The Costs and Economics of Preservation

Neil Grindley
JISC
Brettenham House, 5 Lancaster Place
London, WC2E 7EN
+44(0)2030066059
n.grindley@jisc.ac.uk

ABSTRACT
Given that preservation is now a fairly well-described problem, it should, in theory, be possible to calculate with a reasonable degree of accuracy what costs are likely to accrue to an organisation that has responsibility for the long-term stewardship of digital assets. This paper will introduce and describe some of the work that has been carried out over the last 5 years to help institutions and research groups to understand both the cost and the economics of preservation, and to examine the difference between those concepts. It will also describe ongoing phases of work that are being funded in the UK by JISC that are attempting to further advance understanding in this area and where possible apply or implement previously theoretical approaches. Some indication will also be given as to where collective international effort may be of universal benefit.

Keywords
Preservation, costs, economics, models

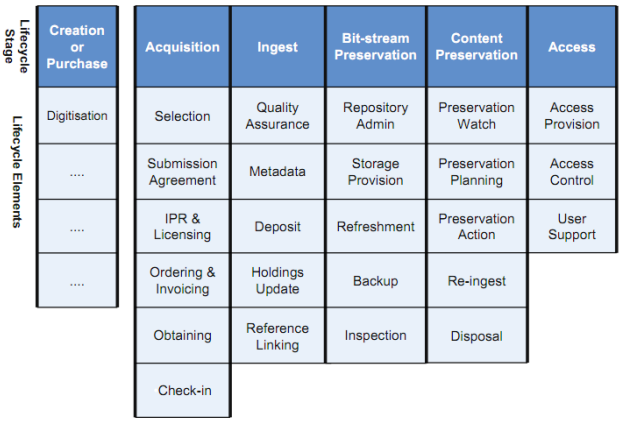
1. INTRODUCTION
In the last five years, some groundbreaking work has been done relating to the costs and economics of digital preservation. The LIFE project[1] undertaken by University College London and the British Library devised and refined a lifecycle costing model for digital objects which incorporates a generic preservation cost component and a costing tool. The Keeping Research Data Safe (KRDS) project[2] examined this same issue but specifically with a focus on the long-term management of research data. In the US (with some UK involvement) the Blue Ribbon Task Force on Sustainable Digital Preservation and Access[3] (BRTF) spent two years analysing the economic conditions under which a variety of digital object types might best be maintained for future utility.

The purpose of this paper is to look at the various different ways that these three initiatives are currently being followed up and to propose future actions and reactions in response to them. The three follow-on activities are all being funded by JISC in the UK and have not yet been widely disseminated or discussed, either in the UK or internationally.

iPRES2011, Nov. 1–4, 2011, Singapore.

2. THE LIFE PROJECT
The LIFE project (Lifecycle Information for e-Literature) began in 2005 and consisted of three phases of work to investigate the possibility of defining an entire lifecycle model for a digital object and to then relate the parts of that lifecycle to the likely management and maintenance costs that might be incurred by the owners or keepers of the digital asset in question. The original context of this work and the initial focus was on estimating the cost of large homogeneous collections of materials, such as might be looked after by a national library or a large research intensive university. As such, the resultant model and tool may be more suited to a certain types of collections management procedures rather than others. Despite any perceived limitations, however, it is clear that the wider community valued this attempt to estimate retrospective and future costs and the LIFE project final reports from all three phases of work have been extensively downloaded and referenced.[4]

Figure 1 The LIFE model[5] (c.2010)



The influence of the project can also be measured by follow-on work and the Danish National Library and Archive have used the LIFE project model (in a somewhat adapted form) for their own purposes.[6] The LIFE-SHARE project[7] in the UK, based at University of Leeds, has also picked up on the LIFE modeling work and has used it to investigate the skills and strategies required for managing end-to-end digitization processes, including preservation of the created content.

At the end of phase 3 of the LIFE project, a functional tool was produced, based on a series of complex Microsoft Excel spreadsheets, that meant that users could input parameters and figures into a form, and then view costs over various timescales

and with various degrees of precision (depending on the level of detail of the original parameters) to support decision making about the cost of storing and managing digital materials over time.

2.1 Piloting the LIFE Model

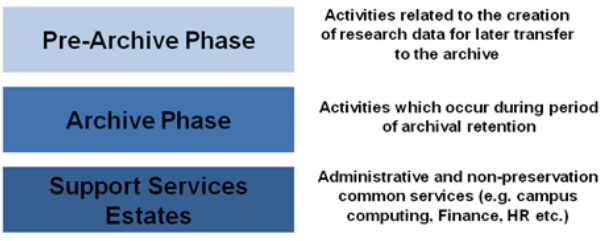
To follow up and properly exploit the 3 phases of work on the LIFE model and the tool, JISC commissioned HATII (University of Glasgow), working under the auspices of the Digital Curation Centre, to take the LIFE tool out into the UK HEI community in order to check whether a tool that was initially devised and developed with large document-type collections in mind could be applied in the university context where a different scale and scope of materials might require analysis, and where different input data (in particular relating to salaries and overheads) may be apparent.

As stated by the project page on the DCC website, it is anticipated that "the participating HEI repositories will benefit from a greater understanding of their day to day running costs and may even be able to identify inefficiencies in their current processes. In the longer term, this increased understanding of actual costs may inform strategic planning and policy development at the institution. The cost data provided by the targeted repositories, once anonymised, will have the potential to enrich the LIFE model for all subsequent users of the tool and will help to provide more accurate cost estimates for a broader group of organisation types."[8] For the conference presentation itself, it will be possible to give a summary of the results of this short pilot phase as it is scheduled to deliver the final report around about the time of writing this submission. One of the purposes of presenting this work to the IPRES audience will be to elicit support for further engagement with the tool and to create opportunities for further international collaborative work around the topic of cost modeling.

3. KEEPING RESEARCH DATA SAFE (KRDS)

The first phase of KRDS[9] took place in 2007 and was one of the early pieces of work that JISC commissioned in the area of research data management. Subsequent to this JISC established a substantial progamme of work[10] to support this activity but it was clearly seen from the outset that the approach taken with the LIFE project could usefully be extended to cover information defined as 'research data', and that the challenges associated with this form of information were different and discreet enough to require separate investigation.

Figure 2 Highest Level of the KRDS Activity Model[11]



The KRDS work was led by Neil Beagrie (Charles Beagrie Ltd.) and was carried out in collaboration with partners including OCLC and the UK Data Archive (see KRDS web page for the full list of contributors and partners). Following two phases of KRDS project work and some further funded activity to produce discreetly bundled related material (i.e. a fact sheet; a user guide; and detailed and summary activity models), an additional collaborative activity occurred in conjunction with an existing project called I2S2 (Infrastructure for Integration in Structural Sciences), based at the University of Bath. The objective of the I2S2/KRDS project was "to test, review and promote combined use of the Keeping Research Data Safe (KRDS) Benefits Framework and the I2S2 Value Chain Analysis tools for assessing the benefits of digital preservation of research data." This collaborative work makes more explicit the work relating to the benefits (as well as the costs) of managing research data that KRDS project began to seriously address in its second phase, and which the original I2S2 project engaged with at the outset as part of proving the value of integrated research infrastructure.

3.1 The Costs Observatory

In the course of presenting the conclusions from the second phase of the KRDS work, it was suggested to JISC (by Neil Beagrie) that some consideration should be given to the establishment of a 'costs observatory' that would facilitate the gathering, processing, analysis and dissemination of appropriate costs information relating to the management of long-lived data. The motivation for this suggestion originated from the experience of trying to collect authentic, useful and comparable cost information. It proved to be an extremely challenging task, particularly devising ways of comparing the data across different types of organisations, and one of the most prominent conclusions was that it would be far easier and more effective to setup a method of capturing cost data going forward than to try and retrofit comparison schema to diverse information sources.

To examine if the concept of a costs observatory was a workable idea, JISC commissioned a short (10 week) consultation and scoping study from Key Perspectives Ltd. during the period May – July 2011. According to the text of the invitation to tender, the "principal target outcome [of the currently imagined 'costs observatory'] would be to influence strategic planning and policy formation within institutions and enable them to make wiser, more realistic and cost effective decisions about managing information."[12] The detailed objectives (of the proposed observatory) were to:

- pro-actively seek and collect costs information relating to the short, medium and long term management of digital materials and data
- develop capability and status as a trusted broker of sensitive and confidential financial information
- analyse the financial data and produce reports and recommendations for universities and colleges (HEI's), funding bodies and strategic agencies on issues to do with the costs and economics of managing information
- support the UK HE sector with determining its existing and predicted Information management costs
- monitor and identify relevant economic, legislative and environmental issues
- liaise and co-ordinate with relevant service and information providers

Key Perspectives Ltd. did some analysis and scenario-building work and consulted with various representatives from the UK HE community on the efficacy of the proposed 'costs observatory', and then presented their conclusions to JISC in a report. In relation to one of the principal concerns laid out in the ITT, i.e. the scope of data to be collected – or to put it another way – the type of information (e.g. research data, administrative information, systems data, student records, learning and teaching materials, etc) that the observatory would gather, the report concluded that the focus would sensibly be on research data. This conclusion was arrived at through a combination of logistical possibility; declared community requirement; most pressing urgency; and territorial availability (i.e. it is not an area addressed by existing services in the UK). Whilst the ultimate conclusion to the question of the requirement and utility of this proposed service was a cautious endorsement, the report strongly questioned its overall feasibility (at least in terms of the way that the observatory was envisioned in the original ITT).

The purpose of presenting this work at IPRES is to offer the broader community an opportunity to comment on the costs observatory concept. To facilitate this, further detail will be provided about the conclusions of the Key Perspectives report.

4. THE BLUE RIBBON TASK FORCE ON SUSTAINABLE DIGITAL PRESERVATION AND ACCESS

The third and final strand of work to be included in this paper is an activity that was initiated by the National Science Foundation and the Andrew W. Mellon Foundation in the U.S., but was also supported by a number of other funders including JISC.[13] The purpose of the Task Force was to:

- Conduct an analysis of previous and current models for sustainable digital preservation, and identify current best practices among existing collections, repositories and analogous enterprises
- Develop a set of economically viable recommendations to catalyze the development of reliable strategies for the preservation of digital information
- Provide a research agenda to organize and motivate future work in the specific area of economic sustainability of digital information

The Task Force was convened over a two year period and delivered a significant and influential report in February 2010 that was widely referenced and nominated for the 2010 DPC Preservation award.[14] One of the features of this work that distinguishes it from the preceding projects, but also makes it nicely complementary, is that the focus is not on the 'cost' of digital preservation, but is more to do with the economic factors and strategies that may determine whether it will be possible to sustain digital information in accessible and comprehensible environments for the foreseeable future.

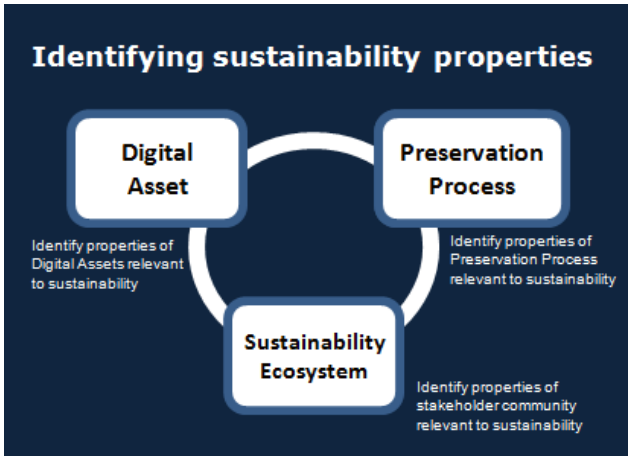
One of the features of the report is that it reframes some of the imperatives of digital preservation into an alternative (economic) language, where the laws of supply and demand, and some more specific language such as describing digital materials as 'depreciable durable assets', and discussing their 'non-rival' nature in terms of presenting a 'free-rider problem' offer a new type of terminology for understanding the challenges associated with managing information.

Whilst this report makes essential reading for a wide range of organisations dealing with a diverse array of data types, it presents the results of two years of detailed and deep thinking into a complex area. After two additional dissemination events in Washington and London, it was apparent that some form of synthesis work was required to present both the conclusions of the BRTF work itself and the subsequent discussions about it.

4.1 The Economic Sustainability Reference Model (ESRM)

In discussion with one of the BRTF panel members (Chris Rusbridge) about the possibility of commissioning some synthesis activity, it became apparent that an alternative idea had been suggested by another panel member (Brian Lavoie - OCLC) to create a different kind of summary of the BRTF conclusions. Building on the approach taken with the OAIS reference model (open archival information system ISO 14721:2003) Lavoie and Rusbridge suggested that a similar (but necessarily different) approach might be taken with the economic framework first outlined in the BRTF report, and that any resulting graphical depiction or conceptual model might not only act as a more concise and immediately descriptive synthesis of the BRTF work, but may also represent a useful and flexible community tool around which an ongoing discussion about economic sustainability might be based. From the outset, it was envisaged that if the framework received community endorsement, then it might provide a foundation for the kind of standards development process that the OAIS reference model underwent.

Figure 3 The current top level components of the ESRM[15]



At the time of writing, a draft version of the ESRM is still in preparation and the only public exposure the idea has had was at a workshop that took place in Tallinn, Estonia in May 2011, in conjunction with the Aligning National Approaches to Digital Preservation Conference.[16] A report from this event is still forthcoming but in summary, the delegates in attendance approved of the approach and endorsed further work to develop the reference model. IPRES represents another opportunity to demonstrate the latest iteration of the model and to elicit feedback about its likely usefulness and relevance to organisations facing

genuine (rather than theoretical) finance-related problems when preserving their digital assets.

5. OVERALL AIMS

Whilst this paper references six distinct (but more or less related) activities for presentation in a fairly short space of time, it should be acknowledged that the three initial activities (LIFE Project, KRDS project, and the BRTF initiative) ought to represent familiar territory to a lot of the IPRES attendees, many of whom will have extensive knowledge of the published literature on preservation. The objective therefore would be to address these activities with cursory descriptions (enough for those not familiar with them to understand their principle purpose) and then to move rapidly onto describing, and where possible evaluating, the new work that has been commissioned to follow up and build on the earlier work.

As noted above, when introducing all three new areas of work, it will be useful to provoke comments, opinions and discussion from the IPRES delegates to feed into the planning and implementation of next phases. It is not yet apparent to JISC whether and how further funding should be directed at any of the three projects and although it seems highly likely that an ongoing investigation into the economics, costs - and perhaps particularly – the benefits of digital preservation, would be appreciated by the broader community, detailed scoping is required. This presentation to IPRES could be an important part of that process.

6. REFERENCES

[1] Lifecycle Information for e-Literature. http://www.life.ac.uk/

[2] Keeping Research Data Safe (KRDS) Project. Phases I and II. http://www.jisc.ac.uk/publications/reports/2008/keepingresearchdatasafe.aspx

[3] The Blue Ribbon Task Force for Sustainable Digital Preservation and Access: http://brtf.sdsc.edu/

[4] During 2008, the University College London (UCL) e-Prints Repository recorded that the total number of downloads for the two LIFE project outputs available at that time was 1588. This meant that the items were the 9th and 11th most popularly downloaded items in the repository that year. http://discovery.ucl.ac.uk/past_stats/annual-2008.html#top50

[5] Diagram taken from presentation given by Brian Hole (British Library) at the Preservation and Archiving Special Interest Group (PASIG), Madrid, July 2010, available at: http://www.life.ac.uk/3/docs/Hole_pasig_v1.pdf

[6] A recent description of Danish work on the cost of digital Migration is published in the International Journal of Digital Preservation. http://www.ijdc.net/index.php/ijdc/article/viewFile/177/246

[7] LIFE-SHARE project, University of Leeds, http://www.leeds.ac.uk/library/projects/lifeshare/

[8] DCC LIFE Project web page: http://www.dcc.ac.uk/projects/life

[9] Keeping Research Data Safe: http://www.jisc.ac.uk/publications/reports/2008/keepingresearchdatasafe.aspx

[10] JISC Managing Research Data Programme: http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx

[11] Beagrie, N., (2011), Keeping Research Data Safe (KRDS) User Guide, p.5: http://www.beagrie.com/KeepingResearchDataSafe_UserGuide_v2.pdf

[12] A JISC Closed ITT (invitation to tender) was circulated on 16th May 2011 to a limited number of consultancies

[13] The Blue Ribbon Task Force on Sustainable Digital Preservation and Access: http://brtf.sdsc.edu/

[14] Digital Preservation Coalition Digital Preservation Award: http://www.dpconline.org/advocacy/awards/581-2010-digital-preservation-award

[15] Lavoie, B., Rusbridge, C., slide shown at the ESDI Roundtable Meeting, 26/05/11, Tallinn.

[16] Economic Sustainability of Digital Information (ESDI) Roundtable Event, 26th May 2011, National Library of Estonia, Tallinn, Estonia, http://www.educopia.org/events/ANADP/ESDI_Roundtable