

Understanding the costs of digitisation

detail report

CC403D006-1.0

27 October 2009

Cover + 47 pages

Dr Max Hammond
Dr Claire Davies

curtis+cartwright 

Curtis+Cartwright Consulting Ltd

Main Office: Surrey Technology Centre,
Surrey Research Park, Guildford
Surrey GU2 7YG

tel: +44 (0)1483 685020
fax: +44 (0)1483 685021
email: postmaster@curtiswright.co.uk
web: <http://www.curtiswright.co.uk>

Registered in England: number 3707458

Registered address:
Baker Tilly, The Clock House,
140 London Road, Guildford,
Surrey GU1 1UW

Executive summary

- 1 This document has been prepared by Curtis+Cartwright Consulting Ltd on behalf of the Joint Information Systems Committee (JISC). This document is the detailed output of a study to synthesise the experiences of a range of digitisation projects to provide JISC and the digitisation community with an evidence base to support funding allocation, project planning and project and programme management. Case studies are drawn from five digitisation projects that supported this study, and links are provided to other resources that provide supporting information.
- 2 Digitisation projects are unique, and many elements are interdependent. It is not possible to provide a formula (or even approximate figures) to cost a generic project. This report highlights areas which are likely to require significant resource to complete, and suggests approaches which will minimise the risk of a project running into difficulties, and maximise the efficiency with which it can be conducted.
- 3 Guidance in this report is based on lessons from these case study projects, but the lessons are limited to those that affect the cost or efficiency of a project.
- 4 The projects that supported this study ran into relatively few unexpected problems. This probably reflects the increasing maturity of the digitisation community; many issues have already been identified and synthesised into guidance elsewhere. Nonetheless, a few specific issues have continued to cause difficulties:
 - For large-scale digitisation projects (of around £0.5m and greater), a full-time project manager is necessary.
 - Recruiting staff can be challenging. Digitisation projects are typically short-term, and some regions do not have an established pool of staff available to fill these contracts. Difficulty in recruiting staff will have knock-on impacts on project schedule and budget.
 - University administration may have a different view to project directors of the grade that staff working on digitisations projects should be appointed to. This can cause delays in recruitment.
 - Collections that are not well understood (poorly indexed, heterogeneous) are difficult to digitise, and present particular problems when outsourcing digitisation. Effort spent upfront investigating the collection is helpful in reducing uncertainty.
 - It is important to consider the tools that will be used to catalogue collections, in detail and upfront. If digitisation is being outsourced, this includes the tools that the supplier will provide to the project. Inadequate tools will significantly hamper workflow throughput.
 - Clearing Intellectual Property Rights (IPR) is very challenging. Undertaking IPR clearance will require substantial effort and expertise – but may be worthwhile nonetheless.
 - Digitisation projects usually end by transitioning to a service to make the content available online. The design of this service, both technical and aesthetic, should be considered in detail early in the project, as it will affect many other decisions.
 - Services that make available collections of significant public interest may attract extensive media attention. A plan should be in place for handling this, both in terms of staff time and technical capacity to handle demand.

This page is intentionally blank

List of contents

Executive summary	1
List of contents	3
Document history	5
1 Introduction	7
1.1 Introduction	7
1.2 How can this document help me?	8
1.3 What is in the document?	8
2 The foundations	9
2.1 Introduction	9
2.2 The planning process	9
2.3 Defining objectives	10
2.4 The organisational context	10
2.5 Content selection	12
2.6 Partnering and sub-contracting	14
2.7 Staff requirements, skills, retention and training	15
3 Project management	19
3.1 Introduction	19
3.2 Project start-up	19
3.3 Ongoing project management	20
3.4 Information management	21
3.5 Quality Assurance	21
4 Content capture	23
4.1 Introduction	23
4.2 Logistics	23
4.3 Pipeline	23
4.4 Capturing text	24
4.5 How much metadata can be added during capture?	24
5 Metadata generation	27
5.1 Introduction	27
5.2 Planning	27
5.3 Generating metadata	28
6 Procurement	31
6.1 Introduction	31
6.2 Defining requirements	31
6.3 Tendering	32
6.4 Managing relationships	33
7 IPR	35
7.1 Introduction	35
7.2 Rights management	35
7.3 Rights clearance	35
8 Service delivery	37

8.1	Introduction	37
8.2	Website design	37
8.3	Technical considerations	39
8.4	Sustainability	40
8.5	Media Relations	40
8.6	Satisfying funders and partners	41
A	Case studies in detail	43
A.1	Overview	43
A.2	The John Johnson Collection	44
A.3	LBC/IRN Archive	44
A.4	Oldbaileyonline	45
A.5	Archival Sound Recordings	46
A.6	Freeze Frame	46

Document history

Version	Date	Description of Revision
0.1	20 August 2009	Internal draft
0.2	20 August 2009	Combined separate authors, draft for internal review
0.3	21 August 2009	Draft for review by client
0.4	15 September 2009	Interim version for review, awaiting further input from projects
0.5	15 October 2009	Draft for approval by client
1.0	27 October 2009	Release version

This page is intentionally blank

1 Introduction

1.1 Introduction

- 1.1.1 This document has been prepared by Curtis+Cartwright Consulting Ltd on behalf of the Joint Information Systems Committee (JISC). Since its inception in 2004, the JISC digitisation programme has funded a wide range of work to make available unique, hard-to-access material. This document is the detailed output of a project to synthesise the experiences of a range of digitisation projects to provide JISC and the digitisation community with an evidence base to support funding allocation, project planning and project and programme management. Not all the projects investigated here were JISC-funded.
- 1.1.2 Digitisation projects are distinct, and it is not possible to provide a formula (or even approximate figures) to cost a project. For example, two of the projects considered here are the LBC/IRN Archive, and the British Library Archival Sound Recordings (ASR) projects. Both of these projects were digitising audio recordings, but whereas the LBC/IRN collection is a homogeneous set of similar media carriers in good condition, with an existing (albeit partial) catalogue, ASR had a massively diverse collection on a range of media (some of which required conservation). There were no major problems with the intellectual property rights in the LBC/IRN collection, whereas the ASR materials had varied and unknown rights held in them, some of which were held commercially.
- 1.1.3 Attempting to compare these two projects quantitatively is unhelpful – the numbers could be generated, but without full consideration of the context, they would be meaningless. Each collection has specific content, condition and location – and digitisation must be planned as a unique effort.
- 1.1.4 This report highlights areas which are likely to require significant resource to complete, and suggests approaches which will minimise the risk of a project running into difficulties, and maximise the efficiency with which it can be conducted.

Acknowledgements

- 1.1.5 This project has greatly benefitted from project staff giving up their time to support this work, and contributing to the case studies in this document. In particular, the following individuals' input has been invaluable:

Individuals	Project
Richard Ranft Peter Findlay	Archival Sound Recordings
Heather Lane Naomi Boneham	Freeze Frame
David Tomkins	John Johnson Collection
Robert Shoemaker	Oldbaileyonline
Stanley Peters	LBC/IRN Archive

- 1.1.6 Further information about these projects can be found at Annex A.

1.2 How can this document help me?

- 1.2.1 In the last decade, millions of pounds of public funding have been available to investigate and experiment with digitisation and online services. This funding has resulted in the creation of a vast quantity of digital material, the deployment of a considerable ICT infrastructure, and the development of a significant body of expertise in the sector.
- 1.2.2 In the current climate of budgetary constraints and pressure to demonstrate value for money, there will be less funding available and a focus on delivering projects that make best uses of the resources available. Therefore, it is imperative to capitalise on the expertise in the sector to enable future digitisation projects to be conducted more efficiently (lower costs and effort), more effectively (better results) and with fewer problems and unforeseen complications.
- 1.2.3 This document is based on the experiences of completed digitisation projects. It will aid you in planning and costing your digitisation project, to help keep your costs down, help you anticipate potential pitfalls and manage procurements more effectively. It will also help programme managers with budgets for content digitisation to allocate money and draft and assess tenders more easily.

1.3 What is in the document?

- 1.3.1 This document contains information on potential costs (High, Medium and Low), good practice, key things to consider and common pitfalls at all stages of a digitisation project – project start-up, content selection, content capture, generation of metadata, sub-contracting, project management, service delivery, *etc.* It also shares the experiences of other digitisation projects and provides links to useful sources of further information.
- 1.3.2 This document does not contain a formula into which you can input details of your collection and output the cost of the project – there is no standard digitisation project, and such predictions are unlikely to be accurate. Each digitisation project is different, from the actual content and how you choose to digitise it, through to IPR issues and ease of staff recruitment.
- 1.3.3 Within this document, “factor profiles” are used to compare some of the options. These reflect the likely cost and effort required to overcome successfully. They do not imply that high-cost activities should necessarily be avoided, but they should be highlighted early in the planning process for careful consideration.

High	Options that are likely to require substantial expense or effort to successfully implement. If these are not considered carefully, there is the risk that they will lead to substantial cost overruns
Medium	Options which will require an intermediate level of effort or expense
Low	Options which generally require less effort or expense

Table 1-1: Example factor profile

2 The foundations

2.1 Introduction

2.1.1 This section starts by providing an overview of the process of planning a digitisation project. It then provides some information on things it is useful to consider upfront (as early as the bid writing stage), including:

- defining objectives for the project (sub-section 2.3);
- considering the organisational context (sub-section 2.4);
- selecting the content to be digitised (sub-section 2.5);
- using partners and sub-contractors (sub-section 2.6);
- staff matters: requirements, skills, retention and training (sub-section 2.7).

2.1.2 Many of these elements will be expanded within the remainder of the document.

2.2 The planning process

2.2.1 Planning and costing a digitisation project is generally guided by the requirements of funders. Some funders may have more rigorous requirements than others, and you may also be competing with other organisations/departments for limited funding.

2.2.2 How much planning you can do at the bid writing stage, and how much will need to be done once the project is underway (see "Detailed Planning", p.20) will depend on your particular project. It can be tempting to plan every detail of a project at the very start, but this is not always possible and can result in nugatory effort, particularly if the bid and project manager are not the same person.

2.2.3 In all circumstances, the more that is known about the collection to be digitised, the better you will be able to plan the project. If the content is poorly understood (and this is not recognised upfront), it may end up costing more money and taking more time to complete the project, or you may not be able to digitise all of the content you had hoped to.

2.2.4 For example, if the content is in boxes – how many items are in a box, what state is it in, how homogeneous is the collection? Often the people managing the bid and project are not the ones who know the most about the collection. If you don't have this kind of information about your content, find out if anyone else does – *eg* the curator. If not, either take some time to understand it at the bid writing stage or plan the project to take this into account - *eg* allocate a contingency budget, plan a project review halfway through.

2.2.5 For further information on content selection, see sub-section 2.5.



Prior understanding of a collection helps planning

The **LBC/IRN Archive** consists of a known number of audio tapes of known format, which had extant catalogue information (although not at sufficient detail for the desired outputs of the project). Tapes were selected for the project with the aid of the catalogue, based on clear criteria, and were individually identified and bar-coded for automatic processing. This level of existing information meant that there was a good understanding of the volume of material to be digitised, and it was simple to track the process.

In contrast, the **John Johnson Collection** consists of a range of printed ephemera separated into boxes of similar materials. Although the material was known to be relatively homogeneous (all printed material), there was no item-level index of the majority of the collection, and only a rough estimate of the quantities of material present. Some initial counting exercises provided what turned out to be surprisingly accurate estimates.

It was not possible to know in advance the number of scans that would be required, and the digitisation contractor charged per scan rather than per item. The project adopted a workflow that catalogued each item before digitisation, to aid tracking and to highlight potential copyright issues at an early stage. The uncertainty in how much of the collection could be catalogued persisted to the end of the project.

2.3 Defining objectives

2.3.1 The key objective of the content digitisation will determine how the project should be conducted. For example, the key objective might be one of:

- **preserve fragile content** – *eg* in this instance you might need to focus on high quality scans but limited metadata generation;
- **improve availability of content** – *eg* the service functions you would like to offer users (*eg* search facilities) will guide the metadata you need to generate;
- **meet a funding body objective** – *eg* this may influence the content you select to digitise;
- **develop an in-house capability** – *eg* in this instance you may wish to invest in equipment, train staff and build a body of in-house expertise.

2.3.2 If you are working with partners (commercial partners, institutions and other public sector organisations), it is important to consider what their key objectives are as well – they may be different to yours. For example, if it is a public-private partnership, the commercial partner may be providing their services at a discounted rate in return for being able to make money from the content in the future – this will influence, for example, the metadata that needs to be generated.

2.3.3 You should also consider how you will measure if the project has met its objectives. This may comprise defining measures of success such as number of digitised items, the number of users to a website in the first 6 months *etc.*

2.4 The organisational context

See also: 8.4 – Sustainability

2.4.1 The support of your organisation will be a crucial aspect of successfully bidding for and completing a digitisation project – you will have to work within the constraints of its IT, legal

and HR departments. They will be responsible for major decisions such as the grades at which you recruit staff, the level of Full Economic Costing (FEC) requested, and establishing consortium agreements and sub-contracts.

- 2.4.2 Internal organisational bureaucracy is frequently cited as one of the major “headaches” of a digitisation project – particularly as you will likely be dealing with multiple departments (HR, IT *etc*). It can take significantly longer to get things done than anticipated, so early engagement and negotiations are crucial. Project managers have frequently found that negotiating and prompting internal departments is a very time consuming part of their role.
- 2.4.3 It is also important in your funding proposal not to underestimate how long activities such as staff recruitment can take – further information about this can be found in sub-section 3.1.



Be prepared to defend your staff grade decisions

Several projects had difficulty persuading their host institutions to recruit staff at the grade that they felt was appropriate. In both cases below, the institutions judged that the staff roles were technical rather than research, and graded the positions as appropriate, despite funding being available for the higher grades.

The deputy director of **Freeze Frame** (who had effective responsibility for the project) felt strongly that it would have been inappropriate to recruit at the lower grade, and argued her position forcefully with the host institution. After several months’ delay, the host institution accepted the higher grade, and recruitment could begin. The project deputy director felt that the staff she was able to recruit at this grade were a key factor in the success of the project.

The **LBC/IRN Archive** remained unable to recruit staff at the grade for which they had funding. This contributed to difficulties in filling posts with skilled staff, which had knock-on effects on the degree of quality assurance which the project needed to undertake.

Producing a business cases

- 2.4.4 Your organisation may require you to produce an internal business case (in addition to the funding proposal) to justify why the funding is being sought, how the project would form a good strategic fit with the organisation, how it will transition to a service, and what resources (time and money) are being committed. This may be an important prerequisite for getting your funding proposal approved.
- 2.4.5 Forming a business case does not necessarily imply that the service developed needs to be commercial – rather it should show how it aligns to the host institution, and how it will be implemented, and sustained long-term.
- 2.4.6 The Office of Government Commerce has developed guidance on business cases for public-sector services, which may be a helpful starting-point (see Further Information).

Capitalising on previous experience

- 2.4.7 Your organisation may have some previous experience of conducting digitisation projects. This may mean that the organisation is familiar with the procurement route, and that there is a body of expertise within your organisation that you could benefit from.

2.5 Content selection

See also: 2.3 – Defining objectives

- 2.5.1 Most readers of this document will have a collection in mind that they wish to digitise (this may be driven by the objectives of your project - see sub-section 2.3), but many will have to select groups from that collection due to budgetary or other constraints. Selection of content should be given careful consideration as there are various factors (discussed below) which will affect the cost of the project and must be considered upfront.

Physical characteristics

- 2.5.2 The physical characteristics of your content will influence your estimations of the content for digitisation, the methods used to capture the content, the preparation time required and ultimately the cost of the project. For example, if you are using sub-contractors, they may charge you for under- and over-estimations.
- 2.5.3 Digitisation of time-based materials such as analogue audio and video are likely to prove more challenging than "spatial" media (images *etc*). This is largely due to accessibility, *eg* analogue audio is very difficult to estimate in terms of hours without real-time playback, because carrier length is often a poor indicator of audio duration.
- 2.5.4 A particular question is whether content will require conservation before digitisation – particularly when it will be shipped to an external sub-contractor.

High	Fragile material Time-based media (video, audio)
Medium	Text-based resources
Low	Images

Table 2-1: Physical characteristics factor profile

Homogeneity

- 2.5.5 Homogeneous content is likely to prove easier (and less time consuming and costly) to digitise than heterogeneous content.
- 2.5.6 The digitisation of homogeneous content can generally be run as a single project, where the main obstacles are discovered at the beginning and the content estimations (*eg* number of scans, number of hours) will be fairly accurate. With heterogeneous content, you may find that you have more difficulty estimating the content, and will encounter new problems with each different type of media. In this instance, it may be helpful to define a set of sub-projects which each deal with the different types of media.

High	Mixed formats with specific technical requirements (<i>eg</i> audio carriers)
Medium	Common format, mixed materials (<i>eg</i> all photographic negatives, but varying size)
Low	Uniform materials (<i>eg</i> newspapers)

Table 2-2: Homogeneity factor profile

**Heterogeneous collections can be challenging**

The **Archival Sound Recordings** projects faced a broad range of challenges related to the variety of material they were attempting to digitise. The collections included everything from ambient sounds, through pop music to oral histories, recorded on an assortment of media, using a range of techniques. This led to technical, content, and IPR issues varying widely across the collections.

The ASR projects treated each of their collections essentially as separate projects due to the great variations between them. Different collections could be subjected to different degrees of automation in the digitisation process, and needed to be presented differently on the output website.

It is inevitable with a research collection that many of the problems with artefacts only become known when the material is digitised or catalogued. However, the broad scope of ASR increased the difficulty in estimating the content prior to bidding for funding, which presented challenges keeping all facets of the project on schedule as problems only became clear later.

Intellectual Property Rights (IPR)

- 2.5.7 Institutions and other public sector organisations are often custodians of content protected by copyright and related rights (*eg* performance rights). Some of the collection you wish to digitise may be protected by such rights, and you are obliged to seek permission for providing online public access to this content.
- 2.5.8 Rights clearance is a very time-consuming, and thus expensive, activity. Furthermore, it may be necessary to make financial offers to rights owners to clear the rights. Consequently, many institutions and organisations are deterred from digitising copyright-protected content despite much of this material having high academic, cultural and historic worth.
- 2.5.9 It is important that you determine whether your content is protected by copyright. This will enable you to decide between:
- Avoiding copyright-protected content, and thus avoiding the problem.
 - Digitising copyright-protected content, in which case you will need to make detailed plans about how to approach this. Further information about clearing rights, and external references, can be found at Section 7.



Plan an approach to IPR early

Freeze Frame opted to digitise only material for which they held the IPR, and a very small range of material for which they could easily obtain permission (due to existing close relationships with the rights holders). Taking this decision upfront removed the issues of rights clearance from the project.

Archival Sound Recordings, on the other hand, saw their project as a key “test case” in clearing rights to use their collections online. They adopted an assertive approach to clearing rights, employing rights clearance specialists, developing systems to document the clearance process, creating a range of licences to use, and taking a risk-balance approach throughout. This required significant effort, the formation of an IPR board for the project, and taking external legal advice – but enabled the delivery of important collections online.

Demand

- 2.5.10 If the primary objective is to provide a service to the sector (and wider markets), careful consideration of the potential users of the service is critical. This is particularly important when considering the sustainability of the service – you will likely be required to justify the service’s existence to secure future funding, and this may be determined by number of users and quality of user experience.
- 2.5.11 It is therefore important to research (and demonstrate) who would want to use the content, how they would want use it, and if there is a wider market for the service. For example:
- what courses might the content support (higher education, further education and schools)?
 - what research groups might make use of the content?
 - is there a public interest? can this be exploited to generate income for the service?
 - is the collection rare and of special interest to a niche group of users?
- 2.5.12 Information about how they might use the content should inform other areas of the project such as the design and functionality of the website, the way in which the content is presented and the associated metadata.
- 2.5.13 The Strategic Content Alliance has produced The Guide to Researching Audiences which is a good starting point to help you to understand how to go about this research (see Further Information).

2.6 Partnering and sub-contracting

See also: 2.5 – Content selection, 6 – Procurement

- 2.6.1 Many digitisation projects are conducted as partnerships (across the public and private sectors) and make use of sub-contractors. You may choose to work with an external organisation to undertake specific activities (*eg* content capture, metadata generation, website design and build, sales) or to provide additional resources, funding and expertise. There may also be mutual benefits from working with other organisations, which is often the case in public-private partnerships – you may get free (or reduced cost) services if the private company benefits from the investment (*eg* by being able to sell the service to a wider market).

- 2.6.2 It may be obvious from the outset who you wish to collaborate with or sub-contract, or you may have to do some research – the partners employed by projects in the JISC Digitisation Programme may be a good starting point. In all instances, it will be very specific to your organisation, your collection and the objectives of your project.

Investment in equipment

- 2.6.3 Inherently related to the decision of whether to partner/sub-contract with external organisations is the decision of whether to invest in equipment. Depending on the route you choose, you may wish to request funding for equipment (*eg* scanners) to be used and retained by your organisation or your partner/sub-contractor.
- 2.6.4 Investing in equipment is a good option if you wish to develop an in-house capability such that you can digitise additional content once this project has been completed. You would also need to consider how to retain the people skilled to operate the equipment – this is often difficult if staff are on short-term contracts. Engaging partners/sub-contractors to operate equipment on-site and transfer knowledge might be possible, which avoids the issue of staff retention.

2.7 Staff requirements, skills, retention and training

See also: 3.2 – Project start-up

Project manager

- 2.7.1 Project managers often have responsibility for multiple projects at once, or juggle their day-to-day responsibilities with managing a project. However, if you are running a project on the scale of those undertaken within the JISC Digitisation Programme (*ie* £0.5M and greater, 24 months) you will need a full-time project manager. The scale of this task should not be underestimated, and whilst the project manager may be able to undertake some specific project activities, they should not be committed to unrealistic targets.
- 2.7.2 To increase the chances of success of your project, it is important to recruit a project manager with a broad skill base, which includes an understanding of the technical aspects of the project.

Project staff

- 2.7.3 Some projects will be conducted by teams that are already constituted, or by existing staff, but many require a team to be created from scratch to bring in new expertise and additional resources. In the latter case, you will need to allocate time and budget to recruiting and training staff (for more information about staff recruitment see sub-section 3.1).
- 2.7.4 The calibre of staff you recruit will be a major factor in the successful completion of the project. Whilst some elements of the job may be repetitive, projects often require staff with a specialised background and who will be competent and engaged in the project. If recruiting, you may wish to take the following into consideration:
- **Availability:** What is the availability of appropriately skilled staff in your area? It can be difficult to recruit skilled staff on short-term contracts, and it is even harder outside of London and other major cities. This may influence your decision to partner or sub-contract.

- **Grade:** You will need to decide what grade to recruit staff at, and be able to justify and agree this with your organisation – they may have different ideas to you! A common argument can be projects wanting to recruit at researcher grade, and organisations wanting to recruit at technical grade. You are less likely to be able to recruit skilled staff at a technical grade.
- **Skills and training:** You will need to determine what training staff will require at the beginning of the project, and how this will be done. You may have to invest more time in training staff at the beginning of the project if you have recruited at lower grades.
- **Retention:** Retaining staff throughout the duration of the project is important to minimise the time and effort spent recruiting and training any additional staff. Many of the jobs within a digitisation project can be repetitive, so consider how to make the job more varied and interesting – *eg* splitting roles, rotating staff. If you would like to retain any staff after the project has closed, you will need to plan for this – could your digitised collection earn any income to retain a staff member? Consider also that staff on fixed-term contracts may well be looking for alternative employment toward the end of their contracts, and may not wait until they are unemployed.



Staff recruitment may not be easy

Digitisation projects have staffing requirements that are distinct from most roles in a university: specific skills are required, but contracts are typically limited to 18-24 months.

The **LBC/IRN Archive** had difficulty recruiting staff to undertake the cataloguing and segmentation of their audio files. This was in part due to the project's location – there is not the same culture of temporary work in Bournemouth as one might find in larger cities, and it was not possible to attract people to the area for a short-term contract (especially given the staff grading issue discussed previously). As a result, significant extra effort was required to assure the quality of the work undertaken by the cataloguing staff.



Further Information

The LIFE Project has developed a methodology to model the digital lifecycle and calculate the costs of preserving digital information for the next 5, 10 or 20 years. < <http://www.life.ac.uk/>>

Handbook on Cost Reduction in Digitisation

<http://www.minervaeurope.org/publications/CostReductioninDigitisation_v1_0610.pdf>

Good practices in digitisation <<http://www.minervaeurope.org/bestpractices/listgoodpract.htm>>

Collections Link - planning a digitisation project

<http://www.collectionslink.org.uk/digitise_my_collection/digi_implement/digi_planning>

Low cost digitisation projects (from the perspective of developing and transition countries)
<http://www.eifl.net/cps/sections/services/knowledge/lcdp/downloadFile/attachedFile_f0/Low-Cost-Digitization-Report.pdf>

Guidelines for digitization of paper-based documentary heritage
<<http://www.minervaeurope.org/interoperability/digitisationguidelines.htm>>

Digitizing Collections: Strategic issues for the information manager, Lorna M. Hughes, 2003

Digital Preservation, Marilyn Deegan and Simon Tanner, editors, 2006

JISC Digital Media provides a range of materials that are relevant,
<<http://www.jiscdigitalmedia.ac.uk/crossmedia/>>
<<http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/potential-sources-of-funding-for-digitisation-projects/>>

The Office of Government Commerce provides extensive advice on developing business cases and managing projects <http://www.ogc.gov.uk/documentation_and_templates_business_case.asp>

The Strategic Content Alliance Guide to Researching Audiences <<http://sca.jiscinvolve.org/audience-publications/>>

This page is intentionally blank

3 Project management

3.1 Introduction

3.1.1 Good project management is key to a successful project, but it is also an area that is commonly underestimated and undervalued. This section addresses the key costs, effort and pitfalls associated with project management. It covers the following activities:

- project start-up, including staff recruitment and detailed planning (sub-section 3.2);
- ongoing project management (sub-section 3.3);
- information management (sub-section 3.4);
- quality control (sub-section 3.5).

3.2 Project start-up

Staff recruitment

See also: 2.7 – Staff requirements, skills, retention and training

3.2.1 If you do not have a team in place for the project, staff recruitment is likely to be your first activity. Recruiting staff takes time. Typically, 3-6 months is required (following funding allocation) to recruit staff within a university setting, assuming that there are no administrative delays. When a funding body has set a fixed timescale for a project, it can be tempting to take an optimistic view of the time required – but this will just store up problems for later in the project.

3.2.2 Although it may be tempting to recruit project staff and a project manager in tandem, it is good practice to recruit a project manager first, and involve them in the recruitment of the remainder of the project team.



It takes time to recruit staff

The initial project plan for the **John Johnson Collection** anticipated that staff would be available from the project's start date. In reality, the project manager started 3 weeks into the project, and the remainder of the staff were in place 2 months after that (one month to recruit, and one month's notice in their previous positions). This led to a backlog against the scheduled cataloguing and preservation throughput, which the project staff had to work hard to make up over the remainder of the project.

Facilities

See also: 2.4 – The organisational context

3.2.3 Other early activities are likely to include moving to the project premises, setting up IT equipment and procuring equipment (eg scanners).

3.2.4 The time required to procure equipment is often under-estimated - there can be a significant time lag (eg 3-6 months) between thinking about the equipment that you require and it being operational. For example, you will need to develop a specification for the equipment, source it, negotiate a price, wait for delivery, install it and trial it.

Detailed planning

- 3.2.5 Depending on how much planning was conducted during bid writing (see sub-section 2.2), additional planning at project start-up may be beneficial – particularly if the project manager has been newly recruited and is not familiar with the details of the bid. For example, detailed planning of the logistics, Quality Assurance (QA) procedures and information management will help the project run more smoothly. It can also be helpful to familiarise yourself with the content lifecycle, object handling procedures and trial various activities so that you are aware of potential pitfalls early in the project, and to help you achieve higher throughput of content when the project is underway.
- 3.2.6 Front-loading the detailed planning can be a very good use of the lag time at the start of the project when you are waiting on equipment, recruiting staff and negotiating with suppliers.



Detailed planning can increase efficiency

An unexpected administrative delay at the beginning of the **Freeze Frame** project gave the project manager and deputy director an additional 3-month period before the project staff began work. They used this period to increase the level of detail in the project plans, developing workflows, preparing systems etc. Although unplanned, this additional time was hugely beneficial, allowing a very rapid start to work once staff were in place, and only minor changes to systems during the later course of the project.

3.3 Ongoing project management

See also: 2.7 – Staff requirements, skills, retention and training

- 3.3.1 Once the project is up-and-running there will be a considerable overhead for ongoing management, covering activities such as:
- staff management;
 - managing relationships with contractors and partners;
 - liaising with your institution (*eg* contractual matters, budget);
 - reporting requirements (progress updates, interim reports, steering group meetings *etc*);
 - workflow management;
 - quality control;
 - general administration;
 - PR (if your project is of high public interest, do not underestimate the time that may be required to deal with journalists around the launch of the service).
- 3.3.2 As noted in sub-section 2.7, for projects the size of those in the JISC Digitisation Programme, a full-time project manager will be required. Underestimating this level of effort is a common pitfall for digitisation projects. Not recognising the necessary level of project management effort upfront can lead to problems completing the project on time and budget, and the need to back-fill positions within the institution because additional effort is being devoted to the project.

3.4 Information management

See also: 2.7 – Staff requirements, skills, retention and training, 5 – Metadata generation, 7.3 – Rights clearance

- 3.4.1 One of the most challenging aspects of any digitisation project is keeping track of the content, the progress of the project and the information generated during the project, for example:
- the physical location of the content (*eg* in archives, in transit, with a sub-contractor);
 - the progress of each item of content (*eg* digitised, metadata generated);
 - the digital files created (*eg* master file, low resolution files, playback copies);
 - associated information (*eg* rights clearance status and supporting evidence);
 - metadata;
 - the QA status (*eg* has it been signed-off).
- 3.4.2 It is important to keep this information together for each item. For simple projects, this information can be managed in spreadsheets. However, for more complex projects you may need to develop your own system for tracking and managing the information (*eg* by creating your own database or using your library management system). How this is done, and who is responsible for it should be considered at the start of the project.



Plan your information management

Archival Sound Recordings identified the need to develop a “digitisation log”, but did not agree an approach to tracking the workflow with their external digitisation contractor. Spreadsheets were created, but these were not sufficiently detailed to be of use during the process. This led to complexities in identifying materials and tracking progress of work carried out externally. To manage the data properly, the project would have needed to develop a tracking database, shared with the contractor.

The **John Johnson Collection** developed a tracking system to ensure that each item could be tracked through cataloguing, conservation, transit, digitisation, return and reintegration with the collection. The item data was drawn from the catalogue record. This enabled easy exchange of data between partners and the digitisation contractor, and reduced the risk of items being lost or misfiled. The system worked well, and contributed to the smooth progress of the digitisation.

3.5 Quality Assurance

- 3.5.1 Quality Assurance (QA) is an important aspect of any digitisation project, and particularly when using external suppliers for content capture and website development. Failure to properly check the quality of a supplier’s work can have severe implications, particularly if you are contractually bound to do so within a certain timeframe – the supplier may not be amenable to re-doing the work for free. It is also important to apply QA procedures to internal processes *eg* to ensure that archives are created properly, and that content is returned intact to its original location.
- 3.5.2 The QA procedures implemented will be specific to each project, but you should consider at the start of any project how you will do it, how much QA you will do, when it will be done, who will do it and who will sign it off. In general, the outputs of manual processes (*eg* text rekeying) will require more QA than for automated processes. It is good practice to trial your QA procedures with early batches of outputs – is the process right, is it enough, is it too much? This is particularly important when you are trialling new methods and processes.

- 3.5.3 It can be useful to generate “product descriptions”¹ at the outset (*ie* at the procurement stage) which outline your minimum standards for the digitised content, from which you can set quality targets and ultimately evaluate whether the digitised content is satisfactory.
- 3.5.4 It is also useful to incorporate the QA status of items into your information management procedures (see sub-section 3.4).



Test your QA approach

The **LBC/IRN Archive** had planned to undertake “spot-check” QA on 10% of the metadata records they generated. They undertook some exploratory work to investigate the extent and accuracy of the legacy catalogue data they held, but it was only once the project was well underway that they could fully define the QA method – partly due to the quality of staff hired.

They found that due to several factors (including the difficulty in recruiting staff, the quality of the legacy catalogue data they were working with, and the lack of controlled vocabulary in the input tool), 100% of the output needed to be reviewed in order to maintain quality and consistency. This tenfold increase in the amount of QA required had significant implications for the project, requiring the recruitment of additional part-time staff and the extension of the project timescale.



Further Information

The infoNet Project management infoKit provides some project management advice
<<http://www.jiscinfonet.ac.uk/infokits/project-management>>

The Budgeting and Costing infoKit provides some further ideas
<<http://www.jiscinfonet.ac.uk/infokits/project-management/budgeting-costing>>

The Office for Government Commerce provides extensive information on project management
<<http://www.ogc.gov.uk>>

¹

Product descriptions should provide a clear statement of the purpose of the product, and describe the technical processes, requirements and quality criteria for the product.

4 Content capture

4.1 Introduction

- 4.1.1 This section covers the lessons and issues identified relating to the process of digitising content. By considering aspects such as logistics, the digitisation pipeline, automation and text entry upfront, you will be able to budget for the content capture more accurately. It is relevant for both in-house and outsourced digitisation – even if you outsource, you will need to consider all of these areas to both do the procurement and manage your sub-contractor.
- 4.1.2 The information in this section is necessarily at a high-level as the details will be specific to each project and its content. There are clearly significant differences between different media, but also between different collections in the same medium.

4.2 Logistics

See also: 2.5 – Content selection

- 4.2.1 It may be possible to move artefacts for digitisation (whether this is within the same building or to a remote location). Whether this is feasible or desirable depends on the collection, and should be considered upfront in the project planning process.

High	Unique collection – impossible to replace Fragile materials Bulky materials High (financial) value International shipping
Low	Easily-replaceable collection Well-packed for transport Uniquely identified with machine-readable identifiers Robust materials

Table 4-1: Logistics factor profile

- 4.2.2 If materials are shipped, the workflow adopted must track artefacts throughout the process. This is clearly important for logistics (and potentially, liability and insurance), but may also be necessary or helpful in performance management.

4.3 Pipeline

See also: 6 – Procurement

- 4.3.1 If you are outsourcing the content capture to a sub-contractor, your contract with them may stipulate a minimum and maximum throughput of content, outside of which you may be charged additional costs.
- 4.3.2 You will need to manage carefully the pipeline of work such that you do not incur these additional costs. A carefully organised, methodological approach to sending out content, and dealing with returns, will help you keep track of the artefacts.

4.4 Capturing text

- 4.4.1 For some resources, it may be appropriate to capture all the text in the source artefacts as electronic content, for example to enable full-text searching. A number of factors affect how well this can be undertaken:²
- scanning method;
 - nature and condition of the originals;
 - nature of the printing;
 - content;
 - output requirements.
- 4.4.2 There are two approaches to this: rekeying, or Optical Character Recognition (OCR). Both methods have advantages and disadvantages. Typically, accuracy can be increased for both techniques through additional expenditure – it is necessary to decide what level of accuracy is required – and how to define accuracy (per character? Per word?).
- 4.4.3 When rekeying text, the artefacts (or the scans of the artefacts) are viewed, and the textual content is manually entered. Often, a process of double rekeying is used, where each artefact is rekeyed by two separate individuals and the two versions are compared – any discrepancies can be investigated. This work is essentially data entry, and is typically well suited to outsourcing and potentially to offshoring.
- 4.4.4 As rekeyers are fundamental to the quality of the output, and the progress of the project, it is important to ensure that contractors engaged to undertake rekeying are appropriate for the project. Contractual mechanisms must provide appropriate flexibility; for example by letting a larger number of small contracts sequentially, rather than letting the whole task as a single contract.



Consider the order when rekeying

Oldbaileyonline rekeyed the text in a set of scanned images of trial proceedings, out-sourcing the text-entry to companies based in India. The project started with the oldest materials, and worked through chronologically. This meant that the first material that the rekeyers worked on was also the most challenging. With hindsight, the project team felt that it would have been more sensible to begin with the easiest material (which is the most recent in this case), gradually building up skills and experience at the rekeying contractors.

- 4.4.5 OCR is the automated processing of images to identify and digitise text, using specialist software and computer facilities. OCR is best suited to artefacts consisting of text typeset in a plain Latin font; it is much less effective for ornamental typefaces or handwriting.

4.5 How much metadata can be added during capture?

See also: 5.3 – Generating metadata

- 4.5.1 If it is possible to generate metadata during the capture process, this is a natural step to take. This is particularly relevant if material is being rekeyed, where the rekeyers can create some forms of metadata with relatively little overhead.

² See *Deciding whether Optical Character Recognition is Feasible*, Simon Tanner 2004 <http://www.odl.ox.ac.uk/papers/OCRFeasibility_final.pdf> [accessed 17 August 2009].

- 4.5.2 For material that is being digitised “mechanically” – scanned or recorded – some classes of metadata may be automatically generated, or completed in bulk (*eg* type of original, sequence within original collection). Opportunities for automatic metadata generation should be actively considered.

**Further Information**

IMPACT Conference: Optical Character Recognition in Mass Digitisation
<<http://www.ariadne.ac.uk/issue59/impact-2009-rpt/>>

Deciding whether Optical Character Recognition is Feasible
<http://www.odl.ox.ac.uk/papers/OCRFeasibility_final.pdf>

This page is intentionally blank

5 Metadata generation

5.1 Introduction

- 5.1.1 Creating metadata manually is a time-consuming process, often being dependent on skilled staff. For museum collections, it is often tightly linked to the process of cataloguing the collection. It is important to consider upfront what the metadata will be used for – it is not appropriate (for example) to generate extensive metadata for a collection in order to “do the job right”, where full-text searching may be a better choice.

5.2 Planning

See also: 2.3 – Defining objectives, 8.2 – Website design, 8.3 – Technical considerations

- 5.2.1 The creation of metadata must be linked to the objectives of the project, and to the approach that will be taken to delivering the outputs. Whereas it may be tempting to create extensive metadata records for each artefact, this will be an expensive process and must be carefully considered. Metadata should typically be collected for a particular reason: how will users benefit from the information? This needs to be right upfront – it’s not usually possible to go back and add missing data from a collection.



Only create metadata for a reason

Oldbaileyonline created textual records that are well suited to full-text searching, but decided to dedicate significant effort to manually “tagging” records (identifying and categorising the defendant, punishment, etc). This allowed statistical analysis of the data that it is not possible to undertake reliably using full-text search.

The decision to create extensive metadata was taken based on a careful analysis of the user requirements – it was clear that the potential users of the service would benefit from structured data in addition to the unstructured full-text. The project team (and the users of the service) feel that the investment in metadata tagging has been worthwhile for this resource.

- 5.2.2 Metadata can consist of a wide range of types of information; as well as “descriptive” fields (title, keywords, author, *etc*) an important function of metadata is to store administrative and technical information about the artefact (filename to which the metadata file relates, physical details about the original artefact, details of the capture process *etc*).
- 5.2.3 Some file formats can contain embedded metadata (for example, an MP3 audio file can contain information about the artist, album, year *etc*), but for the more complex requirements of typical digitisation projects, metadata is usually handled separately to the artefact. Managing complex metadata records can be difficult – data may come from several sources before being combined into the final record, and this must remain associated with the digital artefact, and will probably need to be updated at several stages of the digitisation workflow.
- 5.2.4 XML is now the standard format for recording metadata, and a range of profiles is available. METS is a widely used standard, which allows a range of other schemata to be embedded within one master file. METS itself is a meta-standard – a specific profile must be used for different types of artefact.³

³ <<http://www.loc.gov/standards/mets/>> [accessed 15 September 2009]

Selecting schemata

- 5.2.5 In a mature domain, where similar collections have been captured, there are likely to be a range of metadata schemata already defined which might be appropriate for re-use. Defining a new schema is a time-consuming and complex process, and prior work should be taken as a starting-point (or a complete solution) wherever possible. The collection might already be in a catalogue – it is necessary to consider whether the current schema will be appropriate for the needs of the collection during digitisation and delivery online.

High	Novel or unique material, no appropriate schema for describing artefacts
Medium	A range of schemata available No consensus on applicability Good but not ideal alignment with collection
Low	Existing schema, already widely-used for similar collections Good understanding of scope and applicability

Table 5-1: Metadata schema factor profile

5.3 Generating metadata

See also: 3.4 – Information management

- 5.3.1 When a collection is not thoroughly catalogued prior to the project, this will occur naturally during the process of metadata generation, but not necessarily in one stage. For example, an artefact may be catalogued at a very basic level prior to digitisation, and this catalogue record enriched afterwards, working from the digital copy.
- 5.3.2 The staff skills required for cataloguing will vary depending on the nature of the collection – is cataloguing a routine data entry task, or does it require curatorial expertise? In general, any metadata that can be generated automatically or in bulk should be.

Tools available

- 5.3.3 It is important to consider upfront the mechanism by which metadata (including catalogue records) will be created, as this has significant implications for the progress of the project. This includes the tools that will be used, and the workflow that will be applied; getting these right during project planning will decrease the risk to the project.
- 5.3.4 If digitisation is being sub-contracted, the contractor may be providing the cataloguing tool – the usability of this tool should be a criterion for selecting the contractor, and should be tested prior to contract award. Some suppliers will be flexible, and be willing to develop their tools during the project to meet the needs of the client.



Get the data tools right from the outset

The **LBC/IRN Archive** used a tool provided by their digitisation contractor to attach metadata to sound files. This tool was very basic, and did not provide functions such as controlled vocabularies, spell checkers, drop-down selections etc. Customisation of the tool to meet the requirements of the project was not considered during contract award, and may have been prohibitively expensive to commission later. This led to significant issues with consistency and accuracy of input data, and greatly increased the effort required for data entry and QA.

The **John Johnson Collection** opted to continue entering data using the catalogue system which contained the existing collection catalogue. This was legacy software, running in DOS, which may have reduced the throughput of the cataloguers. Early in the project, a later windows-based version of the software was trialled, but it did not have many of the customisations that had been made to the DOS version. Adopting the Windows software could have been beneficial in terms of cataloguing throughput if this decision had been made (and the necessary customisations implemented) prior to the start of the project.



Further Information

A JISC metadata template for digitisation projects <<http://deposit.depot.edina.ac.uk/124/1/jisc-digitisation-metadata-template-2007.rtf>>

This page is intentionally blank

6 Procurement

6.1 Introduction

6.1.1 This section builds on the basic information set out in sub-section 2.6 (the foundations: partnering and sub-contracting) to consider more detailed aspects of the procurement of services, covering:

- defining requirements (sub-section 6.2);
- tendering (sub-section 6.3);
- managing relationships with sub-contractors (sub-section 6.4).

6.1.2 The process of procurement needs to be carefully planned to ensure that it can be completed in a timely manner. The British Library's Archival Sound Recordings project conducted a large and complex procurement, and many of the lessons from that exercise are incorporated within this section.⁴

6.1.3 A particular problem for grant-funded projects is that they must estimate the cost of digitisation during application for funding, but the true cost may not be known until the tendering process has been completed, particularly for complex collections. A number of approaches may be taken to reduce the risk of a poor estimate, including benchmarking against other digitisation projects, applying an appropriate costing tool, and informally approaching vendors.

6.2 Defining requirements

6.2.1 If you are planning to procure elements of the project from partners or sub-contractors, it is important that you understand your own requirements, and that these can be written down and understood by partners and suppliers who will be responding to your tender. So-called requirements engineering is a complex process, but one which is necessary to successfully complete large-scale procurements.⁵

6.2.2 This work could be included in the detailed planning at the start of the project (see sub-section 3.2), but alternatively an initial benchmarking or development activity could be undertaken within the project itself. This may or may not be conducted by the primary supplier. Preliminary work on the technical requirements and standards will save time in liaising with the supplier during the main procurement.

6.2.3 Requirements should be expressed in terms of functional requirements rather than technical approaches wherever possible – this allows the bidders to suggest alternative or innovative approaches, and to play to their own strengths. Do not assume that you understand the process of digitisation better than potential bidders do.

Estimating content

6.2.4 Understanding the nature of the material to be digitised is obviously of particular importance when the digitisation is out-sourced. Potential bidders will need to know the volume of

⁴ *Archival Sound Recordings Project: Lessons Learned*. British Library Digitisation Programme. Version 10.2, 5 September 2007.

⁵ For a good introduction, see *Writing Better Requirements*, Ian F Alexander and Richard Stevens, 2002.

material, but also details of the physical nature of the artefacts. For collections with specific technical requirements, bidders are likely to require access to representative samples in order to prepare a proposal; this must be planned and facilitated.

- 6.2.5 Where collections are heterogeneous (or composed of several sub-collections), each segment may need to be considered separately. Allocating effort upfront to estimating and understanding of the collection can yield benefits later, as the risk to suppliers is lower so they may be able to offer lower prices.

6.3 Tendering

- 6.3.1 Procurements on the scale typically conducted for digitisation projects are likely to be above the threshold for European Union procurement, so public bodies (including universities, museums etc) will be required to follow the EU process.

- 6.3.2 A detailed discussion of the process is beyond the scope of this document, but an important choice will be whether to follow Open, Restricted, Negotiated or Competitive Dialogue procedures. Open and Restricted procedures are appropriate where the requirement is well defined, whereas a Negotiated procedure is more likely to be appropriate where the requirement cannot be fully defined at the outset of the tender procedure. The Competitive Dialogue procedure is unlikely to be appropriate for digitisation projects (being intended primarily for major Public/Private Partnership (PPP) or Private Finance Initiative (PFI) deals).

High	Negotiated procedure Competitive Dialogue procedure
Medium	Restricted procedure Open procedure
Low	Single-tender action (<i>eg</i> with commercial partner in project or another department in the institution)

Table 6-1: tendering factor profile⁶



Consider your procurement approach

Archival Sound Recordings undertook an EU Negotiated Procedure to identify a digitisation contractor. This procedure allows short-listing of suppliers who are then given equal opportunity to develop their proposals in negotiation with the client, before presenting a Best And Final Offer (BAFO). For ASR, this took 10 months (rather than the expected 6), but allowed the project to consider a variety of diverse approaches to their complex requirements.

This procurement went well, but in hindsight the project team felt that it would have been more effective to spend more effort upfront to define the requirements more clearly – perhaps by procuring specialist support through an open competition. This would make the larger procurement for the bulk of the work quicker and easier, and may have reduced the overall price, as the opportunity was less risky for the contractors.

“Negotiating over a contract for highly technical digitisation activities is complex and time-consuming and offsets some of the benefits of competitive tendering. [...] there may be too many risks and dependencies to manage when undertaking digitisation of complex archival materials.”

⁶

Note that this is the effort/cost involved in the tendering process; it does not consider whether each approach would provide the best value for money in the product procured.

6.4 Managing relationships

- 6.4.1 It is important to consider during the procurement process the approach that will be taken to managing supplier relationships during the project. Some elements will require a contractual basis, such as performance metrics and change procedures.
- 6.4.2 Appointment of a named individual as the project manager on the supplier side should be a contractual requirement. Having a single individual responsible for delivery (and accessible to the client) is important in establishing good working relationships – and these are essential for successful delivery of the project.
- 6.4.3 A formal change procedure should be agreed and incorporated within the contract. This provides a clear mechanism by which either party can request a change to the agreed work, and helps to ensure that both parties' understanding and approach to the work remain aligned throughout the project.

Special problems of intra-institutional collaboration

- 6.4.4 In cases where the digitisation "supplier" is another department within the institution, particular issues arise. In most cases, there is a weaker contractual basis – it is unusual to draw up formal contracts between departments within an institution. This relationship must be carefully managed; although setting up a contract may not be a sensible use of resources, having an agreed, written statement of the work to be undertaken is essential (as well as a named project manager – see above).



Further Information

JISC Digital Media maintain a list of commercial digitisation suppliers
<<http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/digitisation-services/>>

This page is intentionally blank

7 IPR

7.1 Introduction

- 7.1.1 The approach to Intellectual Property Rights (IPR) is fundamentally tied to the objectives of a project, and must be considered early (see section 2.5). Two elements must be considered – the rights that the project (*ie* the institution holding the collection, in most cases) holds in the content, and the rights that the service will grant to its users at the conclusion of the project.
- 7.1.2 There is extensive advice available on managing IPR, and this section will not reproduce that content – rather, the implications for the management of digitisation projects are discussed.

7.2 Rights management

See also: 8.3 – Technical considerations

- 7.2.1 When considering IPR in content, it is necessary to consider the use to which the material will be put. This will affect the approach taken to rights clearance, and may affect technical decisions on how to deliver the service. The complexity of this problem will be very dependent on the collection, but useful questions are:
- Who will have access to the final service? (*eg* UK education, libraries, anybody)
 - What restrictions will be placed on use of the service?
 - Will there be a chargeable service, or will it be delivered free?
 - Can international access be controlled?
- 7.2.2 It is important to consider whether the planned approach to service delivery will accommodate differentiation of access rights (*eg* provide the service free of charge to UK HE, but charge overseas).

7.3 Rights clearance

See also: 2.7 – Staff requirements, skills, retention and training, 3.4 – Information management

- 7.3.1 Projects that have opted to undertake rights clearance should carefully consider the staff skills, and workflows required to complete this successfully. Rights clearance is a specialist task, and will usually require specialist staff.
- 7.3.2 It may be possible to use rights clearance agencies or contractors, but the experiences of projects that have used these have not always been good; digitisation projects are outside their normal lines of business.
- 7.3.3 For high-risk collections (see box below), IPR clearance is likely to be a complex process requiring significant effort, requiring investment of senior management time and specialist expertise to establish workflows and manage the process. For these collections, it is essential to maintain records of the rights-clearance process – especially when a project is adopting an assertive, risk-managed approach to rights clearance.

High	Commercial materials Mixed collections, where individual artefacts have different rights Collections with the IPR status is unclear or unknown
Medium	Non-commercial materials Research collections
Low	Collections where the IPR are already fully held by the project Collections which are out of copyright

Table 7-1: Rights clearance factor profile



IPR clearance can be a significant challenge

Archival Sound Recordings saw IPR clearance as a key objective of their project, and committed significant effort to this activity. The project took a risk-management approach to IPR, so it was necessary to retain detailed records not only of clearances granted, but also of attempts to identify and contact rights holders, in order to demonstrate due diligence. The project developed a database to maintain information on the rights status of the artefacts. It was difficult to appoint staff or contractors with the appropriate skill set for rights clearance – one contractor was released from their contract due to poor performance.

During the project, the host institution imposed additional requirements on IPR clearance, which created new requirements for the project, and increased the cost and complexity of undertaking rights clearance, already a difficult challenge for project management since the copyrights were controlled externally to the project by many different individuals and agents. The complexity of the rights clearance activities generated delays in the project.

8 Service delivery

8.1 Introduction

8.1.1 Delivery of the digital content as a service is a typical objective for many projects. Accordingly, the project is not just about digitising content; it is about designing and developing an attractive, usable service for the target audience. It can be very easy (and in the end, time consuming) to forget this!

8.1.2 This section addresses the key activities and pitfalls with service delivery, covering:

- website design (sub-section 8.2);
- technical considerations (sub-section 8.3);
- sustainability of the service (sub-section 8.4);
- satisfying partners (sub-section 8.6);
- media relations (sub-section 8.5).



Engage a broad range of users when designing the service

Amongst other user-engagement activities and usability testing, **Archival Sound Recordings** established a small user panel (12 people) at the outset of the project, which gave valuable feedback on user needs, although sound recordings were not yet well embedded within teaching, learning and research. At the outset of the project, use cases were loosely defined – the intent was to develop use cases as the project progressed. In order to get broader representation, an online user community was set up with over 60 participants.

8.2 Website design

8.2.1 Your website will be the main link between your content and your users, and is fundamental to the uptake of your service. Designing a website that is useful, informative, well presented, and accessible is challenging.

8.2.2 Website design is a common element of a project to sub-contract to an external supplier, and in these circumstances, it is important to work closely with your delivery partner to ensure that they deliver a product that meets your needs. A close working relationship is also important if a different team within your organisation is designing the website – they may not have the same contractual obligations, and may be less pressured to deliver an output with which you are happy.



Work closely with your delivery partners

Freeze Frame did not begin planning outputs until the second half of the project. The intent was to store the captured material in the institutional repository and to generate a new front-end using open-source software. Developing the delivery system turned out to be significantly more involved than anticipated – the repository software was not well suited to presenting resource packages in a way that would appeal to the project’s anticipated users. Working closely with the repository software provider and another technical services department at the university, the project overcame these problems under time pressure, to launch the service successfully.

Oldbaileyonline were similarly focused on the capture of material, with less attention paid initially to how to present it at the conclusion of the project. They had agreed to work with a technical services department at their host institution to develop the output site, but at the outset of the project did not have a clear plan for how the material would be presented, or how much the development would cost. As a result, some elements (such as the design of the website, as distinct from the information system behind it) had not been budgeted for.

The project’s major delivery partner was another organisation within the same university, and the relative informality of relationships between university departments had some disadvantages as well as benefits – expectations and responsibilities were not initially clearly defined. More rigorous relationship management procedures were implemented (including minuted meetings) and the project proceeded well.

User-centred design

- 8.2.3 Developers enjoy taking advantage of new technical possibilities – but this may not be best for your service. It is more beneficial to design the website around the requirements that exist within the target audience for your service, and then investigate which technologies could help meet these requirements. This will, of course, need to be balanced with meeting the objectives of your organisation and those of any partners’. However, bear in mind that if your website is not useful to your users and is consequently not used, this will prevent you from meeting the objectives of your organisation!
- 8.2.4 You will need to consider who the service is for and how they will use it, and use this information to focus your requirements (see also sub-section 6.2 – defining requirements). The design of your website will be informed by the audience research undertaken at the beginning of the project (see sub-section 2.5 - demand), and you will probably need to do some further audience research to inform the development of prototypes and content. For example, you could:
- Explore your users’ tasks and goals by conducting a survey.
 - Investigate specific problems or issues in focus groups.
 - Conduct iterative user testing to develop your interface to improve usability. This could involve you watching users undertaking representative tasks to identify problems with information flows. Three iterations are often recommended.
 - Analyse the use of any “beta” site by monitoring web statistics in conjunction to speaking with users to understand their experiences and behaviours.
- 8.2.5 This can be a time consuming activity, and you will need to consider what you can do within the time and expertise you have available. Audience research may appear daunting, but do not be put-off – it can be very valuable, and many techniques can be adapted to non-experts and done on small budgets. You may wish to use a market research company for some

elements of the research (*eg* to conduct a survey) – they may be able to undertake some tasks more efficiently and effectively than you would be able to.

- 8.2.6 It is important to start thinking about audience research early in your project as it may impact other areas of your work. For example, the required website functions will guide the metadata that you generate – you may be planning to generate more than is necessary.
- 8.2.7 See the Strategic Content Alliance’s audience guidance for further information (<http://sca.jiscinvolve.org/audience-publications/>).

8.3 Technical considerations

- 8.3.1 Several specific technical issues should be considered during project planning – although the delivery of the service may be outsourced or transferred to a third party, the project must consider the technical delivery when forming these agreements.

Hosting

See also: 2.4 – The organisational context

- 8.3.2 Consider who will host the service, and on what basis it will be funded. Specific issues include:
- How will content be presented? Is it an archive, a structured museum display, or a library? Different approaches will require different technologies, and may influence choices about which metadata to collect, and what formats to present digital content in.
 - Digitisation projects frequently create large volumes of data (many terabytes). Does the raw, hi-res data need to be available online, or could it be stored in near-line or offline storage, or even destroyed?
 - Will the hosting platform be able to handle the level of traffic which may be generated for a high-profile service? (See also sub-section 8.5)
 - How will future updates to presentation and/or content be managed?

Authentication

See also: 7.2 – Rights management

- 8.3.3 Consider the approach to authenticating users. Although the UK Federation is becoming established as the primary way of authenticating UK HE users, it is not yet ubiquitous. FE users in particular may not have UK Federation access. Will non-education users be granted access? In what circumstances? How will access be granted and managed? Any commercial access to the service will require management of contracts, and relating these contracts to the authentication information.
- 8.3.4 An important consideration is that rights holders may permit the release of their materials to educational users, but not more broadly. How will you reassure them that you will comply with their terms?

Discoverability

- 8.3.5 Even if the content will only be available to a closed group of users, it is likely to be beneficial to make content discoverable from the broader web. This can typically be achieved by

exposing metadata publicly. Several of the projects followed this approach – although their content was restricted to the UK HE community, they exposed the metadata to search engines. This allows users to discover content through public search engines, and may enable broader user of the service.

8.4 Sustainability

See also: 2.3 – Defining objectives, 2.4 – The organisational context

- 8.4.1 A wide range of digitisation projects are funded in the education and research sector, not all of which can be supported by funders long-term. It is important that you consider early in the project how you will make your service financially sustainable after project funding has ended. This is particularly important in times of budgetary constraints, when support you are relying on from your host institution may not be as readily available.
- 8.4.2 A common model is to rely on a combination of support from the host institution (*eg* to provide web hosting) and generated revenue (*eg* to support staff salaries). There are a range of models for generating revenue, and will depend on your specific content and service. There may be opportunities to offer value-added services that you can charge for (*eg* prints of high quality images) or to open up the content to wider markets at a cost. It may be necessary to conduct additional audience research to determine if these are viable models for your service.
- 8.4.3 It is also important to note that some host institutions may not realise how much support they are providing to these services – it is often not transparent and is absorbed within other activities. If the support is not explicitly provided, you should not rely on it in the long-term.
- 8.4.4 A report by the Strategic Content Alliance (SCA) on sustaining digital resources is particularly useful to help you consider your options.⁷

8.5 Media Relations

- 8.5.1 If you have a collection that is of significant public interest, it is important to be aware that the launch of your service may capture the interest of the media. Whilst this is an excellent opportunity to publicise your service, it can be very time intensive giving interviews and attending events (perhaps taking up 1 – 2 months of time).
- 8.5.2 As this will happen in the latter stages of a project, it can interfere with any planned final tasks and project shut-down. This will be particularly problematic if the project manager and project staff are not being retained. It is therefore important to scope the interest of the media early in the project and plan your resource allocation.

⁷

<http://sca.jiscinvolve.org/files/2009/07/sca_ithaka_sustainingdigitalresources_report.pdf>

**Be prepared to deal with public interest**

Freeze Frame digitised a collection of images with exceptionally broad appeal, which generated a significant degree of public interest. The degree of media attention resulted in two months being spent by the project staff dealing with press enquiries full-time. Although the university press office was very helpful in managing the media, there is always a desire to have the content experts available for interview. This unexpected workload pushed back all other project work – although the publicity for the project was a clear demonstration of its value!

8.6 Satisfying funders and partners

- 8.6.1 When concluding a project, it is important that you revisit the original objectives of the project and ensure that you have met not only yours, but any partner objectives. If you have not met your objectives, this may not be a negative outcome as you may have learnt some useful lessons during your project. However, any lessons should be documented so that other projects do not encounter the same pitfalls.
- 8.6.2 It is also important to fulfil any reporting requirements to funders – this may improve your chances of obtaining funding in the future. Some funders may require projects to undertake self-evaluation, or to support external evaluation. External evaluation activities may take place some time after the project has been completed.

**Further Information**

JISC Digital Media have some guidance on ensuring that digital collections remain sustainable
<<http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/sustainability-of-digital-collections/>>

Using the entwined themes of the requisite organisational and technological infrastructure and dedicated resources, this tutorial addresses the establishment of a viable digital preservation program.
<http://www.icpsr.umich.edu/dpm/dpm-eng/eng_index.html>

Market Facing Digitisation considers commercial approaches
<<http://www.slideshare.net/nickpoole/market-facing-digitisation-1535437>>

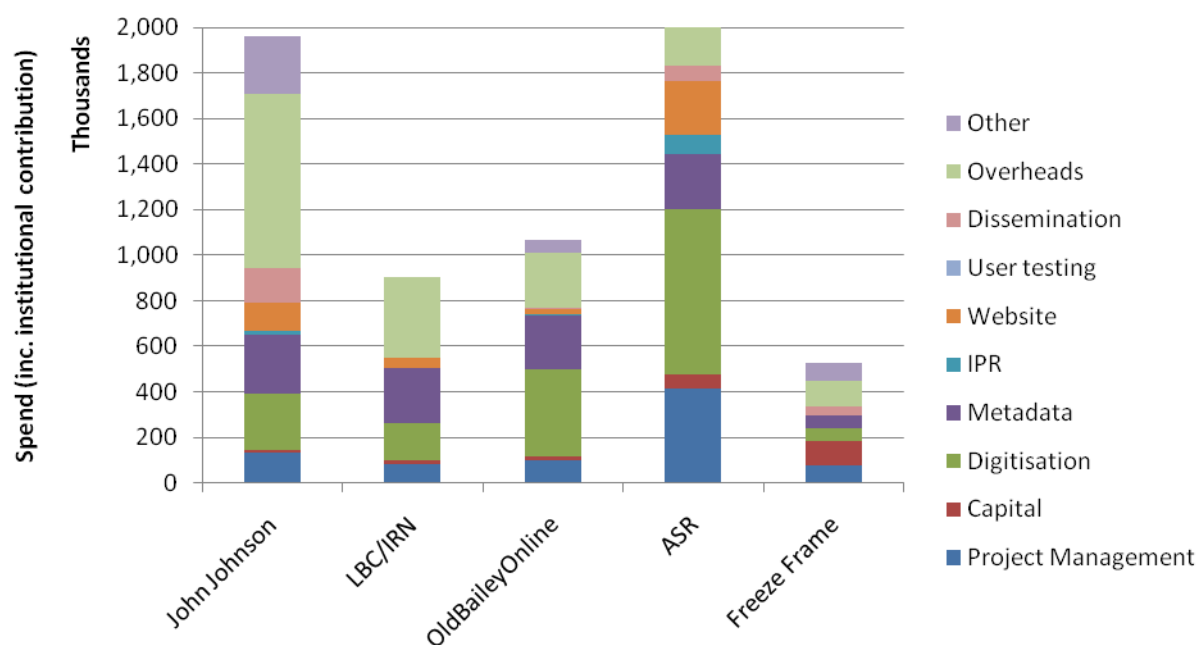
The DCC Curation Lifecycle Model provides a high-level overview of the stages required for successful curation and preservation of data <<http://www.dcc.ac.uk/lifecycle-model/>>

This page is intentionally blank

A Case studies in detail

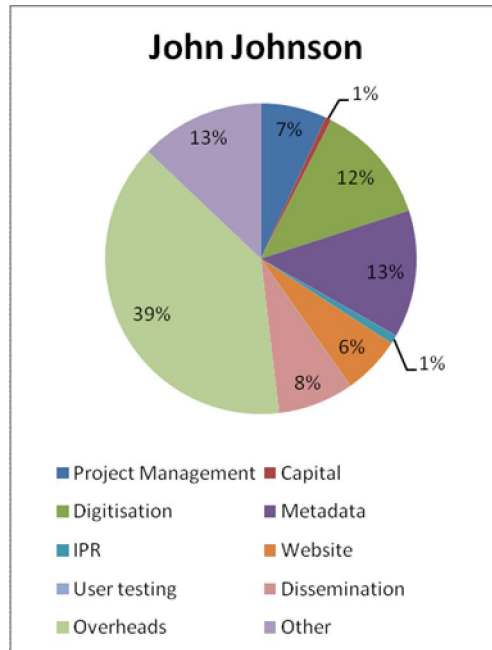
A.1 Overview

A.1.1 The projects investigated by this study digitised widely differing materials, and quantitative comparisons between them are of only limited value. Nonetheless, some cost information is given below, to help give some perspective to the scale of projects, and their overall composition. The breakdown of expense is the “best guess” of project managers; no projects undertook detailed recording of time and effort.



A.2 The John Johnson Collection

A.2.1 The *John Johnson Collection* comprises some 1.5 million items of printed ephemera dated from 1508 to 1939. It therefore covers a wide range of printing and social history and is of worldwide importance. At the start of this project the Collection was only partly catalogued and only 15,000 items were accompanied by digital images. The aim of the digitisation project were to make a much broader selection of these important and valuable documents available more widely, particularly to the UK scholarly community. The new digital resource was to be supported by the creation of extensive, carefully controlled metadata and a variety of finding and searching tools to facilitate its full exploitation.

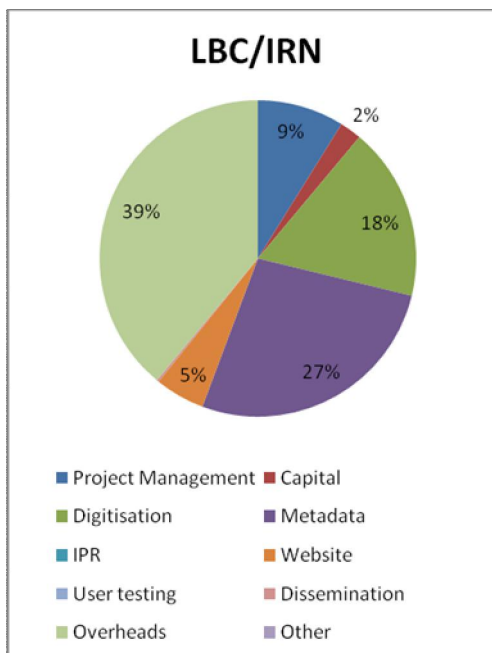


A.2.2 The digitisation was a collaboration between the Bodleian Library (University of Oxford) and ProQuest, a private sector company that specialises in the creation and delivery of electronic resources for the education sector.

A.2.3 For further information see the website: <<http://www.bodley.ox.ac.uk/johnson/>>.

A.3 LBC/IRN Archive

A.3.1 The aim of Bournemouth University was to digitise the LBC/IRN Archive and create a complete online searchable database. The records of early commercial radio were held on 10-inch reel tapes in a bespoke storage facility, along with catalogue information that provided supplementary detail referencing the audio. The archive covered the period 1973 – 1996 and contained recordings relating to news, current affairs features and dramas. The objectives were to select relevant material, create a catalogue based on existing information, digitise the tapes, then to place the audio and catalogue on a website available to Higher Education (HE) and Further Education (FE), <<http://radio.bufvc.ac.uk/lbc/>>.



A.3.2 The digitisation work was contracted to a supplier through the EU tender process. The information on the card index and legacy computer was converted into a single file. The digitisation supplier provided each tape and its associated information, along with specialised software that allowed each audio clip to be segmented. A team of cataloguers was employed by Bournemouth University to enrich the information for each segment. The digitisation

supplier provided the information in a recognised structure, along with its enriched information and the associated audio clip.

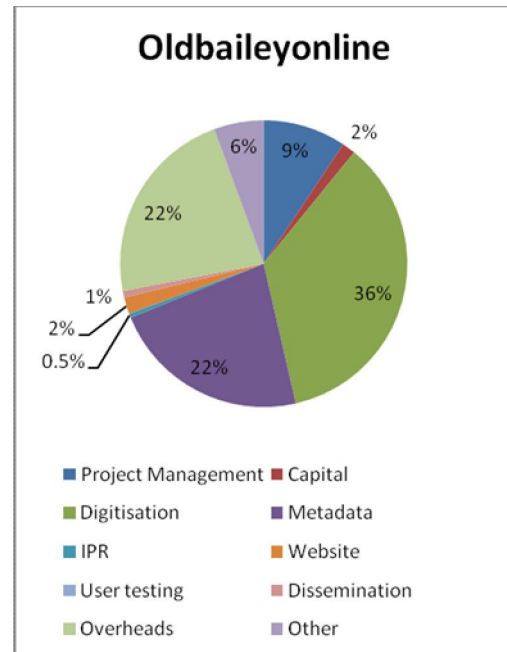
A.4 Oldbaileyonline

A.4.1 The *Old Bailey Proceedings* are the reports of almost 200,000 trials for serious crimes which were held at the Old Bailey, London's central Criminal Court between 1674 and 1913. Initially they were published as a popular magazine, widely read by Londoners of all social classes keen to know the details of the latest grisly crimes. Over time, the *Proceedings* became more serious and more comprehensive and the audience narrowed.

A.4.2 The trial records are an exceptionally compelling resource. They provide a rich and detailed record of the lives of ordinary Londoners. As well as being of interest to academic historians the material is also interesting and attractive to a much wider audience: family and local historians, schools and the general public.

A.4.3 The project (a collaboration between the University of Hertfordshire, University of Sheffield and the Open University) received money for the digitisation from the Big Lottery Fund New Opportunities Fund (under the *Digitisation of learning materials* programme) and the Arts and Humanities Research Council (AHRC). The digitised resource provides access to the text and scanned images of the trial records, marked up and tagged to provide powerful searching tools. In addition the website provides historical background materials and resources for schools.

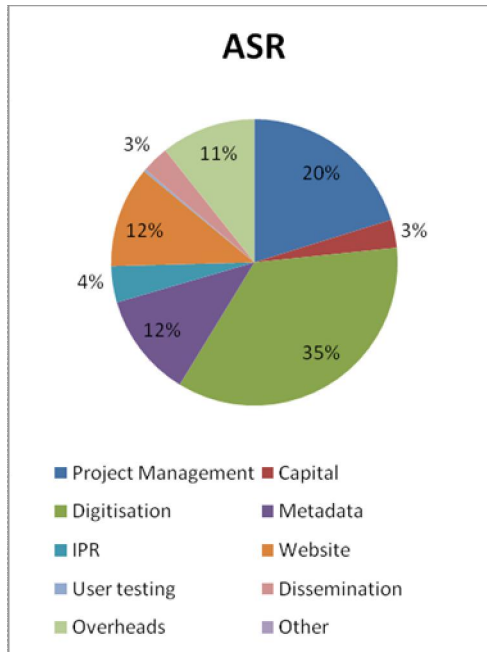
A.4.4 Following receipt of further funding, the website was redesigned, updated and extended in April 2008 to include the whole of the *Proceedings* and texts of *Ordinary's Accounts* ("richly detailed narratives of the lives and deaths of convicts") which have been linked to the relevant trials. For further information on the oldbaileyonline, see the website: <<http://www.oldbaileyonline.org>>.



A.5 Archival Sound Recordings

A.5.1 The British Library's Archival Sound Recordings website, conceived as a way of increasing access to the Sound Archive's extensive collection, went live in autumn 2006. A project to develop the website was funded by JISC, with the UK's Higher Education (HE) and Further Education (FE) communities as the primary audience. Academic institutions have access to a wide range of downloadable content and a small proportion of content is available to other users (including the general public) as streamed files which cannot be downloaded. For more information on the British Library Archival Sound Recordings, see the website: <<http://sounds.bl.uk>>.

A.5.2 The website was developed on the basis that its content would be used by academics, teachers, students and researchers. In the past audio has arguably been undervalued by the academic community in comparison with other primary sources. The appointment of an Engagement Officer signals a desire to address this under-utilisation of audio.

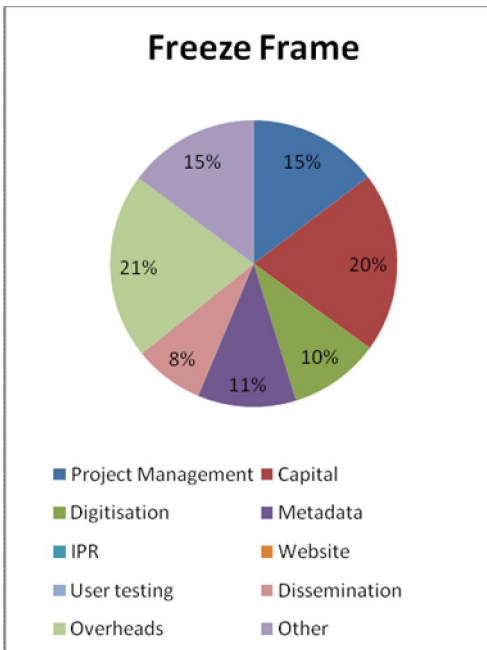


A.6 Freeze Frame

A.6.1 The Freeze Frame project set out to conserve many of the historical photographic negative collections held in the Scott Polar Research Institute (SPRI) readily available to researchers and others without the need to travel to Cambridge in person. Due to the fragile nature of much of the SPRI photographic collections, access was severely limited. The Institute's oldest photographs are daguerreotypes, a significant number are on glass plates, while other more modern negatives are, by their very nature, difficult to view. Research access to these collections has hitherto been negligible due to their format.

A.6.2 The digitisation process has ensured that each image is preserved in its present condition for future generations to view. High-resolution tiff files are stored as digital preservation copies while smaller jpegs may be provided for research access.

A.6.3 A dedicated team was recruited to carry out the digitisation, metadata creation and to produce the educational resources to stand alongside the images. While the Franklin, Fiennes and Ponting images may be some of the most evocative, particular highlights have been the



expeditions of the 1930s which mapped both the Arctic and Antarctic and whose photographs document both life in the Polar Regions and the development of science and technology in these hostile environments.

A.6.4 For more information, see the project's website <<http://www.freezeframe.ac.uk>>