

JISC Project Report: Digitisation Programme: Preservation Study April 2009

***The final public report completed by the Digital Preservation Coalition,
the Digital Archives Department of the University of London Computer
Centre (ULCC) and Portico.***



Contents Page

Foreword and acknowledgements.....	3
1. Executive Summary	4
2. Introduction	7
2.1 Context.....	7
2.2 Background	7
3. Methodology	9
4. Common findings	11
4.1 Recommendations for institutions	11
4.2 Recommendations for funders	14
5. Recommended approach to analyse preservation.....	16
6. Findings from individual projects	32
6.1 Introduction.....	32
6.2. The Framework of the Project Reviews	34
6.3 Individual Projects.....	35
Archival Sound Recordings 2 (ASR2).....	35
Cabinet Papers, 1915-1978.....	37
A Digital Library of Core E-Resources on Ireland	39
Freeze Frame.....	41
InView: Moving Images in the Public Sphere.....	44
The John Johnson Collection: An Archive of Printed Ephemera	46
Pre-Raphaelite Resource Site	49
Welsh Journals Online.....	52
UK Theses Digitisation Project	54
Glossary	57
Appendix A. Interview Questions	61

Foreword and acknowledgements

This report looks at the measures for digital preservation undertaken by 16 of JISC's funded digitisation projects. To help the reader pressed for time, we describe here the structure of the report. It is recommended that **all** readers read both sections 1 and 2.

The executive summary in section 1 draws together the most significant of the report's findings, including individual project risks, general areas of concern and common strengths, together with recommendations for future action.

Section 2 sets the context for the report.

Section 3 describes the methodology used in the project.

Section 4 discusses common findings, positive and negative, which can be synthesised from the analysis of individual projects and makes general recommendations for action by JISC and institutions.

Section 5 recommends an approach which might be taken by future projects to analyse preservation requirements together with several (fictional) worked examples.

Section 6 looks at the findings from a small group of projects that have kindly agreed to share their experience with the wider community.

The appendices contain background information such as the questions used to guide the project interviews.

Throughout the body of the report, projects are generally referred to by capital letters A to P. Appendix B indicates which project is allocated which letter (this key is omitted from the public version of the report.)

Thanks are due to all those who gave generously of their time to help in the production of this report, most of all to the project staff on the digitisation projects covered in the report. They were open and cooperative and did their best to supply us with the information we requested following our visits to them. Thanks also to the staff at the JISC executive, Alastair Dunning and Neil Grindley, for their guidance at critical points during the execution of this study and their support in helping us contact project staff. Our thanks also to the members of our own organizations who supported the authors of this report in information collection, critical comment and analysis and administrative support.

We trust that this report will help JISC, the DPC and the projects themselves spread examples of good practice and improve on current practice where weaknesses have been identified.

The Project Team:

At the DPC	At ULCC	At Portico
Frances Boyle	Silvia Arango Docio	Eileen Fenton
Carol Jackson	Kevin Ashley	Amy Kirchoff
William Kilbride	Kate Bradford	
	Patricia Sleeman	
	Anya Turner	

1. Executive Summary

Over the last decade there has been a rapid growth in funding for digitisation – indeed the term ‘mass digitisation’ has now entered common parlance. The JISC has funded two phases of digitisation activity and the growth in e-content has impacted on many areas of research, learning and teaching in the UK. This investment needs to be protected, a requirement which JISC recognized in the original call for proposals and in the project planning protocols to which projects were obliged to adhere: the former included a request for projects to outline their preservation plans as part of JISC’s assessment of proposals; the latter obliged project managers to draft an ‘Exit and Sustainability Plan’ for each funded project. JISC has followed through on this commitment by commissioning this study.

Scope of the study

This study has concentrated on the measures taken to protect the investment in the digitisation itself, by looking at measures taken to protect digital masters and the metadata associated with them. JISC’s funding has also been invested in the creation of access mechanisms. We were not asked to examine the preservation of access mechanisms but some observations on them are inevitable and are presented in section 4. Preservation of access mechanisms also features within one related recommendation. In addition, we did not examine the wider sustainability of project outputs, as this was out of scope. It is apparent, however, that those projects whose institutions have strong sustainability plans also tend to protect their digital masters, since long-term sustainability will often require returning to those digital masters to produce new derived content. It is the institutions which are critical here – project staff can help guide the institution, but cannot themselves deliver the commitments necessary.

The JISC digitisation projects

The study has found common examples of good practice at project-specific level and more generally. There is much exemplary practice, some of which is highlighted in the sample project studies in section 6 and in the accompanying case notes. The best projects have made effective use of existing institutional practice and infrastructure and have planned for preservation, along with access, from the outset.

Even so we note that some projects within the programme run at risk such that if prompt action is not taken, the content or associated services may not be accessible to their designated community, and the value of the investment therefore lost. Some are struggling with preservation issues which are not well understood in the digital preservation community in general. Almost all demonstrated a willingness to improve – notably project H took prompt action when the initial version of this report indicated high risks for its material, and has succeeded in mitigating those risks to a significant extent, without great additional expenditure.

The high-risk projects are N, I, J, K and G. The risks vary in severity. With project N, there are questions about access to material, and its protection, after 5 years. With others, there are risks associated with information being held in multiple systems or institutions without clear responsibility and workflows for coordinating updates. We have specifically identified these projects as at significant risk.

Digitisation and institutional risk management

Where shortcomings exist, they are often best addressed at institutional level than at project level, but their resolution does not generally require a great additional expense.

Perhaps surprisingly, the study has found that large and respected institutions do not necessarily achieve the best results (see for example projects J and N), although from the

examples presented here it seems they are more likely to do so (and have less excuse for not doing so, since they generally have generic preservation infrastructure available.) Multi-partner projects run the risk of not being clear about preservation responsibility, and hence having no one adopt it without even realising that this has happened.

Four principal themes surfaced through analysis of the preservation plans of the digitisation projects that relate the maturity of institution to the likely success of their digitisation efforts. These are the need for preservation policies; collection management procedures; robust preservation infrastructures; and sustainability. In short, institutions or consortia which have clarity in these four areas considerably reduce the risks associated with long term access to digitized collections. In contrast, institutions or consortia which are uncertain or unclear on these four issues amplify their risk profile needlessly.

Funding for digitisation: lessons learned

There was not a shared understanding of preservation requirements amongst projects and there was not a shared acceptance by institutions that they had a responsibility (and an interest) to protect content beyond the 5-year funding timeframe. Clarity on these points is critical.

For example, JISC funding for digitisation came with the condition that institutions are required to keep materials freely available for five years after the funding ends. Institutions do not all recognize that this is a minimum period for providing access. Instead they tend to interpret this as a maximum period for protecting content. This period is very close to the typical life of any specific IT infrastructure, and therefore the timeframe after which some preservation action (such as migration to new storage media) would need to take place. At least one project (project N) does not have any clear plan for the content or any service associated with it beyond this 5-year period. Therefore a conspicuous risk is introduced as an artefact within the granting process. JISC could do more to protect its rights in the event that an institution does not wish to continue responsibility for service or content after the 5-year period is up. This will be of interest to funders in other sectors and is highlighted within the recommendations. The issue is examined in more detail in section 4.

Lack of clarity about what constitutes preservation is also a concern: terms like sustainability, legacy or curation offer subtle nuances of focus within the preservation community but can hinder comprehension by those outside. The study proposes a straightforward definition of digital preservation which we recommend as a useful starting point for projects and funders alike. The definition is:

Digital preservation is the series of management policies and activities necessary to ensure the enduring usability, authenticity, discoverability and accessibility of content over the very long-term. The key goals of digital preservation include:

- usability – the intellectual content of the item must remain usable via the delivery mechanism of current technology
- authenticity – the provenance of the content must be proven and the content an authentic replica of the original
- discoverability – the content must have logical bibliographic metadata so that the content can be found by end users through time
- accessibility – the content must be available for use to the appropriate community

Core Findings

The most significant findings were:

- External examination (through a study such as this, or an audit process) can change practice for the better merely by asking the right questions
- Without a written preservation policy, the long-term usability, authenticity, discoverability and accessibility of the digital collection is at risk.
- Without defined collection and content management procedures, particularly where metadata is dissociated from content or is held in multiple locations, the long-term usability, authenticity and discoverability of the digital collection is at risk.
- Without maintaining digital collections on a suitable digital preservation infrastructure, the long-term usability and accessibility of the digital collection is at risk.
- Without a plan for sustainability, the long-term usability and accessibility of the digital collection is at risk.

In addition, the study team felt that the simple approach outlined in section 5 of this report might prove useful for future projects or be usefully deployed by funders.

Recommendations

Our recommendations fall into two groups. They are stated here, and explained in more detail in section 4.

For institutions and projects:

- 1. Write and Implement a Preservation Policy for Each Digital Collection***
- 2. Define Collection and Content Management Procedures***
- 3. Preserve Content in a Suitable Digital Preservation Infrastructure***
- 4. Identify How the Collection will be Sustained Over Time***
- 5. Recognize that accepting funding for digitisation implies an institutional commitment which outlives the period of funding.***
- 6. Be clear about retention periods for different types of created digital content***

For funders:

- 7. Consider designating or establishing preservation services for content and use appropriate methods to confirm that conditions are being met***
- 8. Articulate more clearly and enforce requirements for the retention of content***
- 9. Make the long-term duty on institutions clearer***

2. Introduction

2.1 Context

This study aimed to understand the digital preservation work being undertaken by projects whose primary focus was not digital preservation. Its observations need to be understood in that light. These projects had a primary aim of using digitisation to deliver wider access to a range of materials for teaching and research. Selection of materials and the end user experience were the highest priority of each of the projects, and it was not our aim to evaluate those aspects of the project's work. (This was, however, covered by the formative evaluation by Glennafric)¹ The interviews and information gathering which supported this study were carried out between December 2008 and March 2009, with one exception which took place in May 2009. For some of the projects this was immediately before or after the launch of public access to their work, and the bulk of their attention and efforts was (rightly) concentrated on that immediate public service.

It is important to understand something of the contexts in which the projects operated. The risk profiles of projects included such weighty matters as user-needs analysis, rights management, technical infrastructure development, staff retention and development, metadata creation and physical risks to fragile originals. Digital preservation risks were only one of a number of risks that projects had to contend with, and are only coherent when perceived in this context.

Many of the risks faced by projects could not be controlled by them directly. This is particularly true of those projects which had included elements of service provision from the Arts and Humanities Data Service (AHDS). AHDS was engaged by several projects as variously a preservation and advisory service. The removal of central funding in April 2008 made it impossible for AHDS to provide these services to projects, and caused several of them to change their plans radically mid project.

2.2 Background

This study is a response to a call from JISC issued in October 2008 as part of the Digitisation Programme. JISC sought to gain a better understanding of the preservation element of the projects within the Digitisation Programme. Its requirements were an internal report, which would identify strengths and weaknesses in individual projects as well as laying the foundation for the second, public report. The public report would be used to identify best practice and common risks. It would be used as a tool by JISC and other organizations to support the preservation of outputs in current and future digitisation projects and identify actions that needed to be taken at the community level to do this. This document is the public report.

The contract was awarded to a bid lead by the Digital Preservation Coalition (DPC) in partnership with two of its members, the Digital Archives Department of the University of London Computer Centre (ULCC) and Portico.

The study aimed to deliver a comprehensive report, a portfolio of case studies and a suite of recommendations which will bring clarity to and inform the JISC about the preservation elements and methodologies adopted in the Digitisation Programme.

The aims and objectives of the project were to:

¹ Formative evaluation of JISC Digitisation Programme, Phase 2 – Glennafric, October 2009. Available from: <http://www.jisc.ac.uk/whatwedo/programmes/digitisation/reports/evalphase2.aspx>

- provide recommendations for mitigating preservation risk, at both funding agency and institutional level
- produce, if possible, a quick reference guide of practical preservation considerations
- provide a portfolio of case studies representative of different models and cross-sectoral collaborations
- provide a review and evaluation of the preservation plans from the JISC funded digitisation projects identifying common trends, assumptions, constraints, risks and gaps in understanding

The DPC intend to promulgate the findings of the public report and aid JISC in the dissemination process, since the outputs are likely to be of interest to a number of DPC members, many of whom are outside JISC's core HE and research community.

The 16 projects we surveyed had responded to JISC's call for digitisation projects in 2006. A small number has experience of previous JISC digitisation funding (or that from other funding bodies) but many were new to work on this scale. For all the projects, access mechanisms – often innovative – were key drivers. Some also developed value-added resources, such as teaching tools, around the digitized content and access mechanisms. For some, digitisation was a way of helping to preserve fragile originals, but preservation of the digital content was not uppermost in the plans of most of them. (Notable exceptions include national institutions already dealing with large volumes of digitized material in mature preservation infrastructures.)

3. Methodology

The study built on the momentum and investment in the JISC Digitisation Programme and complemented other related JISC projects in the digital preservation area. The methodology adopted was a mix of:

- direct engagement with current practitioners through a series of semi structured interviews
- desk research looking at the digitisation and related projects
- critical analysis of the data and information collated from the engagement exercise.

There were seven work packages carried out across four phases: data collection, analysis, collation of information, and production of final outputs. ULCC conducted and oversaw the first phase of data collection. Whilst Portico lead on the second, analysis phase of the study and acted as lead partners on three of the work packages. Their primary role was to study the preservation plans and data from the digitisation projects, to review the sample data, to describe areas of risk and to make recommendations at the project specific and strategic level. The role of the DPC was to co-ordinate and manage the project work.

In the first phase, sixteen site visits were conducted by researchers from ULCC. JISC identified initial contacts at each project. Contact was made with each project and the purpose of the study explained; this allowed the contacts to nominate additional or alternative contacts within the organization if that was more appropriate for the study. Once this was done, interviews were arranged with each partner. We tried to ensure, as far as possible, that each interview involved two staff from the study team, one focussing on the conduct of the interview and one on capture of responses and findings. Structured questions were devised to ensure that the salient information was captured and there was an opportunity for project staff to engage in as wide a discussion as required; the questions were provided in advance to the project contacts, to allow them time for reflection, information gathering, or to bring in other project staff for part or all of the interview. The semi structured interviews covered areas such as relationships, sustainability, technical formats utilised and funding. The full set of questions may be found in Appendix A. As the projects differed in scale, in types of material, and in the nature of the delivery partners, one of the primary aims was to understand whether any of these differences affected the nature, quality and risk associated with preservation plans and practices adopted by the projects.

Following the interviews, each project was also asked to supply sample metadata or data structures relating to the preservation aspects of the digitized objects; the broad nature of what was required was specified by the analysis partner, but were left non-specific in recognition of the very wide range of technical infrastructures and systems (as well as the wide range of material) in place at each project.

The evaluation and analysis phase looked at the collated data from a range of perspectives. The full reports of the visits were captured on a template and shared with all partners. The team at Portico looked at common issues shared by some or all projects, and those which were particular to a given project. As well as the narrative evaluation a risk analysis and a specific examination of digital objects and related metadata in a preservation context was undertaken. The results of the analysis produced two sets of outputs: project specific findings, which are presented in section 6, and generic findings, which appear in section 4. These then led to the recommended future actions which form section 5 of this report. To ensure a common

understanding a definition of digital preservation is included; it is felt that this was a necessary baseline for the evaluation of risks.

The original plan was then to produce a public report with generic findings and four case notes, chosen both for their variety and for the positive lessons which others can draw from their work. In fact, we found much to celebrate in the project studies and agreed to present as much of it as possible in the public report. Thus, a selection of the detailed project reports appears in section 6 of the public report. The 4 case notes were also produced.

Those selected for case study treatment are the Welsh Journals Online project, the Cabinet Papers from the National Archives project, the British Library's Archival Sound Recordings 2 and the Freeze Frame project, as each approached the preservation challenges in a different way. The Case Notes are available online from the DPC website at <http://www.dpconline.org/>

4. Common findings

Our recommendations include some for projects and institutions, and some for JISC as funders. A number of strong themes surfaced through analysis of the preservation plans of sixteen JISC digitisation projects, including the need for preservation policies, collection management procedures, robust preservation infrastructures, and sustainability. Some of the institutions participating in the survey are succeeding in these four areas, while other institutions need to improve how they address these risks. We address these themes below by offering strategic recommendations to any institution intending to preserve its digital collections for the long-term.

4.1 Recommendations for institutions

1. *Write and Implement a Preservation Policy for Each Digital Collection*

Without a written preservation policy, the long-term usability, authenticity, discoverability and accessibility of the digital collection is at risk.

In this study, only 2 projects could point to an (institutional) policy. Some had strategies out of which policies would later emerge; 7 overall had policies ‘in development.’ The others had no policy and no expectation of one.

Each digital collection should have a policy to identify procedures and stakeholders. The policy should be approved by the stakeholders or their representatives. This policy should address:

- what content is being preserved and for whom
- the objective of the preservation
- who is responsible for preservation of the files and related metadata
- sources of financial support for the ongoing preservation
- how the success of preservation will be measured

Once agreed, the policy should be implemented.

2. *Define Collection and Content Management Procedures*

Without defined collection and content management procedures, the long-term usability, authenticity and discoverability of the digital collection is at risk.

Projects I, J and L (amongst others) have potential risks which arise either from metadata stored in multiple systems with multiple partners, which calls into questions where the authentic copy is, or potential difficulties with attaching metadata to master copies of content.

Collection and content management is an area of great risk for long-term access and it is too often ill-defined. This is an equal opportunity risk and is not specific to either small or

large institutions or projects. For example, while the decision made by a large project at a large institution not to keep the archival content and the delivery content synchronized may be appropriate, it is not documented. Alternately, a small project at a small institution had given significant thought to ongoing collection management and identified staff to provide that ongoing management.

Content and collection management procedures in the following areas should be documented:

Initial Collection Creation: The procedures to create the digital collection should be well documented. While an automated workflow is preferred, a manual content management process is acceptable, if it is well documented.

Ongoing Collection Maintenance: Collections require ongoing preservation maintenance (for example, regular fixity checks that check for content corruption and completeness checks that check for content loss). In addition, collections require ongoing maintenance of their intellectual content (for example, correcting inaccurate descriptive metadata). The person who is responsible for preservation maintenance of the content is unlikely to have the domain knowledge necessary to maintain the intellectual content of the collection. The procedures and staff responsible for ongoing preservation maintenance and intellectual content maintenance must be documented.

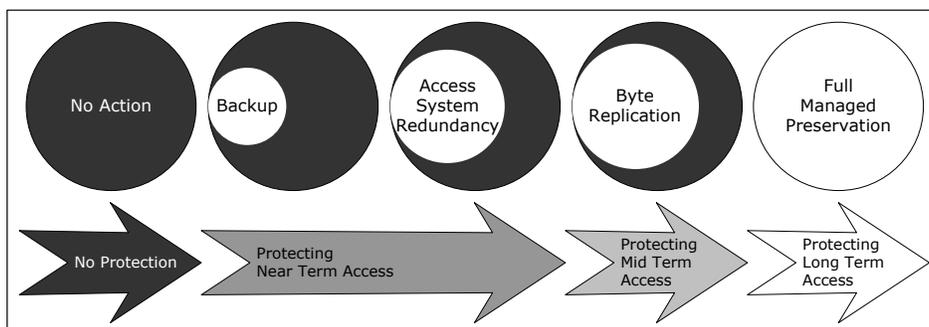
Moving the Collection: The content and its metadata will need to be shipped from one location to another at some point – perhaps migrated between platforms or replicated to a third party preservation service. The processes and procedures for how the master copy of the metadata and master copy of the content files will be packaged together as a unit and moved must be defined. If the party responsible for the collection believes this is not necessary, then that decision and its rationale must be documented.

3. Preserve Content in a Suitable Digital Preservation Infrastructure

Without maintaining digital collections on a suitable digital preservation infrastructure, the long-term usability and accessibility of the digital collection is at risk.

Projects I, G and N (amongst others) all face potential problems in this respect.

Long-term digital preservation cannot be solely accomplished through backup, access system redundancy, or byte-replication. While each of these can offer some short and mid-term protection to content and are often elements of a long-term preservation plan, they are not sufficient.



Many institutions cannot support long-term, actively managed, digital preservation on their own, as it requires ongoing monitoring of the preservation, technological, and designated user community; extensive content management skills; and dedicated staff.

Wherever possible, institutions should consider collaborating to build a digital preservation infrastructure within a parent organization or between organizations or using a third party digital preservation service. The community must continue to determine appropriate ways to identify trustworthy third party digital preservation services. While self-assessments such as Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) or Assessing Institutional Digital Assets (AIDA) and external assessments such as the Center for Research Libraries' audit of repositories and archives using the Trustworthy Repositories Audit and Certification – Criteria and Checklist (TRAC) provide ways to measure the development and trustworthiness of preservation services, even services that appear trustworthy and secure can lose funding (for example, the Arts and Humanities Data Service). This reality requires the community to continue to address how content will be moved from one preservation service to another over time and how to determine the security of different preservation services.

4. Identify How the Collection will be Sustained Over Time

Without a plan for sustainability, the long-term usability and accessibility of the digital collection is at risk.

Sustaining public access is not an essential pre-requisite for sustaining the master content, but it is a very effective driver for doing so.

As shown in recommendations one through three, digital collections require ongoing maintenance if they are to remain preserved for the designated community for the long-term. The length of time for which a collection should be preserved must be defined at the beginning of the project, so that the necessary resources can be budgeted and funding identified and documented as an element in the preservation plan. This recommendation can be met in many ways, including identifying an ongoing income stream for the digital collection, documenting commitment from the parent organization that the collection will be financially supported over time, designating a successor organization that will take responsibility for the collection should the parent organization fail, etc. Most importantly, organizations need to acknowledge that for digital collections to be preserved, ongoing maintenance is required and a financial commitment to support this maintenance should be explicitly secured.

5. Recognize that accepting funding for digitisation implies an institutional commitment which outlives the period of funding

Short term funding for digitisation implies a long term commitment to provide access, manage and derive value from the digitized content.

Funding agencies provide project funding for digitisation (and often for associated metadata creation or enhancement) to help institutions expose content to a wider audience and to develop innovative ways of accessing and using the content. Project funding is time-limited, but there is an expectation that the significant investment made –

much of which will have come from the institution itself – will continue to deliver value after the period of project funding, or after the minimum time period during which access must be provided. Not all institutions are recognising the obligation on them to secure content and metadata in perpetuity or at least for periods much longer than 5 years.

6. *Be clear about retention periods for different types of created digital content*

Not all of the digital content created with project funding warrants long-term retention.

Some publicity or teaching and learning materials may have much shorter useful lifetimes. Some projects argued strongly that this meant they did not need to take account of them in their digital preservation policies. We agree that some materials do not need to be kept forever, but strongly recommend that they are still brought within the scope of a policy, which can then be explicit about the short retention periods to be applied to such material. Such decisions should not be taken by default, and funders and institutions should be seen to agree on them.

4.2 Recommendations for funders

These recommendations have been written primarily for JISC, but some could be seen to apply more widely to any funders of digitisation activities.

7. *Consider designating or establishing preservation services for content and use appropriate methods to confirm that conditions are being met*

If funders reserve the right to take back content then the funder will also need viable mechanisms to check that conditions are met and impose the sanction as appropriate. Funders may withhold funding until such times as projects are able to identify their long term plans, may wish to designate or establish their own preservation services, and may need an independent agent to validate that conditions are being met. These services could be delivered by 3rd parties.

Some projects (such as those in national libraries and archives) are able to make effective use of in-house preservation services whose responsibilities are clearly divided from those of access providers. Others have worked effectively with external providers of preservation services such as JSTOR and CeRch. These providers cannot necessarily accommodate all content types or volumes and are not obliged to provide services to any project which requests them. Content might be more effectively secured if funders could establish relationships with existing providers, or establish new providers, that would allow them to recommend their use to projects in receipt of funding, unless the projects could demonstrate that they had an acceptable alternative.

JISC already has a condition of grant that material must be offered back to it if a project is no longer able to protect it. In reality, JISC does not itself want to have to deal with the content directly in this eventuality nor does it necessarily want to audit the process directly – it is likely that it would wish to find another institution or service to accept the content and provide access to it or to report on compliance with this condition of grant. Also, experience shows that attempting to reuse content at this late stage – when an institution no longer wants it, and when expertise has probably been lost – presents great practical difficulties. Consequently, access to escrow services and other independent audit services would be advantageous to funders.

8. Articulate more clearly and enforce requirements for the retention of content

Funders and projects should be explicit about the expected life span of different types of content.

This recommendation relates to (6) above. Funders might wish to define periods when inviting proposals, or may simply ask each project to propose such periods for agreement by funders. At present, it is unclear whether teaching and learning materials (for instance) are expected to be afforded the same long-term protection as digital masters.

9. Make the long-term duty on institutions clearer

Funders should be explicit about their expectations for long term access and embed these more effectively within terms of the grant.

This recommendation relates to 5 and 8 above. Some projects (such as K, N and I) report that their institutions do not recognize any responsibility to keep content beyond the end of project funding. One said explicitly ‘we have not been funded to preserve this content in perpetuity.’ We believe that funding is meant to help institutions provide access to material which they should want to protect in the long term in any event. JISC already expects significant institutional contributions to project funding to demonstrate institutional commitment, but it is clear that this is not enough. In addition to committing to access over 5 years following funding, institutions should also be required to demonstrate a longer-term commitment to protecting the investment in content creation.

5. Recommended approach to analyse preservation

This section considers practical recommendations for meeting digital preservation needs. As a useful starting point the following definition is offered up as a useful starting point for the digitisation community.

Digital preservation is the series of management policies and activities necessary to ensure the enduring usability, authenticity, discoverability and accessibility of content over the very long-term. The key goals of digital preservation include:

- usability – the intellectual content of the item must remain usable via the delivery mechanism of current technology
- authenticity – the provenance of the content must be proven and the content an authentic replica of the original
- discoverability – the content must have logical bibliographic metadata so that the content can be found by end users through time
- accessibility – the content must be available for use to the appropriate community

“Digital preservation” can seem amorphous. The time horizon is long and the steps that will need to be taken over this time period are, by virtue of the length of time under consideration, unclear.

The way to address this long time frame and lack of clarity is through active and ongoing management of the collections. Before attempting to draft a formal preservation policy for any given digital collection, we encourage all projects to answer questions designed to elicit the current and planned content management processes for the collections. We have found in our survey that many institutions have not given full consideration to questions of how long the content should remain available and to whom or what process will be used to update content once the collection has been built. Answering a small set of practical questions will clarify the nature of the long-term digital preservation needs of institutions’ collections. Some institutions and projects may conclude their digital preservation planning after answering the questions. Other institutions may determine through this initial analysis that they need to develop more robust preservation policies and plans.

As projects and institutions mature, they may need more robust preservation planning. Below are some resources to aid you:

- Planning Tool for Trusted Electronic Repositories (PLATTER – <http://www.digitalpreservationeurope.eu/>) from DigitalPreservationEurope provides a series of guidelines for developing SMART goals for repositories (Specific, Measurable, Assignable, Realistic, and Time-related goals) through specific policies.
- The Digital Preservation Training Programme (DPTP - <http://www.dptp.org/>) is another practical way to develop the skills and knowledge necessary to devise an appropriate digital preservation plan.
- Assessing Institutional Digital Assets (AIDA - <http://aida.jiscinvolve.org/>) is a self-assessment tool for describing institutional readiness and capabilities for digital preservation and can be used by an institution to place itself on a preparedness scale – which can then inform the institution about which areas of digital preservation it needs to further develop.

- The Council on Library and Information Resources (CLIR) report on “Preservation in the Age of Large-Scale Digitisation” (<http://www.clir.org/pubs/abstract/pub141abst.html>) discusses a manageable framework for thinking about how preservation needs should permeate digitisation projects from the beginning.
- When institutions are ready to begin the task of drafting a digital preservation policy, they may refer to the “Digital Preservation Policies Study” produced by Charles Beagrie in October 2008 (<http://www.jisc.ac.uk/publications/publications/jiscpolicyfinalreport.aspx>) for guidance.
- There are a number of repository checklists and risk assessment tools that can help institutions identify weakness in their long-term collection management or preservation plans:
 - *Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist* (<http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>) from the Center for Research Libraries and OCLC provides a checklist of criteria for establishing a trusted repository.
 - *Catalogue of Criteria for Trusted Digital Repositories* (http://www.langzeitarchivierung.de/downloads/mat/nesstor_mat_08-eng.pdf) is another such checklist from nestor (Network of Expertise in Long-Term Storage of Digital Resources).
 - The *Digital Repository Audit Method Based on Risk Assessment* or DRAMBORA is a guided risk self-assessment (<http://www.repositoryaudit.eu>).
- Finally, for projects ready to delve deep into digital preservation, the *Digital Preservation Handbook* (<http://www.dpconline.org/graphics/handbook/>) provides in depth guidance.

As an initial step, however, we recommend that institutions answer the following twenty questions to identify the locations of all their content, for whom it was created and why, the length of time it must be available, and how it will be maintained. Following the questions are some exemplars.

Practical Questions to Answer – the Basis of Digital Preservation Policies and a Plan

Who: Identify the key players involved with long-term preservation of the targeted content.

1. Who is writing the policy and plan?
2. Who will use the content in the short and long-term?
3. Who has responsibility for maintaining the intellectual content of this collection (e.g. making corrections to metadata or content files)?
4. Who has responsibility for maintaining the bytes of the files in this collection (e.g. identifying and fixing corrupted files)?
5. Who approved this policy and plan?

What: Describe or characterize the collection and content.

6. What is the content and from where did the content originate?
7. What file formats, including metadata formats, are present?
8. How many items are in the collection? How large is the collection on disk?

Where: Document the locations of all the copies of the content and metadata.

9. Where is the master copy of the descriptive metadata kept?
10. Where is the master copy of the content files kept?
11. Where are all the copies of the content, including backups, and how are the copies of the content related?

When: Document the targeted preservation timeframe and impact of loss.

12. How long should the content be available for use?
13. If the content is irretrievably lost, what are the repercussions?

How: Document how the key content management and preservation tasks will occur.

14. How will the collection be created (perhaps draw a diagram of the workflow)?
15. How will the collection be maintained (perhaps draw a diagram of the workflow)?
16. Do you expect the content files to be migrated in the future?
17. May the content files be deleted? Added to? Updated?
18. May the descriptive metadata be deleted? Added to? Updated?

19. How will you track who did what and when to the content, if this is important to your organization?
20. How do you associate the master copy of the descriptive metadata with the master copy of the content files and how will you move these two items around together?

Illustrations of Answers to the Practical Questions

Project 1

A University department has a special collection of primary source documents that has grown over time. The department hires a project manager for the duration of the creation of the digital collection to coordinate the digitisation of the content and creation of descriptive metadata. With the help of the University IT department, they place the content into an institutional repository and make it available for use on-line.

Project A - Digital Preservation Policies and Plan

Who: Identify the key players involved with long-term preservation of the targeted content.

1. Who is writing the policy and plan?

The digitisation project manager in the Department of Lake Studies at the University of Lorem Ipsum.

2. Who will use the content in the short and long-term?

The content should be made available for use by anyone in the world.

3. Who has responsibility for maintaining the intellectual content of this collection (e.g. making corrections to metadata or content files)?

The University of Lorem Ipsum IT department has responsibility for the ongoing maintenance of the collection in the institutional repository. If corrections are suggested through user feedback, the IT department should contact the Department of Lake Studies administrator who will then discuss the correction with the Department Director and approve or disapprove it. The IT department will make the changes in the institutional repository.

4. Who has responsibility for maintaining the bytes of the files in this collection (e.g. identifying and fixing corrupted files)?

The University of Lorem Ipsum IT department will ensure that the content files do not become corrupted.

5. Who approved this policy and plan?

Ms. Jones, Director of the IT Department, University of Lorem Ipsum

Mr. Challa, Director of the Department of Lake Studies, University of Lorem Ipsum

.

What: Describe or characterize the collection and content

6. What is the content and from where did the content originate?

The content is digitized postcards, letters and other ephemera. A large portion of it was donated to the Department in 1965 by Mr. Smith. The faculty of the Department of Lake Studies has added to the collection since that time.

7. What file formats, including metadata formats, are present?

The content has been digitized as TIFF images (300 dpi, 48 bit colour). The descriptive metadata is first captured in the Department of Lake Studies catalogue (which is used to describe the analogue content in the collection, as well). The images are referenced by filename in the catalogue record. The catalogue records and TIFF images are exported to the IT department and are placed in the institutional repository.

8. How many items are in the collection? How large is the collection on disk?

There are 4000 images in the collection covering 1000 postcards, 2500 letters, and other ephemera. It is approximately .5 TB.

Where: Document the locations of all the copies of the content and metadata.

9. Where is the master copy of the descriptive metadata kept?

The master copy of the descriptive metadata is kept in the Department of Lake Studies catalogue.

10. Where is the master copy of the content files kept?

The master copy of the content files is kept in the institutional repository maintained by the IT department.

11. Where are all the copies of the content, including backups, and how are the copies of the content related?

The institutional repository also has a copy of the metadata; however it is a derivative and not as robust as what is held in the Department of Lake Studies catalogue. The catalogue has monthly full backups and weekly incremental backups that are housed in the IT department's machine room. The institutional repository also has monthly full backups and weekly incremental backups. In addition, it has monthly backups to tape which are sent off-site.

When: Document the targeted preservation timeframe and impact of loss.

12. How long should the content be available for use?

The content should remain available for use for at least 50 years.

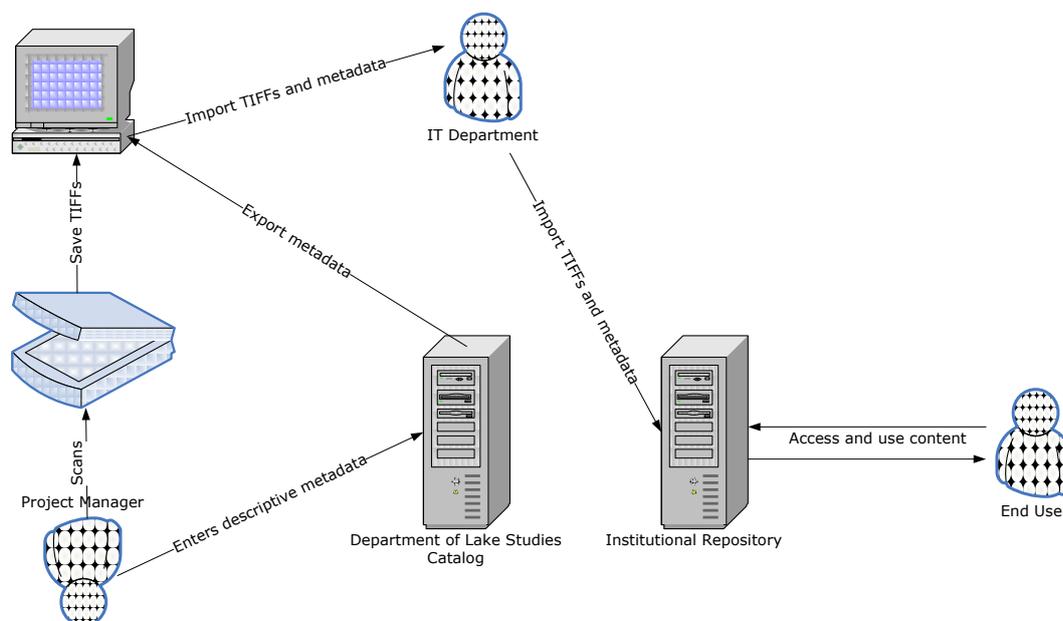
13. If the content is irretrievably lost, what are the repercussions?

There are none.

How: Document how the key content management and preservation tasks will occur.

14. How will the collection be created (perhaps draw a diagram of the workflow)?

The Project Manager or an intern in the Department of Lake Studies will scan the artefacts and create TIFF images. These are temporarily saved to a local computer. The same person will enter descriptive metadata into the Department of Lake Studies catalogue. At the end of every month, the Project Manager exports the metadata from the catalogue to the image file server in Excel format. The IT department captures the Excel file and the appropriate TIFFs and places them in the institutional repository. The repository turns the TIFFs into JPGs on the fly when requested by an end user. The images on the file server at the Department of Lake Studies will be deleted after they are successfully loaded into the institutional repository.



15. How will the collection be maintained (perhaps draw a diagram of the workflow)?

The Department of Lake Studies will not be performing regular maintenance of this digital collection. If a metadata update needs to be made, the IT department may contact the Department of Lake Studies and the changes will be made manually at both the institutional repository and the Department of Lake Studies catalogue. If the Department of Lake Studies initiates a correction, they will contact the IT department to synchronize the updating of metadata.

16. Do you expect the content files to be migrated in the future?

If it is necessary to migrate the files within the next 50 years so that the collection remains usable, yes.

17. May the content files be deleted? Added to? Updated?

The content files may be deleted and updated – though it should be rare and only in the case of an error. The collection will be closed when the project is completed and no additional content files will be added.

18. May the descriptive metadata be deleted? Added to? Updated?

The descriptive metadata may be updated. It should not be deleted, though a note may be made that the digitized file(s) to which it refers has been deleted. As the collection will be closed when the project is completed, we do not anticipate entry of new metadata records.

19. How will you track who did what and when to the content, if this is important to your organization?

It is not important and will not be tracked.

20. How do you associate the master copy of the descriptive metadata with the master copy of the content files and how will you move these two items around together?

The metadata record in the institutional repository is not complete. The reasons for any need to ship the content must be analyzed. For certain purposes, an export from the institutional repository may be sufficient. If a master copy of the metadata must be exported with the master copy of the images, then the IT Department will need to coordinate with the Department of Lake Studies to merge the metadata in the catalogue with the images in the institutional repository. There is no automatic way to do this.

Project 2

A large library has digitized old analogue video recordings. The analogue version of the video recordings is secure in the institutional video vault. The library does not have the skills or desire to provide access to this content and has therefore shipped a copy of the content to a third party access service that specializes in delivery of digital video recordings. That third party service has agreed to provide access to the content for at least 10 years. The third party access service creates smaller delivery files from the master copy of the content provided by the library and then deletes its copy of the original. The library is maintaining a preservation copy of the original digitized recordings in its robust, institutional archive.

Project B - Digital Preservation Policies and Plan

Who: Identify the key players involved with long-term preservation of the targeted content.

1. Who is writing the policy and plan?

Mr. Joe Kline, Director of Library Video Services, University of Smithtown

2. Who will use the content in the short and long-term?

The UK HE and FE community

3. Who has responsibility for maintaining the intellectual content of this collection (e.g. making corrections to metadata or content files)?

The staff of the library video services department is responsible for updating the metadata and content files within the institutional preservation service.

4. Who has responsibility for maintaining the bytes of the files in this collection (e.g. identifying and fixing corrupted files)?

The University of Smithtown's IT department is responsible for maintaining the institutional archive and will provide required ongoing preservation maintenance to this content.

5. Who approved this policy and plan?

Ms. Adelaide Bovie, Director of the Library, University of Smithtown

Mr. Muhammad Bishara, Director of Information Technology, University of Smithtown

What: Describe or characterize the collection and content.

6. What is the content and from where did the content originate?

The University of Smithtown has long been a centre for film and video studies and over the decades, the university library has developed an extensive collection of analogue videos that are now out-of-copyright.

7. What file formats, including metadata formats, are present?

The videos masters are in WAV format. The metadata is in a proprietary framework and uses a qualified Dublin Core for the descriptive metadata. Submaster files, also in WAV format, are created from the original masters. These are each a clip from the original.

8. How many items are in the collection? How large is the collection on disk?

There are 500 master files, with just under 5000 submasters. The collection is approximately 1.5 Tb.

Where: Document the locations of all the copies of the content and metadata.

9. Where is the master copy of the descriptive metadata kept?

The master copy of the descriptive metadata is in the institutional archive.

10. Where is the master copy of the content files kept?

The master and submaster copies of the content files are in the institutional archive.

11. Where are all the copies of the content, including backups, and how are the copies of the content related?

The access provider has a copy of access derivatives of the content and the metadata. This is not tied back to the master copies at all – though if needed, it could be traced through original file name. The access provider is responsible for its own backup regime.

At the University of Smithtown, the master content files, submasters and metadata are all held within the institutional archive. This service is on RAID 5 servers with a 9.99% uptime guarantee. Disk snapshots are made to an off-site, University owned machine room nightly and weekly full backups are written to tape. The tapes are kept on-site for one month and then moved off-site for storage for 3 months.

Within the institutional archive, this content is all filed as the “Library Video Collection”.

When: Document the targeted preservation timeframe and impact of loss.

12. How long should the content be available for use?

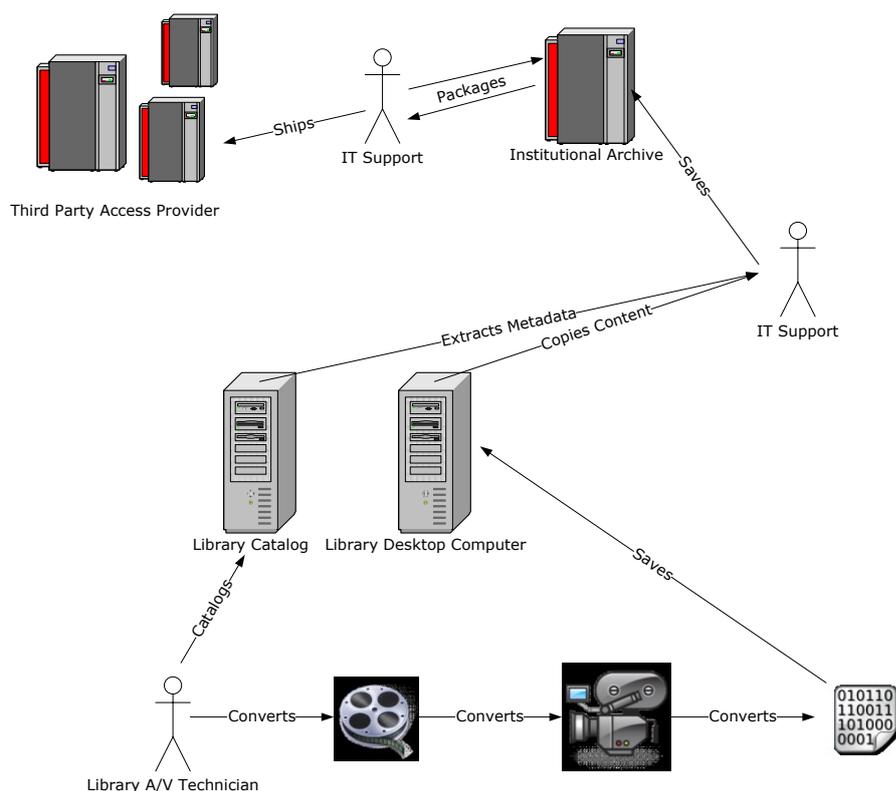
The content should be available for use for 10 years.

13. If the content is irretrievably lost, what are the repercussions?

The content would have to be redigitized from the analogue. This will be possible, as the analogs are secure, but it would be expensive.

How: Document how the key content management and preservation tasks will occur.

14. How will the collection be created (perhaps draw a diagram of the workflow)?



15. How will the collection be maintained (perhaps draw a diagram of the workflow)?

The IT Department does not yet have robust toolsets for editing the metadata or updating the content files within the institutional archive, however they will be developed over time. In the mean time, should metadata need to be updated it will be updated within the library catalogue and then the library staff and IT staff will coordinate on updating the preserved content. The same is true if content files must be updated.

16. Do you expect the content files to be migrated in the future?

Yes. The IT staff has committed to this and it is expected as part of the institutional archive.

17. May the content files be deleted? Added to? Updated?

The content files will not be deleted from the archive, they may be updated.

18. May the descriptive metadata be deleted? Added to? Updated?

The descriptive metadata may be update over time.

19. How will you track who did what and when to the content, if this is important to your organization?

The metadata structure in the institutional archive has the PREMIS concepts of events included and an event record will be made for every update.

20. How do you associate the master copy of the descriptive metadata with the master copy of the content files and how will you move these two items around together?

The institutional archive keeps the two together.

Project 3

A national library has a significant collection of books published in the mid-19th century on acidic paper. It is digitizing this collection in advance of the books disintegrating. The library has a content management system that will allow it to provide access to the content and is outsourcing the preservation of these digitized materials.

Project C - Digital Preservation Policies and Plan

Who: Identify the key players involved with long-term preservation of the targeted content.

1. Who is writing the policy and plan?

Mr. Jason Jackson, Manager of Digital Collections, the National Library

2. Who will use the content in the short and long-term?

The general public.

3. Who has responsibility for maintaining the intellectual content of this collection (e.g. making corrections to metadata or content files)?

The Digital Collections department of the National Library.

4. Who has responsibility for maintaining the bytes of the files in this collection (e.g. identifying and fixing corrupted files)?

The Third Party Preservation Service.

5. Who approved this policy and plan?

Dr. Meredith Jones, Director of the National Library

What: Describe or characterize the collection and content.

6. What is the content and from where did the content originate?

The content was digitized from the brittle and crumbling collection of 19th and 20th century books owned by the National Library. The library has developed a project plan which lays out the order in which different subjects and years will be digitized. Please contact the Manager of Digital Collections for further details.

7. What file formats, including metadata formats, are present?

The final product is one PDF file per book with its corresponding MARC record from the library catalogue.

8. How many items are in the collection? How large is the collection on disk?

The collection is currently 1000 books and is approximately 500 Gb. This project is ongoing and we estimate that the library has over 35 miles of shelves of books to digitize. The project is budgeted for the next 5 years and we anticipate digitizing 500 books a year.

Where: Document the locations of all the copies of the content and metadata.

9. Where is the master copy of the descriptive metadata kept?

The master copy of the descriptive metadata is in the libraries access system. While the data originated in the library card catalogue, that data is *not* considered the master descriptive metadata.

10. Where is the master copy of the content files kept?

The master copy of the content files is kept on the libraries "S" drive. This is also known as `shareddrive-s.nationallibrary.net`. The access system or content management system is run on Fedora. The large TIFFs and OCR files that are used to create the derivative PDFs are not within Fedora, but each book record in Fedora does point to the TIFF and OCR files in their home location on the "S" drive.

11. Where are all the copies of the content, including backups, and how are the copies of the content related?

A snap shot of the S drive is taken nightly and placed on a machine within the same machine room.

The master content files have monthly full backups and daily incremental backups to tape. The tape jukebox is held on a separate machine room off-site.

The access site (which includes the master copy of the descriptive metadata) has monthly full backups and daily incremental backups to tape and to cloud storage. In addition the access site is fully synchronized with a machine in a separate machine room off-site – live fail-over can occur and has been tested.

The long-term preservation of this content is being managed by the Third Party Preservation Service, which holds a complete copy of the PDFs, TIFFs, OCR, and metadata records within its fully replicated archive.

When: Document the targeted preservation timeframe and impact of loss.

12. How long should the content be available for use?

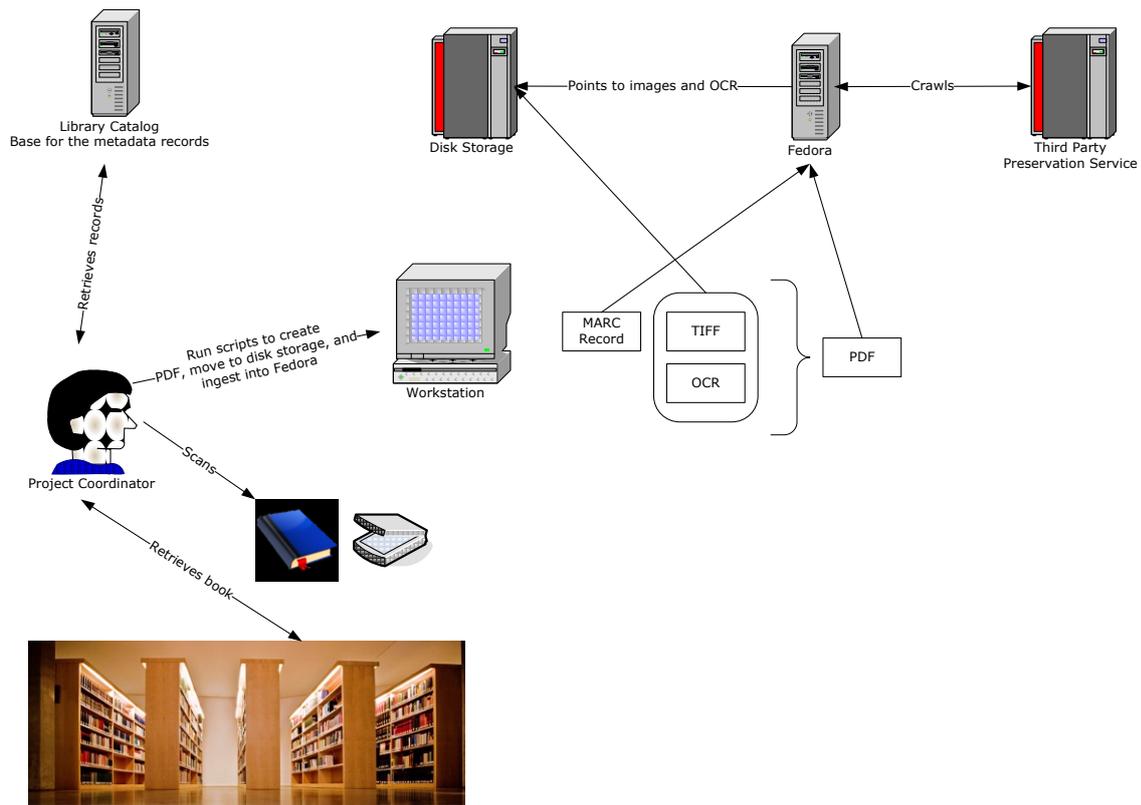
The content should be available for use forever.

13. If the content is irretrievably lost, what are the repercussions?

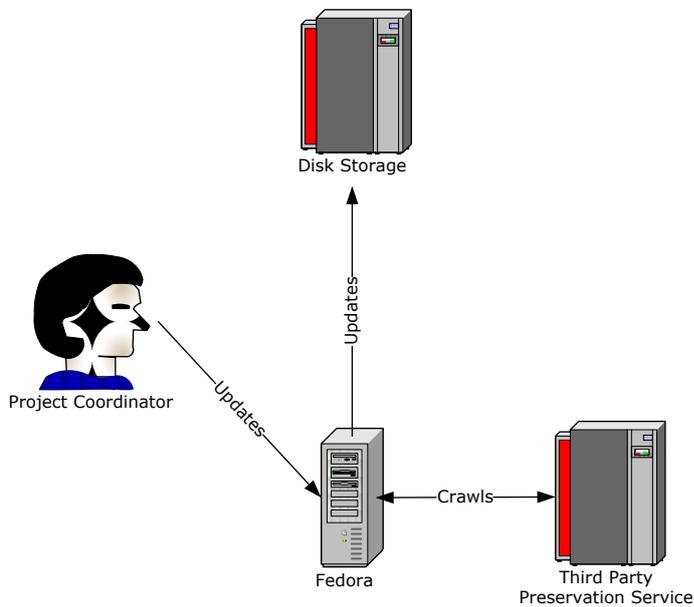
The repercussions are large. The paper cannot be redigitized, it is too fragile. Our only copy of this content is the digital version.

How: Document how the key content management and preservation tasks will occur.

14. How will the collection be created (perhaps draw a diagram of the workflow)?



15. How will the collection be maintained (perhaps draw a diagram of the workflow)?



16. Do you expect the content files to be migrated in the future?

Yes, the Third Party Preservation Service will migrate the content as needed over time.

17. May the content files be deleted? Added to? Updated?

Content files may occasionally be deleted if it's necessary for clean-up or copyright issues. The collection will grow with time.

18. May the descriptive metadata be deleted? Added to? Updated?

Yes, descriptive metadata may be deleted, increased, and updated.

19. How will you track who did what and when to the content, if this is important to your organization?

The Fedora system will track what was changed when and by whom.

20. How do you associate the master copy of the descriptive metadata with the master copy of the content files and how will you move these two items around together?

It is tied together both in Fedora and in the Third Party Preservation Service.

6. Findings from individual projects

6.1 Introduction

The project team reviewed the preservation plans and implementations of sixteen JISC digitisation projects between October 2008 and April 2009. The information itself was collected in a short period between December 2008 and February 2009 (one project was reviewed later, in May 2009.) The reports which follow should be read in that light. Many projects were still actively developing access mechanisms and had not yet established longer-term requirements. For a number of projects, significant changes took place after our review, sometimes as a consequence of the review. These changes are noted where they have been drawn to our attention.

In general, the project team was encouraged to find a high level of commitment and energy across the 16 projects. Keen to pass on lessons that they have learned, a number of project teams have generously permitted the publication of the reports that were produced as a result of our reviews. Consequently, although originally intended only for internal consumption, nine of the sixteen reports are included in this section. The case studies are not anonymized – but nor is it possible to relate these reports to the letters by which they are referred to elsewhere in the report without access to Appendix B (excluded from the public report).

The team first identified risks to preservation and recommendations for each project and then identified the critical preservation themes or strategic risks to preservation. The risks were identified because they impacted an institution's ability to digitally preserve the content.

Our review was based upon the following definition of digital preservation: Digital preservation is the series of management policies and activities necessary to ensure the enduring usability, authenticity, discoverability and accessibility of content over the very long-term. The key goals of digital preservation include:

- usability – the intellectual content of the item must remain usable via the delivery mechanism of current technology
- authenticity – the provenance of the content must be proven and the content an authentic replica of the original
- discoverability – the content must have logical bibliographic metadata so that the content can be found by end users through time
- accessibility – the content must be available for use to the appropriate community

The projects reviewed covered the spectrum of extremely at risk to secure in regard to their long-term preservation and accessibility. We have specifically identified the following five projects as at significant risk such that if action is not taken within the next several years, the services provided by this content may not be accessible to their designated community. These projects at significant risk are N, I, J, K and G. H was originally in this category, but took action after our initial visit which greatly reduced risk. The risks vary in severity. With project N, there are questions about access to material, and its protection, after 5 years. With others, there are risks associated with information being held in multiple systems or institutions without clear responsibility and workflows for coordinating updates.

These at risk projects all can be similarly characterized as independent projects that have neither integrated into a preservation infrastructure within their organization, collaboratively, or with a 3rd party digital preservation service, nor have allocated staff to maintaining the content post funding.

There are also projects in the middle of the spectrum, such as project D. While this project appears to be integrated into an organizational preservation infrastructure, it does not have ongoing staffing commitments or even staff who can currently access and manage the content.

Alternatively, projects such as O, E, B, and A appear to have solid preservation plans in place.

6.2. The Framework of the Project Reviews

The project-specific risks and recommendations follow a template. Each section starts with a table that provides a brief summary of areas that are key to assessing the preservation status of the projects. (An explanation of the summary table is below in Table 1.) Below the summary table in each review is a paragraph describing areas of preservation that the project has addressed well. Following this, the one to four most significant risks to preservation for the project are detailed, along with recommendations. Terms that can be found in the glossary are italicized throughout.

Table 1: Template Summary Table

PROJECT	Name of the project.
1. INSTITUTION	The institution which houses the project.
2. PARTNERS	Any partners involved in the project.
3. URLS	The URLs of key documentation or websites related to the project.
4. PLATFORMS	Most projects rely on more than one platform to manage their digitisation, content management and workflows, delivery and preservation. An enumeration of the platforms and their type and purposes can be helpful in evaluating the complexity and probable sustainability of a project.
5. STAFF	The number of staff currently working on the project and future staffing estimates, if known.
6. CONTENT	The type of content the project manages.
7. METADATA	The formats and type of metadata captured and the framework in which it is placed.
8. SIZE	The number of files and size of the content maintained by the project.
9. COMMUNITY	The designated community for the content (an identified group of users).
10. WORKFLOW	The way in which the data flows through the various systems and steps of the project.
11. TIMEFRAME	The planned length of time the content will remain available.
12. BACKUPS	The types of backups of the platforms and content.
13. PRES APPROACH	The project's stated approach to <i>digital preservation</i> .
14. PRES POLICY	Answers the question: Does this project have a defined preservation policy?
15. SUSTAINABILITY	Answers the question: Does this project have a defined sustainability plan?
16. AVAILABILITY	Answer the question: Is the content currently available on-line to the designated community?
17. CAN SHIP	Answers the question: Is the project capable of packaging and shipping the content?

6.3 Individual Projects

PROJECT	Archival Sound Recordings 2 (ASR2)	
1. INSTITUTION	The British Library (BL) Sound Archive,	
2. PARTNERS	Other BL digitisation projects - for advice	
3. URLS	http://sounds.bl.uk/	
4. PLATFORMS	<i>Platform</i>	<i>Platform Type</i>
	MARC to METS, outsourced	Unknown
	BL Archival Sound Recording platform	Web server + Apps (delivery)
	BL Digital Library System (DLS)	Content Management System for preservation
5. STAFF	Part-time (ongoing commitment)	
6. CONTENT	Nine collections of audio recordings (4,200 hours) including: environmental recordings, music, spoken word, and oral history	
7. METADATA	British Library Application Profile – Sound (BLAPS) which enables linking between objects and relationships between tracks or segments of an interview, it is METS based. Catalogue entries derived from MARC records. METS records viewable on the web.	
8. SIZE	4,200 hours and 7000 images at 20 TB	
9. COMMUNITY	UK HE & FE, but many files also available to public	
10. WORKFLOW	Master (high quality PCM WAV) created. Two derivatives: transcoding to MP3 and WMA. Creation of an uncompressed playback copy with noise reduction and restoration as needed. Access copies are derived from playback copy if it exists.	
11. TIMEFRAME	DLS content in perpetuity, access through 2018	
12. BACKUPS	DLS backups are very secure – duplicate DVD-Rs, 3 set of mirrored server stores at 3 locations and the hard drives have self-checking functionality. Also, commercial dark archive for tape storage.	
13. PRES APPROACH	Master copies and clean playback copies in community standard; access copies rederived over time as needed	
14. PRES POLICY	Under development	
15. SUSTAINABILITY	Reliance on BL	
16. AVAILABILITY	Yes	
17. CAN SHIP	No copies received, but we believe they can ship archival packages	

Contact: Richard Ranft; richard.ranft@bl.uk

Preservation Positives:

The ASR2 project is very transparent. The delivery and preservation systems are separate, but connected. The preservation system is the British Library's (BL) Digital Library Service (*DLS*) and the delivery system is the BL's Archival Sound Recording (ASR) website. The ASR website can be rebuilt from exports of the archival package from *DLS*. This ability to transport an archival package is uncommon and extremely positive. Indeed, the full *METS* files, reading for the stout hearted, are available within the delivery site. The BL has captured significant preservation metadata, including provenance data about the analogue source of the audio file and technical metadata about creation of the master digital version of the audio files. Another strong positive is that the BL has committed four part-time staff members to the ongoing maintenance of this project.

Preservation Risks and Recommendations

1. *Ongoing Updates to Intellectual Content*: Although not part of the original project plan², the BL created all *METS* records in ASR2 in-house and successfully built skills require for *METS* schema refinement and *METS* records preparation and editing.
 - *Risk to Authenticity* – Content management is problematic for nearly every project. With the ASR2 project at BL, it is not clear how updates to the metadata of a preserved recording as encoded in the *METS* would take place. BL does not have a workflow to update or recreate the descriptive metadata in the *METS*.
 - *Recommendation* – Document how descriptive metadata or other preservation metadata will be updated and ensure there is a process to allow for the metadata and *METS* files on the ASR delivery site to be rederived from updated master in the *DLS*.

² appendix C2 of which says “*METS* compilation cannot be resourced from within the library and it is proposed that this service be procured by an external supplier

PROJECT	Cabinet Papers, 1915-1978	
1. INSTITUTION	The National Archives (TNA), Strategic Development Department	
2. PARTNERS	transmediatraining - http://transmedia.co.uk/ - OCR and digitisation	
3. URLS	http://www.nationalarchives.gov.uk/cabinetpapers/default.htm	
4. PLATFORMS	<i>Platform</i>	<i>Platform Type</i>
	Preservation server	Complex preservation system
	Web Site Content Management System	Web server + apps
	<i>MODES Catalogue</i>	<i>MD catalogue</i>
5. STAFF	~11	
6. CONTENT	Minutes and memoranda (digitized from microfilm)	
7. METADATA	<i>METS, MIX, e-Government Metadata Standard, DC</i>	
8. SIZE	50,000 objects and ~800 GB	
9. COMMUNITY	Researchers, learners and teachers	
10. WORKFLOW	Manual: Microfilm sent to contractor. Scanned and OCRed. MD keyed into Excel and images. Images, OCR, and MD are wrapped in METS and placed on TNA preservation server and delivery renditions (PDFs and catalogue entries) are derived.	
11. TIMEFRAME	Currently free access. Will remain free to HE and FE students for at least five years and likely to remain so for A-Level	
12. BACKUP	Masters on servers and DVD, tape copies off-site and within TNA disaster recovery plan	
13. PRES APPROACH	Migration of formats and refreshing of hardware	
14. PRES POLICY	TNA's digital surrogates' policy developed in 2007	
15. SUSTAINABILITY	Have the technology to charge some user groups for use & commitment from TNA to maintain the content for 10 years	
16. AVAILABILITY	Yes	
17. CAN SHIP	Not yet, but we expect they will be able to ship archival packages	

Preservation Positives:

The National Archives brings much experience with the preservation of print materials to its treatment of digital materials. They have a *digital preservation* policy in place and have implemented technology to support it. TNA, in fact, has a very robust preservation system. TNA understands that different needs drive the delivery and preservation of content and therefore has a separate system for each of these activities. One of the challenges TNA has faced is packaging content in a reusable format and the fact that they realize this is a necessity and a challenge is very positive. It is also very encouraging that TNA will be tightly

packaging its preservation master images and its preservation master metadata together through *METS*, meeting one of the key requirements of preservation – the ability to move the master copies of content and metadata together from one organization or space to another.

Preservation Risks and Recommendations

1. *Content Management*: The TNA manually coordinates the movement of content through its digitisation, delivery and preservation processes using the *PRINCE2* project management system. Manual processes inevitably leave room for manual errors and such manual steps will make coordinating updates to the delivery and preservation objects difficult.
 - *Risk to Authenticity* – The manual management of the digitisation and update processes could cause future updates to the content to be incomplete and not performed at both the delivery site and preservation site and prevent any updates from being tracked in the preservation metadata of the preserved object.
 - *Recommendation* – 1) Rigorously define how the delivery and preservation systems are coordinated and how updates and content will flow from one system to the other. Ideally, updates will propagate automatically, however a well-defined manual process with quality control checks is acceptable. Or 2) define a preservation policy that defends why updates at the delivery or preservation site do not need to be propagated to the other copy of this content.
2. *Preservation of Teaching Materials*: For the Cabinet Papers project, the TNA has developed a series of study and contextualization packages. As these packages are not part of the digitized papers and metadata, they are not preserved on the TNA preservation server.
 - *Risk to Discoverability of the Minutes & Memoranda and Risk to Usability, Discoverability, Authenticity, & Accessibility of the Teaching Materials* – The TNA invested in developing these valuable teaching resources and were the delivery site to collapse, these materials would be lost (unlike the digitized content which could be rebuilt from master copies on the TNA preservation server).
 - *Recommendation* – Develop a preservation policy to address these types of secondary sources or teaching materials. Such a preservation policy might determine that the delivery site redundancy is sufficient to preserve the teaching materials, however without the exercise to identify the preservation needs of these materials the TNA cannot judge if it is meeting the needs.

PROJECT	A Digital Library of Core E-Resources on Ireland	
1. INSTITUTION	The Queen's University at Belfast Library and the Centre for Data Digitisation and Analysis (CDDA) (within the School of Geography, Archaeology and Palaeoecology)	
2. PARTNERS	JSTOR ³ - delivery & preservation Linen Hall Library, the Public Record Office of Northern Ireland, the Library and Information Systems Council Northern Ireland, Belfast Public Library, the National Library of Ireland in Dublin, the Royal Irish Academy, University College Dublin, the University of Ulster, the University of Southampton, the University of Leeds, Indiana University Purdue University Indianapolis, the University of California Berkeley, the California Digital Library, the University of Portsmouth, the Department of Culture, Arts and Leisure NI, University College Cork, University College Dublin, and Trinity College Dublin – content providers	
3. URLs	http://www.jstor.org/action/showJournals?selectDiscipline=355902696&browseType=discipline&x=3&y=13	
4. PLATFORMS	<i>Platform</i>	<i>Platform Type</i>
	MD tracking	Bespoke tracking software
	JSTOR - delivery	Web server + web apps
	JSTOR – archive	Replicated file system
5. STAFF	2 PIs 6 @ CDDA 3 @ IS JSTOR	
6. CONTENT	100 key journals, 205 monographs and 2,500 manuscript pages from core Irish Studies collections	
7. METADATA	JSTOR adaptation of NLM Journal Archiving and Interchange DTD (http://dtd.nlm.nih.gov/archiving/), JSTOR adaptation of the NCBI Book Tag Set (http://dtd.nlm.nih.gov/book/), PREMIS for technical metadata	
8. SIZE	2.1 million objects (3 objects / asset)	
9. COMMUNITY	UK – HE, FE, General Public Academic community	
10. WORKFLOW	QUB identifies and procures content. CDDA scans images, captures technical and high level descriptive metadata in a tracking sheet, and performs OCR. Content is sent to JSTOR, detailed descriptive metadata is compiled and the content is put together into JSTOR's packaging. Workflow at QUB is managed through a web based tracking tool which is	

³ Note, on January 1, 2009, Ithaka and JSTOR merged to form a single organization. The new organization, named Ithaka (see the announcement at <http://www.portico.org/news/012509.html>), was and is the organizational home of Portico and thus Portico and JSTOR are now within a single organization. Although we do not believe the organizational change has impacted our assessment of this project, we draw attention to this merger in the interests of full disclosure.

	also where initial metadata is captured. In addition RS staff compile journal title histories and issue-level metadata. Digital content offered back to content provider libraries/journals – none have yet taken it JSTOR will continue to add current issues for live journals.
11. TIMEFRAME	25 years
12. BACKUPS	CDDA backs digitized material up onto Queen’s ‘archives’ system which is kept off site. CDDA also has DVD copies of the material. JSTOR has its own backup systems.
13. PRES APPROACH	Migration – rely on JSTOR (note, the project had wanted to deposit archival copies at AHDS)
14. PRES POLICY	No
15. SUSTAINABILITY	Contract with JSTOR for 25 years
16. AVAILABILITY	Yes – with original estimated page count of 500,000 exceeded to 720,000
17. CAN SHIP	We did not receive copies, but JSTOR moves content between its data centers and could ship content if asked.

Preservation Positives: This project is well run. Queen’s University at Belfast understands the complexity of providing a delivery service and of providing a preservation service and has chosen to partner with JSTOR, a 3rd party archive with expertise in both areas. They also relied on guidance from JSTOR on the digitisation and indexing, which ensures they have the highest possible digitisation and metadata capture standards. Although the initial plan of preserving master copies with AHDS could not be completed when AHDS funding was withdrawn, the project developed a successful alternative.

Preservation Risks and Recommendations

1. *Preservation Policies:* JSTOR is a sound preservation choice and their preservation strategy is available via their website. A substantial number of preservation policies are encoded in the license agreement between Queen’s University at Belfast and JSTOR.
 - *Risk to Usability & Accessibility* – Preservation relies solely on JSTOR.
 - *Recommendation* – Develop and publicly post a set of preservation statements for the project. These project specific statements can be derived from the statements in the licensee agreement and provide reassurance and clarity to the UK HE and FE community on the specific preservation practices that will be applied to the content.

PROJECT	Freeze Frame	
1. INSTITUTION	Cambridge University, Department of Geography, Scott Polar Research Institute (SPRI)	
2. PARTNERS	None	
3. URLs	Delivery Site: http://www.freezeframe.ac.uk/ Project Site: http://www.spri.cam.ac.uk/resources/freezeframe/ DSpace Site: http://www.dspace.cam.ac.uk/	
4. PLATFORMS	<i>Platform</i>	<i>Platform Type</i>
	SPRI file server	File System
	SPRI Picture Library Catalogue	Metadata Management System
	<i>DSpace</i>	Content Management System
	Website (CARET)	File System
5. STAFF	~ 7 for the duration of the project	
6. CONTENT	Glass plates - colour and BW, Photographic negatives from 1845-1980	
7. METADATA	Catalogued at item and collection level. Using <i>SPECTRUM</i> , <i>ISAD(G)</i> , <i>METS</i> , & <i>DC</i> – recorded in a <i>MODES</i> XML record and put in SPRI Picture Library <i>MODES</i> catalogue	
8. SIZE	20,000 images & 1.56 TB	
9. COMMUNITY	<i>HE</i> and <i>FE</i> are main audience but open to all.	
10. WORKFLOW	Not automated	
11. TIMEFRAME	<ul style="list-style-type: none"> • <i>DSpace@Cambridge</i> will byte store for 50 years • SPRI expects the content to be available “indefinitely” • There is no lifespan designated for the Freeze Frame delivery website 	
12. BACKUPS	DSpace and CARET offsite backups; metadata and images are backed up internally at SPRI and within the Department of Geography	
13. PRES APPROACH	To be defined	
14. PRES POLICY	Using Cambridge University’s – to be redefined in 2009	
15. SUSTAINABILITY	Rely on Cambridge University – could sell high resolution images	
16. AVAILABILITY	Yes, as of March 5, 2009	
17. CAN SHIP	No – sent XML samples of either <i>DSpace</i> or Website <i>MD</i> , but not full preservation packages and not master metadata	

Contacts: Naomi Boneham; Scott Polar Research Institute, University of Cambridge;
nab37@cam.ac.uk

Preservation Positives: The Freeze Frame project has integrated some of its master content into the institution wide preservation system – *DSpace@Cambridge*. In terms of maintaining the images for the long-term, this is a solid preservation choice, as institutional support for *DSpace@Cambridge* reduces the chance of this content becoming lost or orphaned. It also addresses some sustainability concerns around preservation, as Cambridge has committed to maintaining the content preserved within its *DSpace* instance.

Preservation Risks and Recommendations

1. *Content Management*: The Freeze Frame project relies on multiple content and metadata management systems and the staff manually moves content between the systems. There is a significant separation between the master metadata, which is maintained in the SPRI Picture Catalogue, and the master image files, which are maintained in the *DSpace@Cambridge*, and yet another separation between the website and *DSpace@Cambridge*.
 - *Risk to Authenticity* - The complexity of the entire application stack is a risk to maintaining the content for the long term. For example, there is no clear process to propagate an update to metadata from the Picture Catalogue through *DSpace* and out to the website. Similarly, there is no process to update an image in *DSpace* and propagate it out to the website and back to the low resolution image used by the Picture Catalogue.
 - *Recommendation* – 1) Determine whether all the systems are truly required. For example, is the website necessary or could access be driven off of *DSpace@Cambridge*? The best preservation occurs when the preservation and delivery systems run separately while remaining coordinated so that updates can automatically flow between them. If Freeze Frame determines that all the existing platforms are required, then they should rigorously define how updates and content flow from one system to another and how a representation of the object in one layer is related to a representation of the object in another layer. 2) Tightly couple the master copy of the object to the master copy of its metadata, so that the metadata and its corresponding object can be packaged together and shipped externally as a single unit. This coupling is best managed by software so that it is relatively automated; however robust documentation can also serve this purpose.
2. *Post-Project Maintenance & Sustainability*: The Freeze Frame project staff have employment contracts for the duration of the project. Although *DSpace@Cambridge* has committed to maintaining for the long-term any content deposited within it, long-term maintenance of a project requires domain knowledge. For example, if the descriptive metadata for an image were identified as incorrect after the completion of the project, then either the *DSpace@Cambridge* staff or the permanent SPRI staff, neither of which may have been involved in the initial project, would need to manage the update process.
 - *Risk to Discoverability & Authenticity* – The risk is that although the images of the collection and metadata may be preserved for the long term within *DSpace@Cambridge*, there will be no long-term maintenance of the intellectual value of the content.
 - *Recommendation* – SPRI should define how its permanent staff will preserve the intellectual content of the digitized collection for the long term and ensure that there is a process established with *DSpace@Cambridge* to coordinate the ongoing maintenance of the collection.
3. *Preservation Policies*: The Freeze Frame project is relying on the yet to be developed Cambridge University preservation policy.

- *Risk to Accessibility & Usability* – 1) Freeze Frame may have different needs and desires than the university as a whole. 2) The master copy of the Freeze Frame metadata, which is key to the long-term preservation of the content, is located in the SPRI Picture Catalogue, which may not be covered by the Cambridge University preservation policy.
- *Recommendation* – Develop a set of preservation policies (they need not be extensive) for the Freeze Frame master images, master metadata and access site before the development of the University wide preservation policy. These project specific policies can influence the University wide policy as it is developed or inform SPRI that they need to find a supplemental or alternate preservation partner. These policies need not be technical and could instead describe service levels. For example, the master images and metadata must be available for the long term, access to the master images and metadata may be asynchronous, the website can be derived completely from *DSpace*, etc.

PROJECT	InView: Moving Images in the Public Sphere	
1. INSTITUTION	British Film Institute (BFI)	
2. PARTNERS	The Parliamentary Recording Unit, BBC, ITN, Open Media, Northern Region Film & Television Archive, South West Film and Television Archive, the Media Archive for Central England, and the East Anglian Film Archive – content providers and advisors Website hosting/development service	
3. URLs	http://www.bfi.org.uk/inview	
4. PLATFORMS	<i>Platform</i>	<i>Platform Type</i>
	<i>Artesia</i>	Image + <i>MD</i> database for digital inventory
	TEC-REC	Content management system/database for archive inventory
	<i>Drupal</i>	Website content management system
	SIFT	BFI Summary of Information on Film & Television Database that holds <i>MD</i>
5. STAFF	12 on project, with long-term management by collections and BFI IT staff	
6. CONTENT	600 curated hours of public record films, parliamentary coverage, national news broadcasts, and campaigning films; also paper documents and other contextual materials	
7. METADATA	Based on <i>Dublin Core 1.1</i> , <i>MODS</i> , and <i>CEN.BT TF 179</i> (European Cinematographic Standard specific to film)	
8. SIZE	2000 primary source film and video works, 7500 scanned documents, 500,000 thumbnail stills at approximately 17 <i>TB</i>	
9. COMMUNITY	UK HE & FE sector	
10. WORKFLOW	Film selected, film transferred to <i>HD</i> video tape, <i>HD</i> video tape is transcoded to sub-master files from which the web versions are derived. Most curation occurs after creation of the sub-master files.	
11. TIMEFRAME	In perpetuity. If BFI fails, expect Higher Education Funding Council for England to take over.	
12. BACKUPS	Digital assets are held on a RAID server with <i>LTO</i> tape backups. The <i>HD</i> tapes (master copies) are stored in BFI vaults to approved international best practice standards.	
13. PRES APPROACH	Migration of submasters. Delivery versions recreated from submasters.	
14. PRES POLICY	In draft form	
15. SUSTAINABILITY	Government funded and will not charge. If funding disappears, they will discard content.	
16. AVAILABILITY	Not as of March 8, 2009 (but by 30 September 2009)	
17. CAN SHIP	Unknown – no samples sent	

Preservation Positives: The BFI has successfully brought its long history with preservation of analogue materials to bear on the *digital preservation* needs of the InView project. While they use a number of systems to manage the content and workflows, they understand that this complexity can introduce errors. BFI has a long-term plan for managing the intellectual content of the InView data, including both a process to coordinate updates to metadata and staff for whom maintenance of this content will become an ongoing task.

Preservation Risks and Recommendations

1. BFI did not provide a sample archival package, which raises the question of whether or not they have the ability to package their digital submaster files and archival metadata together into an exportable archival package. (They have subsequently informed us that they package digital sub-master files and encoding metadata together on its LTO4 preservation version within a standard TAR format. Each JISC D10/50I MPEG file has its metadata stored alongside it in a separate XML file. A BFI-specific DTD (Document Type Definition) is available for metadata validation.)
 - *Risk to Usability & Accessibility* – It is key that any organization engaged in the long-term preservation of digital content have the ability to move the content from within its own organization to another entity. Without this basic ability, it is not possible to promise long-term preservation of the material, as it is completely dependent upon the technology of the moment for management. (BFI is working on this issue as part of a government-funded national heritage strategy.)
 - *Recommendation* – Develop a process whereby archival metadata can be paired both with the submaster images and the *HD* tapes for export.

PROJECT	The John Johnson Collection: An Archive of Printed Ephemera	
1. INSTITUTION	Bodleian Library, Oxford University	
2. PARTNERS	Bodleian Library – content (lead partner) ProQuest - delivery Capital Micromedia Total Network Solutions - digitisation and a (subcontractor of ProQuest) Oxford University Computing Services (OUCS) – file storage and backup	
3. URLs	http://johnjohnson.chadwyck.co.uk/marketing.do	
4. PLATFORMS	<i>Platform</i>	<i>Platform Type</i>
	ProQuest	Web server + Apps
	Hierarchical File Servering System (<i>HFS</i>)	Oxford backup solution
	Digital Asset Management System (<i>DAMS</i>) – under development	<i>Fedora</i> content management on top of Sun Honeycomb archival hardware
	Allegro DB	Metadata database
5. STAFF	~10 FTE3 at Bodleian & ~6 at ProQuest	
6. CONTENT	The John Johnson collection Collection of printed Printed ephemera Ephemera from 1508 to 1939	
7. METADATA	<i>DMD</i> in Allegro (<i>MARCish</i>) and <i>TMD</i> in images – will move to <i>METS</i> , <i>MODS</i> , and <i>PREMIS</i>	
8. SIZE	65,000 items & over approximately 150,000 images and <i>OCR</i> at ~4TB	
9. COMMUNITY	The full service is available to the UK <i>HE</i> & <i>FE</i> , public (via UK public libraries) and schools The metadata is available internationally.	
10. WORKFLOW	Bodleian manages workflow. Bodleian catalogs and conserves selected items. Content is sent to Capita where it is scanned, <i>OCR</i> ed and <i>QA</i> ed. Capita sends content to ProQuest via a hard drive. ProQuest extracts selected material and sends the hard drive on to Bodleian. Bodleian <i>QAs</i> and then preserves the content on <i>OUCS</i> ' Hierarchical File Servering System (<i>HFS</i>). Managed Processes managed via a bespoke tracking application	
11. TIMEFRAME	In perpetuity – 5 years at ProQuest, could be extended or Oxford could deliver	
12. BACKUPS	One copy on <i>HFS</i> . One copy on <i>DAMS</i> at an off-site location. One copy on tape off site.	
13. PRES	Migration	

APPROACH	
14. PRES POLICY	No – under discussion
15. SUSTAINABILITY	Bodleian committed
16. AVAILABILITY	Yes
17. CAN SHIP	No – no true archival packages yet, waiting on <i>DAMS</i>

Contacts: David Tomkins; University of Oxford; david.tomkins@bodley.ox.ac.uk

Preservation Positives: Bodleian understands the difficulty of long-term preservation and that it requires an organization devoted to the content for the long-term, as well as a preservation infrastructure to support the work. Bodleian is well on its way to building a Digital Asset Management System which, when paired with ongoing archive maintenance functions, should be able to preserve content for the long term. In addition, Bodleian has chosen to track its workflow through a bespoke tracking application, rather than simply via spreadsheets and checklists – again showing its deep understanding of the complexity of data management and how manual processes can easily fail.

Preservation Risks and Recommendations

1. *Content Management*: Ongoing maintenance of content requires attention to both its physical preservation needs (are the bits and bytes safe) and to maintenance of the intellectual content of the objects (did we misspell the author's name). The John Johnson Collection delivery and preservation copies of the content are separate, which is good preservation practice, but they are not coordinated. There is no apparent way for updates made to content at ProQuest to propagate back to Oxford, or vice versa.
 - *Risk to Authenticity* – Should the descriptive metadata or structural relationship of the content need to be updated in the future, there is no way to coordinate the updates between the two organizations. However, Oxford see themselves as holding the master copies and believe that ProQuest should not be amending metadata unless amendments are supplied by Oxford, so there is no need to be able to coordinate updates in the reverse direction. This is potentially a risk to authenticity, because it will not be possible to determine from analysis of the archival version of the content what changes were made and when – although the master copy should always be treated as the authentic one.
 - *Recommendation* – 1) Rigorously define how the delivery and preservation systems are coordinated and how updates and content will flow from one system to the other. Ideally, updates will propagate automatically, however a well-defined manual process with quality control checks is acceptable. Or 2) define a preservation policy that defends why updates at the delivery or preservation site do not need to be propagated to the other copy of this content.
2. *Lack of a Preservation Infrastructure*: There are two JISC-funded digitisation projects at Oxford in this review – this project, the John Johnson Collection, and the First World War Poetry Archive. These two projects do not share a preservation infrastructure although they clearly work together to the extent that the institution makes possible. Bodleian has already determined that *HFS* is not an ideal preservation infrastructure, as a preservation infrastructure should include activities such as monitoring of file formats, packaging of metadata and object files together, fixity and completeness checks, and other ongoing archive maintenance tasks. *DAMS* could be a digital preservation solution once fully implemented.

- *Risk to Usability & Accessibility* – Until *DAMS* is complete and the John Johnson content is archived within it, the content is at some small risk of loss, although it is likely to be well-protected in HFS.
- *Recommendation* – Complete *DAMS* and move the archival version of the John Johnson Collection into it. In addition, *OULS* should proactively reach out to other Oxford University projects and aid in getting other at risk content into *DAMS*.

PROJECT	Pre-Raphaelite Resource Site	
1. INSTITUTION	Birmingham Museums & Art Gallery (BMAG), Birmingham City Council (BCC)	
2. PARTNERS	University of Birmingham, Barber institute of Fine Arts - supplied the Academic Advisory Group Digital Alumna - audience research Tak! - website BCC IT Department – information technology support	
3. URLS	http://www.preraphaelites.org/ (live website post launch)	
4. PLATFORMS	<i>Platform</i>	<i>Platform Type</i>
	<i>MINISIS</i>	Collection management system
	Tak! (delivery)	Web server + apps
	<i>Extensis Portfolio</i>	Image management and cataloguing system
5. STAFF	~8 for duration of funding, no staffing plans post-funding	
6. CONTENT	Prints and drawings on paper; paintings on canvas; stained glass; costumes; letters; wood blocks; tapestries; and art objects	
7. METADATA	<i>SPECTRUM, DC, MODS</i>	
8. SIZE	3,000 images and 375 GB	
9. COMMUNITY	UK HE & FE and the international general public	
10. WORKFLOW	Photos taken. Basic metadata captured. Photo files are processed and copies made and retouched. Files saved to a project folder following a naming convention based on <i>MINISIS</i> Object ID. At completion, a low resolution <i>JPG</i> version is created. Once a month the low resolution <i>JPGs</i> are transferred to the <i>MINISIS</i> drives. Metadata is exported from <i>MINISIS</i> to a database for use on the website. <i>Portfolio</i> is used to manage the images. Workflow is managed manually through tracking tasks in Word and <i>MINISIS</i> .	
11. TIMEFRAME	3-5 years	
12. BACKUPS	Local on-line backups and 2 sets of quarterly backups off-line and off-site. <i>JPGs</i> backed up to CD.	
13. PRES APPROACH	None	
14. PRES POLICY	In development	
15. SUSTAINABILITY	Unknown	
16. AVAILABILITY	Not on its own website as of March 11, 2009 – some images at BMAGIC; since made available at http://www.preraphaelites.org/	

Preservation Positives: Though a challenging project, the content appears to have been well managed by a small number of staff. BMAG, having lost data for earlier projects, has implemented a robust set of back-up processes for both the images and the metadata. The workflow for the project was manually managed; however, the images on the file system are managed through the *Portfolio* content management system, thereby reducing the chances of loss.

Preservation Risks and Recommendations

1. *Delivery and Archive Synchronization:* All projects require ongoing maintenance and need attention to both their physical preservation needs (are the bits and bytes safe) and maintenance of the intellectual content of the objects (did we misspell the artist's name). For the Pre-Raphaelite project there is no apparent way for updates made to the metadata in *MINISIS* at BMAG to propagate out to the website.
 - *Risk to Discoverability & Authenticity* – BMAG will undoubtedly keep its catalogue up to date and correct errors as they are found. However, if those updates do not propagate to the delivery site, the content at the delivery site will become out of synchronization with the master copy of the data, and it will not be possible to track changes made in one place to the other. Over time, this will impact the proven authenticity of the objects and whether or not they can be discovered by end users.
 - *Recommendation* – 1) Define how the delivery and image management systems are coordinated and how updates and content will flow from one system to the other. Ideally, updates will propagate automatically, however a well-defined manual process is acceptable. Or 2) define a preservation policy explaining why it is not necessary for updates at the delivery or preservation site to be propagated to the other copy of this content.
2. *Post-Project Sustainability:* BMAG has dedicated staff to the creation of this resource for the duration of the JISC funding, however they have not planned to dedicate any staff to the ongoing maintenance of the resource. The staff interviewed for this study specifically commented that they were uncertain who would conduct website maintenance and review after the funding time period ended.
 - *Risk to Authenticity, Discoverability, Usability, & Accessibility* – BMAG has a limited staff and even fewer staff dedicated to *digital preservation*. BMAG also has highly limited technical abilities, with an IT consultant on site only 10 days a year. Support from the Birmingham City Council appears to be very uncertain.
 - *Recommendation* – Define with BMAG management how the systems managing the digital content and the content itself will be funded and maintained over time.
1. *Preservation Policies & Digital Preservation Infrastructure:* One of the strategic risks identified through the review of all the digitisation projects is that smaller projects and smaller organizations should not attempt to develop their own preservation infrastructure. They should leverage collaborative relationships with other like minded organizations or departments within their own organization or develop a relationship with an existing 3rd party digital preservation service. BMAG appears to have robust back-up capabilities, but the entirety of a preservation infrastructure includes activities such as monitoring of file formats, packaging of metadata and object files together, fixity and completeness checks, and other ongoing archive maintenance tasks. BMAG has no provision yet to fully preserve its content.

- *Risk to Usability & Accessibility* – Without a stronger plan for preservation of the master image files and metadata files, BMAG cannot assure their long-term ability to manage or even find this content in the future.
- *Recommendation* – 1) Define a preservation policy for the content, both the master content files and the master metadata records. 2) Partner with other like minded institutions to develop a collaborative preservation archive, thereby sharing both the responsibility and cost for maintaining the content, or rely upon a 3rd party digital preservation service.

PROJECT	Welsh Journals Online	
1. INSTITUTION	National Library of Wales (NLW)	
2. PARTNERS	Welsh Higher Education Libraries Forum (WHELF) – advice and material Jouve Group – French OCR vendor	
3. URLS	http://welshjournals.llgc.org.uk/content/home (delivery - in beta, no content)	
4. PLATFORMS	<i>Platform</i>	<i>Platform Type</i>
	VITAL (VTLS)	Content management system for delivery based on <i>Fedora</i> and SOLR software
	Digital Archive (Quantum)	Hardware and software archival solution
5. STAFF	>12 (11-12 funded others participate), portions of staff allocated for ongoing maintenance post funding	
6. CONTENT	All academically significant Welsh periodicals published since 1900	
7. METADATA	<i>METS</i> frame containing descriptive metadata (extracted from the custom system that was populated by hand from catalogue <i>MARC</i> records). <i>DMD</i> at article & page level. Technical metadata stored in Digital Archive. <i>TEI, METS, MIX, textMD, PREMIS, WAI, MODS</i>	
8. SIZE	50 titles & 400,000 pages of text at ~1 TB	
9. COMMUNITY	UK <i>HE</i> & <i>FE</i> , the international general public	
10. WORKFLOW	Title selected. Metadata extracted, created and QAed. Title, volumes, issues, pages prepped in workflow. Pages scanned and checked against <i>MD</i> . Images sent to <i>OCR</i> contractor. <i>OCR</i> incorporated into <i>TEI</i> . Delivery <i>PNGs</i> are created. Images and <i>TMD</i> files are deposited in Digital Archive. Rights data added to <i>METS</i> . <i>METS, TEI</i> , and delivery images are ingested into <i>VITAL</i> . Internal processing managed by a workflow program.	
11. TIMEFRAME	In perpetuity	
12. BACKUPS	Master <i>TIFFs</i> are preserved on optical disc in a juke box at the NLW Digital Archive. Three other copies all considered part of the Digital Archive: one in a local tape library, one off-line copy on site and another off-line copy off site. <i>PNG</i> derivatives are stored in the Digital Archive tape libraries (not the optical disc). The <i>METS</i> and <i>OCR (TEI)</i> are held in <i>VITAL</i> on redundant mirrored servers. Work in progress is backed up weekly.	
13. PRES APPROACH	Migration	

14. PRES POLICY	NLW <i>Digital Preservation Strategy</i> Project specific <i>Digital Preservation Plan</i> under development
15. SUSTAINABILITY	NLW has committed ongoing staff time to maintenance
16. AVAILABILITY	Not as of March 10, 2009
17. CAN SHIP	Sent XML, not an archival package -- however we believe they could send an archival package

Preservation Positives:

Welsh Journals Online has one of the stronger preservation plans of the JISC Digitisation Programme projects. They are one of the few projects that supplied the review team with a preservation plan -- other projects reference such plans, but did not (and, perhaps could not) provide the plans to the review team. NLW maintains tight control over their production workflow, with the process managed through a bespoke workflow system. This helps insure that content is not misplaced and metadata misassigned. In addition, NLW has committed to ongoing staff support of the system beyond the length of funding. This was reflected in the preservation plan, and, in the survey of NLW staff, several commented that through the process of creating this plan they realized they must commit 25% of the time of several staff people to support the ongoing access and preservation of this content.

Preservation Risks and Recommendations

- *Content Management of METS & TEI*: NLW has chosen to preserve its master images in their robust Digital Archive. However, the *METS* files which contain descriptive and preservation metadata and the structural information about how to construct archival information packages and the full-text *TEI* files created by merging of the *OCR* text and descriptive metadata are not preserved in the digital archive. The *METS* and *TEI* files exist only in *VITAL*, the content management system that manages the delivery site and provides an entry point into the Digital Archive.
- *Risk to Discoverability, Accessibility, and Authenticity* – The *METS* files are an intrinsic element of the entire archival information package for these newspapers. Without the *METS* files, the preserved image files become a data store without an entry point. The *TEI* files could be recreated by merging together a re-*OCR* of the image files and the metadata from the *METS*, and as such are not as key to long-term preservation as the *METS*, although the costs of having to redo that *OCR* work and recreate the *TEI* files should be considered.
- *Recommendation* – Place the *METS* files in the NLW Digital Archive, alongside their image counterparts. NLW should also consider the costs of preserving the *TEI* files in the Digital Archive against the costs of having to recreate them in the future if a catastrophic disaster were to damage *VITAL* and its backups beyond recovery.

PROJECT	UK Theses Digitisation Project	
1. INSTITUTION	The British Library (BL)	
2. PARTNERS	Research Libraries UK (RLUK – formerly CURL); Cranfield University; University of Warwick; University of Glasgow; University of Edinburgh; Robert Gordon University; University of Birmingham – staffing for retrieval, shipment and refiling of theses	
3. URLs	http://ethos.bl.uk/ http://www.ethos.ac.uk/	
4. PLATFORMS	<i>Platform</i>	<i>Platform Type</i>
	DLS	BL's preservation infrastructure
	ETHOS	Delivery system based on EPrints
	Catalogue	Cataloguing system for metadata
5. STAFF	4 for duration of the project, ongoing staffing plan tailored to expected demand with contingency plan to increase resource if required.	
6. CONTENT	Circa 11,000 theses digitized	
7. METADATA	Unknown	
8. SIZE	Unknown	
9. COMMUNITY	UK HE & FE, the general public	
10. WORKFLOW	The full process was not described.	
	<p>Metadata records are imported into EThOS from the BL catalogue and University Institutional Repositories. Theses are scanned to create <i>TIFF</i> files and then <i>OCR</i>ed. The <i>TIFF</i> files and <i>OCR</i> are combined into PDF files. Theses are also harvested from Institutional Repositories – these theses can be multi-file, multi-format. The metadata records are linked to the PDF files via links on the Ethos database. Each thesis has an 'archive' record which contains the link to the thesis metadata and the location of the pdf and any additional content on the storage drive. Access copies and master copies are on separate servers.</p>	
11. TIMEFRAME	Perpetuity	
12. BACKUPS	Ethos frontend web application server - every day. Ethos backend operations server which includes pdf/Eprints store.– every week Regular full backups of the archived TIFF/pdf filestore.	
13. PRES APPROACH	Migration or emulation as needed	
14. PRES POLICY	Reliance on the BL digital preservation strategy and future policy	
15. SUSTAINABILITY	Yes	
16. AVAILABILITY	Yes	
17. CAN SHIP	Unknown	

Many aspects of this project are shared with Archival Sound Recordings 2 and the British Newspaper Digitisation projects, also led by the British Library: so certain of the risks and positives are generic. Generic markers of long term risks can be traced within the evidence presented to the project: these are worth describing for what they illustrate about projects rather than any specific risks to the Thesis Digitisation project. These markers include the following:

- The British Library is a large institution which receives large quantities of material from third parties and is able to support specialised preservation staff. This is welcome but it has the inevitable consequence that although questions about preservation processes were easily answered by specialist staff it was harder for the project team to obtain a comprehensive description of production processes. Preservation is addressed throughout the life of a project, and thus knowledge of early workflows and content management assists preservation. In this case, that might include understanding the source of the metadata and image files and how those are combined. In practice it means that the BL's preservation specialists have confidence in the work of colleagues and that quality is assured by processes in other sections of the institution.
- There is an apparent risk associated with the separation of metadata and content which on closer inspection is purely notional. In most cases, the master copy of the resource discovery metadata originates from the BL's catalogue, which is an authoritative source in its own right. This means there is a separation between the master copy of the resource discovery metadata and the master copy of the images. Such a separation is usually a marker of long-term risk to preservation, because it is not clear where the master copy of the metadata is located (the catalogue or in a record with the image or in the content management system) and because, should the master copy of the metadata reside in the catalogue, it is less likely to be preserved within the digital preservation infrastructure. However, in this case the cataloguing metadata predates the digitisation and is itself an authoritative source. The two record copies are strongly linked – each contains the unique ID of the other – and the unique ID of the digital item is contained within the EThOS record. This will significantly mitigate against this risk.
- There is a marginal short term risk to digital objects while they are queued for ingest to the Digital Library System, a risk which has been mitigated by providing multiple backups. The objects consist of the *TIFF* scans, *OCR* text, PDF files (which are a combination of the *TIFF* scans and *OCR*), and metadata. Neither the survey nor the project plan describes how all of these elements are combined and maintained for the long-term. There is mention that the content will be placed in the *DLS* in two years, but no explanation of what content will be placed in the *DLS* (PDF files only? PDF files plus original images and *OCR*? All of the above with metadata records?). This two year delay could present both an immediate risk to the content, should the backups proved to be insufficient, and a long-term risk, should institutional knowledge be lost over the course of those two years. However, these risks are small in absolute terms – 2 years is a relatively short time in which to lose institutional knowledge, for instance. In addition, the *DLS*'s function of providing 'preservation in perpetuity' for the BL's digital collections and the processes which surround it are likely to provide a high degree of assurance for this content once it is ingested into the *DLS*.
- EThOS is an ongoing service and therefore it is important that the project identify how content added to EThOS over time will be copied to *DLS*.

- The project necessitates a degree of short term risk because there are numerous partners – but that is offset against the long term benefit of a single co-ordinated set of preservation actions. There are many partners involved in this project including the Digital Preservation team at the BL, the EThOS project, a digitisation vendor, and the libraries that provide the theses to the BL for digitisation. In general, projects where the development and maintenance are distributed among a large number of partners are at greater risk to long-term preservation than are projects with a smaller number of partners, as the chance of misunderstandings are higher. Set against this, the centralised storage function in an institution with a national remit, along with storage of local copies at institutions, is ameliorating the risk of institutional failure.

Although markers that can predict risk to long-term preservation are present in this project, it is likely that the BL has mitigated the problems these markers often identify. Further information on the early, current, and future content management processes of this project is needed to make a full risk determination.

Glossary

AHDS: The UK Arts and Humanities Data Service – funding was withdrawn in April 2008 and some services were taken over by Centre for e-Research (CeRch) at King’s College London.

ALTO: An acronym for Analyzed Layout and Text Object – it is an XML schema that supports encoding OCR-recognized text and the position of that text on the source image at the word level. It is often encoded within METS and in such instances it is referred to as METS/ALTO.

Artesia: A commercial digital asset management system from Open Text.

CEN.BT TF 179: A shorthand notation for the Cinematographic Works Standard metadata framework being created under the auspices of the Task Force 179 of the European Committee for Standardization (CEN). It has been since superseded by CEN.BT Technical Committee 372.

CMS: A content management system – it is software designed to allow organizations to manage their digital objects. It sometimes has a hardware component, as well as a software component.

CONTENTdm: A digital repository system from OCLC – it is available both as a local installation and as an OCLC hosted service and is most frequently used as a hosted service.

Copac: It is a freely available library catalogue with approximately 32 million records and representing the merged holdings of the members of the Research Libraries UK (*RLUK*) - this includes the catalogues of the British Library, the National Library of Scotland, and the National Library of Wales / Llyfrgell Genedlaethol Cymru and increasing numbers of specialist libraries with collections of national research interest.

DAMS: An acronym for “Digital Asset Management System” – it is being built at Oxford University to provide long-term content management to digital content.

DC: See *Dublin Core*.

Digital Preservation: the series of management policies and activities necessary to ensure the enduring usability, authenticity, discoverability and accessibility of content over the very long-term. The key goals of *digital preservation* include: usability – the intellectual content of the item must remain usable via the delivery mechanism of current technology; authenticity – the provenance of the content must be proven and the content an authentic replica of the original; discoverability – the content must have logical bibliographic metadata so that the content can be found by end users through time; and accessibility – the content must be available for use to the appropriate community.

Digital Repository System: Software to enable the collection of content on the web – they are similar to content management systems, but do not enable the creation of robust content management workflows.

DLS: Also known as Digital Library System – it is the software that has been built by the British Library to provide itself with long-term *digital preservation*.

DMD: An abbreviation for descriptive metadata – it is bibliographic metadata that describes an object.

Drupal: An open source content management platform developed for website maintenance.

DSpace: An open source digital repository package.

DTD: A document type definition – it is a specific definition that follows the rules of the Standard Generalized Markup Language (SGML) and provides a specification that accompanies a document and identifies markup elements and the rules for their use.

Dublin Core: A shorthand notation for the “Dublin Core Metadata Element Set”, which is a vocabulary of fifteen properties for use in resource description. It is abbreviated, *DC*.

EAD: The *EAD* Document Type Definition (*DTD*) is a standard for encoding archival finding aids using XML.

EDINA: *EDINA* is the JISC national academic data centre based at the University of Edinburgh – it has a mission to enhance the productivity of research, learning and teaching across all universities, research institutes and colleges in the UK.

Extensis Portfolio: A commercial digital image management system to allow for cataloguing of files, visual organization of files, and drag and drop integration with the operating system.

FE: An abbreviation for Further Education.

Fedora: It is an open source content management platform that enables the storage, access and management of digital content

GB: An abbreviation for gigabyte – it is 1,000,000,000 bytes or 10^9 bytes. A project with content in the gigabytes is relatively small.

GIS: An abbreviation for geographic information system – it is a system that captures, stores, analyzes, manages, and presents data that is linked to location. *GIS* is often used to refer to the data that drives a geographic information system.

HD: An abbreviation for high definition.

HE: An abbreviation for Higher Education.

HFS: An abbreviation for Hierarchical File System – which is a robust file server and backup system maintained by Oxford University Computing Services.

ISAD(G): A standard that provides general guidance for the preparation of archival descriptions. It is used in conjunction with existing national standards or as the basis for the development of national standards.

JISC1: An acronym used for projects that received funding through Phase 1 of the JISC Digitisation Programme.

JISC2: An acronym used for projects that received funding through Phase 2 of the JISC Digitisation Programme.

JORUM: A free online service providing access to teaching and learning resources, for teaching and support staff in UK Further and Higher Education Institutions

JPG: Also abbreviated as JPEG (Joint Photographic Experts Group) – it is the file ending of images using the JPEG method of compression and is often used as a shorthand notation for files of this type.

LTO: An abbreviation for Linear Tape-Open – it is an open standard magnetic tape data storage technology.

MARC: A library standard format for the representation and communication of bibliographic and related metadata in machine-readable form.

MARCXML: A framework for working with *MARC* data in a XML environment.

MD: An abbreviation for metadata – data that describes other data or content.

METS: Metadata Encoding and Transmission Standard -- The *METS* schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium

MIMAS: A JISC and ESRC-supported national data centre providing the UK Higher Education, Further Education and research community with access to key data and information resources to support teaching, learning and research across a wide range of disciplines.

MINISIS: A commercial archive collection management software package.

MIX: An XML schema for a set of technical data elements required to manage digital image collections. The schema provides a format for interchange and/or storage of the data specified in the Data Dictionary - Technical Metadata for Digital Still Images (ANSI/NISO Z39.87-2006). This schema is currently referred to as "NISO Metadata for Images in XML (NISO MIX)."

MODES: A shorthand notation for *MODES* Catalogue System, which is an old cataloguing system designed for special collections and in use by several of the JISC digitisation projects.

MODS: The Metadata Object Description Schema – it is a schema for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications. It can carry selected data from existing *MARC* 21 records as well as enabling the creation of original resource description records.

MP3: An abbreviation for MPEG-1 Audio Layer 3 – it is a digital audio encoding format using a form of lossy data compression. It is a common audio format for consumer audio storage, as well as a de facto standard encoding for the transfer and playback of music on digital audio players. (It should not be confused with MPEG-3 which is a group of audio and video coding standards agreed upon by the Moving Picture Experts Group (MPEG) designed to handle high-definition television signals).

OCR: It is an abbreviation for optical character recognition, which is the recognition of printed or written text characters by a computer. The term *OCR* is often used to label the text files created through optical character recognition.

OUCS: An abbreviation for Oxford University Computing Services.

OULS: An abbreviation for Oxford University Library Services.

PCM: Pulse Code Modulation – it is the usual bitstream encoding format used for *WAV* files.

PNG: An abbreviation for Portable Network Graphics – it is a bitmapped graphics file format endorsed by the World Wide Web Consortium and is expected to eventually replace the *GIF* format. *PNG* provides advanced graphics features such as 48-bit colour, including an alpha channel, built-in gamma and colour correction, tight compression and the ability to display at one resolution and print at another.

Portfolio: See *Extensis Portfolio*.

PREMIS: An acronym used to represent the elements of the *PREMIS* Data Dictionary for Preservation Metadata.

PRINCE2: *PRINCE2* is a generic project management method that covers how to organize, manage and control projects.

QA: An abbreviation for quality assurance – it is often used as a shorthand notation for the staff who perform quality assurance on a project.

RLUK: An abbreviation for Research Libraries UK.

SPECTRUM: A UK and international standard for collections management – it is used by museums and other cultural heritage organizations. It includes a standard format for exchanging object records between different Collections Management Systems, support for rights management, and support for the exchange of User Generated Interpretation through the Revisiting Collections methodology.

TB: An abbreviation for terabyte – it is 1,000 gigabytes.

TEI: A shorthand notation for a set of guidelines created by the Text Encoding Initiative, which is a consortium that collectively develops and maintains a standard describing encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics.

textMD: A XML Schema that details technical metadata for text-based digital objects. It most commonly serves as an extension schema used within the *METS* administrative metadata section

TIFF: Tagged Image File Format (abbreviated *TIF* or *TIFF*) is a file format for storing images, including photographs and line art.

TMD: An abbreviation for technical metadata – it is metadata that describes the technical format of an object.

UKDA: An abbreviation for the UK Data Archive – it is a centre of expertise in data acquisition, preservation, dissemination and promotion; and is curator of the largest collection of digital data in the social sciences and humanities in the UK.

VITAL: A commercial institutional repository product from VTLIS and built on *Fedora*.

WAI: A shorthand notation for the best current practice for embedding accessibility roles and states in HTML documents as defined by the Web Accessibility Initiative (*WAI*) Protocols and Formats working group.

WAV: An acronym for the Waveform audio format (also abbreviated as *WAVE*) – it is a Microsoft and IBM audio file format standard for storing an audio bitstream.

Appendix A. Interview Questions

1) GENERAL	
1-1	Please give a brief definition of digital preservation
1-2	Is access or preservation the priority of this project?
2) ROLES	
2-1a	What is your role in this project?
2-1b	Please describe the skills required for your role
2-1c	Would you benefit from any specific training?
2-2a	What is your organization's role in this project?
2-2b	Do you consider your organization has the right skills for this project?
2-2c	What else is being done within your organization in relation to digital preservation? <i>(Other projects or policy?)</i>
2-3a	Are there multiple partners involved in this project?
2-3b	What are their roles?
2-4a	What are the key roles in this project?
2-4b	Who does each?
2-5a	Please describe the workflow of this project <i>(For each phase if necessary)</i>
2-5b	How do you know when an activity or process has been completed? <i>(Is there a system for the management of the process?)</i>
2-6	Does your organization support this project?
2-7	Who is responsible for the digital preservation?
2-8a	Who will be responsible for the management of the digital resources once this project is finished? <i>(Who will provide long term access?)</i>
2-8b	How was this agreed? <i>(Was it formalised in any way? Contract?)</i>
2-8c	Can they guarantee the future sustainability of the digital resources by providing the necessary infrastructure to deliver, manage and store them?
3) TECHNOLOGY	
3-1a	Who is responsible for digitizing the material? <i>(Is it done in-house or outsourced?)</i>
3-1b	What are the specifications of the contractor's role? <i>(Are they providing digitisation only or other services as well?)</i>
3-2	How are the digital resources being preserved? <i>(Master copies? Back Ups?)</i>

3-3a	How are the master copies and back-ups of the digital resources being stored? (<i>On a server?</i>)
3-3b	Where is the server located? (<i>Offsite? Partner organization?</i>)
3-3c	How do you obtain server space for back ups? (<i>Is there a separate server for back ups?</i>)
3-3d	If the material is being stored internally is there an internal charging policy for back up and server storage? (<i>Hidden costs? If stored externally, does extra space have an additional cost?</i>)
3-3e	What media is being used for storage? (<i>Formats? Tapes?</i>)
3-3f	How many copies of the digitized files are made? (<i>Back ups? Masters?</i>)
3-3g	Who is responsible for creating these copies?
3-4a	Are your content management system; preservation system; and delivery system, the same or separate? (<i>Software? Note: preservation system is unlikely to be in the same system</i>)
3-4b	What are the benefits and drawbacks of this?
3-5a	Are the platforms you are using for this project's content management system; preservation system; and delivery system, purchased or open source?
3-5b	Has it been necessary to modify these purchased open source platforms? (<i>In what way? Who was responsible?</i>)
3-5c	Have you built one or more platforms from scratch?
3-6	How is access provided to the digital resources? (<i>Are they available on-line?</i>)
3-7a	Have you considered how to make the digital resources accessible for the long term?
3-7b	How long will the digital resources be available?
3-7c	How will this be done?
3-7d	Who is responsible for this?
3-8	What plans or approaches will be used to ensure long term access to the digital resources?
3-9a	Are high resolution and delivery resolution files maintained for your material?
3-9b	Are both stored together or in separate locations? (<i>How far apart?</i>)
3-9c	If in separate locations, how are they associated to the delivery resolution files?
3-10	Do you have copies of the digital resources on another continent?
3-11	Does your project have full IT support?
3-12	Are you satisfied that all the technological requirements of this project have been met?
3-13a	What aspect of this project could be improved?
3-13b	What would you do differently next time?
4) POLICY	
4-1a	Please describe the material you are digitizing

4-1b	How is this material being digitized?
4-2	Does your organization have a digital preservation policy in existence?
4-3a	Does your approach fit into an organization-wide approach to digitisation?
4-3b	Has this approach been developed for this project only?
4-3c	What is your approach? (Migration? Emulation? Refreshing?)
4-3d	When do you plan to do this?
4-3e	Is this approach sufficient?
4-3f	Why did you use this particular approach?
4-4a	Does this project fit into the ethos of your organization? <i>(Character? Disposition?)</i>
4-4b	Under which strategic policy of your organization does this project fall? <i>(Policies concerning collections? e.g. Is enhancing the access to collections driving this project?)</i>
4-5a	Do you have a preservation strategy or sustainability model for this project?
4-5b	When was this created?
4-5c	Was it created specifically for this project or for a different reason?
4-5d	If created specifically, how did it develop?
4-5e	What event triggered it?
4-5f	Who was responsible for the development? <i>(Which partner if multi-partner project?)</i>
4-5g	Does it link into another strategy?
4-5h	If the preservation strategy has been re-used, was this practical?
4-6	At what stage in the project did the preservation strategy evolve?
4-7a	Where is the preservation strategy located? <i>(Documentation on Internet?)</i>
4-7b	Do others know where to find it?
4-7c	How was it disseminated? <i>(How was the preservation strategy promoted and publicized? e.g. Intranet? Newsletter?)</i>
4-8a	If you don't have a sustainability model for long-term preservation, what do you have instead?
4-8b	Is this sufficient?
4-8c	How did you decide on this?
4-9a	Has the preservation strategy changed over the course of the project?
4-9b	Were these changes outlined in the original project plan?
4-10	Has it been necessary to react to risks identified during the project? (Staff losses?)
5) METADATA	
5-1	What kind of metadata have you collected for each file?
5-2	What is the purpose of the metadata sets you are collecting? <i>(Access or preservation?)</i>

5-3	Did you use a metadata standard? (<i>Which one?</i>)
5-4	Who collects the metadata?
5-5	What is the procedure for collecting metadata?
5-6	How do you know when it has been collected? (<i>Does your content management system have a specific process? Will there be an empty field?</i>)
5-7	Where is the metadata in relation to the digital resource? (<i>In the content management system?</i>)
5-8	Is this metadata useful for preservation? (<i>Why?</i>)
5-9	Please describe how preservation metadata and access metadata differ
5-10	How much crossover is there between access metadata and preservation metadata, in this project?
5-11	Which metadata formats do you use? (<i>NLM, MODS, MARC, METS, etc</i>)
6) CONTENT AND FORMATS	
6-1a	How many objects are currently in the collection?
6-1b	How many objects will be in the collection by the end of this project?
6-2a	Currently, how many GB is the collection?
6-2b	Does this include high resolution master files, and low resolution delivery files?
6-3	Please estimate what percentages of the collection are image files, audio files, video files, text files etc
6-4	Do you use different formats, metadata standards and preservation approaches for the different digital resources in your collection?
6-5	How did you decide which formats to use?
6-6a	Will content be added to your collection indefinitely?
6-6b	Is there a date after which no new content will be added?
6-7a	Will the digital resources in your collection be updated?
6-7b	How frequently?
6-7c	Will the metadata be updated, or the content files?
6-7d	Will content be deleted from your collection?
6-7e	How are the updates and deletions tracked?
7) RISKS	
7-1a	What risks have you encountered during the course of this project?
7-1b	How did you respond?
7-2a	Are there any potential risks to the sustainability of your digital resources, from within your organization or externally?
7-2b	Have these risks been documented?
8) RIGHTS	
8-1	Who holds the rights to your newly digitized material? (Copyright? User rights?)

	Licenses?)
8-2a	Do you charge for access to your material?
8-2b	Will this change over time?
9) POST PROJECT	
9-1	Please discuss any post project plans