

Assessing Preservation

University of London Computer Centre

This case note has been created to illustrate digital preservation practices within digitization projects, sharing honestly the challenges of establishing digital continuity from projects with a limited lifespan. It is selected from the JISC Digitization Programme which ran between 2004 and 2009. This programme has provided digital access to collections of international significance which would otherwise be inaccessible. In doing so it has also generated an opportunity to learn about the critical success factors in digitization, such as ensuring long term access from short term projects. This case study looks at the methodology the University of London Computer Centre used to assess preservation planning within the programme. It therefore provides a simple methodology for assessing preservation readiness for any digitization project.

Introduction

The Digital Archives and Repositories Team of University of London Computer Centre (ULCC) has a long history of supporting and training staff in digital preservation. Until recently it was responsible for the long term preservation of significant quantities of government data and it runs a popular and successful Digital Preservation Training Programme (<http://www.dptp.org/>).

In 2009, ULCC was asked to work with 16 JISC-funded digitization projects to help them understand and address the challenges of long term preservation for the variety of materials produced by each project. Preservation planning for digitized collections has historically been neglected so it was recognized early on that the process could be of wider advantage. As is often the case, the 16 digitization projects were primarily designed to increase access and, although preservation of digital outputs had been clearly stipulated in the grant, it was going to be a new aspect of the process for many of these projects.

Assessing Digital Processes

At the core of our analysis of the digital preservation plans was a survey and assessment of the various projects and their digital collections. The very diverse nature of the projects and the varied levels of experience with preservation caused us to start with very few assumptions about existing knowledge of preservation. It was also clear that a face-to-face approach would help both parties. Consequently, we undertook sixteen

Digital preservation represented as a three legged stool supported by technology, resources and organization. (Kenny and McGovern 2003).



interviews in the institutions where the collections were held. This allowed a much greater degree of latitude in conversation than would have been possible by other means and it meant the interviewees could show us how they worked. Interviewees were assured about the privacy of their answers, were provided written copies of their answers to check and correct mistakes before they were analyzed.

The project team took as their starting point a straightforward model of digital preservation originally developed by Nancy McGovern and Anne Kenney of Cornell University and now based at the University of Michigan in which robust preservation is a 'three legged stool'. Each leg of the stool - technology, resources and

Key Message

Appropriate and timely examination of the digital preservation plans of digitization projects can have a lasting impact. Projects may not know or understand the risks they run. Simple assessment can help them identify and address these risks sooner rather than later.

organization – needs to be stable and reliable. Weaknesses or shortcomings in any one of these three areas are disruptive to the other two.

Although there are now several survey tools designed to test the robustness of digital preservation facilities, to assess the value of digital collections, or to assess institutional readiness, none of these seemed to fit with the precise task in hand of assessing whether digitization projects were taking appropriate steps to ensure long term access.

Survey

The interviewees responded to questions in eight different areas: roles; technology; policy; metadata; formats; risk management; rights; and post-project planning. Interviewees were provided with copies of the questions in advance in order that they could consider their answers. Although each interview proceeded differently, each one addressed the same key topics.

Roles

We asked who led the project and who was responsible for digital preservation in the long term:

- What is your role in the project?
- If this is a multi-partner project, can you explain the different roles of the partners?
- Who is responsible for digital preservation?
- What is your organization's role in the project?

Technology

Here we considered the technical aspects of the project, how digital objects were captured and by whom. For many, the capture processes were outsourced so it was important to understand the workflow between partners and to highlight strengths and weaknesses.

- Where is your material being digitized?
- Who does it? In-house or outsourced?
- What kind of agreement do you have between the partners in the project?
- What exactly does the contractor do?
- How and when do digitized outputs change hands?

Policy

The policy section was where the most direct questions were asked about digital preservation within the context of the organization and the project. This section turned out to be the most challenging for some of the participants, perhaps because there was no common idea of what digital preservation meant. It was very useful to find out why and how a digital preservation strategy evolved and whether the project has helped shape it. These are precisely the questions which need to be addressed when designing digitization projects.

Interviewees responded to questions in eight different areas: roles, technology, policy, metadata, formats, risk management, rights and post-project planning

- Does your approach to digital preservation form part of an institution-wide approach or is it special to this project only?
- Does this approach work?
- Is there a digital preservation policy in existence?
- In your opinion is the digital preservation policy likely to succeed?
- Why have you taken this approach? Why have you ruled out other approaches?
- Has this project fitted easily into a greater institutional structure of any kind?
- Does this project fall within a strategic policy or vision for the institution?

Metadata

Questions about metadata gathered information about what and how metadata was captured. It was also important to understand the content management system in place for the project and for ultimate delivery to users. Metadata decisions made at the beginning of a digitization project affect the lifespan and longevity of a digitized collection. Good metadata capture in digitization greatly assists the provision of metadata for long term access and preservation.

- What kind of metadata do you collect for each file?
- Did you use a standard for the metadata?
- Who creates metadata?
- How is metadata created?
- How do you check or validate metadata?
- Where is this metadata stored in relation to the original file?

Formats

Different file formats bring with them advantages for capture, preservation and access. Questions were designed to promote clear thinking about these different functions as this is likely to encourage success in the long and short terms.

- What types of digital objects are you working with?
- What is your approach to the different formats that you generate?

- How do your preferred formats, standards and preservation approaches work together?
- Have third parties or suppliers assisted you in any way with digital preservation?
- Has a digitization supplier undertaken any digital preservation work on your behalf?

Risks

An important aspect of the survey was to encourage projects to consider and plan for risks that might emerge in the short and medium term.

- What risks do you perceive to the sustainability of your material?
- At what point in the life of your data are these risks likely to emerge?
- How will these risks be managed and by whom?

Rights

A number of questions elicited information about copyright issues for access over time. Ascertaining who holds the rights to the newly digitized material was very important in order to establish who would be able to repair or restore data that might be lost.

- Who holds the rights to your newly digitized material?
- How do you manage the many different types of right? (Copyright? User rights? Licenses to distribute, replicate or repair?)

Post project plans

Finally interviewees were asked to elucidate their long term plans and prospects for the outputs from their project. Projects are, by definition, mismatched with the long term requirements of digital preservation so it was useful to discover what plans if any existed for the material once project funding had ended. It was also a good opportunity to find out about the roles of the project team and if the expertise built up during the project was going to be retained or dispersed and if so whether any steps were being taken to capture implicit knowledge from the project team.

Prioritizing questions

The survey was drafted in order to make a clear distinction between what was essential to know, what would be useful to know and what would be unnecessary. For example it was essential to know how effectively a host institution supported the project and interesting to know what skills a project team brought with them. Although not directly relevant to the purpose of the survey, we learned a lot about the skills that project teams brought to the programme. Information like this in turned proved to be useful in assessing the longer term prospects of the data created in the



projects. If an expert team was working in isolation from their host institution, and if the host institution had only limited in-house capacity to manage digital outputs, then there was a clear risk to digitized collections which would most likely only emerge after it was too late to be fixed.

A rounded and realistic picture of digital preservation could only emerge if contextual matters were examined and properly considered.

Providing a common understanding

Surveying projects about their long term plans allowed a comprehensive view of the long term prospects of digital outputs from the whole programme and a series of individual reports for each project in turn. The assessment used practices common to survey work to speed things along. All the questions were open ended and we tried to avoid technical terms and acronyms, although the respondents knew well the subject covered. We tried to keep the questions short and simple. It was important not to assume that everyone had the same understanding of the facts or a common basis of knowledge. So, in addition to an explanation of commonly used abbreviations and jargon, each survey opened with an opportunity for each project to explain what they understood by digital preservation and why it would be advantageous for their projects. This way we could be sure that we understood the matter at hand in terms that the projects also understood.

This process was time consuming but rewarding. Face-to-face interviews with project teams produced a wealth of information for the project to analyze. Where possible, the interviews were done by two staff as this would allow the interview to move more speedily and with greater time for reflection and dialogue. In most cases the interviews acted as an encouragement to further discussion and dialogue. This was almost always beneficial but if left unchecked it could cause over-runs in the assessment.

The survey provided the basis for some specific and practical changes within projects which enhanced their long term prospects and ensured that the programme would provide more long-term impact too.

Reflections

On the whole, projects responded openly and enthusiastically to our survey. There seemed to be a willingness and desire to share honestly the highs and lows of their projects. More importantly, the survey provided the basis for some specific and practical changes within projects which enhanced their long term prospects and ensured that the programme would provide more long-term impact too.

On completion of the survey, each project was challenged to supply a sample of metadata and a sample output from their digitization activities. All three were then analyzed to provide a simple report card for the project that included specific actions necessary to improve the long term viability of the data produced.

In many cases the projects were able to provide thoughtful, credible and comprehensive answers about their long term plans. Key lessons have been packaged as and presented as case studies in digitization.

The analysis also identified five projects which were at risk such that if prompt action were not taken, the content or associated services might not be accessible and the value of the investment lost.

Some of the projects were struggling with issues which are not well understood in the digital preservation community in general. Most of the 'at risk' projects demonstrated a willingness to improve and, because the survey was carried out before the end of the projects, there was in most cases some latitude to deal with questions raised. The initial version of the report prompted one project to rethink radically its long term plans. It has subsequently succeeded in mitigating the

risks which were identified, without great additional expenditure.

The following lessons were derived from the survey which are true of all digitization projects:

- 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.

Because the survey was carried out before the end of the projects there was still latitude for dealing with questions raised. The initial report prompted one project to rethink radically its long term plans.

Anne R Kenney and Nancy Y McGovern, 2003 "The Five Organizational Stages of Digital Preservation," in P Hodges M Sandler M Bonn and WJ Pirce-Wilkin (eds) *Digital Libraries: A Vision for the Twenty-first Century, a festschrift to honor Wendy Lougee*, University of Michigan Library, Ann Arbor

For a complete account of the JISC Digitization Programme Digital Preservation Project and access to the project report see:
<http://www.dpconline.org/advocacy/knowledge-base/594-digitisation-programme-digital-preservation-study>

This case study was prepared by **Patricia Sleeman** of the **University of London Computing Centre**, with the assistance of **Portico** and the **Digital Preservation Coalition**. It was made possible with funding from **JISC** and the support of **projects in the JISC Digitization Programme**. Images courtesy of Nancy McGovern and ULCC. November 2010

