LIBER's Involvement in Supporting Digital Preservation in Member Libraries

Paul Ayris

Director of UCL Library Services,
Vice-President of LIBER, Library Services University College London,
Gower Street, London WC1E 6BT, UK,
p.ayris@ucl.ac.uk

Abstract

Digital curation and preservation represent new challenges for universities. LIBER has invested considerable effort to engage with the new agendas of digital preservation and digital curation. Through two successful phases of the LIFE project, LIBER is breaking new ground in identifying innovative models for costing digital curation and preservation. Through LIFE's input into the US-UK Blue Ribbon Task Force on Sustainable Digital Preservation and Access, LIBER is aligned with major international work in the economics of digital preservation. In its emerging new strategy and structures, LIBER will continue to make substantial contributions in this area, mindful of the needs of European research libraries.

Key Words: digital preservation; research libraries; LIFE; Blue Ribbon Task Force

Digital Preservation: the Challenge

> 'In the wake of the digital revolution, stewardship of learned publications has acquired new opportunities as well as highly complex dimensions. These include fundamental shifts in the relationships between libraries, publishers and researchers. In their traditional role as custodians of society's accumulated knowledge, librarians face new challenges with regard to the access and preservation of digital information.'[1]

This quotation, in a nutshell, captures the essential challenges of digital preservation and digital curation for academic libraries. It is taken from an insightful Report on digital preservation, prepared for the Koninklijke Bibliotheek by Stijn Hoorens and his colleagues at the Rand Corporation in 2007.

For research libraries who are members of LIBER, the challenges of digital curation and digital preservation are very great. LIBER, as a membership organisation, takes its leadership role seriously. The purpose of this article is to describe work in which LIBER is engaged on behalf of its members in the fields of digital curation and digital preservation to address the issues.

LIBER and Digital Preservation

Overview

As the principal consortium of European research libraries,[2] LIBER has taken the demands of digital preservation very seriously. In a collaboration between its Preservation and Access Divisions, LIBER has supported two phases of the LIFE project (**L**ifecycle **I**nformation **F**or **E**-Literature). The purpose of LIFE is to identify, via a lifecycle approach, a methodology for the long-term digital curation of assets. Linked to this is work on a Generic Preservation Model (GPM). This project, sponsored by LIBER, has been generously sponsored by the JISC (Joint Information Systems Committee)[3] in the UK.

LIFE 1 and LIFE 2

LIFE 1 reported in 2006,[4] and LIFE 2 submitted its Reports to the JISC in 2008.[5] In the <u>*LIBER Quarterly*</u> vol. 18, nos 3/4, I looked at the outputs of the LIFE 2 project with particular reference to the digital curation and preservation of the outputs of digitisation projects.[6] This was one very important Case Study which emanated from the project, but there were other findings and Case Studies which form a major contribution to international discussions about the economics of digital preservation.[7]

LIFE 1

The first phase of LIFE made a major contribution to understanding the long-term costs of digital preservation; an essential step in helping institutions plan for the future. The LIFE work models the digital lifecycle and calculates the costs of preserving digital information for future years. Organisations can apply this process in order to understand costs and plan effectively for the preservation of their digital collections.

Run from 2005 to 2006, the LIFE 1 project made a major contribution to understanding the long-term costs of digital preservation. The project team felt that this was an essential first step in helping institutions plan for the future of digital collections. Based on a comprehensive review of existing lifecycle models and digital preservation, the LIFE 1 project developed a lifecycle-based methodology to calculate the costs of preserving digital information for the next 5, 10 or 100 years.

- The LIFE Model broke down a digital object's lifecycle into six main lifecycle stages, identifying the costs of these elements over a specific time period, and thus providing a complete lifecycle cost.
- Due to the lack of work undertaken in the area of digital preservation costing before 2005, LIFE 1 also produced the Generic Preservation Model to develop further the Preservation stage of the model. This work allowed institutions to start to identify and reduce the spikes of cost, as well as the frequency of their preservation actions. In the Generic Preservation Model, key elements of preservation activities were identified and the factors which contributed to their costs were modelled. A spreadsheet tool for calculating the costs for digital objects of varying file formats was also developed as part of the model.

- LIFE 1 undertook three Case Studies to test the models.
- The Web Archiving Case Study considered the costs of the British Library's web archiving activities, which selects and archives around 1,000 web site instances each year.
- The e-journals Case Study was based at UCL Library Services. At the time of the Case Study, 8,668 e-journal titles were logged in a UCL Access database.
- Voluntarily Deposited Electronic Publications (VDEP) housed at the BL[8] provided the final Case Study and involved the analysis of over 230,000 files.
- The three Case Studies proved to be highly effective in highlighting both the types of issues that can be encountered in a digital collection, and the ways in which a lifecycle methodology can be utilised to capture and apply a cost to solving these problems.

LIFE 2

One of the key deliverables for LIFE 2 was to make the LIFE Model and findings more accessible to those institutions wishing to either adopt the model, or to make use of the findings — essentially, to answer the question 'how is the LIFE work useful for our own collections?' The LIFE 1 Case Studies comprised born-digital collections, so a key area of expansion for LIFE 2 was the examination of non-born digital material (The British Library Newspaper Collection Case Study). This Case Study allowed for the comparison of analogue and digital lifecycles and costs and is fully described in an earlier article for LIBER Quarterly.[9]

Institutional Repositories were also addressed in two Case Studies (SHERPA-LEAP and SHERPA-DP). The costs of three Institutional Repositories were modelled to the LIFE work (SHERPA-LEAP Case Study),[10] and digital preservation services were examined through the SHERPA-DP Case Study.[11]

LIFE Model v. 2

The LIFE Model provides a view into the typical processes applied to digital objects throughout their lifecycle by an organisation acting as the custodian of those objects. The processes are loosely organised in a chronological order, from their creation through to eventual access. It should be noted, however, that processes can in practice overlap with each other or be executed in a
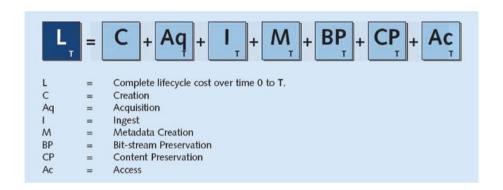
Table 1. Stages of the LIFE Model v. 2.

$$L_T = C + Aq_T + I_T + M_T + BP_T + CP_T + Ac_T$$

L	=	Complete lifecycle cost over time 0 to T.
C	=	Creation
Aq	=	Acquisition
I	=	Ingest
M	=	Metadata Creation
BP	=	Bit-stream Preservation
CP	=	Content Preservation
Ac	=	Access

Table 2. LIFE Model v. 2.

Lifecycle stage	Creation or purchase[1]	Acquisition	Ingest	Metadata creation[2]	Bit-stream preservation	Content preservation	Access
Lifecycle elements	…	Selection	Quality assurance	Re-use existing metadata	Repository administration	Preservation watch	Access provision
	…	Submission agreement	Deposit	Metadata creation	Storage provision	Preservation planning	Access control
	…	IPR and licensing	Holdings update	Metadata extraction	Refreshment	Preservation action	User support
	…	Ordering and invoicing	Reference linking		Backup	Re-ingest	
		Obtaining			Inspection		
		Check-in					

different order. The model aims to capture common processes found in most digital lifecycles. While some processes may not be applicable to all lifecycles, the intention is to provide meaningful placeholders for the majority of typical lifecycle processes.

Stages represent high-level processes within the lifecycle which group related lifecycle processes together. Elements represent the next level down in the analysis of lifecycle processes. They are still relatively high-level, but are focused on a distinct process within the lifecycle. The LIFE Model attempts to describe a standard set of elements to which most digital lifecycles can easily be mapped. Sub-elements represent the specific components of a lifecycle element. At this level of detail, lifecycles are expected to vary considerably from one to another and so the detailed sub-elements that are provided in the full Model documentation are for guidance only.

Table 3. The breakdown of components within the LIFE model.

Lifecycle level	Explanation
Lifecycle	The process from creation to access to preservation for a particular digital object, which can be broken down further into a number of distinct processes.
Lifecyle stage	A high level process within a lifecycle. Provides a way of grouping related lifecycle elements. Processes within a Lifecycle stage typically occur or recur at the same point in time.
Lifecycle element	A distinct and significant lifestyle process that will provide useful costing information for organisations to support planning, evaluative or comparative exercises.
Lifecycle sub-element	A suggested key component of a Lifecycle element. Not significant enough to warrant inclusion as a distinct Lifecycle element.

Lifecycle Level Explanation

Economic Evaluation of LIFE

When the first phase of LIFE was completed, one of the key elements that the team wanted to work on for LIFE 2 was a review of the economic approach used. Professor Bo-Christer Björk from Hanken, the Swedish School of Economics and Business Administration, was brought on board to complete a full independent review to the LIFE approach. The report largely validated the approach taken by the LIFE team.[12] At the same time, it provided a number of recommendations to steer the second phase of the project in the right

direction on key economic issues such as the use of discounting, the role of inflation and costs outside of the lifecycle. The review recommended that all calculations were done using real-term, inflation-adjusted costs. It also recommended that no discounting should be applied.

Case Studies

Three Case Studies were chosen to help refine and review the Lifecycle Model developed in LIFE 1, as well as to expand the testing of the Model to new areas.

The three Case Studies chosen for LIFE 2 were:

- SHERPA-LEAP — Institutional Repositories in the University of London.
- SHERPA-DP — Distributed repository environment for digital preservation of content.
- British Library Newspapers — Digitisation as surrogacy.

A report of the Digitisation Case Study was presented in an earlier issue of LIBER Quarterly, vol. 18, nos 3/4,[13] and so will not be commented on separately here.

SHERPA-DP Case Study

The key finding for this Case Study was that the costs for digital preservation did not vary greatly for differing quantities, as a largely automated process has been established. There were 6,526 objects harvested as part of the process for SHERPA-DP, giving the overall costs highlighted below.

Table 4. Summary of total costs from SHERPA-DP case study.

	Total cost	Cost per object	Annual cost per object
Year 1	£119,801	£18.40	£18.30
Years 1–5	£317,711	£48.70	£9.70
Years 1–10	£530,515	£81.30	£8.10

Stage C Aq I M BP CP Ac Total

There were no costs for Creation or Purchase. Acquisition costs were mostly for the development of the OAI-PMH tool and for integrating the harvester with the AHDS (Arts and Humanities Data Service) repository. Ingest costs were low, since quality assurance was the responsibility of the source repositories: scheduled harvesting using OAI-PMH led to file format characterisation being automated using DROID. The largest cost area was in Bit-stream Preservation, since this included staff elements for system administration and technology monitoring, as well as for storage provision. As with the other Case Studies, Preservation Action was a particularly hard part of Content Preservation to cost, while Preservation Planning and Technology Watch are more consistent across time.

SHERPA-LEAP Case Study

The Year 1 costs per object are summarised below:

Table 5. Repository lifecycle costs per entity (year 1).

Stages	C	Aq	I	M	BP	CP	Ac	Total
Goldsmiths	£3.20	£4.40	£16.80	£1.80	£0.30	£3.20	£0.90	£30.60
Royal Holloway	£0.00	£10.90	£4.10	£6.00	£1.20	£0.70	£1.20	£24.10
UCL	£0.00	£2.30	£6.90	£5.50	£0.10	£0.00	£1.20	£16.00

Overall Repository Operational Conclusions

The variations in costings between the institutions in the LEAP Case Study may be attributed to a number of factors. The narratives show staff on different grades, in differing proportions, working in the repositories. This naturally affects the costings. As the repositories become more stable, staff gradings and roles are likely to become regularised, and comparison across the HE community will become more informative. The studies show that the fact that Goldsmiths handles a range of complex digital materials within its institutional repository structure increases the average handling cost per object.

As with SHERPA-DP, after year 1, the main lifecycle costs are those associated with preservation.

Table 6. SHERPA-DP lifecycle costs per entity (year 1).

Stages	C	Aq	I	M	BP	CP	Ac	Total
Cost	£00.0	£11.40	£0.10	£0.00	£3.00	£2.00	£1.80	£18.40

For SHERPA-LEAP, Bit-stream Preservation costs are based on estimates, both of repository growth and in the technology marketplace. Content preservation will clearly bring costs for the partners in the future, but for the time being those costs are not easily predictable.

This is something that perhaps the Generic Preservation Model can help to answer once it has been further developed and tested. These differences across both the SHRERPA-LEAP repositories and the other Case Studies lead to questions as to whether or not LIFE can yet be used for inter-institutional comparison when the collections themselves are so variable. This is one of the reasons why the context of the Case Studies is so important, and it is critical not simply to take the lifecycle costs at face value.

There is also the question of time and resources taken up to identify these costs in the first place. Each of these Case Studies needed considerable time spent on them, both internally within the institutions in question and externally by the LIFE Team. It would be fair to say that each of the Case Studies took a much longer timeframe to develop than originally anticipated. This should not be under-estimated by other institutions thinking of performing similar costing studies.

Stages C Aq I M BP CP Ac Total

For each of the Case Studies the effort was certainly worthwhile, allowing the institutions to gain a greater understanding of their own costs and processes. As noted by the team in the SHERPA-DP Case Study, it certainly helps to have a business requirement for determining costs, but applying the LIFE model to different institutional settings is recommended to all with an interest in digital curation and preservation.

Overall Repository Strategic Conclusions

LIFE 2 identified a number of strategic conclusions which could be drawn from the repository Case Studies:

- The SHERPA-DP Case Study shows that a 3rd-party preservation solution is possible for digital repositories in the UK.
- As an automated service, SHERPA DP could offer significant cost savings when increased quantities of digital objects are processed.
- For SHERPA-DP, the largest cost area was in Bit-stream Preservation, since this included staff elements for system administration and technology monitoring, as well as provision for storage (including equipment renewal) and offsite duplicate storage.
- The variation in costings identified in the SHERPA-LEAP case studies reveals that the rollout of institutional repositories in the UK is still in its infancy.
- The costing figures prepared by the SHERPA-LEAP partners are not yet robust enough for definitive conclusions to be drawn; it would be too simplistic to make comparisons between institutional costs at this stage.
- Digital preservation is yet to become embedded as a concept in the Higher Education community. This presents a major challenge in advocacy for the global digital preservation community.
- In the SHERPA-LEAP Case Studies, it is suggested that after year 1 the main lifecycle costs are those associated with preservation. However, Bit-stream Preservation costs are based on estimates, both of repository growth and in the technology marketplace. Content Preservation will clearly bring costs for the partners in the future, but for the time being those costs are not easily predictable.
- The Goldsmiths Case Study suggests that higher costs may currently be associated with managing complex digital materials at an institutional level.

LIFE 3?

LIBER is minded to take the work of the LIFE Team to a third phase, with the production of a refined Generic Preservation Model (GPM) which contains a yet more accurate modelling of the workflows and costs around digital preservation. From the lifecycle work which LIFE has undertaken in the LIFE Model, the LIFE Team is keen to produce a toolkit which can be used by

content creators, libraries, researchers submitting research grants, or senior decision makers to predict what the true costs of the long-term curation of a set of digital objects is. If this LIBER-sponsored project can achieve this, it will have made a real contribution to the global digital preservation agenda.

Blue Ribbon Task Force on Sustainable Digital Preservation and Access

To mirror the ground-breaking work of the LIFE project, LIBER is also involved in a US-UK Blue Ribbon Task Force on economically sustainable digital preservation,[14] through the membership of LIBER's Vice-President on behalf of the JISC (Joint Information Systems Committee).

The goals of the Task Force are to:

- Conduct an analysis of previous and current models for sustainable digital preservation, and to 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 organise and motivate future work in the specific area of the economic sustainability of digital information.

In terms of economic sustainability, the Task Force has taken some care to define what it thinks this means:

- The set of business, social, technological and policy mechanisms that

    o encourage the gathering of important information assets into digital preservation systems.
    o support the indefinite persistence of digital preservation systems enabling access to and use of the information assets into the long-term future.

Economically-sustainable digital preservation requires:

- Recognition of the benefits of preservation on the part of key decision-makers.
- Incentives for decision-makers to act in the public interest.
- A process for selecting digital materials for long-term retention.
- Mechanisms to secure an ongoing, efficient allocation of resources to digital preservation activities.
- Appropriate organisation and governance of digital preservation activities.

The Blue Ribbon Task Force on Sustainable Digital Preservation and Access was created in late 2007. Until the end of 2009, the BRTF-SDPA will explore the challenge of sustainability with the goal of delivering specific recommendations that are economically viable and of use to a broad audience, from individuals to institutions and corporations to cultural heritage centres.

The BRTF-SDPA is funded by the National Science Foundation and the Andrew W. Mellon Foundation in partnership with the Library of Congress, the Joint Information Systems Committee (JISC) in the UK, the Council on Library and Information Resources, and the National Archives and Records Administration.

In December 2008, the BRTF-SDPA issued its first-year Report and the remainder of this section will be devoted to an overview of the contents of

that Report.[15] The Task Force will issue its final Report at the end of its second and final year in late 2009. The Final Report will focus on practical recommendations and models for economically sustainable digital preservation in academia, the government sector, and private enterprise.

Interim Report of the BRTF-SDPA

Most would agree that digital information is fundamental to the conduct of modern research, education, business, commerce and government. Comparatively little agreement exists, however, concerning access to, and the preservation of, valuable digital information. The BRTF-SDPA has identified two grand challenges in this respect:

- Who is responsible?
- Who should pay?

The Blue Ribbon Task Force's Interim Report explores fundamental issues and challenges associated with economically sustainable digital preservation and access. The report stresses that:

- the digital preservation and access problem is urgent

The Report urges that access to data tomorrow requires decisions concerning preservation today. It also makes clear that viable digital preservation strategies require attention not only to technical, legal, and social issues, but to economic issues as well.

The Report identifies a number of systemic challenges to economically sustainable digital preservation and access. They include:

- *Inadequacy of 'one-time' funding models* (e.g. research grants or contracts) to address persistent long-term access and preservation needs.
- *Poor alignment between stakeholders in the digital preservation and access world and their roles, responsibilities and support models*. For example, creators, users, and stewards of digital information may be different groups with different funding models.
- *Lack of institutional, enterprise, and/or community incentives* to support the collaboration to enforce sustainable economic models. Many

institutional and community cultures dis-incentivise the common formats, standards, and the hardware/software compatibility needed for digital preservation.

- *Complacency that current practices are 'good enough.'* Both 'carrots', in the form of recognition that access to information is an investment in current and future success, and 'sticks', in the form of penalties for non-compliance, accounting of explicit opportunity costs, or costs of lost information, are needed.
- *Fear that digital access and preservation is too big to take on.* Digital preservation is a big problem, but not insurmountable. Solutions may be as manageable as including a 'data bill' as an explicit, fixed part of an institution's business model.

The Interim Report stresses that institutional, enterprise, and community decision makers must be part of access and preservation solutions. Without their participation, it will be nearly impossible to build the critical foundation of digital information on which the modern world depends.

Some of the Unknowns

The Interim Report highlights some of the factors that are unknown and I give below the final section of that Report, which sums up the issues.

Much is still unknown and in some cases has gone largely unmentioned. To some degree, and as this section of the Interim Report will discuss further, these ambiguities are inherent in any system that seeks to manage resources for the indefinite future. Still, it is possible to suggest some of the places where we can anticipate changes that affect the way that libraries, archives and museums must plan for the management and economic sustainability of their digital collections and that has not been captured in the work thus far.

1. Survivability

First, and most concretely, it is surprising that more attention has not been paid to the economic aspects of threat models to survivability, especially in the wake of both 9/11 and Katrina, and the associated risks or exposures to those threats. Indeed, the 2003 report *Building an Electronic Records Archive*

at the National Archives and Records Administration: Recommendations for Initial Development, by the Computer Science and Telecommunications Board of the National Research Council[16] specifically recommended, 'The risks of the various possible causes of data loss — such as malicious acts, natural disasters, software bugs, human error, and hardware failures — should be assessed and used to make informed engineering cost-benefit trade-offs'.[17] A subsequent letter report in 2003 encouraged the US National Archives and Records Administration (NARA) to 'specify an explicit threat model be developed early in the ERA's life cycle,' noting that a draft specification for follow-on work to the 2003 study 'makes occasional mention of measures that might help to avert threats … but it includes no overall requirement that the system be capable of surviving an attack or incident'.[18] In a second report, NARA briefly addressed the importance of threat modelling and threat countering in the context of a general discussion of record integrity and authenticity.[19]

On the one hand, it is well understood that storage repositories should be backed up routinely, replicated in geographically-distinct locations, and synchronised regularly, and these costs have, in some cases, been accounted for. For example, the LIFE 2 Model does allow for back-up, replication and synchronisation, and a partnership involving NARA, the University of Maryland and the San Diego Supercomputer Center proposed a model for a persistent archive that addressed risk management and disaster recovery as well as technology evolution.[20] However, there has been no analysis of the economic issues addressing, for example, what the optimum number of replication facilities would be when balanced against the probable occurrence of various kinds of natural or man-made disasters. Basic geographic dispersion of data may well not protect against events such as electromagnetic pulse.

Much data loss is due to human error; a very large number of attacks are carried out by insiders. And archives and libraries have often been targets in overt or covert wars. Consequently, there is every reason to expect that this will be the case with digital archives of key cultural materials. So the threats are real at the level of the trans-institutional system but highly unpredictable for any given element in that system. Any model for sustainability and for the costs associated with it must take such unpredictable considerations into account, if only to allow for contingency budgeting.

2. Recovery

Related to the question of failure is the question of recovery. Again, substantial work could be done through case studies of massive failures that might provide some parameters to be used for the suggested contingency budgeting. In the case of Katrina, for example, a close analysis might be undertaken to parse the steps and associated costs of re-instantiating the massive systems that were wiped out in hospitals, banks, and so on. While such efforts may not have been technically considered preservation, they are instances of recovery in the wake of disaster and might contribute to putting dimension around the vague problem of recovery and to acknowledging the importance of contingency planning as part of managing digital assets over the long term. In the near term and as a purely practical matter, *organisations should have contingency budgets and provision for recovery*. Also, we need to abandon the belief that recovery is a routine process that leads to perfect re-constitution; in real world cases, there will often be extensive damage assessment, attempts to reconstruct or re-verify data, and sometimes recovered data will be of questionable quality, but may be used anyway because it is all that is left.

Recovery clearly involves more than simply buying equipment, re-installing programs, and copying data onto the correct place, all items that may show up in the accounting systems as investments. In some instances, having to re-build occasions, revisiting existing workflows, a different organisation emerges as a result of the process. This is one reason why an environmental disaster like an oil spill can look like economic growth. New resources and investments may be brought to bear in a local economy where none previously may have existed so the size of the local economy looks like it has grown. Implementation of new technology is more than simply exchanging one system for another. Word processing applications, spreadsheets, and e-mail systems are well-known examples of the ways that new end user technologies altered workflows, and the learning curve that may be specifically associated with a new technology can have ramifications well beyond the specific purpose for which it may have been intended, thus affecting the overall health and efficiency of the organisation. Such costs and benefits have been traditionally difficult to capture and have not been reflected in the material identified to date or in the testimony before the Task Force by experts who are already managing digital collections although, clearly, all

of them routinely migrate their collections to new hardware and software environments.

3. End Users and Institutions

It is likely that similar changes should be expected as tools to enable preservation are developed and become part of end users' workflows, much as has been envisioned in the development of the PDF standard. Indeed, *Nature* has already called for such awareness in its 4 September 2008 editorial in a special issue on data, noting, 'Researchers need to be obligated to document and manage their data with as much professionalism as they devote to their experience'.[21] The editorial calls for greater support for such endeavours, noting that the number of publicly-funded databases with preservation responsibilities is relatively small: 'Universities and funding agencies need to provide and support curation facilities, tools and training'.[22] Again, modelling the system from workbench to archival repository and its economic implications, including the value of the data that is part of the flow, have not been addressed. Reducing complexity by widespread adoption of standards might be one way in which costs might be rationalized and even reduced. But the first step requires a change in behaviour and then understanding and modelling that behaviour so that the economic dimensions can be understood.

Institutions are embedded in society and culture and the way that policies are formulated, understood and implemented reflect the tenor of the milieu. These, too, have implicit risks that affect the ability of institutions to manage their collections. Societal trends that affect appraisal, selection, and access can be anticipated but not quantified.

4. Privacy

One obvious shift concerns privacy and its tension with access. Changes in the understanding of privacy are immense as recently illustrated in an issue of *Scientific American* (Vol. 299, no. 3, September 2008) devoted to the topic and clearly will have ramifications for the way video and social networking sites are collected, archived, preserved, and ultimately made available. There exists a fundamental tension between protection of personal privacy and personally-identifiable information and certain kinds of epidemiological research, particularly as search technology advances

and it becomes possible to identify individuals from pools of data in which identifying information has supposedly been removed — and indeed was removed given the state of the art at the time the data were processed. There exists a cluster of competing, legitimate concerns, namely: research that requires contextualizing highly granular information in social groups, the desire to protect individuals' confidential information, advances in technology, and the laws and regulations attempting to govern those relationships. Clear guidance is lacking and perception and societal values will change, resulting in an inherently unstable equilibrium that inevitably leaves the management of the collecting agency vulnerable, as the museums that have custody for anthropological collections are already discovering.

The degree to which a custodial institution may be affected by these societal shifts is likely to vary depending on mission, regulatory context and the nature of the material. The larger point is that digital collections are targets for a wide range of legal and public relations attacks and are likely to become victims of physical attacks, whether from natural disasters, random electronic surges, or outright malice. The frequency of these challenges and the costs of dealing with them can be unpredictable but very large. The outcomes of litigation are also unpredictable and potentially life-threatening for otherwise sustainable preservation strategies. Several of the speakers from whom the Task Force heard acknowledged that their revenue streams may be precarious. The risk is that the unpredictable elements — threats from natural disaster, changes in perceptions of value, accidents and malice — will tip an institution from viable to failing.

5. Organisation

Much of the cost modelling that has been done for preservation has focused on trying to quantify the relatively predictable cost factors. Substantial progress has been made and our ability to parse the challenge has become more refined. Serious problems remain, particularly when we think about preservation across long periods of time, notably, the unpredictable but inevitable: the 'black swans' to use Taleb's term (2007)[23] — the very low probability events, the high-cost legal challenges, the threats that were not considered in the threat model but came to pass anyway. Some of this can be handled by contingency planning and the development of contingency budgets and strategic reserves; other parts

may lend themselves to insurance approaches. Yet another question is the choice of appropriate scale of preservation activities: they need to be big enough to have flexibility to respond to challenges but not so large that their failure is catastrophic.

But it seems clear that there is no substitute for a flexible, committed organisation dedicated to preserving a corpus of material. This organisation must be able to make choices and devise strategies to deal with unexpected problems of all types. If necessary, it can conduct triage and make compromises. Modelling predictable lifecycle costs and arranging funding streams to support these costs is necessary but is clearly not going to be sufficient. As implied in the definition of economic sustainability that guides this study, the design of appropriate organisations, the economic implications of the organisation as more than the sum of a series of flows, and the organisation's placement in legal, public policy and cultural settings are clearly going to be key to achieving long-term sustainability of digital collections.

Next Steps

LIBER is clear that digital curation and digital preservation are key items on the agenda of research libraries.

As such, LIBER will:

- Support LIFE in continuing to seek funding for a Phase 3 of its development — to finalise the Generic Preservation Model (GPM) and to construct a predictive tool for the costing of digital curation.
- Hold a Workshop entitled *Curating research 2009: e-merging new roles and responsibilities in the European landscape. A workshop on long-term digital preservation*.

  o The Workshop is being held on 17 April 2009 at the Koninklijke Bibliotheek/National Library of the Netherlands, The Hague.
  o This Workshop aims to develop a basic understanding of the issues presented by the long-term digital curation and preservation of resources which are (to be) deposited in institutional and subject-based repositories — both within research institutions and research communities. It will highlight the

> state of the art in digital curation and will cover best practices, including possibilities for outsourcing.
>
> o   Target groups are policy makers and managers of digital objects within libraries and research institutions, e.g. research librarians, directors of research institutions, repository managers and middle management; publishers are also invited.
>
> o   The website is at www.kb.nl/curatingresearch

- Continue its input into the Blue Ribbon Task Force, through the contribution of the LIFE project, into Year 2 and the Task Force's final Report.
- Ensure that digital curation is embedded in the new, emerging LIBER strategy which will be presented to LIBER participants at the LIBER Annual General Conference in Toulouse in 2009.

> o   See http://liber2009.biu-toulouse.fr/

Notes

---

[1] Stijn Hoorens, Jeff Rothenberg, Constantijn van Oranje-Nassau, Martin van der Mandele, Ruth Levitt, *Addressing the uncertain future of preserving the past: towards a robust strategy for digital archiving and preservation* (Cambridge: Rand Corporation, prepared for the Koninklijke Bibliotheek, 2007), linked at http://www.rand.org/pubs/technical_reports/TR510/

[2] See http://www.libereurope.eu

[3] See http://www.jisc.ac.uk

[4] See http://eprints.ucl.ac.uk/1854 and http://eprints.ucl.ac.uk/1855

[5] See http://eprints.ucl.ac.uk/11758 and http://eprints.ucl.ac.uk/14110

[6] See http://liber.library.uu.nl/publish/articles/000264/article.pdf

[7] In what follows, I am grateful to be able to draw on the contributions of the LIFE Project Team, particularly Richard Davies, Rory McLeod, Paul Wheatley and Helen Shenton at the British Library.

[8] For VDEP at The British Library, see http://www.bl.uk/aboutus/stratpolprog/legaldep/index.html

⁹ See note 6 above

¹⁰ See http://www.sherpa-leap.ac.uk/

¹¹ See http://www.sherpadp.org.uk/index.html

¹² See B.-C. Björk, *Economic evaluation of LIFE methodology: research report. LIFE Project* (Unpublished Report commissioned by the LIFE 2 team), at http://eprints.ucl.ac.uk/7684/

¹³ See http://liber.library.uu.nl/

¹⁴ See http://blueribbontaskforce.sdsc.edu/

¹⁵ For the full Report, see http://blueribbontaskforce.sdsc.edu/biblio/BRTF_Interim_Report.pdf. For the text which follows, I am deeply indebted to the members of the Task Force and the team which compiled the final version of the Report and the supporting materials. The members of the Task Force are Francine Berman [co-Chair], Director, San Diego Supercomputer Center and High Performance Computing Endowed Chair, University of California San Diego; Brian Lavoie [co-Chair], Research Scientist, OCLC; Paul Ayris, Director of Library Services and Copyright Officer, UCL (University College London); G. Sayeed Choudhury, Associate Dean of Libraries, Johns Hopkins University; Elizabeth Cohen, Academy of Motion Pictures Arts and Sciences, and Stanford University, Stanford, CA; Paul Courant, University Librarian, University of Michigan; Lee Dirks, Director of Education and Scholarly Communications, Microsoft Corporation; Amy Friedlander, Director of Programs, Council on Library and Information Resources; Vijay Gurbaxani, Director, Center for Research Information Technology and Organization, and Professor, Information Systems and Computer Science, University of California at Irvine; Anita Jones, Professor of Engineering and Applied Science, University of Virginia; Ann U. Kerr, Vice Chair, International IEEE Mass Storage Systems and Technology Committee, and Consultant, AK Consulting; Clifford Lynch, Executive Director, Coalition for Networked Information (CNI); Dan Rubinfeld, Robert L. Bridges Professor of Law and Professor of Economics, University of California at Berkeley; Chris Rusbridge, Director, Digital Curation Centre, University of Edinburgh; Roger Schonfeld, Manager of Research, Ithaka; Abby Smith, Historian and Consulting Analyst to the Library of Congress; Anne Van Camp, Director, Smithsonian Institution Archives.

¹⁶ See Committee on Digital Archiving and the National Archives and Records Administration, Computer Science and Telecommunications Board, & National Research Council of the National Academies, *Building an electronic records archive at the National Archives and Records Administration: recommendations for a long-term strategy* (Washington, D.C.: National Academies Press, 2003).

[17] Committee on Digital Archiving and the National Archives and Records Administration, Computer Science and Telecommunications Board, & National Research Council of the National Academies. [2005a]. *Building an electronic records archive at the National Archives and Records Administration: recommendations for a long-term strategy* (Washington, D.C.: National Academies Press, 2005), p. 83.

[18] See Committee on Digital Archiving and the National Archives and Records Administration, Computer Science and Telecommunications Board, & National Research Council of the National Academies, *Building an electronic records archive at the National Archives and Records Administration: recommendations for a long-term strategy* (Washington, D.C.: National Academies Press, 2003), p. 88.

[19] See Committee on Digital Archiving and the National Archives and Records Administration, Computer Science and Telecommunications Board, & National Research Council of the National Academies. [2005b]. *Building an electronic records archive at the National Archives and Records Administration: recommendations for a long-term strategy* Washington, D.C.: National Academies Press.

[20] See R. Moore, J. JaJa and R. Chadduck, (n.d.), A pilot persistent archive: partnership between SDSC, UMD, and NARA (n.d.). See http://www.archives.gov/era/pdf/it-conference-jaja-workshop.pdf

[21] See *Nature* (4 September 2008). Editorial. *Nature,* p. 455.

[22] See *Nature* (4 September 2008). Editorial. *Nature*, p. 455.

[23] See N.N. Taleb, *The black swan: the impact of the highly improbable* (New York: Random House, 2007).