erpa

Information Society Technologies

Cost Orientation Tool

September 2003

ELECTRONIC RESOURCE PRESERVATION AND ACCESS NETWORK

Introduction

Cost is one of the main criteria for people responsible for digital preservation when it comes to funding sustainable preservation infrastructure and related activities, and hence there is a need for understanding the scope, the different perspectives, and essential factors of inherent costs. These costs have to be assessed against the value of digital objects and the benefits of their preservation or the risks of losing them. This cost orientation tool will help to think through the costing issues involved in digital preservation. This survey of cost factors does not provide costing information. There is still a lack of sound costing information to build upon.

General considerations

Mostly, initial investments and ongoing costs are considered to be the two main categories of costs. New initiatives may require high investments because of the innovative nature of digital preservation facilities and the fact that in most cases no infrastructure is available.

There may be different levels of investment, though, depending on the ambition and the urgency. Initially and as a start it is possible to establish a temporary, but rather safe and stable environment for preservation with relative little funding. In later stages this can evolve and be expanded to a more sophisticated infrastructure. Sustainability, both technically and economically, of the established preservation environment therefore can also be seen as a combination of investments and ongoing costs.
The fact that information technology (IT) for preservation is still not mature and very much evolving is an important factor in defining preservation policies and deciding upon inherent investments.

The following aspects have to be taken into account with respect to costs for digital preservation:

- Costs include all activities from the creation and capture of the digital objects to their disposal (either destruction, transfer to another institution or long term preservation).
- Preservation is ongoing and active for the whole retention period of an object
- Preservation is typically a long-term venture that needs to be financially sustainable.

Finally, experience with digital preservation is still rather little, and insight into costs is often based upon experiences that are overhauled by new developments and therefore not always applicable or reliable any more. In the area of data curation centres, however, models are emerging.

An overview of cost factors

In the table cost factors are identified that should be taken into consideration. They should and can be integrated in the existing business context.
The factors are arranged around people, digital objects, laws and policies, standards, methods and practices, technology and systems, and organisation. Each of them has its own set of factors that will influence costs aspects. These cost factors are shortly indicated and described. Some considerations are given and if necessary interdependencies with other processes or cost influences are highlighted.
Although it still requires a kind of formula to be able to calculate the costs, this may be an approach to get a better view on costs. Also a distinction can be made between issues at a macro level, such as the technical infrastructure, staff etc., and issues at a micro level that regard each task or step in digital preservation processes.

Factors/issues		Cost impact	Considerations
Objects	Influence on creation	Less influence means mostly higher costs	Within an organisation it is more possible to influence creation of documents/information objects then from outside the organisation. Policies, standards, and procedures can help their creation and management. There are different levels of influence from none to complete.
	Existing	Existing objects require more work to prepare for ingest and storage	If objects already exist and for instance are made with proprietary software it requires more effort to prepare them for ingest. Also there is the risk of having no or insufficient contextual metadata.
	Complexity	The growing complexity of objects entails more maintenance	Complex objects often are multimedia objects or compound objects consisting of different types of formats. They can also have dynamic behaviour.
	Preservation period	Long term retention entails longer maintenance	It is important to keep objects no longer than required.
	Appraisal/value	The intrinsic value of the information objects has to be established. Consequent and sound appraisal also helps to reduce costs.	This goes hand in hand with identifying the risk of losing the object. Consequent appraisal policy helps in keeping only the necessary.
People	Skills	Specific skills are required	Examples of required skills are: - understanding of the nature of digital objects. - Understanding of available preservation methods - Etc. Available skills have to be multidisciplinary
	Quality	Well trained, skilful and experienced people re expensive, but will reduce overall costs	Quality is difficult to define and very much depending on the requirements and the business context (i.e. preservation) it is applied. This may include a sufficient number of people in order to achieve the set quality criteria.
	Training	Training needs to be up-to-date and adequate depending on the job-function. This requires funding.	Rapid developments in IT and new insights in digital preservation and strategies require continuous training

	Experience	With a growing level of experience, less inadvertent accidents will happen.	The discipline of preserving the information is leading above technology. However, without technology no preservation.
Standards	Standards	Use of **standards** will lower the effort of own development and, at the same time, facilitate long-living solutions. Format standards help in reducing the maintenance and ease migration procedures	Standards will not exist forever and there are many standards to choose from. Standards however help to reduce costs. The area of metadata requires special attention: in general one could say less metadata, less costs. Standardisation in metadata (both in syntax and semantics) also enables interoperability and effective interoperability enables easy and efficient information exchange, interaction, and business.
Practices	Workflow	Needs to be coherent and consistent	Procedures and preservation strategies are coming together in a workflow. Helps to shape good time planning
	Operation	Costs include people, material. Adequate and efficient software tools will minimise human intervention and accelerate processes.	Operation and performance have to be evaluated and audited on a regular basis.
	Processes	Each of the processes should be cost effective	This entails an analysis what steps have to be taken in each process and an assessment of the inherent necessary skills and tools (suitability of resources). Monitoring and evaluating the outcome in relation to resources spent is necessary.
Systems, Methods and technologies	Preservation method	Each type will have a different costing-profile. It may be necessary to employ **multiple** preservation methods in parallel.	Types of methods may be migration, use of XML or other standards, procedure for harvesting or capture, UVC-method, and so on. Reasons for a suite of methods may be a heterogeneous collection of different object types, or the necessity to enhance the probability of success in preservation.
	Validation of methods	**Validation of** the potential success of a method is necessary before using it.	Monitoring and validation has to be performed on an ongoing basis. It will support the quality of preservation of objects. At the moment preservation methods are still rather in their infancy and will change

	Sustainability	Methods should be chosen with the idea that they survive several generations of IT.	Essential for reducing costs. It is obvious that methods will change over time with new insights. In that case it should be easy to adapt them in practice.
	Portability	Methods chosen for preservation should allow easy portability to other or new system platforms	This goes for the method itself as well for the objects they support.
	Components	Technical infrastructure has to be implemented and maintained	What components are necessary is dependent on - The level of ambition - The complexity of the objects - The chosen preservation methods - Topology On occasion proprietary systems may appear cheaper, but in the long run they are more expensive (e.g. suppliers first want to establish client dependency).
	Maintenance	Maintenance of systems both in the sense of keeping them operational and in keeping them up-to-date	
	Operation	Costs include people, material. Adequate and efficient software tools will minimise human intervention and accelerate processes.	
	Flexibility	For maintenance over time and through changes in IT. Rapid change in IT requires flexibility in adaptation.	It has to be possible to add new components at low cost, to change (update) and to migrate them.
	Facilities	Location, security, safety, back-up (redundant storage)	
	Class of preservation	Distinction can be made between preservation of bitstreams and of functionality (which includes bitstream preservation)	Preservation of bitstreams is less costly and simpler to achieve, than preservation of functionality. The latter may include different levels of preserving functionality (e.g. preserving full functionality requires emulation of the original technical environment).
	Modularity	In the maintenance and replacement of (parts of) the infrastructure modularity may help to be cost-effective	
Laws and policie	Legislation	Organisations and their business may be subject to specific **legislation** that need to be accounted for.	These may include accountability, access, copy right and privacy issues.

	Policy	Each organisation should have a preservation policy to enable consistent and cost-effective management. Regular monitoring of policies and activities is necessary.	Policies have to be embedded in and connected to the business activities (including risk management). They updated with new insights and the evolution of technologies and preservation methods. Internal or external audits may help to keep policies adequate.
Organisation	Relationship building	Co-operation with other organisations may lower costs through synergies and economies-of-scale.	Co-operation can be achieved by collaborative research, joint ventures, sharing technical tools, infrastructure, as well as knowledge. More commercially approach is to outsource activities.
	Capacity building	Organisations should build the capacity for adequate preservation with respect to people, policies, practices, methods and technologies.	Capacity building depends on the needs of the organisation and its role and responsibilities.
	Responsibilities	Clear identification and assignment of responsibilities help to avoid misunderstanding and failure of procedures and systems.	This includes identification of all people involved.

How to collect cost information?

It will not always be easy to single out costs for preservation. It will differ for types of organisations as well as for different types of digital objects. For cultural heritage organisations it will be core business and therefore much easier to identify. For companies, such as pharmaceuticals, or for government administration it will not be core business and it may be more embedded in business processes.
At this moment, however, information of costs is lacking, because of the state of infancy in which the area of digital preservation still is, because of the rapid developments in IT, and because no sound cost models exist. Collecting costing information on a regular and continuing basis is therefore needed. The table above may be used to identify the different categories for this.

How to apply this tool?

This instrument is meant to get a better picture on the cost aspects involved in digital preservation. It does not provide calculation methods (or formulas). Every organisation will have to identify its own needs that will be dependent on the business context.
A possible approach may be:

1. identify business needs and scope of preservation (policy and risk questions)
2. identify laws (regulatory environment)
3. identify types of digital objects that will be created and need to be preserved (as well as how long they need to be preserved)
4. identify consequences for people and organisation
5. identify methods, standards, tools, technologies, systems to be used.

After having done this information may be collected about the costs involved for each of the components by inquiries with suppliers, similar institutions or organisations, special interest groups or networks and so on.

Some references

Kevin Ashley: Digital Archive Costs: Facts and Fallacies. Paper at the DLM Forum 1999. http://europa.eu.int/ISPO/dlm/program/abst_ka_en.html

Mary Feeney: The digital culture: maximising the nation's investment - Chapter 5: Estimating the costs of digital preservation. A synthesis of JISC/NPO studies on the preservation of electronic material. National Preservation Office, 1999. http://www.ukoln.ac.uk/services/elib/papers/other/jisc-npo-dig/

Stewart Granger, Kelly Russell, and Ellis Weinberger: Cost elements of digital preservation (version 4); October 2000.
http://www.leeds.ac.uk/cedars/colman/costElementsOfDP.doc

Brian F. Lavoie: The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making. White Paper; April 2003. http://www.oclc.org/research/projects/digipres/incentives-dp.pdf

Shelby Sanett: Toward Developing a Framework of Cost Elements for Preserving Authentic Electronic Records into Perpetuity. In: College & Research Libraries v.63,n.5; September 2002; p. 388-404.; ISSN 0010-0870.

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

Eugene Dürr, and Kees van der Meer: Emulation and Conversion: Organisational and Architectural Overview - way of working, costs, methods. Report at the E-Archive project; v1.2, December 2001. http://www.library.tudelft.nl/e-archive/Documenten/Resultaten/roquade2.pdf

Vanessa Griffin, Kathleen Fontaine, Gregory Hunolt, Arthur Booth, David Torrealba: Cost Estimation Tool Set for NASA's Strategic Evolution of ESE Data Systems. Paper presented at the CNES Symposium "Ensuring Long-Term Preservation and Adding Value to Scientific and Technical Data"; November 5-7, 2002. http://sads.cnes.fr:8010/pvdst/DATA/5-8_griffin.pdf

Tony Hendley: Comparison of Costs and Methods of Digital Preservation, Chapter 5: Developing a Cost Model. British Library Research and Innovation Report 106; 1998; ISBN 0 7123 9713 2. http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html

Making of America IV (The American Voice 1850-1876): Assessing the Cost of Conversion. University of Michigan, Digital Library Services; July 2001.
http://www.umdl.umich.edu/pubs/moa4_costs.pdf

Victor Zlotnicki, *SEEDS: Some thoughts on Data Management for NASA missions.* Presented 7 February 2002, Earth and Space Sciences Division and Physical Oceanography DAAC, Jet Propulsion Laboratory
http://lennier.gsfc.nasa.gov/seeds/WP_Zlotnicki.doc