and paste information from one document to another

Project Organisers	
Skill/Competency	Roles Which Already Use This Skill
Organisation of project	Higher degree students; engineers, mathematicians; scientists
Team building/leading skills	Stage manager; DSM; Team leaders in sport or similar areas; director; producer; editor
Assessing others for roles	HR personnel; audition panellists
Quick assimilation of information	Stage manager; DSM; stage crew (inc. Fly crew etc.);
Ability to understand technical requirements	Stage manager; architects; engineers
An eye for detail	Costumiers; make-up artists; creative artists
Communication skills	Team leaders in sport or similar areas; stage manager; DSM
Methodological approach to tasks	Scientists, stage crew; mathematicians; engineers

Table B2: Skills/Competencies for Project Organisers

Scanner Operators	
Skill/Competency	Roles Which Already Use This Skill
Ability to differentiate fine detail in an image	TV cameraman; photographic technician; photographer; lighting designer; lighting technician; creative artists; film editor; VT editor; TV racks operator; CGI operators;
Ability to balance colour in an image	TV racks operator; VT editor; film editor; CGI operators; creative artists;
Ability to see the features that comprise an image	TV cameraman; photographer; lighting designer; creative artists; film editor; VT editor; TV racks operator; CGI operators;
Self Monitoring	Sports coaches/instructors;

Table B3: Skills/Competencies for Scanner Operators

Metadata Writers	
Skill/Competency	Roles Which Already Use This Skill
Ability to see/differentiate detail in text	Copy editors; proof readers; academic writers; creative artists; tutors/teachers/lecturers; graphic designers; computing code checkers; journalists
Ability to carry out repetitive task while still paying attention to detail	Regular participants in sport;
Self Monitoring	Sports coaches/instructors;
Written communication skills	Copy editors; academic writers; journalists; tutors/teachers/lecturers;
Broad general knowledge	Pub quizzers; primary level teachers;
Methodological approach to tasks	Scientists, Stage Crew; Mathematicians; Engineers; Higher Degree Students

Table B4: Skills/Competencies for Metadata Writers

B3.3 Digitisation Policy

There is advice available from many of the sources listed in Section B7 on what to include in a digitisation or preservation policy if you do not have one already. Some also include templates that could be adapted to suit your circumstances. As well as these The National Archives (TNA), the National Library of Wales (NLW), the National Library of Scotland (NLS) and the British Library (BL) have their documents available for public scrutiny on their Web sites if you want to compare yours to theirs.

In general the contents of a digitisation policy are very similar to other policies and should include:

* A statement to indicate why digitisation is being carried out – is it to improve access to the collections, or to aid preservation of the collection?

* A statement on how the archival materials are chosen – take the contents for this section from the risk assessment questions used in section B3.4 if appropriate.

* A statement on the scope of digitisation activity – do you aim to digitise everything or only carry out digitisation activities when there is time to spare from other projects

* A statement to indicate how this policy fits in with other policy documents, in particular the preservation policy – you don't need to state how the two fit together.

* A statement to indicate who is responsible for the digitisation of archival materials

* A date of acceptance, and

* A signature by the relevant person to show its acceptance

Having written your policy it should be made available for viewing.

B3.4 Selection of Archival Materials for Digitisation

When archival materials are being assessed for digitisation it can help to use a formula to provide a consistent basis for comparison.

Table B5, below, is a simplified example of a table based on a standard risk assessment table which has been completed using the photographs and script from the imaginary collection as examples. The 'Questions to Consider' and the number values allocated can be changed to fit different collections and situations by asking questions that are relevant to you. For example it may be that the script is of high value to you because of the person that made the markings on it; or it may be that the markings lessen the value because the script is not in pristine condition.

The higher the number in the final column the higher the risk to the materials and therefore perhaps greatest consideration should be given to digitising these materials. This table, Table B5, illustrates how deciding what to use to calculate the risk needs to be considered carefully in relation to the values of the individual archive as it may skew the outcome of the calculation away from the desired conclusion. Faced with the result in Table B5, you may decide that the photographs should be given priority because their value to the collection is higher.

Questions to Consider	Type of archival materials in collection	Type of risk to archival material (i.e. what will cause damage?)	How likely is this to happen? 1= not likely 5 = very likely	Number of these materials in collection	Value of materials to archive. 1= low, 5 = high	Total for materials (= pink x yellow x blue)
Example	Annotated Script	Rips and tears due to use	5 – paper rips quite easily	1	5 – unique, but not used often	5x1x5 = 25
	Photograph – recent plastic film	Deterioration due to use	5 – material accessed a lot by users	5	5 – used extensively	5 x 5 x 5 = 125

Table B5: Risk Assessment of Archival Materials

Another view of this process is illustrated in Figure B2 which shows some of the questions that could be asked about the collection photographs. The white background boxes are the improvised answers to the questions in the blue boxes above. By following the questioning process through you can see how they provide a consistent background for assessment of the photographs and other archival materials in the collection.

Figure B2 and Table B5 are just examples of possible questions to ask yourself and how to rate their importance. In Figure B2 the questions being asked are in boxes with a blue background and the answers given in white. The green and orange ones show how the risk is worked out from the answers in white.

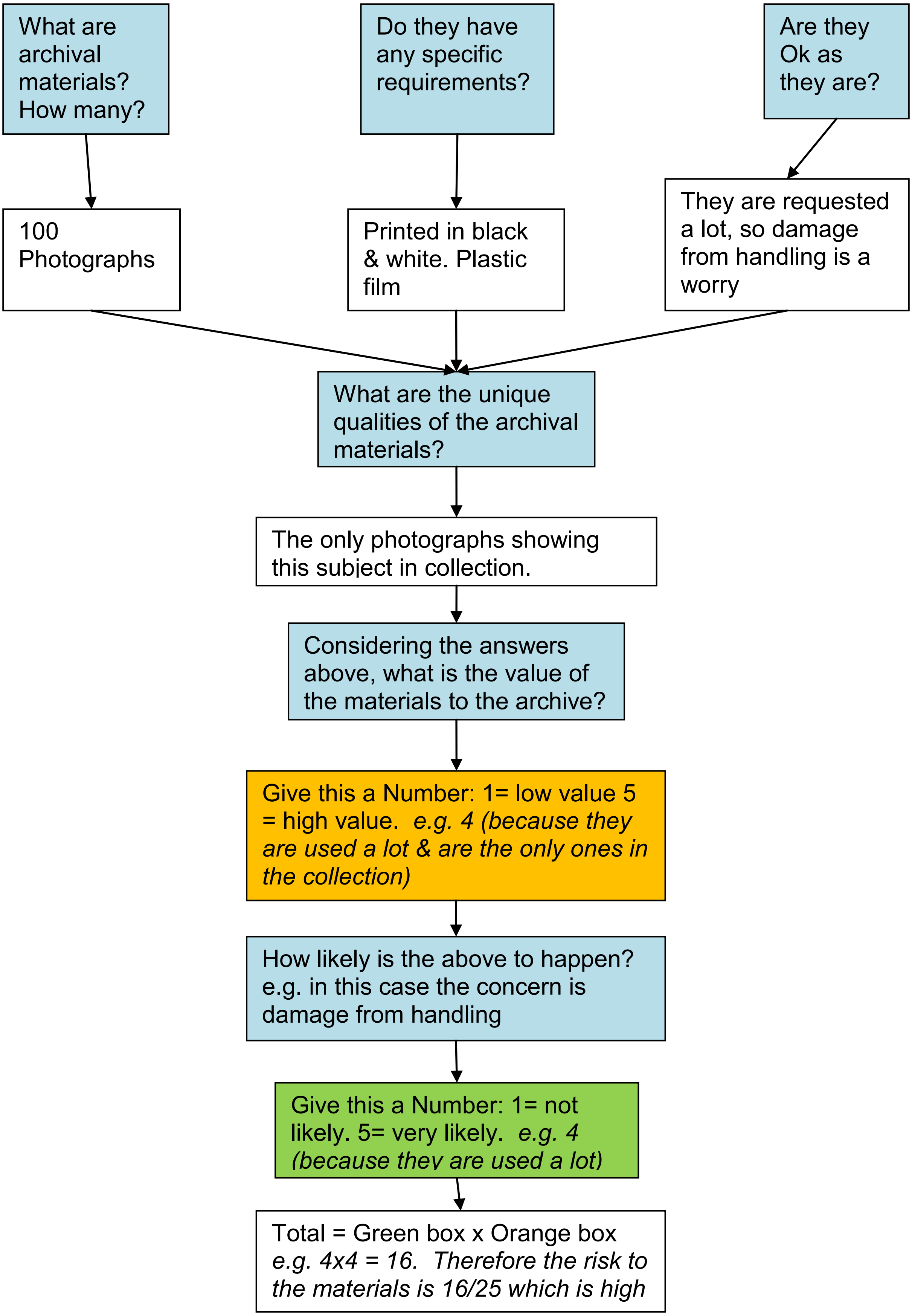


Figure B2: Decision Tree View of Risk Assessment Table

The assessment of materials for *preservation* was the subject of a project whose outcome, PLATO, can be found on the Open Planets web site (http://www.ifs.tuwien.ac.at/dp/plato/intro.html). PLATO is meant to be used to assess the current state of digital files and to then indicate what needs to be done to ensure they are preserved for future access using the means already decided upon by the user. It also makes assumptions about the technical capacity of those using the tool – check this. If you feel this is appropriate for your archival materials then this is the tool for you.

There is a list of some of the risks you may like to consider in Table B30. These lists are not exhaustive as they are only intended to provide a starting point for thinking about your assessment.

B3.5 Allocate an Individual Identifying Number

Having decided which archival materials are to be considered for digitisation you now need to be sure of identifying the exact materials again. To do this each item of archival material needs to be given its own individual identifying number, if it doesn't already have one.

The number needs to be unique to be sure of identifying the correct item of archival material each time and its associated digital objects. If these numbers are associated with the archival material beyond the life of a digitisation project, they become important for their continued management. This management can include functions such as when conservation treatments become due or flagging up when associated digital objects need to be checked for consistency.

How numbers are allocated and recorded is down to you. If all your archival materials are already individually identified then you can use these numbers; perhaps the numbers can be printed out by creating a report if they are held in a database or maybe you need to devise a system for creating them and write them by hand on paper.

The individual identifier is the 3 letter code for the collection + 5 digits e.g. if the first group of items written down in the imaginary collection are the photographs the first will be labelled dpt00001, the second dpt00002 etc. The three letter code can be taken from the title initials of the collection, but it needs to be unique. If you keep a list of the initials used, and add to it each time a project is started, you won't accidentally re-use them.

For photographs the individual identifying number should be written on the reverse using a soft pencil. This should also be away from any markings in case the back is scanned too. Other archival materials should be marked in an appropriate way so as not to damage them. For slides the identifier should be written onto a sticker stuck onto the outside of the slide box. The sticker is outside because they can cause damage to the slides if stuck on the inside.

B3.6 Obtain Copyright

Digitisation is copying and needs permission before it is carried out. Placing a digital object on the Web is publishing and also needs permission. Because of this, before any digitisation is carried out it is necessary to obtain permission from the copyright holder and the owner of the rights to a work to carry out the process.

When items were signed over to a collection, obtaining permission to digitise was not generally considered until recently. So a first step in seeking consent should establish if permission to carry out any processes of this kind was explicitly mentioned in any accession agreement (a contract that relates to when the archival material first came into the collection). This may mean checking the original document for each accession to the collection if individual rather than generic contracts were used.

It can be difficult to determine whether items are in copyright. Using publications like Tim Padfield's book (listed in section B7.2) can help you to determine if your items are in copyright. Further help may be available from sources such as JISC mailing lists, if you are a member. Having ascertained who owns the copyright on analogue materials in order to digitise them you need to get permission from the copyright owner. For owners other than yourself, contact your copyright adviser for help in writing to them. If the copyright holder cannot be discovered then it cannot be assumed that copyright does not exist.

It may also be necessary to contact people featured in archival photographs to seek permission to digitise them if suitable permissions were not obtained when the photograph was taken. Again contact your copyright adviser for advice.

While obtaining these permissions is in progress, processes up to the point where digitisation begins, such as sending archival materials to receive any conservation treatments needed, can be carried out. This work is carried out at this point because conservation treatments can take some time to complete and the effects of handling during digitisation may make any defects worse.

The process of obtaining pertinent rights can take time and, if relevant to the project, can be added to the selection of materials risk assessment table. The time taken to obtain copyright can be relevant if, for example, funding is only available for a specific time and obtaining copyright could take too long.

B3.7 Where Will Digital Objects be Stored?

Storing digital objects takes space somewhere and has expenditure associated with it. It therefore makes sense to use the most efficient and cost effective way to store and maintain the digital objects you are creating.

B3.7.1 Calculate Storage Requirements

The amount of storage needed for digital objects depends upon a range of factors from the type of file the digital objects are stored as, to the file size and any security requirements. For instance, one of the recommended file types for storing the archival, or reference copy of a digital object is a .TIFF. This is an uncompressed format so it takes up more storage space than, for instance, a JPEG file of the same digital object which compresses the file data.

During the project you will create three digital objects in total from the archival material. These are:

- The archival digital object
- The viewing digital object, and
- The thumbnail digital object

The archival digital object is the largest of these. The viewing and thumbnail digital objects each have a smaller size than the archival digital object, but it is important to include these in your calculations. For the sake of calculating storage requirements you may wish to assume that the file size of the viewing digital object will be approximately 2/3rds the size of the archival digital object and the thumbnail about a quarter (if desired, the exact file size can be set using some imaging software).

If you really are starting from scratch you will need to allow for somewhere to store the data, the metadata, you are collecting about the digital objects you are creating. Where this is depends upon the storage and delivery system you have chosen. Perhaps because the amount of storage needed for these files is relatively small, this is not covered widely.

To work out your storage requirements you either need to decide the maximum file size before commencing digitisation or, if this is not possible, the expected file size for flat objects can be worked out by knowing how big the archival materials are. This size, the number of pixels per inch (ppi) the scanner will be set to, and the quality of the digital object you are hoping to create are then used to calculate the resulting file size. The first step is to work out the area of the archival material in inches:

Area (in inches) = Length × Breadth

Then multiply this by the number of pixels (the scanner settings give the number of pixels the digital object will have):

Number of Pixels in the digital object = (Area in inches) × (Number of ppi)

There is a fixed number of 3 bytes per pixel file space required for a 24 Megabit colour digital object file (the usual choice for colour). Therefore:

Number of Bytes of storage required = Number of pixels in digital object × Number of bytes per pixel

Equation B1: Storage Requirements Calculation.

Note: If metric measures are preferred then replace the number of inches above with the equivalent number of centimetres.

It is suggested that you write these equations into a spreadsheet and use this to calculate the requirements for each project. From a management point of view this will mean that you can also keep track of how much data you should have. If you then discover that the amount of data is very different to that anticipated it can be an indication of a problem.

In CPR this is currently done simply using a spreadsheet (or workbook).

B3.8 What Storage Should you Use?

Where you store your digital objects depends upon your aims and the options open to you at the time. If the objective is to keep illustrations of the archival materials for insurance purposes, then they do not have to be kept somewhere they can be accessed immediately via a permanent link. If the long term aim is to provide greater access to the collection there is general agreement that the archival digital object and a copy of the metadata should be kept somewhere away from the main storage system and the viewing and thumbnail digital objects and metadata on the system for access.

B3.8.1 Which Physical Storage Media to Choose

In choosing storage media a number of criteria need to be considered in order to arrive at the most appropriate compromise. Some general points include:

- the cost,
- any constraints from IT systems already in place,
- any constraints the storage system chosen could place on the digital collection
- the available storage conditions of the archival materials and
- any specific needs of the chosen storage medium.

And two general thoughts:

- storage media of all types should be handled carefully as any damage will lead to accelerated degradation of the media and therefore the digital objects on them
- It is not always possible to meet the exact storage requirements of different storage media, but whatever the media, it is generally fluctuating conditions that accelerate the harmful effects of heat and humidity changes.

The choices made will influence the future of the digital object which in turn affects the digital collection.

As has already been indicated using unsuitable storage conditions can considerably shorten the life of the medium or affect the digital object. Even if storage conditions are perfect for the chosen media there is always the risk of accidental damage. You can try and avoid any losses from this kind of mishap by keeping more than one copy of the digital object, if possible, on a different medium to the digital objects in regular use.

Selecting the storage option most appropriate for your project means; balancing the affordable, the ambitions for the digital objects and any external factors such as interoperability with established systems. Having established the requirements for storage the next step is to assess the options open to you. One way of carrying out this assessment for yourself is to carry out a similar exercise to when you assessed the archival materials to be digitised.

For a recent project in CPR these four options were considered:
* University controlled hard-drive space.
* External hard-drive (in Department).
* Own repository module independent of University (e.g. DSpace).
* Externally hosted and maintained repository.

A summary of some points that were balanced to make the final decision are in Table B6.

To ensure the longevity of the digital objects the storage media should be refreshed periodically and within the manufacturers recommended time span. By having regular checks it is hoped that any problems will be spotted before they cause significant problems.

Part of choosing the right storage is to use an appropriate file format for the type of digital object being created. Different file formats are used because they are often designed to work best with a specific file type. The most appropriate for each media type however is not always clear. A selection of accepted formats and their uses is in Section B7.4.

Option	Positives	Negatives
University controlled hard-drive space	1. Back up of files performed automatically by University 2. Daily management of access managed automatically by University 3. University carries out regular disaster recovery tests as part of its' management of digital objects 4. In case of disaster University experience could mean digital objects are recovered when perhaps it could be expensive for department to discover the means to do so	1.Control of digital objects is distributed between department and University as department would decide upon exposed content
Own repository (e.g. Dspace) module independent of University	1. Department retain control of digital object(s) 2. If formatted to use OAI-PMH interface has richer search capabilities as well as exposure through the network	1. Department have to manage and carry out content back up routines 2. Department need to manage long term maintenance e.g. carry out regular crash recovery tests 3. Interface between internal and external system would need managing 4. Department need to establish and pay for a link to the Web 5. Significant commitment of time and resources required to implement successfully 6. No current community of expertise within Department 7. Some repository software requires a considerable level skill and a wide range of ability to install e.g. SQL writing, XML schema writing and web design
External Hard-drive (in department)	1. Department retain control of distribution of digital objects to Web or other outlets 2. Department regulate access to digital object	1. Department have to manage daily back up of files 2. Department need to manage long term maintenance e.g. carry out regular crash recovery tests 3. Department need to establish a link to the Web
Repository externally hosted and maintained (e.g. eclap)	1. Department control data exposure through upload choices made 2. See list of advantages at: http://www.jiscinfonet.ac.uk/infokits/repositories/benefits 3. Possibility of International exposure of content	1. Continued commitment to community development to ensure continued success of exchanges 2. No control over who signs up to community 3. If project fails it is not yet clear how content could be recovered

Table B6: Table of Some Storage Options & Points Considered in The case of CPR

Some file formats for digital files are widely accepted as being suitable for storage e.g. for uncompressed image files .TIFF is often used, but more recently JPEG2000 (jp2) has gained ISO recognition and promises both lossless and lossy compression in the same file format. The advantage of using this lossless format over .TIFF files is the resulting file size; it is much smaller thus leading to equivalent reductions in the space required for storage.

Another consideration is the software that is used to access digital objects. The current position of support and access available for specific versions of software can be evaluated using the PRONOM database associated with DROID at the TNA web site.

Up to date advice on the suitability of the chosen media for your files should be sought as this a rapidly changing field. Some of the sources of advice listed in Chapter 15 of this guide could be a good place to start.

While assessing your options for storage, attention should also be paid to the expected means of recording the digital objects and metadata onto the storage and playback medium. For instance, if magnetic media are used, then the proximity of the heads and tape mean that it is easy for small particles to affect the tape adversely.

Marking the physical storage media for identification is sometimes overlooked but writing onto a DVD or CD with the wrong kind of pen can impair the ability of the disc to repel the effects of temperature and humidity. In turn this degradation affects the length of time it can successfully be used to store digital objects. Using labels with glue that attacks the protective layer on a disc can have the same effect. It is for this reason that in CPR stickers to identify slides are stuck on the outside of the box

B3.9 Check the Physical Workspace

You take a lot of care to keep your collections in conditions that don't accelerate decay due to age, so why not take as much care over the environment that you digitise in? The physical conditions of a workspace are important to those carrying out the digitisation as well as for the continued safe keeping of the archival material. If a seat in front of a computer is uncomfortable then it will be difficult to concentrate on the fine adjustments needed when creating a digital object. This can affect the quality of the digital objects being created. Creating a suitable digital object can take longer through the need to repeat scans if they didn't collect enough of the available data in the first pass. This costs in terms of both time and money.

Another major consideration is the lighting in a room. When entering a room JISC recommend that the lights are turned on and the curtains or blinds drawn so that natural light is kept out, no matter what the outside conditions. This is because natural light levels will vary throughout the day. So by excluding it and using an artificial light source, the light levels around the scanner will be kept even. By keeping the light in the room even, it will not influence the ambient light falling onto the bed of the scanner, or the monitor. Changes in the light on a monitor affect how the eye sees a digital object and therefore the changes made when adjusting it. Equally the positioning of lights in a room should be considered when placing equipment as reflections also alter the appearance on a screen.

When choosing a room try to select one, or have it re-decorated, with neutral or light coloured walls. This is because the walls reflect colour and influence the appearance of colours in the archival material and on the screen.

B.3.10 Carry Out a Test

Ideally this is a test of the whole procedure to check the process collects all the required information. This is particularly important if there are fragile or very sensitive materials to be digitised that require special handling. Ironing out any difficulties that occur during the preparation phase of a digitisation project can mean avoiding damage to sensitive or fragile materials.

B3.11 Management of Digital Collections

Having created a set of digital objects and metadata, there is now a need for a means to use and manage them. How this is achieved depends upon the individual circumstances of a collection because what is right for one, may not suit another. This is because the contents of collections and the chosen access methods can be unique.

Through using the structures suggested throughout this guide you have already built in most of the steps needed for managing a collection sustainably. In order to manage the collection in the long-term you now need to use these systems differently. The risk assessment you did for selecting the materials could now be utilised to provide information to manage the digital object and the associated archival materials. By building in quality assessment you have ensured that the digital objects are in the best shape possible to survive. The metadata you have compiled can be used to manage the collection by flagging up when digital objects ought to be checked for consistency. This is done to ensure that they aren't allowed to degrade and become useless. One way you can do this is by creating a report using a spreadsheet. The same process can also be used to indicate when the storage media should be checked for consistency, when the archival materials are due to have archival procedures carried out on them etc. In other words, all the data you have collected during the digitisation process now become your tool to manage and ensure the survival of the digital objects for the long-term. There is more on this in Section B6.

B3.11.1 Systems for Managing Digital Objects

When deciding what kind of system to adopt for managing your digital objects there are several considerations to be assessed in relation to the situation. These include the file types used for storing the digital object, the number of digital objects to be managed and the proposed delivery method of the digital objects. Some practitioners believe that the biggest argument in favour of adopting a system to manage digital objects is that digital collections need to be managed in a much more active way than documentary collections. This is because in addition to the degradation of the storage media for digital objects there are also the issues of the availability of a suitable playback media, and maintenance of the links between the metadata and digital objects to be considered.

B3.11.2 Using a File Name Structure for Organisation

Using a structure to create file names will help the organisation of digital objects, especially if that structure reflects any structure already in place. Using an agreed convention to name files will help when storing them before they are moved to a more permanent home. The structure will then help with finding the right digital object when it comes to associating them with the resource to be used for display.

Structured file names can be used as a temporary control measure in conjunction with a database or spreadsheet while a critical mass is built up. If structured correctly this could then be imported wholesale into the permanent system selected.

B3.12 File Naming

The names given to files do not have any specific conventions, but using a framework that is logical makes it easier for users to implement, and to find files after the project is completed. This is because a logical sequence is easier to understand. A logical progression also makes it possible to discover errors or omissions through automated searching.

Writing the name down however, does have the convention of using lowercase letters in the name and avoiding the characters reserved for system use plus a three digit file type extension.

For our imaginary collection adopting this writing system would give a file name that looks like this: dpt001.jpg

The method for deciding a file naming sequence needs to be decided locally and documented so that it may be followed accurately in the future. This document can become part of a digitisation policy to ensure consistency, depending upon the level the policy is written at.

B3.13 Summary Of Steps so Far

By following these steps through you have:

- Decided which archival materials to digitise
- Obtained copyright to digitise those archival materials
- Worked out roughly how big the file size of each of the 3 digital objects created from each piece of archival material is expected to be
- Decided on which media you are going to store the archival digital object

- Decided on which media you are going to store the viewing and thumbnail digital objects
- Decided which file type you will use for which digital object
- Decided where you are going to store metadata, and
- How the storage will be structured

B.4. In More Detail …. Scanning Phase

B4.1 Outline of Workflow for Scanner Operators

The instructions below are an outline of a digitisation workflow. Any detailed or specific instructions should be taken from the manual for the scanner and software you are using.

- Adjust the lighting and window coverings to suit the room
- Before you start scanning, check the workspace is laid out correctly for you.
- Place the archival materials to be digitised close by so that they can be moved onto the scanner easily.
- Switch on the computer attached to the scanner
- Switch on the scanner and use the software to check the calibration of the scanner is correct.
- Check the scanner is ready to scan. This could just mean looking to see if a green light is showing on the scanner, or it could be more complex.
- Check the list of items and archival materials match
- Check the function buttons on the scanner or software are set up correctly. These can be used to adjust the scan to record the maximum amount of the data possible from the archival materials for the resolution set. Or forgetting to adjust them could be why a scan doesn't quite come out as expected.
- Place archival materials on the bed of the scanner using the guides along the side of the bed, or the side of the scanner bed where no obvious marking exist, to get the archival material level
- Locate, or set up, the folder where archive digital objects are to be saved to named with the 3 initials of the project
- If the software is capable, set up the automatic numbering
- Start the scan preview
- Crop the scan to the edges of the preview as this can affect the scan.
- Adjust the scan to make an exact copy by looking at each preview individually. Ask yourself:
    - Are there any areas where there is no detail, but could or should be?
    - Are the colours the same as the archival material?
- Adjust the preview until these things are corrected.

- ❖ Repeat the adjustments for each preview, if more than one can be done at once. When all the previews are adjusted
- ❖ Do the scan.
- ❖ Save the newly created digital object to the designated folder using the agreed naming scheme
- ❖ Capture any required metadata
- ❖ Repeat this until all the items in the batch have been scanned.

B4.1.1 What You are Aiming to do

When digitising archival materials, whatever the intended use, you are trying to capture as exact a copy, with as much detail as possible, within the limits of the equipment and any other reproduction parameters. So if you are producing a digital object for archiving and then making copies to be used on a web site, the initial scan to create the archival digital object needs to capture the maximum amount of information available. It needs to do this while making the digital object look the same as the source archival material. For modern colour materials, an exact colour match should take a slight precedence over the amount of detail. For sensitive archival materials capturing the amount of detail takes precedence over a match to the material. This is because for archival material, such as delicate glass negatives, digitisation is generally being undertaken primarily to aid preservation, and not necessarily to increase access. For these frailer archival materials it is important not to mistake good pictures for those that contain a lot of information. It doesn't matter for the archival digital object if the contrast is not good and the picture appears flat if this results in capturing details, such as a lace pattern on a dress, which would otherwise be missed. In this kind of situation, using the scanner controls to acquire greater detail is acceptable.

If a scan does not turn out as expected, or you think it could be bettered, then don't repeat it until the end of the batch. This way you don't lose your place in the workflow and run the risk of missing a piece of archival material out.

Having captured the digital objects it is important to save them where directed so that they can be located again easily. This is important because at this stage it is unlikely that there will be any permanently linked identifying metadata created, so the only reliable place to locate the digital object is where you have saved it to.

B4.1.2 Some Standards that can be Used for Digitisation Projects

These standards may vary for your situation, or from project to project, but the aim of the procedure should be clear. The application of standards is important because decisions made about the digital object will have an impact the workflow, and on the quality of the digital object made.

What	**Standard/accepted parameters**
Scanner Resolution for producing a digital object with the best balance between data capture and resulting file size.	300ppi minimum
File format for archival quality digital objects uses lossless compression	.TIFF; .RAW; .jp2;
File formats for access copies can be compressed. This means smaller file size and therefore less time to download for users	.jpg; .jp2; .PNG
Metadata: 15 element Dublin Core minimum accepted for data exchange	DCMES

Table B7: Generally Applied Standards

B4.2 Scanning Images

A scanner is a piece of hardware that uses software to interpret and transmit data to a linked computer. The data is the transformation of an analogue object to the 1's and 0's of digital code. The types of scanner most used in digitisation projects are reflectance scanners and transmission scanners.

B4.2.1 Reflectance Scanners

A flat bed, or reflectance scanner is designed to be used for scanning non-reflective materials. In other words it works best on things like books or maps.

In general, devices that are designed to carry out this type of operation have a flat space on which the material to be scanned is placed; a light source above to illuminate it and the scanning device above the bed, near the light source. Sometimes there is a glass lid that smoothes archival materials such as newspapers.

Reflectance scanners use sensors arranged in rows and activated by reflected light, which move in a predetermined sequence over the item being scanned. The sensors either collect the information for all three colours in one pass or make one pass for each of the colours in the array. The software in the scanner then collects and processes this data to produce the information on a computer screen.

B4.2.2 Transmission Scanners

Scanners used for film are called transmission scanners because the film needs to be lit and transmitted to capture the positive as a digital object. This is because the image, a negative, needs to be projected to become the positive used for display. Some of these scanners use similar techniques to those outlined for flatbed scanners, others use a block onto which the complete image is projected and captured.

B4.2.3 Technical Adjustments to Scanner and Effect on Digital Object

Adjustments to scanners play a critical part in determining the appearance of the resulting digitised image of an item.

Spatial Resolution

The resolution of a scanner is usually measured in pixels per inch (ppi) and in the literature the optical resolution or interpolated resolution are quoted. Optical resolution is the number of pixels the sensors can record and is normally the first number quoted in manufacturers' details e.g. 1200 x 1200 ppi. The interpolated resolution is the optical resolution combined with that the scanner software can add to it. This results in an increased file size, but not more data from the scanned item in the file. The most appropriate settings for the item to be scanned is worked out according to the formula below and can be adjusted to suit the required output.

$$Required\ scanning\ resolution = Output\ device\ resolution \times \frac{Desired\ size}{Original\ size}$$

Equation B1: Equation to Calculate Required Scanning Resolution of a Scanner.

The generally accepted compromise scanner resolution is 300ppi because this provides the best balance between the resulting file size and the amount of scanned data.

For most of the time in CPR you want to collect as much data as you can because CPR have taken the decision that they want to create archival digital objects that contain more data than the recommended minimum. Therefore you use a higher setting than 300ppi to create the scan and have a larger file to store.

If you have a scanner whose resolution is only quoted for the whole area of the scanner bed then the size of a side of the scanner bed divided by the resolution number quoted gives the resolution in ppi:

$$\frac{Smallest\ number\ from\ quoted\ resolution}{Size\ of\ smallest\ side\ on\ scanner\ bed\ in\ inches}$$

$$e.g. using\ a\ scanner\ bed\ of\ A2\ paper\ size$$

$$and\ a\ quoted\ resolution\ of\ 4800\ x\ 6775\ pixels\ gives:$$

$$Scanner\ resolution\ (ppi) = \frac{6775}{23.4}\ \text{to}\ \frac{4800}{16.5}$$

$$= 205\ to\ 291\ ppi\ (to\ nearest\ whole\ number)$$

Equation B2: Equation to Calculate Actual Scanning Resolution

The resolution is often quoted as two numbers e.g. 4800 x 6775. Care should be taken to ensure that the quoted numbers both relate to the scanner resolution because sometimes the first number is the resolution and the second, how far the scanner head moves to the next scanning point.

B4.3 Digital Cameras

Another method of capturing a digital object is to photograph it using a digital camera. Using this method means that additional equipment in the form of a tripod, and if available, camera studio flash lights are needed.

Studio flash lights are usually used in addition to, or instead of, the flash placed on top of a camera because lighting from two angles can provide more modelling for the archival material. Using only the flash on top of a camera can only provide light from the front and can have the effect of flattening the appearance of the archival material in the digital object. This is not a problem if photographing a map or the pages of an album, but if the archive material is a 3D object, it mis-represents the object to flatten it.

B4.3.1 Is My Camera Suitable?

To work out whether the camera can reproduce the size of digital object you need. Multiply the size of the digital object you want by the number of ppi e.g. for a picture to be the equivalent of a 10” x 8” photograph, scanned at 300 ppi:

(10 x 300) = 3,000 pixels and (8 x 300) = 2,400 pixels

Therefore the minimum resolution for a camera to produce a photograph 10 x 8 is

= 3,000 x 2,400 ppi or = 7200000 or 7.2 Mp

Equation B3: Required Resolution of Cameras

Therefore the camera specification needs to have a resolution that at least equals, or preferably, exceeds 3000 x 2400 pixels to be able to reproduce a digital object that is the equivalent of a 10” x 8” photographic displayed at 300ppi.

B4.4 Adjusting the Digital Object

Having created the archival digital object the next step is to make copies of it for different uses. To do this you manipulate the digital object and save the results as separate files, generally in a different place, so that the archival digital object is kept as it is, where it is. These digital objects are created as separate files so that the archival digital object remains for reference, and to allow you to create further copies as required. To do this you manipulate the digital object using imaging software.

When you arrive at this stage in the process for each item of archival material you should have:

- 1 archival digital object saved to a folder.

For CPR this is in a folder on the desktop

- 1 viewing digital object and

- 1 thumbnail digital object.

For CPR this is again saved to a different folder on the desktop

All the copies should now be transferred to their allocated storage space.

B4.4.1 Software Adjustments and Effects on Digital Objects

All hardware needs software to allow it to 'talk' to a linked computer. For scanners this software has two purposes. One is to transmit the code sequence that the computer can then use to display on its screen an image of the scanned archival material; the other, is to allow adjustments to be made to the scanner to change properties of the digital object being produced. For some film scanners these changes can be expressed as choices between differing film formats e.g. between Kodachrome and Fuji. In other types of scanner and cameras these choices use language that is familiar, but has specific meanings in this context.

Before creating any digital objects the automatic features on a scanner should be examined carefully to ensure that they can be turned off if required. This is because some software compensates for things like scratches to an original. From an archival point of view when digitising an item to allow access, these imperfections are needed because they are part of the defining features that make the item what it is. Changing the appearance could constitute misrepresentation of an item, which is one of the objections of some archivists to the digitisation of items.

B4.4.2 Cropping and Rotation

Some scanners have set areas they cover when scanning archival material e.g. the size of an A4 sheet. These areas can usually be identified by either, markings on the bed of a scanner, or through the shape of a mask that may be fitted, to guide where archival materials should be placed. Even with these guides, archival material can still be smaller than the marked area so it may be necessary to crop the scan area to the edges of the archival material. Ideally this should be done on the scan preview so as only the required scan is obtained. Bear in mind that this can also affect the scan outcome as the scanner may be programmed to adjust the levels obtained by a scan automatically.

B4.4.3 Image Perception

When looking at a picture the viewer often subconsciously judges it. Whether the viewer considers it ‘good’ or not is often due to the innate ability to balance, or to fail to balance, the parts that makes a specific picture.

One of the first things people notice is the relationship between the subject and the edges of the picture. Then, possibly, the balance between light and shadow followed by whether the colours look right to them. People automatically make many allowances for the way these qualities change with the passage of the Sun between night and day. When scanning a picture to create an archival quality digital file, adjustments need to be made to be able to record the largest amount of evidence available in the picture, to do this an understanding of the processes involved is needed.

B4.4.4 Colour Theory

The primary colours used to reproduce all other colours depend upon the method being used to produce them. In subtractive mixing, the three primary colours used are magenta, cyan and yellow. Placing all three together results in black being projected back to the eye. Removing, or reducing one of the colours results in that range of hues being taken away from the original black spot. This is the result of the absorption of particular frequencies of light by the colours used.

In additive mixing the colours are mixed by adding them together. Imagine a white piece of paper. By adding paint to it you subtract from the white light you see reflected back from the paper the colours not in the paint. So by painting blue on the paper, you subtract green and red. The primary colours used are red, blue and green. This is the methodology employed by screen based technologies such as television and computer monitors as well as theatre lighting.

It is accepted that the human eye is good at identifying colour when seen in comparison to others even in differing lighting conditions, but less so when seeing them in isolation. This is because humans build a set of references which enable them to compare colours over time. So humans can recognise differences between shades in varying lighting conditions.

The ability to internally adapt to colour temperature and intensity changes can cause problems when trying to capture an image unless a person acts to bring it to the fore of their consciousness. This is because technologically enabled means of seeing e.g. photographic film or the receptors in a digital camera, do not have this level of adjustment facility. The colours surrounding a subject are also important as they affect the saturation of the perceived hue. Extrapolating from this, digitisation projects should strive for neutrally coloured, evenly and consistently lit surroundings to avoid natural human over compensation for apparently pallid colours in the screen image when viewed next to a strongly coloured wall or changes wrought through natural changes in daylight.

Humans have an array of methods for mentally comparing the colours of objects they encounter. In comparison the reproductive ability of man-made devices is limited to three colours, red, green and blue, and any technological input achieved by its manufacturers. This makes achieving an accurate colour balance difficult but worthwhile pursuit.

B4.4.5 Contrast

The contrast in an image represents either the difference between the blacks and whites or the extremes in the amount of colours present. A way of visualising this is to imagine the shadows created on a very sunny day and a cloudy day. On the sunny day the shadows are much darker than on the cloudy day. This change in darkness is the contrast between the amount of light at the Earth’s surface on a sunny and a cloudy day. On a sunny day it is much harder to stand in the sunlight and see the detail in the shadows compared to a cloudy day. This is why if there is little detail in the dark areas of a digital object, reducing the contrast and re-scanning the archival material might allow more detail to be seen.

B4.4.6 Colour Temperature

The colour temperature of a light source is the equivalent colour of light given off by a black body radiator, an ideal source, at a particular temperature in Kelvin. So the colour temperature of any object is the temperature at which it produces a light that exactly matches that given off by the black body radiator. This is important because it explains why ‘white’ is different for different objects. For digitisation projects the main way it can affect a project is through the colour temperature of the light source used by the scanner. As the element in the lamp heats up, the colour temperature changes. This is why it can be important to turn on the scanner and then wait awhile before you start scanning. Many scanner builders get around this by using cold lamps. The design of these lamps is meant to get around this curve of increasing heat by reaching their highest temperature much more quickly than traditional lamps.

B4.4.7 Tone

The tone of a digital object is where on a range between pure black and pure white the contents of a digital object lies. This range of values are often displayed using a type of graph called a histogram. Ideally it should have a smooth appearance, like a classic histogram, without any spikes. Often in digitisation projects these spikes occur when the black and white points in a scan are not set to true black or white for the digital object being created i.e. this can vary from one photograph to the next. This is an ideal, however, and spikes or clipping can occur and may be the best that can be done with the archival material or equipment available.

B4.4.8 Curves

Curves are a graphical representation of the density of the digital object against the range from black to white in the digital object. In other words they are a mathematical drawing of the range of brightness and contrast values over the whole digital object. They have a distinctive ideal shape which can be used in a digitisation project to examine and adjust a digital object to achieve a more balanced digital object.

B4.5 Quality Assurance of Digitised Images

Quality assurance (QA) should be carried out by someone other than the person who did the digitisation. This is because familiarity can lead the person scanning the digital objects to miss something if they carry out the quality checks as well.

It is suggested that each batch of archival materials digitised, is checked off systematically on some sort of record as it is made and checked. This is because if any mistakes are found it is better to discover them while the materials are still to hand.

Having found a mistake there should be a method for recording this e.g. a sheet to fill in describing the fault. Apart from the obvious of providing written proof of a fault, these records can also prove a useful cross-check for scanner operators indicating where they may need extra training or skills updating.

Table B8 and Table B9 show some of the things to look out for on a quality assurance check.

What to look at	What to look for
Archival/storage copy	Look at areas of where the digital object is white. See if these are white because there is no data or because the scanner wasn't adjusted to capture it.
	Look at areas where the digital object is black. See if this is because there is no data there or because the scanner wasn't adjusted to capture it.
	If not sure, look at places like the inside of ears. If these are white/black it is probably because the digital object doesn't have any data in that area.
Check colours are adjusted properly	Look to see if whole picture has the same colour tones as the original material.
Check digital object is cropped accurately	Look closely at the edges. See if you can spot anywhere where the edges are not quite straight or a different colour. Wobbly edges are often spotted because the cropping isn't quite right and it shows as a change in the colour at the edge of the digital object.

Table B8: Table of the Image QA Points

Questions to ask yourself	What to look at
What is the file size	Is the digital object the correct size
Was the scan done at 300ppi	Is the digital object the correct resolution
What is the file name	Is the file name correct
Was the archival digital object stored as a.TIFF	Is the file format correct
What is the colour of the archival material?	Is it the same as the digital object
Can you see beyond the edges of the digital object onto the scanner bed	Does the digital object need cropping
Is the digital object the right way up	Does the digital object need rotating
Does it look the same as the archival material	Does it look right

Table B9: Digital Object Properties to Inspect

B4.6 Authenticity of Archives

Archival authority comes from trust because, like mathematical formulae, it is backed up by logical processes that can be applied successfully to many situations and provide reliable outcomes. Some archivists feel creating digital objects from archival materials violates some of these basic principles and, therefore, the bond of trust between an archive and its users. The principle most often felt to be violated is that relating to the principle of ‘not changing the item’.

Others see new forms of carrying media as being able to be absorbed through the adaptation of current practice to reflect the emerging new paradigms. In this way authenticity is maintained by adopting the spirit of current practices rather than by applying them as dogma.

B4.6.1 Document Authenticity

In many places, an archival digital object is the copy from which all others are taken. The archival material is scanned in such a way as to capture as much of the detail present as possible. This means making adjustments on the fly to accommodate this. Other copies are made according to known presentational parameters such as formatting all the digital objects as pdfs’ so that the text of an item can be searched or as a .jpg for display on a web site. The creation of an archival digital object and its copies are documented. This documentation becomes part of the metadata for the digital object thus proving they are what they say they are.

B4.6.2 Representation

Representation is not the same as authenticity, although they are inextricably linked through the catalogue of the facility’s holdings. The representation of an archive can affect the authenticity of an archive, but the authenticity of an archive cannot affect its representation. This is because authenticity reflects the integrity of the whole archive whereas representation is about only one aspect of an archive.

So what does this have to do with digitising archival materials? Well if the issues of representation and authenticity are ignored then mis-adjusting digital objects could lead to an unintentional eroding of trust in an archive and therefore why it exists.

B5. In More Detail ..... Metadata Writing Phase

The word metadata has several roots. Metadata in this context is used to refer to the data that is generated when digital objects are made. This, metadata is usually accepted as being further divisible into 3 parts: administrative, descriptive and structural metadata. Alternatively, metadata is described by the functions it fulfils. These can be described as: resource description, information retrieval and the management of information. They are collected because this data is used to fulfil various management functions during the daily life of the digital object. In general as much of the process as possible should be automated as collating it all by hand is expensive, if only in a person's time.

B5.1 Outline of Metadata Collection Process

This table, Table B10, shows a suggestion for an outline of the steps to take to collect metadata for a collection of digital objects.

Stage Description	**Notes**
1. Open files of digital objects and look at them.	This gives you an idea of the type and scope of contents.
2. Open a separate document, text or spreadsheet, which ever you are most comfortable with, and create a table of phrases or words you think you will need to use to describe the digital objects.	In particular make a note of specific terms for CPR this is from the Dictionary of Theatre: Terms by Pavis in CPR. This becomes your consistency sheet.
3. Open the first digital object.	
4. Complete each of the columns on the metadata sheet	This is where your list of terms might come in useful
5. Check the data	
6. Save/publish data as appropriate	

Table B10: Table of Metadata Collecting Process

B5.2 Collecting Metadata

Metadata are collected to allow everyday actions, such as finding a specific digital object, to be carried out during the daily life of the digital object and, if it becomes damaged, to reconstruct a copy.

There are 2 actions that are key when collecting this data:

- Keep it up to date, and

- ❖ Keep checking its accuracy

Keeping metadata up to date is important to ensure accuracy and as much care should be taken in its making as in the making of the digital object itself. As outlined in section B4.6.1 the accuracy of the entry of metadata contributes to the authenticity of the items contained within an archive or library. Accuracy is important because faulty decisions, such as delaying conservation treatments, can be made if it is not up to date. In turn these decisions can lead to unalterable events adversely affecting the collection.

One of the functions of metadata is for it to be used as a substitute for the object e.g. when a catalogue is being browsed. Another of the functions of metadata is to enable the long-term management of the digital object. Metadata assists with this through detailing any processes undergone by the digital object. It can also be used to flag up when there are risks to a digital object e.g. through obsolescence, by comparing the file types to a list, such as that on Pronom.

B5.3 What Metadata to Capture

B5.3.1 Administrative Metadata

Administrative metadata is the data that is needed to manage the digital object and includes preservation metadata. The types of data that are typically collected for this function include scanner settings, the date of capture of the digital object and the file format. This is the section that is updated whenever a process is carried out on the digital object. In a lending situation, this data is also included here.

B5.3.2 Structural Metadata

The structure of archival materials is often obvious to sighted users. For those with visual impairment, or users who are remote from the resource, it is less so. For digital objects the physical structure of the archival material needs to be copied.

The structure of, for instance, a script is a very important part of the archival material and this should be repeated in the digital object so as to recreate this. In recreating the page order of a script as well as recreating the physical appearance the relationship between the pages is also recreated. This relationship also recreates meaning inherent in the archival material in the digital object e.g. if the archival material was in large print or on a postcard rather than, for example, plain A4 paper recreating the meaning recreates the subliminal meanings carried by an archival object such as a postcard.

Structural metadata is also used as an aid to navigation. If the pages in a script are randomly arranged then reading from one page to the next is at best extremely difficult.

B5.3.3 Descriptive Metadata

Descriptive metadata is the kind of information found in a traditional catalogue. It fulfils the same function for a digital object and is often used to link the archival material to the digital object.

Table B11 uses the example of the annotated play script, which is a printed book, as an example to illustrate the 3 types of metadata.

	What it does	**Example**	**Function**
Administrative metadata	Describes the features of an object(s) used to manage the object(s)	Who printed the script and when; where it is kept and if it is in good condition. If the script has had conservation treatments or been loaned.	Used to manage an object throughout its life. Includes all data needed for managing preservation
Structural metadata	Describes the internal structure/arrangement of the object(s)	Describes the arrangement of the item or the set of items e.g. the order of the pages of a script or which page an illustration is in that book	Structure is part of the meaning of an object. Especially important for digital objects as structure doesn't have an enduring physical form. Structure is part of meaning
Descriptive metadata	Describes the appearance of the object(s)	Describes what the script is about, what it looks like, the paper it is printed upon and when etc.	Object descriptions help humans identify an object. Also aids attributing meaning to an object

Table B11: Table of Metadata Types

B5.4 Metadata Standards

As indicated in Table B11, metadata is collected to carry out specific functions to help manage the collection. This is because it was realised that the numbers of items comprising a collection make it impossible to do this by remembering individually what actions need to be taken and when.

Having been collected, for humans to be able to use metadata effectively it needs to be re-arranged. This can be done using individual methods, but it is easier for others to understand if the way data is arranged is consistent. The advent of computer systems made this an arrangement issue as they cannot interpret and re-arrange data in an order to suit the individual, as humans can. To facilitate the exchange of data a series of rules governing the presentation and contents of metadata were drawn. Some of these rules are listed in Table B12. These standards show how the different functions of metadata interact to produce information that can be accessed in different forms and formats depending upon the situation. It also shows how the different standards combine to write a complete metadata schema.

	Has Part(s) Called	**Does what?**	**Example**
Metadata Structure standards	Element sets/schemes	Rules for the way the contents are arranged	Dublin Core Metadata Element Set (DCMES)
Metadata Content standards		Rules for writing the contents down e.g. the way a date is written	Anglo American Cataloguing Rules (AACR2)
Metadata Exchange standards	Formats	Rules for the way the contents are sent or received	Marc21
Metadata Values standards	Thesauri/Controlled Vocabulary	Rules for the words used to describe contents e.g. World War 1914 – 1918 rather than WWI.	Library of Congress (LoC)

Table B12: Table of Metadata Arrangement Rules and Applied Standards

The strength of a standard depends upon the intended end-use of schema e.g. the strength of EAD (Encoded Archival Description) is in structural description. Standards ensure that the aspect of metadata being described is consistent across the database being constructed and for others using the same rules. As usual you need to make choices based on your needs. Metadata is written to the fulfil purposes of resource description, information retrieval, managing the resource being described, documenting ownership and authenticity of the resource and to facilitate interoperability. It follows then that decisions about things such as the level of data collection for the metadata should to be made towards the beginning of a project as metadata is being collected from the start.

There are a number of metadata systems that have been assessed as attaining ISO standards e.g. MPEG-7 has ISO 15938. Which is applicable to a situation needs to be assessed carefully as each has its own foibles and areas of expertise. For instance MPEG-7 was developed to provide metadata for audio visual content. If you have a collection that comprises only one type of material then one of these may be applicable. If not then the assessment becomes more difficult as by satisfying one set of needs you inevitably short-change another. There are some accepted schema that through adoption for particular projects e.g. DCMI is used by OAI-PMH, have gained in general use. The intended future uses of digital objects therefore need to be clearly established before a metadata schema can be accurately assessed and adopted.

B5.5 Writing Metadata

Whatever the reason a particular piece of data is recorded it needs to be useable by users. To facilitate this, metadata has writing rules for certain fields. Which these fields are, and what the rules are, depends upon the particular schema and set of rules in use.

Writing rules generally cover things in written language which could easily be expressed in different ways. For instance do you write ‘Mister Smith’, ‘Mr. Smith’ or ‘Mr Smith’ on an envelope? If the last of these is entered into a computerised system unless it is configured to link all three of these together, a search for ‘Mister Smith’ would show no result.

Having decided how to write the data there are also rules on what to write in the data. This is because when describing a resource using different words could again give no result. For instance, if searching for a piece of music by its name it is helpful to know that in a particular database the title is given in the original language. Thus searching that particular database for “*The Blue Danube Waltz*” would give no result, but searching for “*An der schönen blauen Donau”* would retrieve copies of the required piece.

B.5.6 Where to Get the Data From

The data for each element is collected from different sources depending upon the level of the collection being written about, the format the resource is in, etc.

The level at which these descriptions are made is something that should be discussed and written down at the beginning of the project so that relevant data can be collected throughout. These decisions should be documented as they themselves become part of the metadata being collected. This data has future use because it can serve as a reminder of how the project was conducted if it is some time before the next, and it tells those preserving the collection how it was constructed and why particular decisions were taken.

> The depth at which you decide to collect data will determine where, and how much extra research you need to do to to arrive at your level of description. For instance, should you write in depth about each digital object, or provide a more basic description? If you write in depth have you the time, resources and money to allow researchers to discover any extra information needed?

Metadata should be also arranged so that it makes sense to users by having the same data for each item arranged in the same order for a collection. And, if this is transferred, so that this data also makes sense to the new audience it is being transferred to. Some of these fields will contain regulated text, such as the date, but in the others you are trying to collect and present the data that is unique to an item or a collection. This is because sometimes the differences between items or collections are what a user is looking for. An example from the imaginary collection could be the entries made for the poster. Is it a special edition produced for a specific performance? If so this should be recorded in the metadata. The data collected can be physically part of an item or, in the case of digital objects, linked virtually.

> For the annotated text in the imaginary collection; if the Director's notes are difficult to read they could be typed into the metadata. This is the kind of decision that is made on during the pre-scanning phase in most instances, but can also be the result of an on the spot decision.

The digital objects are copies of the physical items they represent so all the descriptive details are the same apart from those that identify it as a digital object. This is probably the most obvious example of the repetition that can occur when writing metadata. If the metadata storage method allows, linking the data fields so that they automatically carry over repeating data entries would mean that the errors that can inevitably creep in can be avoided. In the case of a digital object there should also be a permanent link to the record of the physical item it represents. The specific information that is needed for a digital object; its file size, format and the creation date (or the date it was modified) can be obtained from the file containing it.

There are a number of ways of obtaining this data. You could open each digital object, right click on it, click on properties at the bottom of the list that opens up and then type the data presented into a database or spreadsheet. Another way of doing the same is to use an automated tool such as DROID. DROID, and its associated database PRONOM, are free tools available for download from the TNA web site.

B5.7 Metadata Storage

Metadata should be stored with the digital object where possible. How it is stored depends upon the sophistication and scale of your collection. If your collection is large enough, using separate collection management software could be a solution. If it is only small, a database, drawn using open source software to allow data exchange, could suffice.

B5.8 Some Metadata Schema Examples

B5.8.1 Dublin Core

The 15 element Dublin Core Metadata Element Set (DCMES) was devised to provide a middle ground between professionally drawn catalogues and search engine indexes to provide a description method for the excluded majority of documents.

This provides some of all categories of metadata and is generally accepted as the minimum needed for effective data exchange between facilities. Adopting it as the metadata scheme will allow for the expansion or transfer of data if needed in the future. The facility for exporting metadata can also mean that metadata can be imported too. This can be desirable if accepting a new collection where the metadata has already been written using the same standard.

The adaptability of DCMES has been demonstrated by its adoption, and refinement by specific communities. In the UK the TNA used DCMES as the basis of the minimum acceptable requirements for its Electronic Document and Records Management systems (EDRMs), a specialist type of computer system for managing records.

<u>B5.8.2 METS</u>

METS stands for Metadata Encoding Transmission Standard and as the name suggests it is a protocol for the transfer of metadata. It is a standard that expresses as a complete schema the metadata on a digital object using XML. This use of XML provides a means for constructing crosswalks to facilitate data exchange. METS is an example of a wrapper. Wrappers ‘wrap’ a digital object and its metadata with code that tells a receiving computer how to understand what is inside and how to be able to make it work.

<u>B5.8.3 AACR2</u>

AACR2 stands for Anglo-American Cataloguing Rules 2 and is a standard for writing catalogues. It regulates things such as whether to place a full stop after the word ‘Mr’ or how names should be written in a catalogue entry e.g. it prescribes that the name of the author who wrote “A Midsummer Night’s Dream” should appear as Shakespeare, W. rather than W. Shakespeare. It is used here purely as an example of the consistency these rules give as it makes it easier for a user to scan the detail in a catalogue entry because they will have a mental picture of what the entry they are looking for should look like. It also enables accurate automated searching for specified entries as machines cannot ‘see’ variation.

<u>B5.8.4 MARC21</u>

MARC (Machine Readable Cataloguing Standard) was established to allow the mechanisation of many of the processes associated with cataloguing. Here it is used as an example standard. After evolving through several different generations

MARC21 was designed in 2000 and subsequently adopted widely. Created originally for traditional bibliographic records it has maintained its industry place through adapting to new requirements through the adaptation and introduction of new fields for description.

B5.8.5 ISAD(G)

The General International Standard for Archive Description is a standard devised to achieve consistent description of content and context through the use of rules for organisation. There are paths for crossing between other standards such as EAD as interoperability was also a design feature.

B5.8.6 EAD

Encoded Archival Description schema was devised to turn into machine readable code, archive descriptions and catalogues. Based on ISAD(G) it uses XML to communicate. Its strength is in describing the structure of archives.

B5.9 Quality Assurance Metadata

Like digital objects, metadata should be checked for quality and consistency. The exact method depends upon the system employed to manage the data, but checking for spelling mistakes and consistency can be as simple as getting someone else to read it. Some of the metadata could be checked using a simple check sum, with errors highlighted.

B6. What Happens Next?

B6.1 When the Digitisation Has Been Completed

After the scanning and metadata writing has been completed and the viewing digital objects have been placed on show, then that's it, right? Well, yes and no.

Yes, it is the end of the active conversion phase for those particular archival materials, but now the long-term management of the digital objects needs to be considered. This requires further decisions on exactly how this will be achieved. This is where digitisation moves into the area of curation or preservation. The foci for these activities are the metadata and storage characteristics. One possible method to consider at this stage is the implementation of a preservation metadata set. This is a defined set of metadata that are used to manage preservation of digital objects. Some of the decisions that will inform this will already have been considered while choosing the storage medium and which metadata is collected.

To ensure that the digital object and metadata are kept in good condition, it is suggested you build a set of checks that are executed regularly, so that they become habit, and, do not slip and become forgotten.

B6.2 Metadata

This is where the metadata gathered throughout the digitisation project comes into its own. The data on the digital objects can be used to prompt reports or to organise conservation treatments on the archival materials.

The metadata itself should be checked periodically for internal consistency. One way of doing this is to extend the reports generated by and used for quality assurance.

B6.3 Storage

Here storage includes the software used for access as well as the file format and physical medium used for keeping the digital objects on.

The general conditions for keeping the physical medium and the file format in good health were addressed while making choices earlier. But they also need a regime of regular testing and replacement to ensure you keep your digital objects and metadata in condition.

The kinds of test you may wish to perform will depend upon the exact set up of your system, but some suggestions are:

* Check the file format of the digital object is not corrupted. The reults obtained from a DROID scan can be used for this

* Check that the software for accessing the file format of the digital objects is not becoming obsolete . Pronom can do this for you

* Check that your digital objects can be accessed from where you expect them to be.

B6.4 Evaluation

When a digitisation project has been completed, and before the next, there should be a process that establishes what worked, and what didn't. This discussion should be recorded so that next time you make new mistakes, rather than the same ones again. The format this takes can be adjusted to suit the situation. For instance, if the team works best by sitting informally and discussing the outcomes, then you should do this. If a more structured approach works best then perhaps the adoption of a questionnaire could be considered.

B7. General Pages

This chapter lists some of the possible some sources of help that are available. These lists are not intended to be exhaustive but should be used as a spring board for your own enquiries.

B7.1 Possible Sources of Help

Table B13 and the bibliography following it list of some of the places where you can find help on all aspects of digitisation.

Type of Issue	**Source**	**Web Address** (checked 16/10/2011)
Sources of general digitisation and preservation advice	Facet are the publishing arm of the UK professional association for librarians (CILIP)	http://www.facetpublishing.co.uk/
	Focal Press publish books that cover topics such as photography	http://www.focalpress.com/
	JISC	http://www.jiscdigitalmedia.ac.uk/
	DCC	http://www.dcc.ac.uk/
	DPC	http://www.dpconline.org/
	TNA	http://www.nationalarchives.gov.uk/information-management/default.htm
	UKOLN	http://www.ukoln.ac.uk/
	Library of Congress	http://www.loc.gov/
	National Library of Wales	http://www.llgc.org.uk/index.php?id=3985
	British Library	http://www.bl.uk/
Professional Associations (log-in needed for some pages)	Chartered Institute of Library and Information Professionals (CILIP)	http://www.cilip.org.uk/Pages/default.aspx
	Information and Records Management Society (IRMS)	http://www.irms.org.uk/
	Archives and Records Association	http://www.archives.org.uk/

Table B13: Table of Possible Sources of Help

B7.2 Some Possible Bibliographic Resources

Bulow, A., & Ahmon, J. (2011). *Preparing Collections for Digitisation*. London: Facet.

Dublin Core Metadata Initiative. (2011). *Dublin Core Metadata Element Set, Version 1.1*. Retrieved 06/06/2011, from Dublin Core Metadata Initiative: http://dublincore.org/documents/dces/

Haynes, D. (2004). *Metadata for Information Mangement and Retrieval.* London: Facet.

Langford, M., & Bilissi, E. (2011). *Langfords Advanced Photography: The Guide for Aspiring Photographers* (8th ed.). Oxford: Burlington: Focal Press.

NSW HSC . (2011). *Information Processes and Technology: Calculating Image File Size*. Retrieved 08/08/2011, from NSW HSC Online: http://hsc.csu.edu.au/ipt/mm_systems/3289/image_file_size.htm

Padfield, T. (2010). *Copyright for Archivists and Records Managers* (4th ed.). London: Facet.

The National Archives. (n.d.). *Caring for your Photographic Collection.* Retrieved 16/03/2011, from TNA: http://www.nationalarchives.gov.uk/documents/archivesconservation_photo.pdf

W3C Schools.com: http://www.w3schools.com/

Western States Digital Standards Group; Digital Imaging Working Group. (2003). *Western States Digital Imaging Best Practices: Version 1.0.* Retrieved 12/03/2011, from Western States Digital Standards Group: http://www.mndigital.org/digitizing/standards/imaging.pdf

B7.3 Some Risks to Archival Materials

This group of tables list some of the things to be considered when handling different formats of archival materials. This list is just a start, and can be expanded as you digitise more formats of archival materials.

Risks to Archive Materials From Digitisation	
Type of Material	**Risk**
Loose papers/vellum/ parchment (all types)	Damaged edges Can have fastenings at edges Size of item could be large e.g. a map Loss of context through being replaced incorrectly into file Folds can be weak May not lie flat May require handling with gloves to prevent damage
Books	May contain blank pages. Need to be kept to maintain context May be difficult to open sufficiently for some scanners – needs software that corrects for curvature of the page to generate results that can be scanned using OCR readers May require handling with gloves to prevent damage from skin oils May need to be placed on a special cradle to prevent damage to binding
Photographs & slides	Need careful handling to prevent damage to original. Gloves should be worn for items not made from modern plastic film to prevent damage from skin oils

Table B14: Table of Risks to Archive Materials from Digitisation

Table of Risks Due to Obsolescence		
Type of material		**Risk**
Photographs	Non-safety film negatives	Examples of problems include vinegar syndrome and items that burst into flames on contact with the oxygen in air
	Slides	Some projectors not being manufactured any more. Therefore risk of material becoming un-useable as intended
	Prints	Prints can be damaged through use, and fewer facilities are available for processing
	Plastic film negatives	Not many film cameras being sold and therefore processing facilities less common
Video	VHS	Few players available. Therefore risk of material becoming un-useable.
	Beta-max	Few players available. Therefore risk of material becoming un-useable.
	C+	Few players available. Therefore risk of material becoming un-useable.

Audio Recordings	½" tape (2, 4 or 8 track)	Few players available. Therefore risk of material becoming un-useable.
	Cassette tape	Few new players available. Therefore risk of material becoming un-useable.
	Record – plastic	Few new players available. Therefore risk of material becoming un-useable.
	Record – Shellac	Few players available. Therefore risk of material becoming un-useable.

Table B15: Tables Showing Formats and Threats to Them

B7.4 Table of File Formats and their Usual Applications

This list contains some of the most popular choices of digital file format for specific media types.

File Format Stem (Name)	**ISO number**	**Media Used For**	**Why**	**Suitable for Master/ Access Copies (M/A)**
.TIFF *(Tagged Image File Format)*		Still Images	Flexible, widely adopted. Lossless storage files therefore can recreate all of saved picture. Specification publicly available. (Langford & Bilissi, 2011, p. 219)	M
.JPEG2000 [JP2] *(Joint Photographic Experts Group)*	ISO 15444-2	Still images	Combination of both lossless and lossy files. Platform independent. Specification publicly available. (Langford & Bilissi, 2011, p. 219)	?/A
. PDF/A-1 *(Portable Document Format)*	ISO 19005-1	Text documents	Open Source, Platform independent. (NDIIPP, 2011)	M/A
.MPEG4 *(Moving Picture Experts Group)*	ISO 14496	Audio Visual	Convergence across technical production paradigms. (Koenen, 2002)	M/A
.MP3	ISO11172-3	Audio	Lossy, compressed audio format. Open standard. Widely used. (JISC Digital Media, 2010b)	A

File Format Stem (Name)	ISO number	Media Used For	Why	Suitable for Master/ Access Copies (M/A)
.WAV		Audio	Uncompressed audio files. (JISC Digital Media, 2010b)	M
.RAW	ISO 12234-2	Stills or Film camera unprocessed files	Stores widest possible range of information from source camera, before processing to useable format (a bit like undeveloped film). Original metadata remains after processing, so changes can be reversed. (Langford & Bilissi, 2011, p. 209)	M
.ODF *(Open Document Format)*	ISO 26300	Documents including spreadsheets and databases/	Specification widely available (Digital Curation Centre, 2011).	M/A

Table B16: Table of File Formats & Common Uses

B7.5 Some Forms That Can be Adapted for Planning a Project

The forms on the next few pages are intended to be used as an aid to planning a digitisation project or as the basis for making up your own.

The fields and values in the forms below are only intended as a guide to the length of a digitisation project because they do not include extra time for problems encountered.

<table>
<tr><td colspan="2">Collection Title: e.g. The Dragon's Playtime</td></tr>
<tr><td colspan="2">Collection Code: e.g. GB3261/CPR/CPR/CLT/The_Dragon's_Playtime</td></tr>
<tr><td>Today's Date:</td><td>Name of Person Filling Form:</td></tr>
<tr><td>Expected Start Date: e.g. 01 Jan 2011</td><td>Projected End Date: e.g. 01 Dec 2011</td></tr>
</table>

Scanner Make & Model: Epson Pro V750 ☐ Bookeye ☐	Computer: iMac ☐ PC ☐	Imaging Software: EpsonScan ☐ BCS-2 ☐
Metadata Collection: CPR Spreadsheet ☐ Eclap ☐	Digital Object Storage: External hard drive ☐ University Network ☐	Image Adjustment Software: Photoshop Elements 9 ☐
Name of Organiser:	Name of Scanner Operator:	Name of Metadata Writer:

<table>
<tr><td>Section Name:</td><td>Section 3 Letter Code (e.g.tdp):</td><td>Format Archival Material:</td></tr>
<tr><td colspan="3">Time to scan: 4 slides on Epson scanner is approx. 15 minutes and for 1 photograph (6"x4") at 300dpi is approx. 5 minutes.</td></tr>
<tr><td>Number Items in This Format:</td><td colspan="2">Time to Scan Format (= No. items X time taken):</td></tr>
</table>

(repeat the above format section for each format of the archival material to be digitised)

Project Digitisation Time	
1. Total Time to create digital objects (scans) = time for format 1 + format 2 etc	(mins)
2. Convert this time to hours by dividing the time in ① by 60	(hrs)
3. For each hour of scanning time in ②, add 0.25 (15 minutes) time to adjust the scan preview	(hrs)
4. For every 10 scans created, add 0.5 (30 minutes) to the time in ③ to create the viewing copies	(hrs)
5. For each hour in ④ add 0.2 (2 sets of 5 minutes) rest from the screen time. This is the total time to create the digital objects.	(hrs)
6. To work out how many days this is, divide by 7.5. 7.5 are the total number of working hours in a day.	(days)
7. It is generally accepted that the time to create metadata is 2x the time to create the digital objects. To calculate this multiply ⑥ by 2	(days)
8. The total number of working days to create digital objects and write metadata is = ⑥ + ⑦	(days)

Table B17: Project Planning Table

Appendix C: **CPR Quick Start Pages**

These pages are written for CPR personnel and are intended to be used as a day to day guide on what to do when to digitise the archival materials in your project. It covers the use of the three scanners available in CPR and writing metadata into their spreadsheet. In short it is a more of a 'do this, do that' set of instructions with brief back ground notes, rather than a treatise on practice. For this reason more information upon the why and wherefore of what you are doing is in the separate '… in more detail' pages in the Generic and Organiser's Guide.

For consistency an imaginary collection is used throughout the guide. This collection comprises:

- a copy of a script for an imaginary play, with 2 scenes, called "Dragons Playtime" annotated with directors notes
- 100 production photographs, and
- 3 digital objects for each photograph and the annotated play script

Figure C1 shows the outline of the workflow being described in this guide.

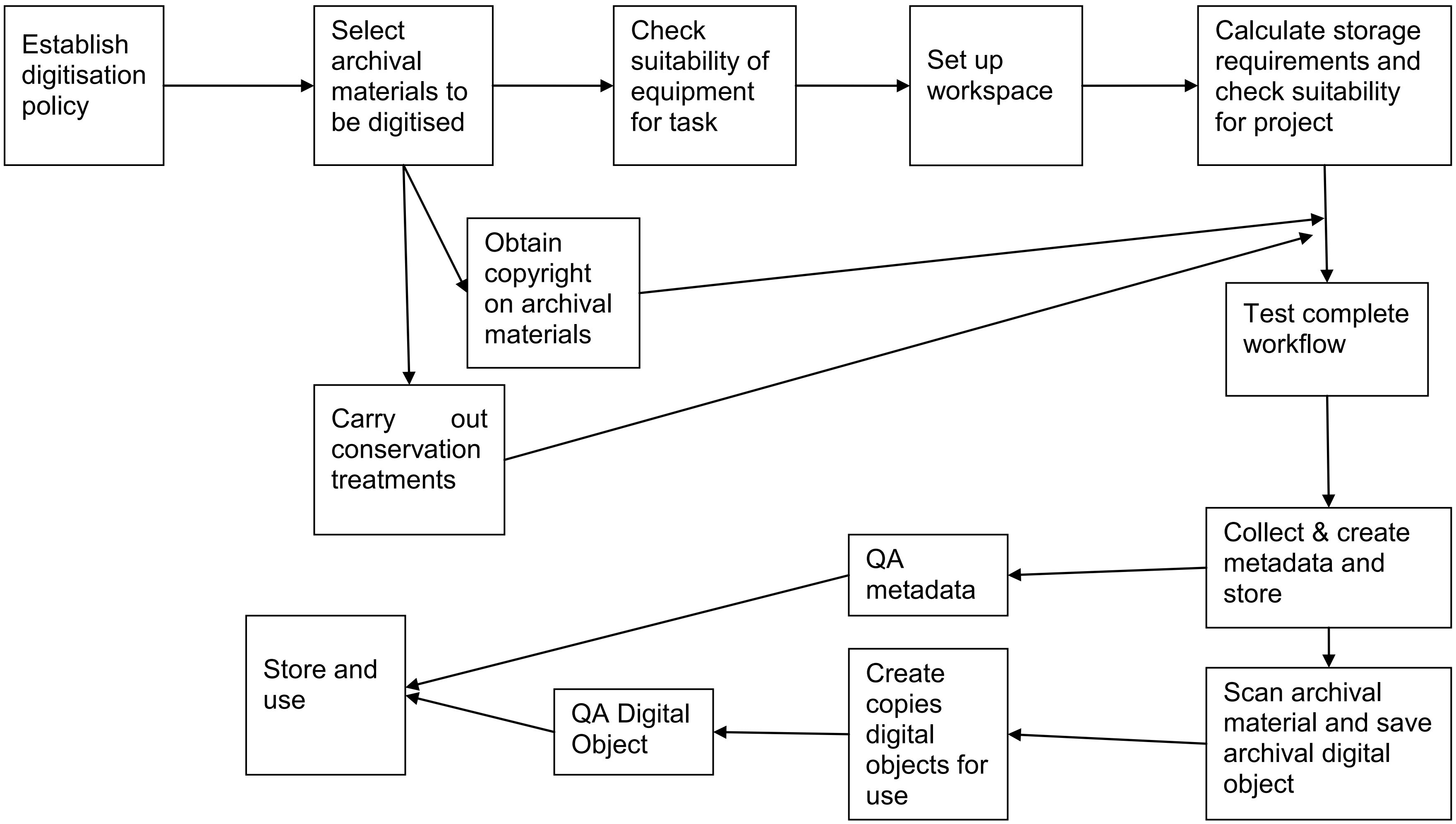


Figure C1: Outline Digitisation Workflow

C1. Quick Guide to ….. Metadata

Metadata in this context it is used to refer to the data that is generated when digital objects are made. They are collected because this data is used to fulfil various functions during the daily life of the digital object and, if it becomes damaged, to reconstruct a copy.

There are 2 actions that are key when collecting this data:

- ❖ Keep it up to date, and
- ❖ Keep checking its accuracy

Keeping metadata up to date is important to ensure accuracy and as much care should be taken in its' making as in the making of the digital object itself. As outlined in the '….In More Detail' pages of the Generic and Organiser's Guide, accuracy is important because faulty decisions, such as delaying conservation treatments, can be made if it is not up to date. In turn these decisions can lead to unalterable events adversely affecting the collection. Another function of metadata is for it to be used as a substitute for the digital object e.g. when a catalogue is being browsed.

C1.1 Writing Metadata

As can be seen from the workflow, Figure C1, the first thing that needs to be done is to record the collection ID, the collection name and its 3 letter code in the spreadsheet.

- ❖ Open the spreadsheet titled 'Metadata' using Open Office. This is currently stored in the folder labelled 'Metadata' on the desktop
- ❖ Check the sheet that is open is labelled 'Archival_Material'.

To begin collecting archival material metadata:

- ❖ Fill in the collection details in the first 3 columns in English.
- ❖ The first column labelled 'dc.identifier' is where the CPR ID is written. This consists of at least 4 sections. See Figure C2 and Table C1, for a guide on what to write where.

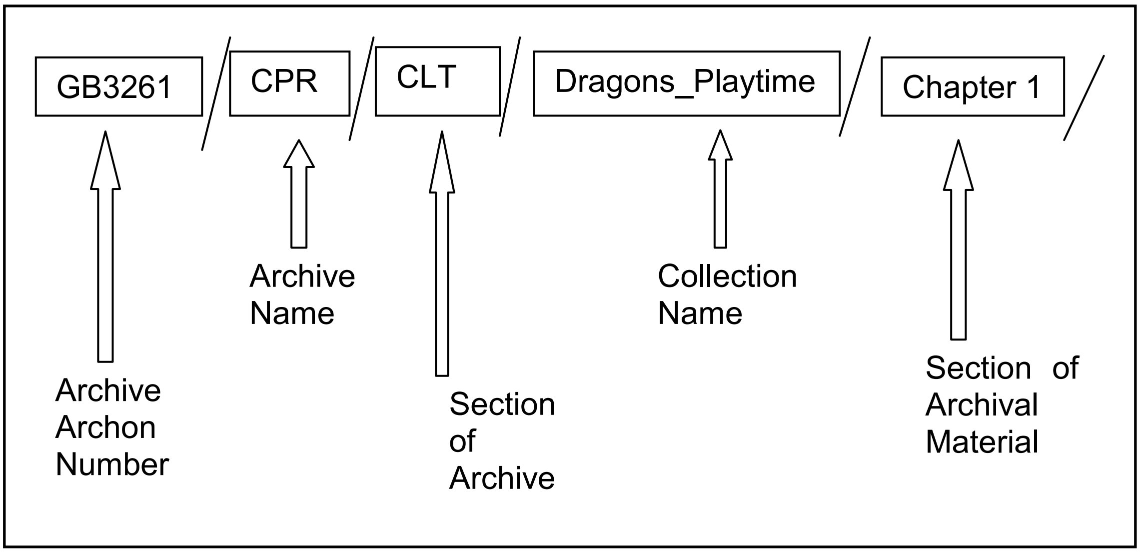


Figure C2: How to Write the dc.identifier

Column (field) 1	Column 2	Column 3
dc.identifier	dc.title	dc.titleAlternative
CPR ID	Collection title	Section title
GB3261/CPR/CLT/Dragons_Playtime	Dragons Playtime (dpt)	Scene I
GB3261/CPR/CLT/Dragons_Playtime	Dragons Playtime (dpt)	Scene II
GB3261/CPR/CLT/Dragons_Playtime	Dragons Playtime (dpt)	Production Photographs

Table C1: First Three Columns of Metadata Spreadsheet

- ❖ The title is typed as it appears on the archival material. After the title include in brackets the 3 letters that will become the stem of the identifier for all formats of the archival material. This can be taken from the title, but needs to be unique. A list of initials used is also in the metadata folder with the example spreadsheet and should be added to each time a project is started.
- ❖ For the example shown, the individual pages in the script would be identified by the individual identifying number given to each page.
- ❖ The metadata spreadsheet is coloured with different colours, Table C2, to make it easier to tell where you are and each row contains the information

about the different formats for 1 piece of archival material or in the case of the script, 1 page.

Colour of Metadata Sheet Section	What is Being Described in Section
Light blue	Archival materials
Yellow	Archival digital object
Purple	Viewing digital object
Pink	Thumbnail digital object

Table C2: Table of Metadata Sheet Colours

- ❖ After the first 3 columns the next one to fill in has a blue background and is labelled 'dc.description.tableOfContents'. Into this field type the individual identifier, or accession number, for each piece of archival material.
- ❖ The individual identifier is the 3 letter code for the collection + 5 digits e.g. if the first group of items written down in the imaginary collection are the photographs the first will be labelled dpt00001, the second dpt00002 etc.
- ❖ The maximum number of scans in a section should be set at a number that fits in with the collection e.g. the 100 photographs will break down easily into sections comprising 20 or 25 file numbers. When entering these digits into the spreadsheet, after the first 25 a gap should be left in the numbering. This is so that:
  - o If it is discovered that an extra scan needs to be inserted during the process it can be placed close to any others it is associated with, and
  - o The digitisation process is broken down into sections that are more easily managed physically.
- ❖ This gap should be at least 5 digits and re-start on a 0 (zero) to make searching for the beginnings and endings of sequences easier later in the process.
- ❖ If writing more than one section, or batch, e.g. 2 chapters of the script leave a gap of 5 rows between them; again to make navigation of the sheet easier. They can be removed after metadata collection is complete.
- ❖ Then move along to the column coloured yellow and labelled 'dc.relation.isVersionOf' and type the same individual identifier into it. To make this easier you can choose to hide the columns in between. To do this,

- Click on the letter at the top of the column you want to hide and hold it down while you drag across the columns to highlight any more. When all the columns you want to hide are highlighted,
- Click on ‘Format’ in the toolbar
- Hover over ‘Hide &Unhide’
- Slide across
- Click on ‘Hide Columns’. This will hide all the columns that have been highlighted.

❖ Copy the same information into the columns with the same heading, but coloured purple and pink,

❖ To unhide the columns
  - Click and drag across to highlight the columns either side of the hidden ones
  - Click on ‘Format’ in the tool bar
  - Hover over ‘Hide & Unhide’ in the menu that appears
  - Slide across to the next menu
  - Click ‘Unhide Columns’. This will make the hidden columns re-appear.

❖ When finished save one copy, using Open Office.

❖ Open MS Excel

❖ Click on the Orange ‘Home’ button top left

❖ Choose the folder symbol marked ‘Open’

❖ And re-open the metadata spreadsheet you have just closed using MS Excel

❖ Click ‘Save As’, and choose the option that allows you to save the document as an Excel spreadsheet

❖ Pass the MS Excel copy to the Scanner Operators so that they know how many scans there are and what to save them as.

❖ If you already have the data, e.g. whether photographs are processed in black and white or colour, continue completing the information in the relevant columns in the Open Office version of the metadata spreadsheet.

For more information on this topic see the In More Detail ....for Metadata Writers Section B5 in the Generic and Organiser’s Guide.

C2. Quick Guide to …. Operating a Scanner

What you are trying to do is create on a computer screen a copy of the archival material that looks as close as possible to the original. This is done by creating a copy that captures as much detail as possible from the archival material and then making two more copies adjusted to look like the original.

Equipment Required:

- A scanner: there are 3 possible scanners in CPR. These are the Epson Perfection V750 Pro Professional Photoscanner, the Canon Canoscan F800 or the Bookeye Color N2
- And its driving software: these are EpsonScan for the Epson, Arcsoft for the Canon and BCS-2 for the Bookeye scanners respectively.
- Plus a computer to view, adjust and store the scans: in CPR these are currently an iMac that drives the Epson scanner and a PC that is used for writing metadata and has the drivers for the Bookeye scanner.

This section has instructions on the operation of all three scanners for the archival materials you are most likely to come across. The first section has an outline of the process and some general tips on how to achieve consistent digital objects. The following three sections have instructions for the individual scanners while the fourth describes how to use Photoshop to produce the additional copies of each digital object.

C2.1 Pre-Digitisation Processes

For all archival materials:

- Before you start scanning, check the workspace is laid out correctly for you.
- Turn off the lights above the computer and close the window blinds
- Place the archival materials to be digitised close by so that they can be moved onto the scanner easily.
- Switch on the computer. On iMac’s this button is on the back on the bottom left. It can be found by sliding the hand underneath and sliding across the back until an indentation is found.

- If the computer has a message window saying that it cannot detect a mouse or key board turn these on too. The keyboard and mouse use 2 AA batteries and operate wirelessly. The button for the keyboard is on the right hand end; and on the mouse when the button is on a green panel is in view.
- Open your email
- Open the metadata spreadsheet that has been sent to your email account using Excel for Mac. This is usually found at the bottom of the screen and is coloured green.
- Create a folder on the computer desktop named with the 3 initials of the project name
- Inside this folder create 2 more. One with the 3 collection initials + 'viewing' and the second the 3 initials + 'thumbnail'
- Switch on the scanner
- When placing archival materials on the bed of the scanner use the guides along the side of the bed, or the side of the scanner bed where no obvious marking exist, to get the archival material level
- Start the scan preview
- Repeat the adjustments for each preview until you are happy. If more than one can be done at once, repeat until all are complete.
- Do the scan for real
- Set up the automatic numbering
- Save the newly created digital object to the designated folder using the agreed naming schema
- Repeat for the rest of the collection

C2.2 Using the Epson Scanner

- Switch on the scanner using the button on the bottom left, front.

C2.2.1 For Photographs:

- Lift the lid and check the document mat (the piece of black plastic with a white squishy side) is in place. If not place this into the lid. All 4 cups slot into place at once i.e. the plate doesn't slide into place from top to bottom as you might expect.

- Place the photograph on the scanner bed. The mark to line-up next to is located in the right hand corner next to the hinge when standing at the short open end.
- Close the scanner lid
- If there is a symbol that looks like an open scanner in the dock on the computer screen (the bar along the bottom of the screen), click on this to open the scanner software. If this symbol isn't there then:
    - Click on the square, blue face symbol (Finder), then 'Applications' and choose 'Epson Scanner' from the list to open the software
- On the 'Epson Scan' page choose the options listed for each of the operations detailed:

Operation	**Choice**
Mode	Professional
Name	Current setting
Original:	
Document Type	Reflective
Document Source	Document Table
Auto Exposure Type	Photo
Destination:	
Image Type	24-bit Color
Resolution	300dpi
Target Size	Original
UnSharp Mask	☑
Rest of choices	Blank

Table C3: Table of Options to Choose for Photographs

- Click 'Preview'
- When the preview scan appears on the screen adjust the digital object using the buttons above the 'Unsharp Mask' box (histogram, tone and color palette) by removing the photograph from the scanner.
- Look at each preview individually and crop the edges as this can affect the scan. Ask yourself:
    - Are there any areas where there is no detail, but could or should be?
    - Are the colours the same as the archival material?
- Replace the photograph the same way round
- Click the button labelled 'Scan'
- A window opens with a series of options for saving the digital object

- Next to the ‘Scan button is a small folder symbol. Click on this
- In the ‘Location’ box choose other and navigate to the folder set up on the desk top at the start,
- Highlight and click on this. Then,
- Click ‘Choose’
- For ‘File Name’ fill in the three letter code + three zeros e.g. dpt000 and
- Set the number in the right hand box to the number of the next digital object e.g. 01. This would give a digital object number of dpt00001for the next scan completed
- For ‘Image Format’ choose ‘Type’ .TIFF and
- Click ‘OK’. This returns you to the previous page. Now
- Click ‘Scan’. It will rescan the photograph using the new settings and save to the designated folder
- The folder it has been saved to will open when the scan has finished. Close this by clicking on the red button with a cross, top left.
- Replace the photograph with another
- Click on ‘Preview’ to start the scanning process again
- When the batch is complete
- Go to section C2.5 on page 275.

C2.2.2 For Slides:

- Decide which side on the slides has the emulsion on. Place the slides with the emulsion side upper most. The side of a slide that has the emulsion on is the side that looks dull or less shiny.
- Remove the document mat in the lid of the scanner so that the top and bottom glass is visible. Store this mat flat and so that the white piece doesn’t become marked
- The slide holder is kept in the box labelled ‘Computer Stuff’.
- The holder has a several symbols on the side that faces upwards in the scanner. One is a square with abc written as though in a mirror and a second symbol next to it with a line through the abc viewed the correct way round.
- This indicates that slides need to slide into the holder with the correct view side, downwards (or emulsion side upwards).

- Carefully load this holder with slides in the correct individual number order. The order the scanner will read the slides is A1, A2, A3 etc. as shown in Figure C3, the mock up of how the scanner copies slides.

Hinge on scanner lid			
D1	C1	B1	A1
D2	C2	B2	A2
D3	C3	B3	A3
D4	C4	B4	A4
ON/OFF Switch			

Figure C3: Diagram of Order Scanner Will Read Slides

- On one side of the holder are 2 tabs, painted with white triangles, that stick out. These triangles marry up, point to point, to two on the side of the scanner bed next to 2 slots.
- Place the little feet on the reverse side of the holder tabs into the slots on the scanner bed.
- Close the scanner lid
- If there is a symbol that looks like an open scanner in the dock on the computer screen (the bar along the bottom of the screen), click on this to open the scanner software. If this symbol isn't there then:
    - Click on the square, blue face symbol (Finder), then 'Applications' and choose 'Epson Scanner' from the list to open the software
- On the 'Epson Scan' page choose the options listed for each of the operations detailed:

Operation	**Choice**
Mode	Professional
Name	Current setting
Original:	
Document Type	Film (with Film Holder)
Film Type	Positive
Destination:	
Image Type	24-bit Color
Resolution	3200 dpi
UnSharp Mask	☑
Rest of choices	Blank

Table C5: Table of Options to Choose for Slides

- Click ‘Preview’
- When the preview scans appear on the screen double click to enlarge them individually. If they are the wrong way round use the buttons labelled ‘Frame’ to adjust them to turn them the correct way.
- Look at each preview individually and crop the edges as this can affect the scan. Ask yourself:
    - Are there any areas where there is no detail, but could or should be?
    - Are the colours the same as the archival material?
- If necessary adjust the digital object using the buttons above the ‘Unsharp Mask’ box (histogram, tone and color palette) by lifting the scanner lid and comparing to the original slide. When satisfied,
- Click the button with lots of small squares to return the scans to their original size
- Ensure each preview has a tick underneath to select it for scanning
- Click ‘Scan’.
- A window opens with a series of options for saving the digital object
- Next to the ‘Scan button is a small folder symbol. Click on this
- In the ‘Location’ box choose other and navigate to the folder set up on the desk top at the start,
- Highlight and click on this.
- Click ‘Choose’
- For ‘File Name’ fill in the three letter code + three zeros e.g. dpt000
- Set the number in the right hand box to the number of the next digital object e.g. 01. This would give a digital object number of dpt00001
- For ‘Image Format’ choose ‘Type’ .TIFF
- Click ‘OK’. This returns you to the previous page.
- It will scan the slides using the new settings and save to the designated folder
- The folder it has been saved to will open when the scan has finished. Close this by clicking on the red button with a cross, top left.
- Remove the slides from the holder and replace in the box taking care to put the correct slide on the correct square
- Close the preview by clicking on ‘Preview’ again,
- Repeat the process until all the slides in the batch are scanned and stored.
- Go to section C2.5 on page 275.

C2.3 Using the Bookeye Scanner

The details for this section are taken from the Bookeye Scanner Handbook and the BCS-2 Handbook where appropriate. There is a hard copy of each of these manuals with the scanner in CPR.

- When entering the room, it has been discovered through experiment that it is better to, close the blinds and turn the lights off. This stops too much light falling on the bed of the scanner and making the digital objects faint.
- Switch on the attached PC first.
- Switch on the scanner by flicking the switch on the rear, right hand-side and then hold down the 'start' button on the front for a few seconds. When the scanner is ready it displays a 'READY TO SCAN' message in the small panel to the right of the buttons (keyboard). The buttons below the scanner bed also light up with the automatic settings shown.

FORMAT	TYPE	PAGES	COLOR	BRIGHTNESS	CONTRAST	OUTPUT	COPIES
A3	FLAT	LEFT	COLOR	AUTO	AUTO	Not lit	1

Table C6: Table of Scanner Default Button Settings

What each of the labels for the each of the functions printed on the scanner means, are shown in table C7 below.

FORMAT	Measured from centre. Marked on bed and along edge nearest buttons.
TYPE	FLAT = *must* touch bar next to buttons and be *placed centrally*
	BOOK = books, has *automatic fold correction.* Must be placed away from bar next to buttons
	FOLDER = scans left & right separately
	AUTO = no functions
PAGES	SINGLE = material scanned as one digital object
	LEFT/RIGHT = one side only is scanned
	BOTH = both sides scanned and 2 digital objects created
	AUTO = no functions
COLOR	COLOR = scans using 24-bit color
	GRAY = scans using 8-bit gray scale
	PHOTO = scans in bitonal and carries out a dithering!!!!!
	B&W = scans using black and white only
	AUTO = no functions
BRIGHTNESS	Automatically controls brightness during scanning
CONTRAST	Automatically controls contrast during scanning
OUTPUT	Choose output size for connected printer
COPIES	Choose how many copies are to be printed

Table C7: Table of Scanner Button Functions

These settings can be used to make changes for individual archival materials before they are scanned; or they can be why a scan does not come out as expected. It should be noted that each time the scanner is turned off it returns to its default settings.

C2.3.1 Check Scanner Adjustment

At the start of a new project check the adjustment of the scanner by using the charts supplied.

- Place the chart with the black and white lines onto the scanner bed
- Click on 'Network Scanner Settings' to get to a screen that looks like Figure C4.
- Click on the 'Advanced Settings' button at the bottom of the page.
- Click on the 'Settings' button and follow the instructions to complete test

C2.3.2 Creating Temporary Storage

Having completed this check, you return to digitising the archival materials

- Create folders for temporary storage of the digital objects on the desk top.
- On the desktop, double click on the icon (picture) for the BCS-2 software. The window should open showing a large space with rows of icons above. Reading from the top there are 3 bars (rows) before the large display space. The lowest toolbar starts with a button labelled 'SCAN'. This row is where you make most adjustments to the scanning process.
- Click on 'Scanner' in the top tool bar and check that there is, or add, a tick next to 'Network Mode' and 'Bookfold Correction' in the list as shown below.

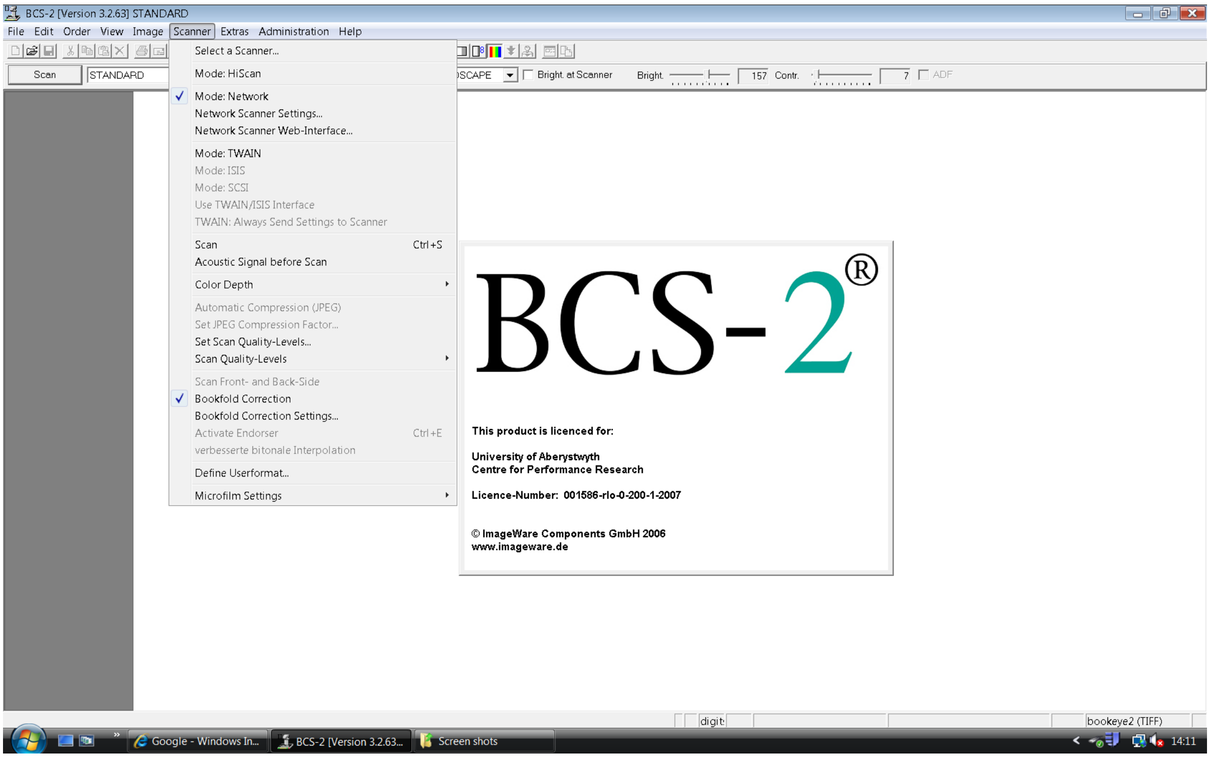


Figure C4: Initial Settings Window

For new projects the next step is to create the settings required for a batch of digital objects.

Setting Up a Profile/Batch Settings

- Click on ‘Network Scanner Settings’ to get to a screen that looks like the one below, Figure C5. In this window ensure that the option button next to.TIFF is ticked in the file format box top right and next to ‘Page Splitting’ choose ‘Auto’.
- Then that in the lower left ‘Mode’ portion ‘Book’ has a tick next to it. Click on OK.

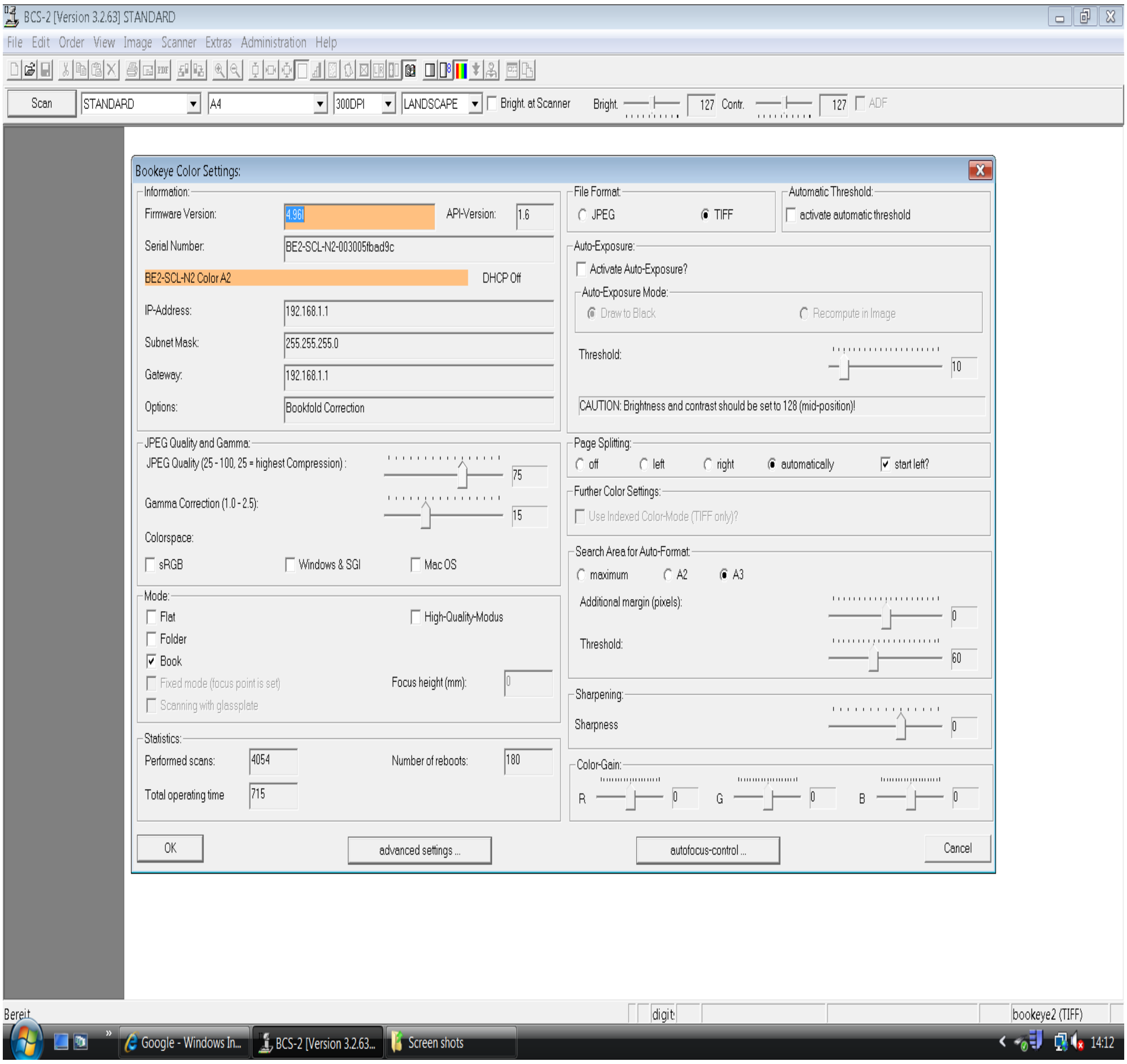


Figure C5: Network Scanner Settings Window

To scan a batch of materials you first need to set up the ‘Options Dialog’. To do this:

- Click ‘ORDER’ in the top tool bar.
- Then in the menu that opens click ‘OPTIONS’. This should open a window with a number of options. Which to choose are shown in the screen shot Figure C6.

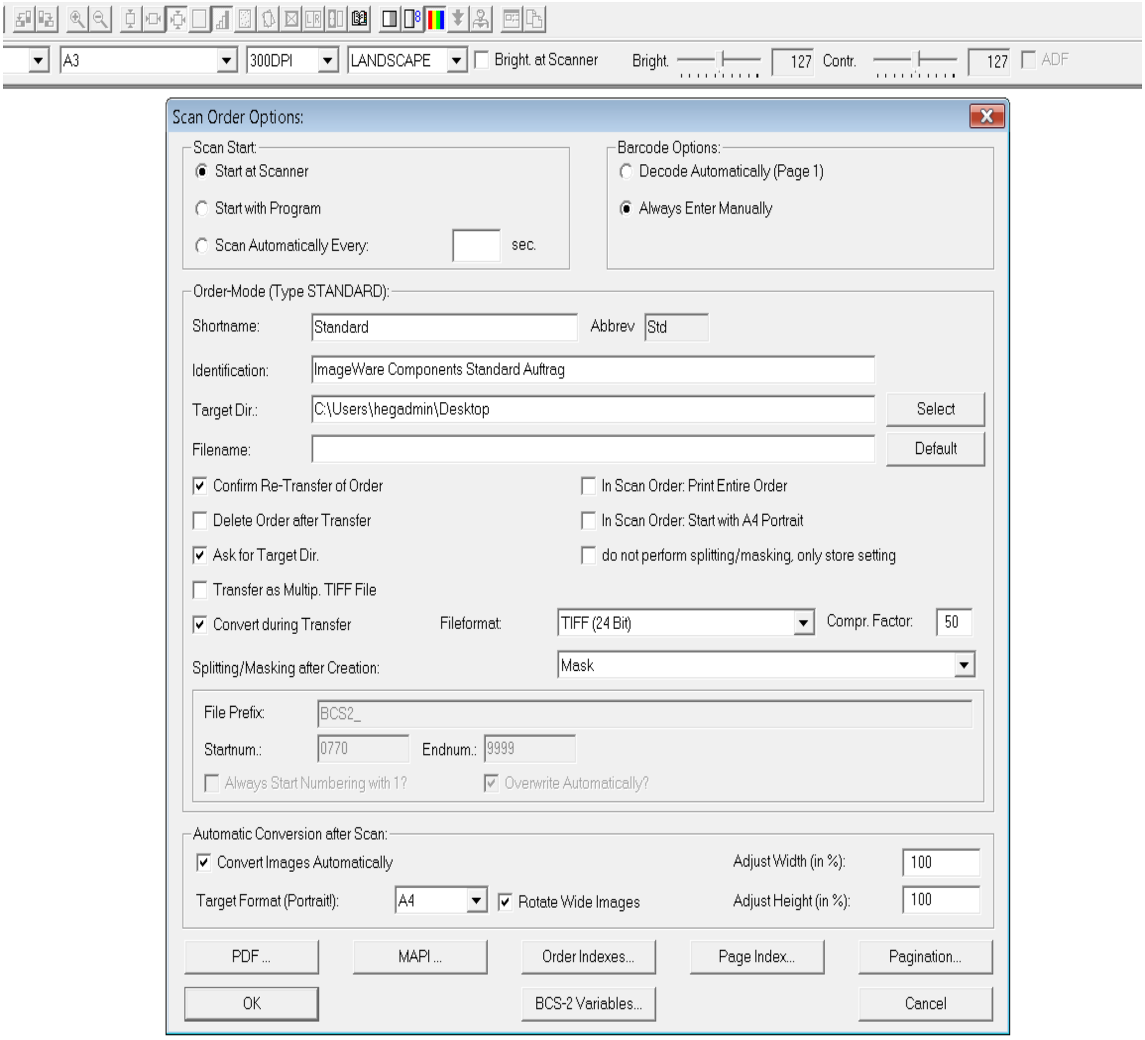


Figure C6: Options Dialog Window

- Then click on the ‘page index’ button at the bottom of the same window. Tick the ‘page index’ box.
- In the description box type the archive file name + the first 3 letters of the production name e.g.
- Tick ‘use prefix’ and add to the formula e.g. Type ‘%Dragon: Act I: Scene II: 3d’ this will add ‘Dragon: Act I: Scene II: 001’ as the file name to the first digital object and 002 to the second, etc.
- If doing a colour scan check that it is selected on the scanner buttons or by clicking on the button in the top grey bar with 3 stripes of colour. For a black and white scan choose ‘GRAY’ on the scanner functions or on the tool bar.

- This is where any other adjustment to the setting buttons e.g. removing all the contrast, discovered during testing should be made.
- Then *click* on 'ORDER' then on 'CREATE'. This uses the settings from above to create an order and the lowest tool bar on the screen changes to include extra boxes showing the order details.
- In the first window that opens, Figure C7, type the project name (e.g. digitisation projects) you have set up in the top box and leave the second blank. Click 'OK'.

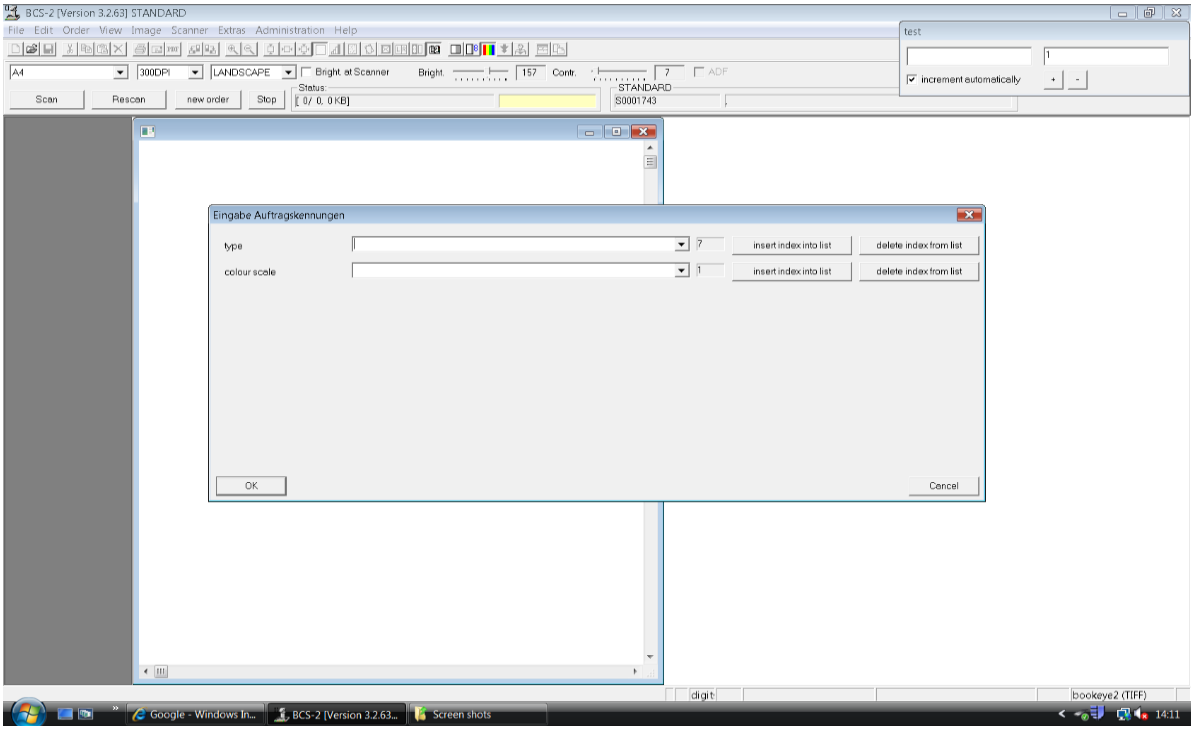


Figure C7: Project (Order) Name Window

- This opens a second window, Figure C8. Fill in the name of the specific batch/order, e.g. Act I: Scene I, you are about to process.

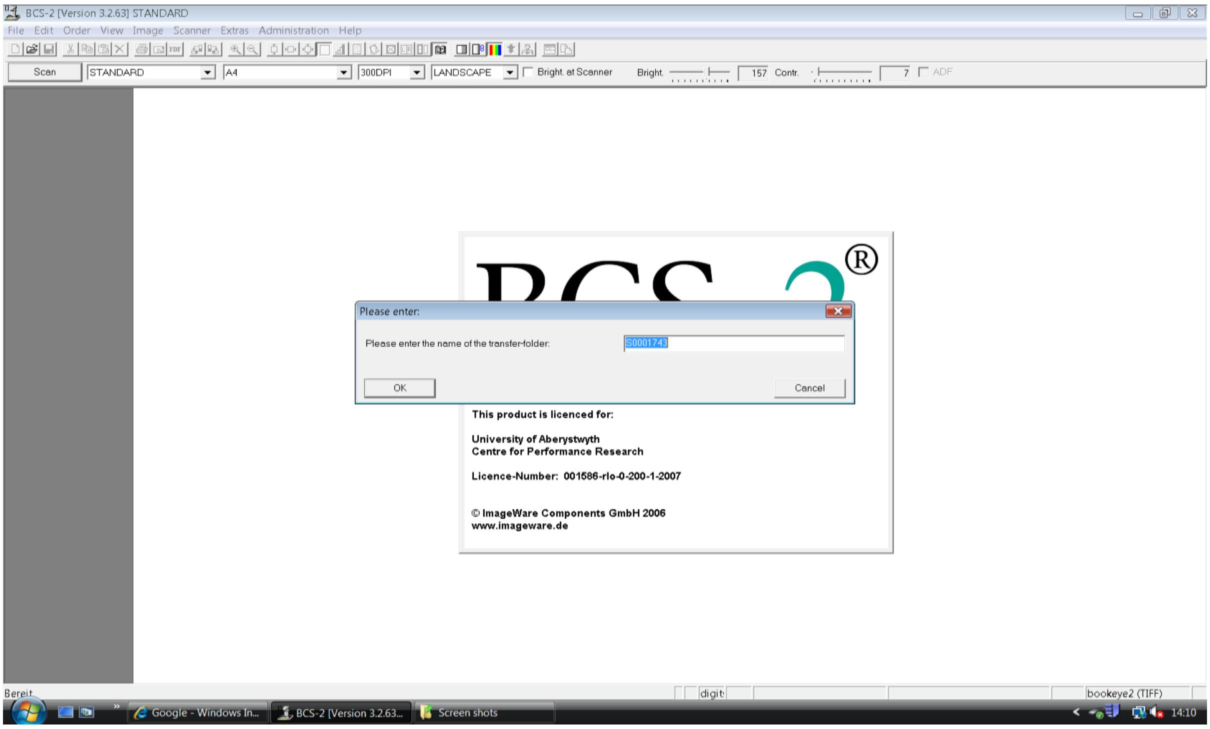


Figure C8: Project Section Window

- Check that the settings on the lowest tool bar are 'STANDARD' 'A4' (unless a larger size is needed) '300DPI' and 'LANDSCAPE'.
- Place the items (and the reference strip if used) in view on the scanner bed and press the 'scan' button on the scanner.
- Repeat this until all the items in the batch have been scanned.
- If a scan goes wrong then highlight it by clicking on it, and then click the delete cross. Repeat it as an individual scan at the end and drag and drop it into place.
- When all the scans are completed exit the order mode by clicking 'STOP' in the top tool bar. This opens the order selection window, Figure C9 which displays all the data on the completed digital objects.

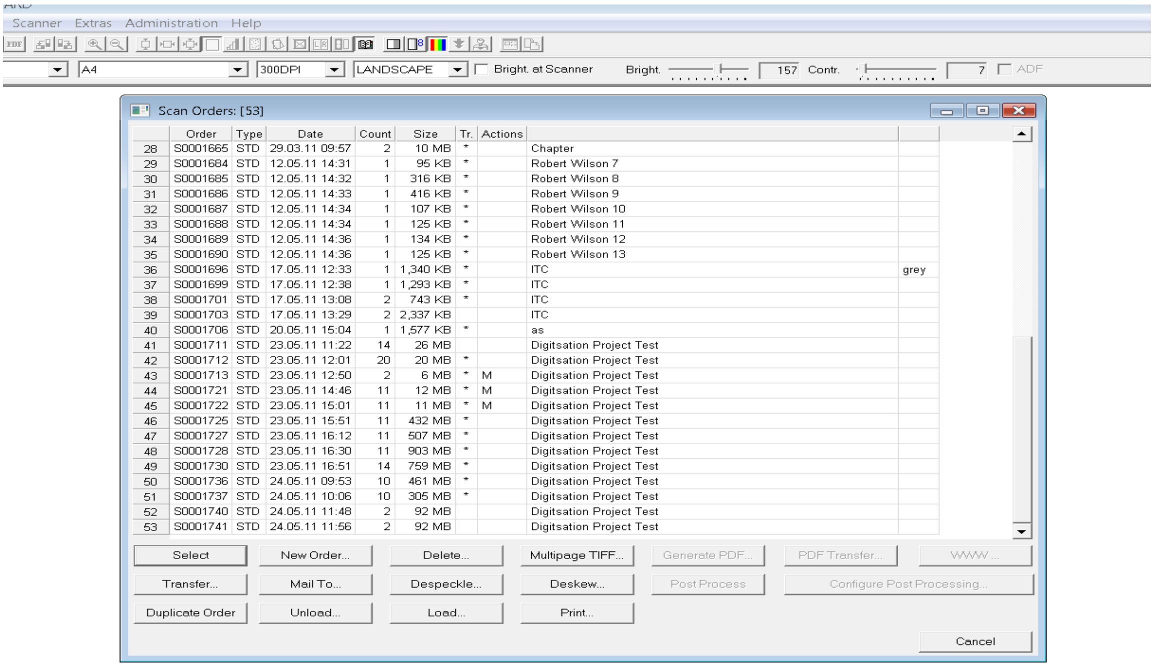


Figure C9: Completed Batch/Transfer Window

- On the left, next to the batch (order) just completed, is a green square. Click on this to highlight the order.
- Then click on the button labelled 'Transfer'. This transfers the digital objects to the folder on the desk top you tell it to
- Close the 'Scan Orders' window and minimise the scanner software window.
- To scan more batches press 'SCAN'.
- The scanner assumes you want to use the same 'TYPE' as before and opens the same window as in Figure C8. Fill in the batch number as before then
- Go back to "Setting Up a Profile/Batch Settings" and repeat until all the archival materials in the batch have been processed and saved. Then
- Go to section C2.5 on page 275.

C2.4 Using the Canon Scanner

For slides:

- Place the slide holder into the slot by the hinge on the bed
- Remove the lid section to reveal a light strip

- Place the slides face down in the holder i.e. so that you are looking at the slide and the picture is correct

For non-reflective archival materials (e.g. Photographs):

- Place the archival material onto the bed of the scanner near to, but not beyond, the arrow at the edge. This arrow marks where the scanning starts from. If you leave a large gap the scanner will not scan the archival material.

For all archival materials

- Click on the 'Windows' button bottom left
- Click 'Computer'
- Click on 'Expansion Drive D;'
- Click 'Photo Studio 5.5' shortcut to open the scanner software
- For the first time that day or after the software has been closed click on 'File' then 'Select Source'
- Click on 'CanoSoft' – this is the one highlighted
- Click on the icon that looks like an open scanner with 'acquire' appearing as a name in the same toolbar (the fourth one from the left).
- This starts the preview scan and opens a new window with a series adjustment options in a box to the right.
- If you have already performed a scan then it will ask to delete the previous previews before it starts the new one.
- Click 'ok'
- When the preview scan is complete the first preview will appear as a thumbnail.
- In the box on the right there are three tabs at the top.
- Click on 'Advanced'. This will take you through to a series of options.
- In the 'Input Settings' box:
    - for slides choose 'Color Positive Film', 'Color' and 'Pixels'. Depending upon previous settings it may want to redo the preview at this point; let it.
    - for photographs choose 'Full Platen'
- In the 'Output Settings' box:
    - Check that 'Resolution' has 2400 selected.
    - And that the 'Output size' is flexible

- In the ‘Image Settings’ box everything is either ‘off’ or ‘none’ except the ‘High Quality’ button which is ‘Yes’
- Use the brightness, contrast and curves buttons below this box to adjust the preview to try to show the most information possible.
- When each preview is competed tick the small box for each one to be scanned. Then,
- Click scan. A window appears indicating how long it will take to perform the scan. Until this is done you cannot do anything else with this program, so you could do something else. When the scan has finished a box with ‘Retouch and Save’ already chosen appears.
- Click the ‘OK’ button.
- The scan(s) open in a window.
- Click on ‘File’ in the top tool bar, then ‘Save As’ and save the file in the appropriate folder using the three letter project code and five numbers (e.g. dpt00001, dpt00002 etc.)
- It is a good idea to save in regularly sized small batches e.g. of four or five.
- Repeat the process until all the batch has been scanned and stored, then
- Go to section 2.5 on page 275.

C2.5 Adjusting the Digital Object For Viewing and Storage

Having created an archive digital object the next step is to make copies of it for different uses. To do this you manipulate the digital object using imaging software and save them as separate files: generally in a different place so that the archival (or reference) digital object is kept as it is, where it is. I have given instructions using Adobe Photoshop Elements, but other software will follow the same principles if not the letter of the instructions.

- Turn on the iMac computer if you have not already done so. To do this feel for the switch for the monitor by running the fingers of your left hand behind the lower edge until a dimple is felt. Press here and the monitor should click on. If you get a message saying that there is no key board or mouse detected, they run on AA batteries and need to be turned on also.

- First set up a new folder named with the three letter code for the project and two sub folders named viewing and thumb that the new digital objects will go into on the desktop
- Open Photoshop from the dock at the bottom of the screen. If the symbol is not there then:
  - Click on the square, blue face symbol (Finder), then 'Applications' and choose 'Adobe Photoshop', then 'Photoshop' from the list to open the software
- Open the first archival digital object using Photoshop by clicking on 'File', then 'Open' and navigate to the desktop and the folder where the digital objects just created have been stored.
- If necessary rotate the digital object by clicking on 'Image' then either 'Rotate' or 'Flip' as appropriate
- If the edges of the digital object need to be adjusted to make it look like the same as the archival material, then
- Click on digital object on the screen, then either
- Click on 'Crop' symbol in the left margin or 'Image' in the top tool bar, then 'Crop'. The software will place a box over the digital object. Use the squares halfway down the sides to pull it to the shape and size required to fit to the edges of e.g. the photographic paper.
- Click on the green tick at the bottom to accept the crop and
- Click 'Save', then OK to the next two windows to overwrite the previous copy of the digital object
- Do this for all the digital object in the batch. Then, to
- Re-size the cropped digital objects to create two smaller digital objects, one named 'viewing' the other 'thumbnail'. The viewing copy is the larger one that appears on the internet and the thumbnail is a very small copy sometimes used as a preview.
- Click 'File'
- Click 'process multiple files'. And into the boxes that appear, Table C8 put:

Viewing Copy	
Action	Choose
Process Files From:	Folder
Source:	Click Browse and choose the folder with the 3 collection initials (has form similar to ‘Clare Wood-Fisher: Desktop:’ then 3 letter name)
Destination	Click Browse and choose the folder with the 3 initials + either viewing (has form similar to ‘Clare Wood-Fisher: Desktop:’ then the 3 letter name: followed by the 3 letter name + viewing)
File Naming:	☑ box labeled ‘Rename Files’
Next row should read:	‘Document’ + into box type _v for the viewing copies
Starting serial #1. Compatibility:	☑ Windows ☑ Mac ☑ Unix
Image Size	☑ re-size Images
Width	600 pixels for viewing copy
Height	Leave blank
Resolution	Choose 96 ppi for viewing copies from the drop down list
Constrain Proportions	☑
File Type	☑ Convert Files to: JPEG High Quality for the viewing copy
	OK

Table C8: Table of Option Choices for Viewing Digital Objects

- Repeat the process for the thumbnail copy using the parameters in Table C9, below:

Thumbnail Copy	
Action	Choose
Process Files From:	Folder
Source:	Click Browse and choose the folder with the 3 initials (has form similar to ‘Clare Wood-Fisher: Desktop:’ then the 3 letter name)
Destination	Click Browse and choose the folder with the 3 initials + either viewing or thumbnail (has form similar to ‘Clare Wood-Fisher: Desktop:’ then the 3 letter name: followed by the 3 letter name + thumbnail)
File Naming:	☑ box labeled ‘Rename Files’
Next row should read:	‘Document’ + into box type _th for the thumbnail copies
Starting serial #1. Compatibility:	☑ Windows ☑ Mac ☑ Unix
Image Size	☑ re-size Images
Width	120 pixels for thumbnail copy
Height	Leave blank
Resolution	Choose 96 ppi for viewing copies from the drop down list

Constrain Proportions	☑
File Type	☑ Convert Files to: JPEG Low Quality for the Thumbnail copy
	OK

Table C9: Table of Choices to Make Thumbnail Digital Objects

- Each of the archival digital objects in the folder chosen should be processed by the software and then appear in the appropriate folder.
- Insert the transfer USB stick,
- Click on the icon when it appears on the desk top to open it and,
- Drag the whole folder containing all the digital objects from the desktop to it to transfer them
- Add to the spreadsheet the name of the person operating the scanner in the columns headed dc.creator for all the digital objects made
- Add the date in the date column. Then,
- Add the scanner and software details into the dc.provenance column for all the digital objects created. Then,
- Save the completed spreadsheet and
- Return the spreadsheet and USB stick with the digital objects to the metadata writer.

C2.6 QA of digitised images

Quality assurance (QA) should be carried out by someone other than the person who did the scanning. It is suggested that each batch is checked off systematically on some sort of record as it is made. Having found a mistake there should be a method for recording this e.g. a sheet to fill in describing the fault. Apart from the obvious of providing written proof of a fault, these records can also prove a useful cross-check for scanner operators indicating where they may need extra training or skills updating.

Table C10 and Table C11 show some of the things to look out for and questions to ask yourself if carrying out a quality check.

What to check	What to look for
Archival/storage copy	Look at areas of where the digital object is white. See if these are white because there is no data or because the scanner wasn't adjusted to capture it.
	Look at areas where the digital object is black. See if this is because there is no data there or because the scanner wasn't adjusted to capture it.
	If not sure, look at places like the inside of ears. If these are white/black it is probably because the digital object doesn't have any data in that area.
Check colours are adjusted properly	Look to see if whole picture has the same colour tones as the original material.
Check digital object is cropped accurately	Look closely at the edges. See if you can spot anywhere where the edges are not quite straight or a different colour. Wobbly edges are often spotted because the cropping isn't quite right and it shows as a change in the colour at the edge of the digital object.

Table C10: Table of Items to Look for on a Scan

Questions to ask yourself	What to look at
What is the file size	Is the digital object the correct size
Was the scan done at 300ppi	Is the digital object the correct resolution
What is the file name	Is the file name correct
Was the archival digital object stored as a.TIFF	Is the file format correct
What is the colour of the archival material?	Is it the same as the digital object
Can you see beyond the edges of the digital object onto the scanner bed	Does the digital object need cropping
Is the digital object the right way up	Does the digital object need rotating
Does it look the same as the archival material	Does it look right

Table C11: Table of a Possible QA Questions

For more detail on these topics see the In More Detail ....for Scanner Operators section B4 of the Generic and Organiser's Guide.

C3. Quick Guide to …. Metadata

After the copy of the metadata spreadsheet has been completed and forwarded by the Scanner Operator, the next step in the metadata collection process is to complete the technical metadata for the digital objects created.

C3.1 DROID

The specific information that is needed for the metadata spreadsheet for a digital object; its file size, format and the creation date (or the date it was modified) can be obtained from the digital object itself. There are a number of ways of obtaining this data. You could open each digital object, right click on it, click on properties at the bottom of the list that opens up and then type the data presented into a database or spreadsheet. Another way of doing the same is to use an automated tool such as Droid. Droid, and its associated database Pronom, are free tools available for download from the TNA web site.

C3.1.1 To Use DROID

- ❖ Transfer the digital objects from the iMac, using the USB, to the external (D;/) hard-drive on the PC
- ❖ Create a folder named with the collection initials on the desktop
- ❖ Click on the DROID folder to open it, then
- ❖ Click on the file named droid (it has MS- DOS Batch File in the ‘Type’ field in the display) and the software opens
- ❖ Click the Green plus sign (add) in the bar above the blank space, and
- ❖ Choose the digital objects by clicking through the menu until you reach the appropriate folder for the archival digital objects.
- ❖ Highlight this and click OK at the bottom of the window. This uploads the folder details to DROID
- ❖ Click the white start arrow. A blue line goes along the bottom of the box showing progress. When the process is completed the start and pause arrows are greyed out
- ❖ To show the results click on the ‘+’ sign next to the chosen folder name. A screen similar to that below appears

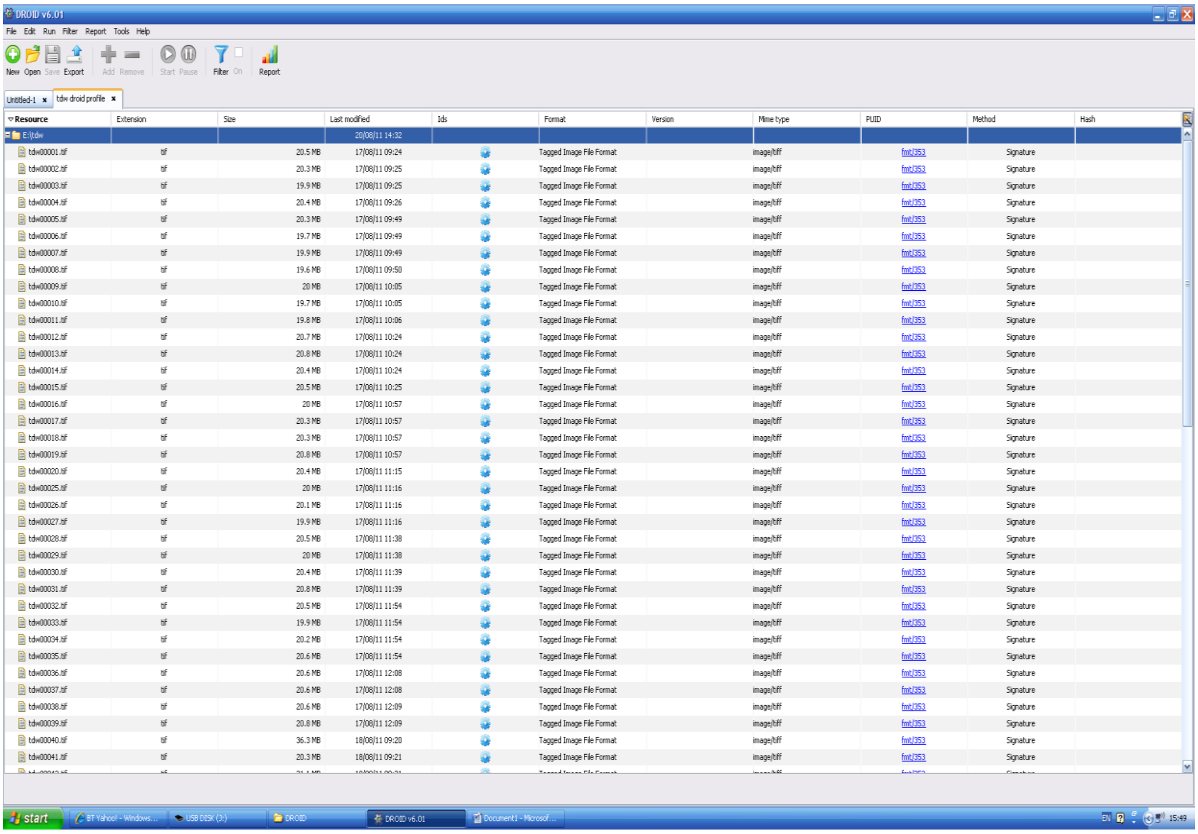


Figure C10: DROID Results Window

- ❖ Click ‘export’ and check the box next to ‘untitled-1’’. Check the button ‘one row per file’ is checked, then
- ❖ Click on the ‘Export profiles ...’ button
- ❖ Select the folder you just created on the desktop to save the report in
- ❖ Type the collection initials + the digital object type (archive, thumb etc.) as the name of the .csv spreadsheet when prompted, and
- ❖ Click ‘save’
- ❖ Repeat if you have more digital objects.

C3.2 Adding the Data from DROID Reports

The DROID reports consist of a series of spreadsheets containing data on the digital objects. These spreadsheets need to be opened and the relevant information copied and pasted into the metadata spreadsheet in the right columns (this section could be done by the scanner operators if required).

- ❖ Open the CPR metadata spreadsheet in the folder on the desktop labelled ‘metadata’ using Open Office

- Click on the folder with the saved DROID reports and open the .csv file for the project you are going to write the metadata for. These open in exactly the same way as an Open Office spreadsheet or MS Excel file. When you look at the spreadsheet you will recognise the contents of several of the columns.
- Look for the column headed 'Method' and check that the files have 'signature' here. If they do not, sort them from largest to smallest using the name column
  - Click on the letter above the column headed 'Name'
  - Click on 'Sort & Filter' then,
  - Click 'A to Z' and because files identified by the 'stem' are likely to be hidden copies and therefore smaller, the files should sort themselves so that all those identified using the 'signature' are in blocks together. Then,
- Highlight the first batch values in the URI column (column 'c' on the DROID output)
- Copy, and
- Paste into the CPR spreadsheet into the dc.relation (DOI of ....) field.
- Repeat the highlight and copy process for the 'size' column, and paste into the dc.format.Extent column on the CPR spreadsheet
- Again, repeat the highlight, and copy process for the extension column and paste into the dc.relation.hasFormat (file format of digital object) column on the CPR sheet
- Repeat these actions for all the files in the collection.
- Save the CPR metadata spreadsheet
- Close the DROID reports.

C3.3 Finishing the Metadata Collection

This is when the more detailed descriptions, if not already completed, are written. It is easier to do this at this stage as the digital object can be used to view the archival material with its attendant extra functionality from viewing software.

- Open a spreadsheet or a word document, which ever you are most comfortable with, and
- Create a list of phrases or words you think you will need to use to describe the digital objects. In particular make a note of specific terms from *Dictionary of*

the Theatre: terms, concepts analysis by Patrice Pavis in CPR. This sheet will allow you to write entries that describe the same thing consistently across this collection.

- ❖ Open the first digital object,
- ❖ Click on the dc.relation column and check that the same digital object opens
- ❖ Complete writing the columns on the spreadsheet. The table below gives a set of rules to use when writing the details into the spreadsheet. Using the table means that the same name will be written the same way throughout and can be found when searched for.

Type of entry	**Description of how to write the entry**	**Example**
Personal Names	Use a person's professional name(s) as they appear on publicity materials for production	USE "Director: Richard Gough" NOT "Director: R. Gough"
Production Titles	Use the title as printed on publicity material for the production	USE "Dragons Playtime" NOT "Dragon's Playtime"
Professional Credits	Use the job title as printed in publicity materials for the production	USE "Director" NOT "Director/Producer"

Table C12: Table of Examples of Writing Rules

- ❖ Save as an Open Office Document
- ❖ Repeat this for the remaining digital objects
- ❖ When the metadata spreadsheet is complete save a copy to the folder labelled 'Metadata' on the external hard-drive.

C3.4 QA Metadata

Like digital objects, metadata should be checked for quality and consistency. Again like digital objects someone other than the writer should do this. How you carry out the checks are very individual and can be made quite sophisticated with, for instance, the spreadsheet being formulated to highlight squares with incorrectly formatted data in them. To help you get started a couple of ideas on how you might achieve this are floated here:

To check for consistency of data entry:

- ❖ Read the data looking particularly for differences in the way data is written into different rows of the same column e.g. dates should be formatted as dd-mmm-yyyy all the way down

- ❖ Check for numbers where there should be text and vice versa
- ❖ Check that the data in the columns make sense e.g. is the description in the correct field, is the entry in the format field likely?

To check the file sizes have been correctly added:

- ❖ Highlight the dc.format.Extent column and
- ❖ Click on 'AutoSum'. This should give a number in the square after the last number.
- ❖ Open the DROID report for the collection
- ❖ Repeat the 'AutoSum' action on the 'Size' column. The two should give the same approximately the same result (the result may not be exactly the same because of the way the numbers are formatted to round up on the different spreadsheets).
- ❖ If there is a large discrepancy there may be a mistake somewhere in the CPR spreadsheet. If this is the case go back and check the data systematically.

As an additional check these 'sums' could be stored on a separate sheet on the spreadsheet. If these are added then the total should go up each time and be roughly equal to the amount of storage used – another possible check.

C3.5 ECLAP Metadata Spreadsheet

The ECLAP metadata spreadsheet needs to be filled so that if archival materials are uploaded to ECLAP the information required is to hand. Much of this is a repeat of that in the CPR spreadsheet and having been checked can be copied across. However there is much that is not as it is particular to performing arts e.g. the names of dressers and make up designers which needs to be researched. There is no reason this shouldn't be investigated whilst digitisation is going on if there is a suitable way to do this. But don't add it until the scanning process is complete as it is better to work on one sheet at a time to avoid mixing up which is being completed. It is saved to the external hard-drive only. Having completed this data it should also be checked using the same QA process as before.

C3.6 Metadata Collection Stages

As can be seen, the tenth action in Table C13 is to complete the ECLAP metadata. The list in this table is a brief overview of the actions you have carried out to collect the metadata of the digital objects created. More information can be found in section B5 In More Detail.....For Metadata Writers

Stage Description	Completed
1. Collect the archival materials together and assess where the logical breaks in sections occur	
2. Write archival material data into spreadsheet	
3. Email copy of the sheet to Scanner Operator	
4. When completed spreadsheet is returned, open files of digital objects and look at them.	
5. Open a separate document, text or spreadsheet, which ever you are most comfortable with, and create a list of phrases or words you think you will need to use to describe the digital objects. In particular make a note of specific terms from the *Dictionary of the Theatre: terms, concepts analysis* by Patrice Pavis: Toronto & Buffalo; UP Toronto,1998 kept in CPR. This sheet allows you to write entries that describe the same thing consistently across this collection.	
6. Open the first digital object.	
7. Open the spreadsheet and complete the description fields	
8. Complete each of the fields on the form	
9. Check the data	
10. Complete the ECLAP spreadsheet	

Table C13: Table of Metadata Stages

C3.7 Changes to Process After the Installation of Repository Software

The planning and digitisation processes are likely to remain unchanged, but metadata writing will probably be affected by the installation of repository software. Whilst the exact changes will depend upon the software installed, it is suggested that you keep the first spreadsheet to control the workflow and pass back the data to be filled by metadata writers. The ECLAP metadata spreadsheet will still need to be filled as it is unlikely that the repository software will be configured to do this, at least initially.

Appendix D: **Example Letters**

D1: Letter to Prospective Interviewees

This letter was sent, as an email, to the professionals who were approached to take part in the research interviews.

Dear XXX

I am a KESS (Knowledge Economy Skills Scholarship) MPhil student based at Aberystwyth University studying the knowledge transfer of digitisation techniques between national institutions and small collections.

I am writing to ask if you would consent to being interviewed as part of this project. I would like to discuss how the digital resources being made may impact upon future scholars in the area of performing arts. If you agree I would be grateful if you could nominate time when this would be convenient for you in the near future.

The interview would be conducted in accordance with University ethics policy, a copy of which is below. Please read this information to help you make an informed decision on whether to take part.

Regards

Clare Wood-Fisher

Student Name: Clare Wood-Fisher (clw09@aber.ac.uk)

Academic Supervisor: Lucy Tedd (lat@aber.ac.uk)

Before you decide whether or not to be interviewed it is important that you understand:

a) why the research is being done
b) what it will involve

Please take time to read the following information carefully. If any of the information below is unclear or if you would like more information about this research project and what it involves, then please contact me.

Part 1 Why am I doing this research/study: what is its purpose?

The project is researching the knowledge transfer of digitisation skills from national institutions to small collections; I am undertaking this research project as part of the requirements for an MPhil. The findings of the project will be used to:

- compile a practical user guide for digitisation of resources in small collections and
- to inform the MPhil thesis.

Part 2 What does the study involve/what is the conduct of the study?

Before you agree to be interviewed, please note the following about this study:

1. **Duration**: The interview should take around 1 hour of your time, although this can be adapted.
2. **Confidentiality:** All the information you give will be treated confidentially. Both the conversation and the information you provide will be completely confidential and treated confidentially.
3. **Anonymity**: All interviews will be anonymous and personal data removed at the transcription stage. No individuals or individual facilities will be identified in the results. Any direct quotes included in the report (that is, quotes of the things recorded in the interview), will be used selectively and anonymously (that is, no one will be able to attribute the words to you).
4. **Data security:** The information will be kept securely, and for only as long as necessary to: a) analyse the research data and b) report on the research and its findings.
5. A full report and a summary of the research findings will be available via the institutional repository, CADAIR (http://cadair.aber.ac.uk).
6. **Consent**: If you consent to take part in this study, I will assume that:
   - you have read and understood the information in this letter.
   - you understand that participation in this study is voluntary and that you are free to withdraw from the study at any time, without giving any reason and without any of your rights being affected.
   - you understand that your responses will be treated confidentially and in confidence.
   - you understand that your responses will be treated anonymously.
   - you allow me to use your direct quotes, anonymised, in the study's report/write-up.

Please contact me (via my contact details listed below) if you have any questions or concerns about the study.

Thank you and I look forward to hearing from you.

Regards

Clare Wood-Fisher
(email: clw@aber.ac.uk)

Llywodraeth Cynulliad Cymru
Welsh Assembly Government
Ewrop & Chymru: Buddsoddi yn eich Dyfodol
Cronfa Gymdeithasol Ewrop
Europe & Wales: Investing in your Future
European Social Fund

D2: Copyright Letter Template

This letter was prepared for the Director of CPR to use when contacting the holders of copyright on archival materials held in CPR's collections. It asks for permission to digitise and explains why. It also indicates how it is expected to display the digital objects created. The letter was written because the copyright holders are relatively few and CPR wanted to ask permission to digitise the complete holdings of a person, rather than to keep writing and asking each time they began a new section of archival materials.

Dear XXX

You may remember that you were commissioned by Cardiff Laboratory Theatre/Centre for Performance Research to take photographs of XXX Production. As the copyright holder, we (CPR) are writing to you to ask permission to digitise and display the resulting images on the web. The aim of carrying out this process is to provide an educational resource for students and academics as well as to enhance access to the collection.

If you are willing for the digitisation to go ahead, please print and complete and the attached form with your preferred copyright notice wording to be published alongside the XXX(e.g. photograph).

Yours sincerely

Copyright Permission Slip

I hereby allow *CPR to digitize the photographs* listed, for the purpose of providing access and educational resources on the Web.

Signed: ____

Date: ___________________

Name (please print): ___________________________________

Preferred copyright wording (if different to © A N Other):

List of copyrighted material granted permission

Appendix E: **CPR Collections' and Appraisal Policy**

This collections policy has been copied directly from the CPR web pages (http://www.aber.ac.uk/en/media/CPR-Collections-Policy.pdf). The numbers and formatting have not been changed to reflect the rest of the thesis. This is the collections policy referred to in Chapter 2.

CPR Collections' Collection and Appraisal Policy

Vision Statement

CPR Collections is an international theatre and performance repository which collects, catalogues, stores and permanently preserves materials for the use and benefit of present and future generations of scholars, theatre practitioners, and the public in Wales and beyond.

1

Name of the Repository:

1.1 Centre for Performance Research Collections (CPR Collections)

Address:

1.2 CPR, The Foundry, Parry Williams, Penglais Campus, Aberystwyth SY23 3AJ.

Identity of the governing body/authority

1.3 Centre for Performance Research Ltd, an Educational Charity (no. 701544) limited by guarantee (Reg. No. 2315790).

1.4 CPR Collections is a division of Centre for Performance Ltd.

1.5 Origins:

The Centre for Performance Research (CPR) was established in Cardiff in 1988, by Richard Gough and Judie Christie. CPR's predecessor was Cardiff Laboratory Theatre, which began in 1974.

CPR is a multi-faceted theatre organisation located and rooted in Wales, working nationally and internationally. CPR produces innovative performance work: arranges workshops, conferences, lectures and masterclasses; curates and produces festivals, expositions and exchanges with theatre companies from around the world; publishes and distributes theatre books, as well as the journal Performance

Research, and houses a resource centre and repository that specializes in world theatre and performance.

2 CPR Collections:

2.1 Resource Centre

The printed collection in CPR's Resource Centre contains 8,000 monographs and journals specializing in contemporary visual and experimental performance; the traditional, 'classical' and folk culture of Asia, Africa and the Americas (especially India, China, Japan and Indonesia) as well as twentieth century theatre, dance, Live Art and performance (especially from Europe, the Americas and Australasia but also including Asia and Africa).

There is also a wide-range of audio-visual material representing instrumental and vocal music from world cultures and documenting the work of contemporary theatre groups and practitioners.

2.2 International Theatre Collection

This collection, deposited by Barbara and John Cavanagh in 2007, consists of 22,000 books, monographs and journals as well as 8,000 further items consisting of posters, prints and other theatre ephemera. The material is in over 20 languages and refers to many more cultures across the world.

New accessions, in line with the CPR Collections Policy, will be added to this collection with the exception of audio visual material which will be added to the Resource Centre.

2.3 Cardiff Laboratory Theatre/CPR Archive

This collection consists of and continues to add the documentation and production and design notes for all of CPR's productions including those performed under t he organisation's previous name: Cardiff Laboratory Theatre which started in 1974. There are 500 audio-visual tapes documenting the proceedings of the many festivals, workshops, lectures and conferences organized by CPR and documentation of the research undertaken by CPR into the themes and subjects around which it organised programmes, workshops and performances.

2.4 Giving Voice Archive The biennual GIVING VOICE Festival began in 1990—with plans to become annual—and the archive continues to document each; from the planning, research and development stages to film and audio and paper records of the talks, workshops and performances.

2.5 Performance Research Archive

The Performance Research archive consists of all the papers, images, and correspondence the journal receives and generates—since its conceptualisation in

1993 and first issue in April 1995—but also contains evidence of the extensive research and specialist conversations generated by its themed issues.

3

Information about the legal status of the repository

CPR Collections have adopted the Standard for Record Repositories as the basis for formulating their collection policy.

4

Information about Scope of, and Limitations to, Policy

4.1 Context

CPR aims to develop and improve the knowledge, understanding and practice of theatre in its broadest sense, to affect change through investigation, sharing and discovery and to make this process as widely available as possible. Its programmes of work combine cultural co-operation,

collaboration and exchange, practical training, education and research, performance, production and promotion, documentation and publishing, and information and resource gathering. CPR Collections plays an essential role within CPR by identifying, collecting, preserving, and making accessible both existing and emerging bodies of thought and practice in international theatre.

4.2 Definitions

CPR Collections collects published works (monographs, series, journals), ephemera (posters, prints, programmes, clippings, photographs etc), records and documents, DVDs, audio recordings and videos, and the unpublished theses and dissertations of Aberystwyth Performance and Theatre Studies postgraduates.

CPR Collections therefore consists of both library and archival items and is best described as a Special Collection.

CPR Collections do not collect microform or microfiche items, or 3dimensional objects or artefacts.

4.3 Geographical Area

CPR Collections will collect items on theatre, dance and performance from around the world - contemporary and traditional, global and local – that emphasize the physical, visual and interactive aspects of human artistic endeavour; but with regard to Great Britain mainly focussed on post 1900 innovative, experimental and multi-cultural forms and manifestations.

4.4 Limitations

CPR Collections will not actively collect records primarily relating to British theatre from periods before 1900 or material that mainly focuses on dramatic work (play scripts, play analysis and biographies of playwrights) from Europe (including Great Britain), Americas and Australasia from all periods.

4.4.1 Acquisitions not covered by the policy

Acquisitions outside the current stated policy will only be made in very exceptional circumstances, and then only after proper consideration by the governing body of CPR Collections itself, and the agreement of any other interested archive authority has been sought.

4.5 Other Institutions

CPR Collections' interest in international—as opposed to British—theatre and performance means it doesn't directly compete with other major British theatre collections and it will seek to avoid competition, conflict and duplication of effort with the V&A, for example, whose collecting policy will sometimes overlap with that of CPR Collections. When advising potential depositors CPR will draw attention to the existence of other repositories with similar or overlapping interests, to ensure that material is offered to the most appropriate institution.

5 Methods of Acquisition

5.1 Accessions

All items accepted by CPR Collections, whether by donation, purchase, transfer, or bequest are subject to the signing of a transfer of title form by the owner or their representative and will be listed on an entry form which is also signed by the owner or their representative. This states the various conditions under which the items are accepted. Only items which in the judgement of the Collections' Director are of sufficient quality for permanent preservation will be accepted. Finding aids may be required with deposits where bulk is likely to render items otherwise inaccessible until processed. CPR Collections will not normally accept archives and records which are of a particularly specialist nature, requiring skills or equipment beyond the office's resources to preserve, exploit or interpret.

5.2 Donations

Once an item is accepted as a donation, it becomes the property of CPR Collections. This is stated on the back of the entry form which the donor is asked to sign when the object is accepted by CPR Collections.

5.3 Transfer

CPR Collections will accept transfers of items from other TNA recognised organisations and institutions which meet CPR Collections Policy requirements.

5.4 Purchases

Items may be purchased from individuals, shops and auctions if they are unique or when it is considered that similar items may not be given, and that they will form a significant addition to the collections. Such items may be sold by CPR Collections if a better specimen is donated to the Collections at a later date.

Purchases are normally made with funds which are donated to CPR Collections, money raised by CPR, or grants. Acquisitions purchased with the assistance of grant aid will be held subject to the terms and conditions of bodies from whom such aid has been received.

5.5 Bequests

Items are generally accepted as part of a bequest, but they will not be accepted if

(1) they fall outside CPR Collections' collecting policy.

(2) the terms of the will impose unacceptable conditions on CPR Collections as part of the bequest.

(3) the items could not be cared for satisfactorily.

5.6 Duplication

Normally, duplicates of items will not be accepted unless they differ in a significant way from an item already in the collection, or form part of a group of related items.

(e.g. when someone gives two identical items together, or gives one item of a type already represented in the collection with other items which were associated with it).

5.7 Loans

Loans of items to CPR Collections are subject to a formal loans agreement, which is outlined on the back of the entry form.

Loans will only be accepted if they are items which fall into the collecting policy, but are not already represented in the collections.

A full description of all items on loan will be sent to the lender, who will be asked to sign an agreement concerning the conditions of the loan. This will include a request to give a reasonable period of notice before the specimen is returned to them.

If a lender should die during the period of a loan, the items will form part of the estate of the former owner and its loan will be negotiated with those responsible for the estate. All efforts will be made to maintain a record of the ownership of loans.

6 Appraisal policy

6.1 CPR Collections shall, in accordance with the wishes and requirements of depositors, evaluate and select for transfer or destruction those documents deemed not to be worthy of permanent preservation, and the intention shall be made clear at the time of deposit.

6.2 By definition, CPR Collections has a long-term purpose and should possess (or intend to acquire) permanent collections in relation to its stated objectives. CPR Collections accepts the principle that, except for sound curatorial reasons, there is a strong presumption against the disposal of any items in the collections.

6.3 CPR Collections will establish that it is legally free to dispose of an item. Any decision to dispose of material from the collections will be taken only after due consideration.

6.4 When disposal of an item is being considered, CPR Collections will establish if it was acquired with the aid of an external funding organisation. In such cases, any conditions attached to the original grant will be followed. This may include repayment of the original grant.

6.5 Decisions to dispose of items will not be made with the principal aim of generating funds.

6.6 Any monies received by the CPR Collections' governing body from the disposal of items will be applied for the benefit of the collections. This normally means the purchase of further acquisitions but in exceptional cases improvements relating to the care of collections may be justifiable. Advice on these cases will be sought from CyMAL and TNA.

6.7 A decision to dispose of an item, whether by gift, exchange, sale or destruction (in the case of an item too badly damaged or deteriorated to be of any use for the purposes of the collections), will be the responsibility of CPR Collections acting on the advice of professional curatorial staff and not of the Director of Collections acting alone.

6.8 Once a decision to dispose of material in the collection has been taken, priority will be given to retaining it within the public domain, unless it is to be destroyed. It will therefore be offered in the

first instance, by gift, exchange or sale, directly to other record repositories subscribed to The National Archives Standards likely to be interested in its acquisition.

6.9 If the material is not acquired by any repositories subscribed to TNA Standards to which it was offered directly, then the archive and repository community at large will be advised of the intention to dispose of the material, normally through an announcement in professional journals and through online professional forums where appropriate.

6.10 The announcement will indicate the number and nature of items involved, and the basis on which the material will be transferred to another institution. Preference will be given to expressions of interest from other repositories subscribed to TNA Standards. A period of at least two months will be allowed for an interest in acquiring the material to be expressed. At the end of this period, if no expressions of interest have been received, CPR Collections may consider disposing of the material to other interested individuals and organisations.

6.11 Full records will be kept of all decisions on disposals and the items involved and proper arrangements made for the preservation and/or transfer, as appropriate, of the documentation relating to the items concerned, including photographic records where practicable.

Information Concerning Access

Notice of every new acquisition and of any restrictions on its access or use will be made public at the earliest opportunity and CPR Collections will systematically supply copies of its finding aids to the National Register of Archives and to users via the CPR web site.

Dates and Policy Review

Date of Board of Directors meeting: 30th July 2009

Date of Policy Review: 2 years from above

The National Register of Archives and CyMAL will be notified of any changes to the Collections Policy, and the implications of any such changes for the future of existing collections.

Appendix F: **Evaluation Questions**

Questions for the Evaluation of Quick Start Guide

Whole CPR Quick Start Guide					
Please answer the following questions based on your impressions of the whole "Quick Start Guide". Where questions have ratings for answers 5 is high: 1 is low					
1. What was your impression of the appearance of the CPR Quick Start Guide?	1	2	3	4	5
2. What did you think of the language used in the CPR Quick Start Guide?	1	2	3	4	5
3. Were there enough illustrations in the CPR Quick Start Guide or would it have benefitted from more?	Yes/No (if 'No' please elaborate)				
4. Do you think that the CPR Quick Start Guide covers the right topics?	Yes/No (if 'No' please elaborate)				
5. Did you feel there were any topics missing from the CPR Quick Start Guide?	Yes/No (if 'No' please elaborate)				
6. Does the layout of the workflow in the CPR Quick Start Guide make sense to you?	Yes/No (if 'No' please elaborate)				

Metadata Sections of the CPR Quick Start Guide					
Please base your answers in this section on your impressions of the Metadata Sections of the CPR Quick Start Guide. Where questions have ratings for answers 5 is high and 1 is low					
7. What was your impression of the appearance of the metadata sections of the CPR Quick Start Guide?	1	2	3	4	5
8. What did you think of the language used in the metadata section?	1	2	3	4	5
9. Were the explanations clear and could you follow them to achieve the task goal?	Yes/No (if 'No' please elaborate)				
10. Does the layout of the workflow in the metadata sections make sense to you?	Yes/No (if 'No' please elaborate)				
11. What changes would you make to the metadata sections?					

Scanning Section of CPR Quick Start Guide					
Please base your answers in this section on your impressions of the Scanning Section of the CPR Quick Start Guide. Where questions have ratings for answers 5 is high and 1 is low					
1. What was your impression of the appearance of the scanning section of the CPR Quick Start Guide?	1	2	3	4	5
2. What did you think of the language used in the scanning section?	1	2	3	4	5
3. Were the explanations clear and could you follow them to achieve the task goal?	Yes/No (if 'No' please elaborate)				
4. Does the layout of the workflow in the scanning section make sense to you?	Yes/No (if 'No' please elaborate)				
5. What changes would you make to the scanning section?					

Table F1: Table of CPR Quick Start Guide Evaluation Questions

Questions for the Evaluation of Generic and Organiser's Guide

Generic and Organiser's Guide	
Please base your answers in this section on your impressions of the Scanning Section of the Generic and Organiser's Guide. Where questions have ratings for answers 5 is high and 1 is low	
1. What was your impression of the appearance of the Generic and Organiser's Guide	
2. What was your impression of the language used in the Generic and Organiser's Guide	
3. Were the explanations clear and could you follow them to achieve the task goal?	Yes/No (if 'No' please elaborate)
4. Were the sections that gave additional explanations on topics from the CPR Quick start Guide clear?	Yes/No (if 'No' please elaborate)
5. Were any sections easy to understand and follow? If so which one(s)?	Yes/No
6. Were any sections difficult to understand? If so which one(s)?	Yes/No (if 'Yes' please elaborate)
7. What changes would you make to the Generic and Organiser's Guide?	

Table F2: Table of Generic and Organiser's Guide Evaluation Questions

Appendix G: **CPR Storage Options Tables**

Prepared for: CPR as part of research to prepare a strategy for digitisation.

Date: 04th July 2011

Validity: This information was prepared in July 2011. While much may remain valid for some years, it would be prudent to check for updates and changes to the given examples as this may affect the basis for decisions.

DSpace has been used as an example of repository software throughout as this is the software used by CADAIR and the majority of WRN (Welsh Repository Network) facilities. Other popular repository software applications include Eprints, Fedora and ICA Atom.

Storage Method	University Controlled Hard-Drive Space
Positives	• Daily back up of files performed automatically by University • Daily management of access managed automatically by University • University Information Services carry out regular disaster recovery tests as part of management of digital objects • In case of disaster AU experience could mean digital objects are recovered when perhaps it could be expensive for CPR to discover the means to do so
Negatives	• Control of digital objects is distributed between CPR and University as CPR would decide upon exposed content
Usual use of storage type	• All digital objects + metadata created by a University (images and text files)
Suitable for long or short term storage?	• Suitable for long and short-term storage because maintenance regime carried out by University should ensure longevity of digital object.
Installation requirements	• unknown
Storage Method	External Hard-Drive (in CPR)
Positives	• CPR retain control of distribution of digital objects to Web or other outlets
Negatives	• CPR have to manage daily back up of files • CPR need to manage long term maintenance e.g. carry out regular crash recovery tests • CPR regulate access to digital object • Interface between University system and CPR system if this occurs • CPR need to establish a link to the Web
Usual use of storage type	• Digital objects of any type + metadata

Suitable for long or short term storage?	• Ok short term if being used for access. • Ok for use as back up copy long term. • Subject to same need for management as any other media.
Installation requirements	• Cost of Hard-drive unknown. Would need to be sourced through University IS. • Installation costs – unknown • Would depend upon arrangement and anticipated future requirements
Storage Type	Own Repository (e.g. DSpace) Module Independent of University
Positives	• CPR retain control of digital object(s) • If formatted to use OAI-PMH, interface has richer search capabilities as well as exposure through the network
Negatives	• CPR have to manage and carryout content back up routines • CPR need to manage long term maintenance e.g. carry out regular crash recovery tests • Interface between University system and CPR system would need managing • CPR need to establish and pay for a link to the Web • Significant commitment of time and resources required to implement successfully • No current community of expertise within CPR • A considerable level skill and a wide range of ability can be required to install e.g. SQL writing, XML schema writing and Web design
Usual use of storage type	• OS repository – all types of digital objects + metadata
Suitable for long or short term storage?	• Used for long-term management and storage of Digital Objects. • Amount of time invested and initial costs usually make this a long-term commitment. • Can be used for short-term storage, e.g. exhibitions, once established
Installation requirements	• Software = Free & Open source • 1 x software to run repository e.g. DSpace • 1 x server to host repository e.g. new server will be needed to run and host the software • 1x cabling e.g. new server needs cables to patch into University network • 1x person to install (DSpace say it takes one person with the technical ability one day. This doesn't include the configuration of the new server) • 1x person to configure new server
Storage Method	Own Repository (DSpace) Module Linked to CADAIR

Positives	• University controls access to contents (security) • Immediate visibility through OAI (Open Archives Initiative - http://www.jiscinfonet.ac.uk/infokits/repositories/technical-framework/oai-pmh) as Cadair is linked to this • If controlled by system such as SWORD only need to deposit once to both Cadair and CPR repository if formatted that way • OAI-PMH interface harvests metadata. Therefore has richer search capabilities http://www.jiscinfonet.ac.uk/infokits/repositories/benefits • Content exposure through WRN throughout Wales • Community of expertise developed within University in relation to DSpace as this is the software used by Cadair
Negatives	• Reliant upon future maintenance of Cadair • Reliant upon future cooperation between IS and CPR • Other constraints are unclear • A considerable level skill and a wide range of ability is required to effect installation e.g. SQL writing, XML schema writing and Web design
Usual use of storage type	• OS repository – all digital objects + metadata
Suitable for long or short term storage?	• Used for long-term management and storage of Digital Objects. • Amount of time invested and initial costs usually make this a long-term commitment.
Installation requirements	• DSpace Software = Free & Open Source • 1 x software to run repository e.g. DSpace • 1 x server to host repository e.g. new server will be needed to run and host the software • 1x cabling e.g. new server needs cables to patch into University network • 1x person to install (DSpace say it takes one person with the technical ability one day. This doesn't include the configuration of the new server) • 1 x person to configure new server • 1x person to write and install link between CPR DSpace and Cadair
Storage Method	Repository Externally Hosted and Maintained (ECLAP)
Positives	• CPR control data exposure through upload choices made • See list of advantages at: http://www.jiscinfonet.ac.uk/infokits/repositories/benefits • Possibility of Europe-wide exposure of content
Negatives	• Needs continued commitment to community development to ensure continued success of exchanges • No control over who signs up to community • If project fails it is not yet clear how content could be recovered
Usual use of storage type	• OS repository – all digital objects + metadata
Suitable for long or short term storage?	• Used for long-term management and storage of Digital Objects. • Amount of time invested usually make this a long-term commitment.

Installation requirements	• 1x person to learn about system and upload content • suitable person within University to sign permission forms before content can be exposed • Commitment to community needed to ensure continued successful operation i.e. if there are technical changes there may be a need to make alterations to IT systems to enable continued access

Summary of Storage Option Tables	
What	Why
External hard-drives	• In combination with University hard-drive/server space because gives more than 1 copy (principle from LOCKSS project). • Keeping external hard-drives off site will mean have geographically distributed copies. • As it is a technically less complex system, it will allow CPR to narrow down on to the management aspects of the system and refine them, so that in the future they will be ready to take on more complex arrangements.
University controlled hard-drive/server space	
Repository software	A consideration for the future because: • Costs – JISC gave £4k to each member of WRN for hardware. This is equivalent today to approx cost of a server. You need 2 for any repository. • Complexity – cannot recommend it at this stage as technically and theoretically complex. Technically – level of skills required for installation, on-going maintenance and routine operations is relatively high. This takes time. Good understanding of theory is needed to be able to combine archives, records management, digitisation and preservation practices with technical ability to carry out the work. In addition need to write it down – this forms part of overall metadata.
Archive software	• If CPR goes it alone, they could consider ICA Atom or similar OS software. This still has same hard and software installation needs as a DSpace type of installation, but seems more out of the box. This is because the compromise is that it is a much simpler system. It doesn't have the ability to do as much as DSpace. • ICA Atom (currently) needs web server [11kb download – no headroom given; using PHP 5.3 or later] and database server. They recommend Apache and MySQL software for these servers [both FOSS]. • Looks to be much simpler to install and run; therefore to manage. Can still export to CADAIR if that becomes an option. • Compromise is that all those I have seen have same look to them.
End aim	• Keep copies on external hard-drives as safety copies – still need management. With a spreadsheet configured using XML should be able to import metadata to both D-Space and ICA atom based system. • In time add repository software when have funds, time and people

Summary of Storage Option Tables	
What	Why
	available. This gives at least 2 copies (3 or 4 would be better) of each file. Means if things do go wrong, can recover quicker than if had to do it all again. • From an insurance point of view, can show taken reasonable steps to safeguard against loss of files. Also have metadata of archival material showing condition, that you had it etc. on a date.

Table G1: Table Summarising Storage Options Presented to CPR