

Note: This is a new section and has been prepared by Deborah Woodyard-Robinson [March 17 2006]

"Delays in taking preservation decisions can (and most often will) result in preservation requirements that are more complex, labour intensive and therefore costly."

- Cedars Guide to Digital Collection Management [1](#)

Through several recent projects and studies we are now beginning to understand what it costs to manage digital material in the long term. See [Exemplars and Further Reading](#) Preparing cost models and estimations is an invaluable task. It combines the initial investments with ongoing costs to inform sensible and economical decision making and provides advice on the total resources required to implement digital preservation.

Calculating the cost of digital preservation is a complex task, but perhaps even more challenging is assessing the value of this work and securing the funding to perform it. Key decision makers must be convinced that the value of the digital assets is equal to or greater than the cost of the services to maintain them in order to establish economically sustainable processes and business models.

Costs

There are too many variables for a single model to be applied to developing digital preservation costs, but there are now several tools and case studies available that can be useful for guidance.

Using an established model as a basis can be helpful but be aware of significant differences in collections and material types, organisation mission and the services they provide as all these aspects of an organisation can have significant effects on their costs.

A standard approach to determining costs is to break down the digital life cycle into processes based on workflow or a system model such as the OAIS Reference Model [2](#). Each stage or process, called a cost event (for examples see table 1 below), is then evaluated for likely cost sources (for examples see table 2 below). Depending on the purpose of the study a total cost

may then be calculated per item, per time period for preservation of all collection material, or per process.

Life cycle management is a sensible tool for allocating costs. Using a structured approach such as this can help identify costs which may not have been considered (e.g. costs of selection, etc) and also reinforces that costs are cyclical and very few are one-off expenses. This is well illustrated by Shenton ³ in the examination of life cycle management of library collections. Stages identified in the life cycle of traditional collections start with selection, acquisitions processing, cataloguing and pressmarking and go through to preservation, conservation, storage, retrieval and the de-accession of duplicates. Similarly the digitised material lifecycle was broken down into selection, checking intellectual property rights, conservation check and remedial conservation costs, retrieval and reshelving costs, capture of digitised master, quality assurance of digitised master and production of service copies, metadata creation cost, access cost over time, and storage costs over time. An observation of this study was that a one off cost such as cataloguing may appear to be a large proportion of the initial cost, however over time it may easily become a smaller cost component than a recurring cost in the life cycle such as providing access.

Similarly the LIFE (Life Cycle Information for E-Literature) Project ⁴ aims to establish individual stages in the born-digital life cycle and examine the cost to provide the full financial commitment of collecting digital materials over the long term. On that basis it also hopes to identify possible cost reductions and potential efficiencies.

It may be helpful to use the OAIS reference model as a guide to enhance and inform the future of the long term digital life cycle. The OAIS discusses many processes that will be needed for long term preservation that may not yet be fully implemented within an organisation.

Table 1: Typical cost events ⁵

Activities

Cost Events

System creation and management activities

- Creating organisational infrastructure
- Creating repository architecture
- Archive administration
- Repository operation
- Maintenance
- Upgrades

Digital material workflow/life cycle activities

- Selection,
- Acquisition,
- Validation
- Creation of digital collections,
- Conversion of deposited material
- Rights negotiation and management
- Resource Description, e.g. Cataloguing
- Metadata and preservation metadata creation
- Storage
- Evaluation and revision
- Disposal/Deaccession

Specific preservation activities

- Technology planning activities such as Technology watch
- Long-term strategies e.g. migration and emulation

Specific access activities

Access to objects
Access to catalogues
User support

Table 2: Typical cost sources [6](#)

Cost Type

Cost Sources

Digital object/data acquisition

Purchase price / licensing cost

Labour

Personnel will include dedicated staff as well as varying proportions of senior management, supervisor,

Technology

Hardware
Software

Level of Requirements (e.g. speed, availability and performance)

Non-Labour operational costs

Facilities and Space (e.g. rent and electricity)

Materials and Equipment

Communications

Insurance

Legal costs

Several factors can have a significant influence on the result of these cost events. The relationship of costs and institutional strategies such as [collaboration](#) , [third party services](#) , [rights management](#)

,
[training](#)

and

[standards](#)

are discussed in the previous sections. Other factors include:

Labour

Undoubtedly the greatest cost in the digital material life cycle today is labour. Therefore the ability to automate or batch process digital materials and to participate in collaboration on research and services will reduce the cost of digital preservation most significantly.

Object types and storage size

The complexity of the material submitted and number of objects acquired generally has more impact on costs than the total storage size. The type and variety of formats accepted into the repository will also affect cost, because for example proprietary formats are likely to be more difficult and expensive to manage in the long term. It may be possible to reduce costs by limiting the formats the repository will accept, or transforming material into a standard common format. This can be done to reduce the number of file types and possibly reducing the storage size. However, it is also necessary to realise that due to storage redundancies required for back up each gigabyte of deposited data requires more than one gigabyte of disk space in repository storage.

Beware of generalising storage sizes for digital file formats. The definition of an “object” must be very specific to make sense of figures. For example an image may be a small low resolution GIF or a large high quality TIFF. When one file size is multiplied by many thousands of objects this can impact storage predictions and costs considerably. Yet the smallest version of an object may not be the most cost effective to preserve if it cannot serve the required purpose, such as in the case of the high quality, substantially larger image.

Repository boundaries

A clear set of guidance documents such as the organisations mission and collection (selection) policies and guidelines will reduce long term cost defining the aim and direction of collections and services for more efficient decision making.

Existence of services that can be shared such as file format registries and technology watch services will reduce long term preservation costs. Availability of software tools for providing automation will also be a key factor particularly for smaller organisations not able to afford to create their own.

Preservation service level

The various levels of preservation service offered by an organisation will also significantly affect cost. Effectively there is a long-term trend for a rapid increase in the quantity of computer storage per unit cost, so the cost of bit preservation over time is declining towards zero. The real costs are in providing staff and access over time (and meeting increasing user expectations of service for this). Therefore a repository only offering bit-level preservation, where the only undertaking is to guarantee storage and delivery of the sequence of bits, will have lower costs than a repository managing full migration paths or emulation solutions.

Timing

Preservation strategies enacted early in the life cycle are likely to be more cost effective than salvage attempts left until technology has already moved on significantly. For example, creating preservation metadata while sources such as the producer are still available is much faster and cheaper than to attempt to divine the appropriate information at a later date. It may also be cheaper for the producer to create such metadata as they are likely to have the information required at hand and find it easier to understand. Similarly solutions to technology obsolescence change with time from easily and quickly solvable while the technology is familiar, to the area of a few specialists or a state of digital archaeology requiring significantly more time and expense to restore.

Another aspect of timing in relation to costs is the period of retention. Expect that materials to be preserved indefinitely will be more costly than those for finite retention. The cyclical nature of digital preservation expenses will likely determine that disposal of material not required permanently in a timely manner will provide a cost benefit, providing the disposal event does not cost more to achieve than the continued preservation expenses.

Business models

As described above methods for assessing the costs of managing digital materials in the long term are becoming clearer. Further to this work is the need to prove the benefit or return of this investment in order to secure adequate funding.

Key stakeholders and decision-makers need to be motivated to contribute to the medium to long term preservation of digital materials. These key stakeholders include the producer, the rights holder, the repository and the consumer, who each may or may not be the same entity depending on the organisation. Each stakeholder will have different interests and require different incentives to actively participate in the preservation process.

In examining the business model a clear focus should be on the end purpose of the archiving process which is to serve the consumers or "designated communities" of current and future users.

Organisations such as the British Library are now exploring ways to enumerate the value of their collections and services which had previously appeared to be unquantifiable [7](#). These new calculations based on "Contingent Valuation" [8](#) prove the worth of supporting their operations to their key stakeholders and funding sources.

One reasonably simple method to assess the value of maintaining digital material in an organisation is to present the case if there were inaction and the assets were effectively lost to the community. What would be the cost of replacing them, recreating them or managing to work without them?

The *espida* Project [8](#) at the University of Glasgow is developing a sustainable business focused

model for digital preservation at an FE/HE institution. They have recognised that digital preservation so far has frequently been funded only by short term projects, yet there is an on-going cost for preservation. There is a lack of commitment to sustainable funding. Therefore there is a need to demonstrate the benefit weighed against the costs and risks. There needs to be a demonstrable return on that preservation investment, even if it isn't directly financial in nature.

To open PDFs you will need Acrobat Reader. [Click here to download Acrobat Reader](#) . **References**

1. Cedars Guide to Digital Collection Management (2002). Executive Summary pp.7.

Available online:

<http://www.leeds.ac.uk/cedars/guideto/collmanagement/>

2. (as done by TNA see DCC DPC workshop notes)

3. Shenton, H. Life Cycle Collection Management *LIBER Quarterly* , 2003, 13(3/4) pp. 254-272

4. The LIFE Project: <http://www.ucl.ac.uk/ls/lifeproject/>

5. This list is derived from cost events described in several studies:

1. LIFE PowerPoint presentation

http://www.ucl.ac.uk/ls/lifeproject/documentation/espida_event.pdf [PDF] 11 February 2005

2. James, Hamish., Ruusalepp, Raivo., Anderson, Sheila., Pinfield, Stephen (2003) [*Feasibility and Requirements Study on Preservation of E-Prints*](#) [PDF] Report Commissioned by the Joint Information Systems Committee

3. CEDARS Guide to Digital Collection Management

<http://www.leeds.ac.uk/cedars/guideto/collmanagement/>

Section 5. Costs: Processes and People pp.19-21

4. Ashley, K. (1999). 'Digital Archive Costs: Facts and Fallacies.' DLM Forum '99.

http://europa.eu.int/ISPO/dlm/fulltext/full_ashl_en.htm

Update 19 March 2008

No longer available - information at

<http://ec.europa.eu/archives/ISPO/dlm/>

5. Arturo Crespo, Hector Garcia-Molina: Cost-Driven Design for Archival Repositories. Joint Conference on Digital Libraries 2001 (JCDL'01); June 24-28, 2001; Roanoke, Virginia, USA.

<http://www-db.stanford.edu/~crespo/publications/cost.pdf> [PDF]

6. This list is derived from cost sources described in several studies:

1. Cost Orientation Tool, ERPANET (Date Created: Sep 2003)

<http://www.erpanet.org/www/products/tools/ERPANETCostingTool.pdf> [PDF]

Update 27 November 2006

Link broken. New location

<http://www.erpanet.org/guidance/docs/ERPANETCostingTool.pdf>

2. Marley, Steve., Moore, Mike., Clark, Bruce (2003) *Building a Cost-Effective Remote data Storage Capabilities for NASA's EOSDIS*

<http://storageconference.org/STORAGECONFERENCE/2003/presentations/B05-Marley.pdf> [PDF]

Paper presented at the Twentieth IEEE/Eleventh NASA Goddard Conference on Mass Storage Systems & Technologies April 7-10, San Diego

Update 19 October 2009

Link broken. New location

<http://storageconference.org/2003/presentations/B05-Marley.pdf>

3. Connaway, Lynn., Lawrence, Stephen (2003) *Comparing Library Resource Allocations for the Paper and Digital Library: An Exploratory Study*

D-Lib

9,

(12)

<http://www.dlib.org/dlib/december03/connaway/12connaway.html>

4. Costs of Digital Preservation, Testbed Digitale Bewaring (Date Created: May 2005) (Netherlands) <http://www.digitaleduurzaamheid.nl/bibliotheek/docs/CoDPv1.pdf> [PDF]

5. Arturo Crespo, Hector Garcia-Molina: Cost-Driven Design for Archival Repositories. Joint Conference on Digital Libraries 2001 (JCDL'01); June 24-28, 2001; Roanoke, Virginia, USA.

<http://www-db.stanford.edu/~crespo/publications/cost.pdf> [PDF]

6. Rosenthal, D. S. H., Robertson, T., Lipkis, T., Reich, V., Morabito, S. (2005) *Requirements for Digital Preservation Systems: A Bottom-Up Approach*

D-Lib

11

,(11)

<http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>

7. *Measuring Our Value*, Results of an independent economic impact study commissioned by the British Library to measure the Library's direct and indirect value to the UK economy

<http://www.bl.uk/pdf/measuring.pdf> [PDF]

8. The *espida* Project:

<http://www.gla.ac.uk/espida/>

[See Exemplars and Further Reading](#)