

Building the Business Case for Digital Preservation

Neil Grindley

JISC

Abstract

This paper will describe work that JISC is funding to build an evidence-base of material that will provide a range of organisations with practical and plausible reasons for investing in digital preservation. The SPRUCE Project is leading on the work and is using hackathon-type events and other enabling measures such as disbursing small grants and providing practical technical advice, in order to better engage with stakeholders and elicit accounts of the real challenges that digital collection-holders face on a day-to-day basis. Complementary JISC-funded work in areas such as 'sustainability', 'value' and 'costs' will allow the SPRUCE project to build a nuanced and multi-faceted series of business cases. Case-studies arising from this project will be valuable in their own right, but the genuinely novel aspect of this work will be the assembly of methods used to arrive at convincing and usable arguments to support engagement with, and investment in, digital preservation.

Author

Neil is a Programme Manager at JISC with responsibility for digital preservation. JISC is an organization that funds and supports technology-related projects and services for the UK Higher and Further Education sector and is influential within and beyond the UK as an innovative agent of change. Neil is also currently a board member for: the Digital Preservation Coalition; the Open Planets Foundation; and the Alliance for Permanent Access. Previously, Neil has worked on projects supporting the use of advanced ICT methods for humanities research and was the IT Manager at the Courtauld Institute of Art.

1. Building the Business Case for Digital Preservation

1.1 The Problem

Based on a great many national and international activities over the last ten or fifteen years, there are reasons to suppose that digital preservation is now a well-described problem capable of drawing on a useful range of resources and tools to support practice. It is also apparent that a number of different international organisations and agencies have made a significant impact with helping practitioners across various sectors tackle the challenges inherent in the long-term management of digital information.

However, whilst much has been achieved, a gap and a challenge persist. Whilst the vast majority of organisations are now confident and willing to acknowledge – through a mixture of intuition and logic - that long-term access to (and therefore preservation of) digital materials is an area where thinking, strategy, and perhaps even policy is going to be increasingly vital, insufficient evidence currently exists to help them fully articulate what the nature, scale and scope of the threat to their digital assets is likely to

be. To put it another way, they don't yet have the means or experience to accurately gauge the *value* that digital assets represent to their organisation, and consequently don't know to what extent digital preservation will serve the long-term financial and strategic interests of the organisation. The question thus becomes, 'what is the business case for digital preservation?'

JISC has a remit to help UK universities and colleges ensure that they are managing their digital assets in the most effective way possible and part of this task is to support practitioners to be able to make the case (either to internal decision-makers or external funders) for ongoing investment into digital preservation activities. JISC has therefore recently funded a collection of projects with the objective of building an evidence base that will support practitioners to make that case. The main purpose of this paper is to describe the assembly of methods by which the empirical evidence is being gathered and a useful by-product is to list some of the stated impacts that the projects have identified, which can in turn be assessed for the benefits that have accrued to the organisations involved.

1.2 Obstacles and Methodology

Any considered process to gather information in support of the business case for preservation meets several early obstacles. Firstly, it is clear that organisations will normally be reluctant to freely admit any failures to properly manage their digital assets. Reputation is an important currency for all organisations, public or private, but where the organisation has a remit to act on behalf of the public, and more especially where it has a trusted role around the stewardship of objects, reputation is a very important issue. So it will be difficult to identify and then subsequently disseminate stories where identifiable individuals in actual organisations have made inadequate attempts to manage their digital objects.

Secondly, it is difficult to gather information in the form of case studies and data from people with the simple proviso that the 'community needs them', or that it will ultimately be of benefit to the whole community that an evidence-base has been amassed. In the main, people need more motivation and more evidence of the near-term benefits to them to really engage with a process such as this.

Thirdly, and rather fundamentally, the fact that the digital preservation community is alert to the dangers of ignoring digital preservation does not mean that the broader community is necessarily convinced either by its requirement as an activity; or by the terminology and methodology proposed by those advocating digital preservation. This limits the potential sources of information from which the evidence base can be built.

To try and tackle these and other obstacles, JISC issued a call for proposals in September 2011¹ that invited ideas for 3 different types of project. The first strand called for a single project to 'inspire, guide, support and enable UK HEI's [and potentially other types of organisations] to address preservation gaps; and to use the knowledge gathered from that support work to articulate a compelling business case for digital preservation'. A mass of anecdotal evidence suggests that professionals working within organisations do not generally want to fill in surveys, or answer questions, or generally meditate *about* preservation. What they really want to do is get on and do something practical about their problems, ideally with recourse to accessible and expert support (on-site if possible) enabling them to tackle some of the specific problems that they are facing in their own institutions relating to the long-term management of information.

¹ "JISC Grant Funding 12/11: Digital Infrastructure Portfolio", last modified August 22, 2011, http://www.jisc.ac.uk/fundingopportunities/funding_calls/2011/07/grant12_11.aspx

The objective, therefore, was to setup a support and enabling project that *actively* engaged and helped stakeholders with problems. This would in turn implicitly facilitate the exchange of information necessary to build the business case evidence base. In effect, the provision of case studies, use case information and stated requirements from stakeholders becomes the *quid pro quo* for the support and assistance given by the project to them. To further enhance the project's chances of getting engagement from stakeholders, a proportion of the grant (the suggested amount was 20%, i.e. £50,000 from the total award offered of £250,000) was designated as an enabling fund from which the project could itself make small awards of up to £5,000 to institutions that wanted to pursue preservation objectives that might usefully serve as exemplars to the rest of the community.

The enabling grants added a fourth and potentially rich layer for the project to gather appropriate material to use in construction of the business cases. Going from the least to the most active engagement the project can gather data via the following routes (or as a by-product of the following activities):

1. Desk research and literature reviews
2. Interviews, discussions and questionnaires (this activity was not ruled out as a strategy)
3. Active support, advice and guidance to organisations with preservation challenges
4. Enabling grants to cover costs of preservation development activities

The second strand of activity in the JISC Call invited proposals for short projects to address a specified digital preservation problem and to do so within a period of 4 months (Nov 2011 – Feb 2012). The purpose was to set number of projects running that would provide the principle support and enabling project with a small existing stakeholder group which it could work with. It was also meant to ensure that early case study material emerged for inclusion into the evidence base, which would test assumptions around what sort of activity provided the most useful evidence, what kind of output and data would be required, and how effective the whole perceived mechanism would be for achieving the objectives of the call. A total of £75,000 was available to fund between 5-7 projects, attracting from between £10,000-£15,000 each.

In addition to helping the support project, these '*Active Case Studies*', were an attempt to navigate around the first of the three obstacles defined above as hindering the gathering of evidence. Rather than just asking an organisation to tell everyone else about a digital preservation problem that they were facing (a passive case study), this measure also offered them a means to specify and implement a solution to their problem. Giving an organisation the means to fix and own the problem they have admitted to is a different proposition from simply asking them to announce that they have a problem.

The third strand of work was less focused on building the business case for preservation but was nonetheless an indirect attempt to address the third of the three obstacles. The purpose of this strand was to enhance the capability of organisations to more effectively include digital preservation as a component within the information management training and development courses and sessions that they organise for staff and students. This would in effect start to disseminate at least a basic level of preservation knowledge to more people within institutions and over time should result in higher levels of acceptance from more people about the principles, and perhaps the purpose (the business case even) of digital preservation. The call text makes the principle even more explicit,

Part of any embedding process for preservation should involve making appropriate aspects of it accessible and relevant to a range of people across the institution. By

drawing out some of the more universally applicable issues around technology obsolescence, media refreshment, bit rot, long-term storage costs and digital forensics, it may be possible to positively influence the behaviour of staff and students within HEI's to not only store their data more effectively, but to actively *manage* it, and therefore become more actively engaged with the value of that data and what it may or may not enable them to do.²

Given the level of funding (£100,000 for 3-4 projects of 9 months duration) this is clearly a limited contribution to what could potentially be a very large change programme, and one which would need a long timescale to measure demonstrable impact. It is however, an attempt to address part of the problem and it also again mentions the word 'value' and expresses the need to convey to stakeholders the idea that part of the challenge of purposefully engaging with preservation is understanding and acknowledging how much *value* digital materials have. This concept features in other areas of work that JISC has recently commissioned and some of this work will be referenced towards the end of the paper – particularly where it has direct relevance to building the business case. However, 12 months on from the invitation of proposals, it is now possible to examine and assess some of the activities that resulted from the call.²

2. Projects

The enabling and support project was taken on by a consortium of organisations led by the University of Leeds and is called the SPRUCE Project (Sustainable Preservation Using Community Engagement).³ The principle idea behind SPRUCE continues on from an earlier JISC-funded project called AQUA (Automated Quality Assurance)⁴ which succeeded in defining and running events that brought together digital content specialists and technical developers to tackle real problems in real time (over the course of a 3 day event). These 'people mashups' allowed the content specialists to bring problems and questions to the event along with problematic or test data, and gave them a chance to sit down with a developer and work on solutions and ideas. Along with plenary discussion and presentation sessions and documentation that was captured as the event progressed on a dedicated wiki, these events received good feedback and more importantly, produced plentiful evidence that this was an effective way to accelerate the rate at which real problems encountered by those with responsibility for digital collections might be examined and tackled by technical digital preservation practitioners.

Over the course of just 2 events, the AQUA project produced descriptions of work on 24 different solutions to problems in seven categories.⁵ Admittedly these solutions were in different states of usability and functionality but it was clear that as an open and shared community activity and resource, there was great mileage in this form of cooperative problem-solving. Thus, the same approach was proposed for the SPRUCE project and the engagement required of stakeholders to build the business case was designed

² "JISC Grant Funding 12/11: Digital Infrastructure Portfolio", last modified August 22, 2011, http://www.jisc.ac.uk/fundingopportunities/funding_calls/2011/07/grant12_11.aspx (p.27)

³ "SPRUCE Awards - funding opportunity for digital preservation", last modified June 12, 2012, <http://wiki.opf-labs.org/display/SPR/SPRUCE+Awards+-+funding+opportunity+for+digital+preservation>

⁴ "AQuA", last modified July 16, 2012, <http://wiki.opf-labs.org/display/AQuA/Home>

⁵ "AQuA Solutions", last modified July 11, 2011, <http://wiki.opf-labs.org/display/AQuA/Solutions>

around further mashup-type events. At the time of writing SPRUCE has run one further event and has added further information about datasets, issues and solutions to the wiki page.⁶

What has become most obvious as a result of the AQUA and SPRUCE project outputs is that they are even more valuable as part of a larger community effort to align similar activities that are being instigated and managed by other entities such as the Open Planets Foundation (OPF),⁷ the SCAPE project,⁸ and the Digital Preservation Coalition.⁹ The information that has emerged from the JISC-funded projects forms part of a larger body of material that has been brought together under the headings of ‘datasets’ (11 categories/60 items), ‘issues’ (19 categories/159 items), and ‘solutions’ (14 categories/80 items), all of which represents a rich treasure-trove of practical work that either practically supplies a working tool to fix a problem or issue, or at the very least flags up the work-in-progress thought processes that have been applied to a challenge, and may indicate directions for future work.

The URL is on a different section of the OPF wiki from the SPRUCE pages. See: <http://wiki.opf-labs.org/display/REQ/Digital+Preservation+and+Data+Curation+Requirements+and+Solutions>

Whilst not yet specifically addressing the detailed objectives of the JISC call to collate and define business cases for digital preservation, it is clear that a great deal of information from a wide variety of sources is being marshalled to declare the type, frequency and scale of problems that organisations are facing in relation to the types of information management problems that digital preservation techniques are designed to tackle. The SPRUCE project is less than half way through its 2 year duration and is due to finish at the end of October 2013.

In the second strand of work, five Active Case Study projects were funded and constitute a diverse range of activities and tasks, all of which have different perspectives that will usefully build the evidence base to justify digital preservation.

2.1 *The Carcanet Email Preservation Project* aimed to capture and preserve an important email archive and represented a good opportunity for the John Rylands Library at the University of Manchester to begin to engage with complex and emerging practice around a critical area of collections management.¹⁰ The final report is instructive in a number of ways and a good account of the various lessons that can be extracted from a short low-budget project. Sections 3.4 and 3.5 (p.18) are relevant for the evidence base.

2.1.2 Immediate Impact [of the project]

- For the staff at Carcanet Press, significant progress has been made towards preserving their digital archive, and they are likely to have increased confidence in the Library’s ability to deal with further digital accessions in future.
- The project has ensured the ‘rescue’ for the archival record of a large body of research-rich material which was formerly at risk of loss; several of the files acquired resided only on the hard drives of Carcanet staff.

⁶ “Digital Preservation Requirements and Solutions”, last modified May 8, 2012, <http://wiki.opf-labs.org/display/SPR/Digital+Preservation+Requirements+and+Solutions>

⁷ “Open Planets Foundation: a community hub for digital preservation, accessed September 5, 2012, <http://www.openplanetsfoundation.org/>

⁸ “SCAPE: Scalable Preservation Environments, accessed September 5, 2012, <http://www.scape-project.eu/>

⁹ “Digital Preservation Coalition”, accessed September 5, 2012, <http://www.dpconline.org/>

¹⁰ “Carcanet Email Preservation Project, Final report”, accessed September 5, 2012, <http://www.jisc.ac.uk/whatwedo/programmes/preservation/carcanet.aspx>

- The project marked the Library's first acquisition of a substantial 'born-digital' archive, thus enhancing its ability to continue collecting modern archives in the digital age.
- It has improved the confidence of Library staff in their ability to acquire and preserve born-digital archive material, marking a vital first step in practical digital preservation.
- The project has contributed to the advancement of strategic goals, both for the Special Collections Division and for the Library-wide Digital Preservation Steering Group.

2.1.3 Future Impact

- The project has resulted in a large digital archive which, although currently embargoed, will form a key resource for future researchers – not just those working in the field of literary studies, but also for historians, sociologists and others.
- The ability to deal with born-digital archives has the potential to enhance the UML's reputation as a collecting institution at an international level.
- The University's central IT Services have expressed interest in preserving institutional emails, and may draw on the experiences and outcomes of the project.
- The project will contribute to the growing dialogue about issues involved in the long-term preservation of email.

2.2 *The Digital Directorate Project* based at the Institute of Education explored existing digital records management practice within the central administrative function of the organisation (the section producing most of the high-level governance and strategic planning records). The project created relevant documentation including a detailed retention schedule, an overview of relevant preservation metadata and procedures for the appraisal of electronic records. A final report is available.¹¹

The immediate internal impact of the project can be summarised as follows (p.8 Final Report):

- It enabled the records management staff to establish new partnerships with key directorate staff (a highly influential group within the organisation)
- It clarified the organisation's internal committee structure which in turn improved information retention decisions
- It opened a useful channel of communication between the records management staff and IT staff in relation to secure storage
- It allowed project staff to become more familiar with the effective use of preservation metadata and the need to adapt standards to meet internal requirements
- It afforded a chance for staff to develop procedures for the appraisal of electronic records, especially in relation to the acquisition of personal electronic archives

Future impact is stated as follows (p.9 Final Report)

¹¹ "The Digital Directorate: Digital preservation of governance records at the Institute of Education- Final Report", accessed September 5, 2012, <http://www.jisc.ac.uk/whatwedo/programmes/preservation/DigitalDirectorate.aspx>

- A better understanding of internal committees will allow a substantial programme of de-accessioning to take place. This will be a large undertaking but will result in a streamlined historical archive and more space
- A much larger programme will now be rolled out across other departments in the organisation based on templates and documentation developed in this short project
- It will have a long-term and positive effect on the way the organisation uses metadata and how it tackles accession and cataloguing procedures
- It will inform a forthcoming project in collaboration with the IT department looking at a permanent solution to the secure storage of semi-current electronic records

It should also be noted that Digital Directorate project was awarded an enabling grant from the SPRUCE project to continue aspects of its work. (Details of all grants awarded to date at: <http://openplanetsfoundation.org/blogs/2012-06-08-spruce-makes-funding-awards-digital-preservation>).

2.3 *The Future Proofing Project* was a collaboration between the University of London Computer Centre (ULCC) and the University of London (Archive).¹² It worked with a simple toolkit of services and software and succeeded in plugging into a network drive to create preservation copies of core business documents that required permanent preservation. It purposefully set out to be a relatively simple intervention that made use of open source migration and validation tools. The case study sought to demonstrate the viability of the approach.

The report is a valuable and clear description of the work, which should be of broad interest to a wide variety of people working in IT environments that are expedient rather than optimal. The impact of the project was as follows (p.34 Final Report):

- The beginnings of a permanent record store
- The creation of a standard University of London position on digital preservation
- A holistic approach that works for many document/record types
- Costs savings, as storage is reduced
- Demonstrating the viability of an alternative approach to EDRM (electronic documents and records management)
- Good for disaster recovery and business continuity - puts set of core documents into a more portable safer place
- Shared drives not abolished, just managed better

2.4 *The Preservation of Publications in Education* (POPE) project¹³ aimed to reverse the effects of link rot on around 5,000 documents that made up the Digital Education Resource Archive (DERA).

¹² “Future Proofing: enabling practical preservation of born-digital records - Final Report”, accessed September 5, 2012, <http://www.jisc.ac.uk/whatwedo/programmes/preservation/futureproofing.aspx>

¹³ “Preservation of Publications in Education (POPE) – Final Report”, accessed September 5, 2012, <http://www.jisc.ac.uk/whatwedo/programmes/preservation/pope.aspx>

The final report outlines the following future steps will be taken as a result of the project work:

- A follow-on project has been specified to OCR image files representing text to render them machine-readable
- Work is being proposed to re-index the preserved publications using the London Education Thesaurus (LET) which would allow it to sit alongside content from other databases which had been classified under the same term
- A Preservation Strategy will be developed and plan for the IOE Library and Archives which will include ensuring that the material which has been preserved in POPE remains viable in the long-term
- The IOE Library has undertaken to ensure that all new publications which are accessioned to the library will also be preserved (where licences allow) in DERA. This is only possible because we have saved the time we spent fixing broken links and can now use it more profitably.

2.5 *The Publishing Online to Preserve Scholarship* (POPS) project¹⁴ grew out of a need for low-cost online journal publishing. Academic staff indicated that they wanted to be able to create online, Open-Access journals and sought advice from Learning and Information Services to see if a local solution was feasible.

Whilst digital preservation was not the principle objective of this project, it nonetheless represented a good opportunity to examine the aspects related to the sustainability of scholarly materials. The use of an ePrints repository platform brings material into an environment that has the potential to facilitate digital preservation and plug into relevant tools and methods. (See the JISC-funded Preserv project).¹⁵

Accounts of the third *Enhancing Capability* strand of work can be found in the four project blogs.

- DataSafe (University of Bristol) - <http://datasafe.blogs.ilrt.org/>
- SHARD (Institute of Historical Research & University of London Computer Centre) - <http://shard-jisc.blogspot.co.uk/>
- PrePARE (Cambridge University) - <http://preparecambridge.wordpress.com/>
- DICE (London School of Economics) - <http://lsedice.wordpress.com/>

3. Complementary Work

Building the business case for preservation is a complex proposition and there are no easily definable boundaries that can be placed around the topic. One of the critical factors that need to be addressed in any consideration of business cases is the concept of sustainability and how to design strategies that lend assurance and permanence to an enterprise. One of the strong themes to emerge from work that JISC has collaboratively supported on the topic of sustainability – both in terms of the economic sustainability of digital preservation and access,¹⁶ and on the sustainability of projects and project outputs¹⁷ - is that it will

¹⁴ “Publishing Online to Preserve Scholarship (POPS) – Final Report”, accessed September 5, 2012, <http://www.jisc.ac.uk/whatwedo/programmes/preservation/pope.aspx>

¹⁵ “Preserv2/EPrints Preservation Plugins 1”, accessed September 5, 2012, <http://files.eprints.org/422/>

¹⁶ This relates to work carried out by “The Blue Ribbon Task Force for Sustainable Digital Preservation and Access”, accessed September 5, 2012, <http://brtf.sdsc.edu/>

be difficult to design any long-term future for digital assets and services if their value proposition is uncertain or ignored. The concept of *value* is significantly linked to notions of ‘cost’ and ‘return on investment’ but also has overlaps and connections with other concepts such as benefits, impact, risk, efficiency, etc.

Another very important factor that is difficult to ignore when tackling this topic are the roles and responsibilities of those implicated in digital preservation, i.e. who needs to take ownership of the decisions that are required to manage and sustain digital assets over time, or to fund and maintain the solutions and services that are used for those management processes. An initiative that has tried to think through the relationship between some of these factors (mainly from the point of view of examining prospects for sustainability) is the Economic Sustainability Reference Model. This is a work-in-progress which is seeking to get broad community validation for a framework that will facilitate more thoughtful and sustainable design of measures to ensure long-term preservation and access.¹⁸

Further work on the reference model, on enhancing cost models for digital preservation, and on the other cost-determinants of digital preservation (e.g. risk, impact, benefits etc.) is planned as part of a forthcoming European Commission funded project that is in the last stages of negotiation. This proposal should allow for intensive work to be carried out over a 24 month period to really help progress a raft of concepts and issues in this field and to join up the effort that a number of projects (including APARSEN¹⁹, TIMBUS²⁰, ENSURE²¹) are already expending on different related parts of the conceptual jigsaw.

4. Conclusion

Building an effective evidence base on which to create convincing business cases for digital preservation will necessitate active collaboration with the community. This will best be achieved by a variety of different approaches spearheaded by an expert engagement project or programme that will lead and inspire people to contribute their stories and evidence. The enabling actions that occur in the process of collecting the evidence to illustrate the value of digital preservation may only be supported by modest financial awards of short duration but they appear to foster a desire within recipient organisations to implement lessons learnt more widely and to embrace more methodical policies and processes for information management across the organisation.

¹⁷ This relates to studies carried out by Ithaka S+R in association with the JISC-led Strategic Content Alliance, accessed September 5, 2012, <http://sca.jiscinvolve.org/wp/category/sustainability>

¹⁸ An early account of activity is available from the “Unsustainable Ideas” blog – accessed September 5, 2012, <http://unsustainableideas.wordpress.com/economic-sustainability-ref-model-page/>

¹⁹ “APARSEN - Alliance for Permanent Access to the Records of Science of Europe Network”, accessed September 5, 2012, <http://www.alliancepermanentaccess.org/index.php/aparsen/>

²⁰ “TIMBUS - Timeless Business Processes and Services”, accessed September 5, 2012, <http://timbusproject.net/>

²¹ “ENSURE - Enabling kNowledge Sustainability Usability and Recovery for Economic value”, accessed September 5, 2012, <http://ensure-fp7-plone.fe.up.pt/site>