



Costs and Benefits of a Shared Digital Long-Term Preservation System

Esa-Pekka Keskitalo

National Library of Finland,
PL 26, 00014 University of Helsinki,
esa-pekka.keskitalo@helsinki.fi

Abstract

This paper describes the cost-benefit analysis of digital long-term preservation (LTP) that was carried out in the context of the Finnish National Digital Library Project (NDL) in 2010.

The analysis was based on the assumption that as many as 200 archives, libraries, and museums will share an LTP system. The term 'system' shall be understood as encompassing not only information technology, but also human resources, organizational structures, policies and funding mechanisms.

The cost analysis shows that an LTP system will incur, over the first 12 years, cumulative costs of €42 million, i.e. an average of €3.5 million per annum. Human resources and investments in information technology are the major cost factors. After the initial stages, the analysis predicts annual costs of circa €4 million.

The analysis compared scenarios with and without a shared LTP system. The results indicate that a shared system will have remarkable benefits. At the development and implementation stages, a shared system shows an advantage of €30 million against the alternative scenario consisting of five independent LTP solutions. During the later stages, the advantage is estimated at €10 million per annum. The cumulative cost benefit over the first 12 years would amount to circa €100 million.

Introduction

Securing LTP requires planning of sustainable models for costs and funding that ensure the usability of information. The way of action must balance costs and benefits over a long period of time.

Digital preservation is characterized by its active nature. According to present understanding, digital materials cannot survive without constant assessment and prevention of risks. Resources, from electricity to highly skilled IT specialists, will be needed permanently.

A better understanding of the funding challenges of the future is urgently needed in order to make informed decisions about the direction of digital preservation — and in order to explain the feasibility, if any, of digital preservation to the decision-makers.

Background

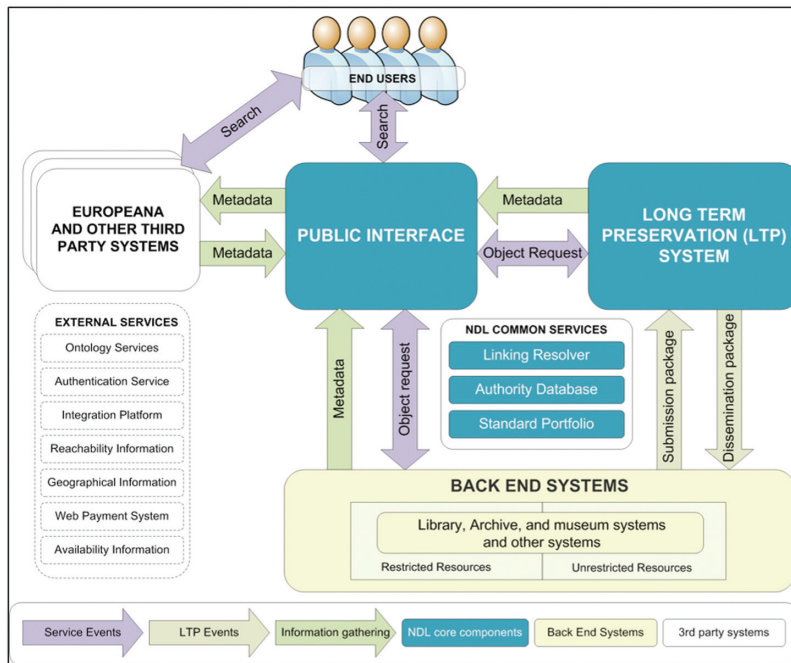
The National Digital Library Project

The NDL Project was launched in the Summer of 2008 by the Ministry of Education and Culture. Its objective is to improve availability and usability of digital materials offered by archives, libraries, and museums; and to develop an LTP solution for these materials. The Project is part of broader efforts to improve the national digital infrastructure and services.

The ultimate target of the NDL is to provide better conditions for research, learning, and general information retrieval, as well as to promote art and creativity. In addition, it pursues better productivity of the participating organizations. It is believed that through shared solutions it is possible to reduce overlapping costs in digitization, management of digital materials, and creating digital services, thus enabling reassignment of resources and more rational use of public funds. The NDL project is also aligned with the European Union's objectives for digitization of cultural materials and scientific information, their digital availability, and their long-term preservation.

The project has two major branches, or sections: the section of 'the public interface', and the section of 'long term preservation' (see Figure 1). This paper focuses solely on the latter [The entire NDL project was presented to the LIBER Annual Conference in 2010 (Hormia, 2010)]. The enterprise architecture of the NDL was published in 2010 (Enterprise, 2010).

Fig. 1: The National Digital Library.



Long-Term Preservation in the NDL Project

The idea of extensive cooperation between the archive, library and museum sectors on LTP goes back several years. In February 2007, the Ministry of Education and Culture set up a working group with the task to draw up national principles of LTP, and to prepare a first draft for organizing LTP, paying special attention to the benefits of a shared infrastructure and centralized

services. The working group reported in January 2008. It found that LTP is a challenge that must be looked at as a whole and that shared LTP systems and services are both feasible and desirable.

The LTP Section of the NDL Project continued along the lines of the 2007 working group. A report was submitted in June 2010 (Long-Term, 2010). The cost-benefit analysis discussed here is an annex to that report. The work on LTP has continued after the report with specialists preparing more detailed plans for the LTP system, its administration, etc. Final decisions on funding were still pending when this article was submitted.

The cost-benefit analysis was carried out by CSC — IT Centre for Science Ltd (www.csc.fi). It is a state-owned company that provides IT support and resources for universities, research institutes and companies. The analysis was supported by a number of surveys and sub-projects carried out by the NDL Project.

The author of this paper was closely involved in the process, being the representative of the National Library of Finland in the LTP Section of the NDL Project organization, and acting as a liaison with the library sector in preparing the cost-benefit analysis and supporting surveys.

Cost Analysis and Benefits Analysis

Economic theory offers many tools for *cost accounting*. The foundation of cost accounting is an understanding, as thorough as possible, of all cost factors and cost effects of a function through its life cycle.

The analysis of *life cycle costs* is an attempt to understand thoroughly the total life cycle of an operation and use the information in comparing options. Life cycle analysis also assists in timing and prioritizing future investments.

Cost-benefit analysis is a tool of financial decision-making. It can be used to compare options and to recognize beneficial and harmful factors in the operational environment. For example in policymaking, economic

and environmental benefits are converted into money values for easier comparison.

The cost analysis conducted in the NDL Project exploited the LIFE model (Lifecycle Information for E-Literature model) (Life, 2008) with modifications. The LIFE model has been created for the purpose of estimating storage costs of digital collections.

The Standard Cost Model for Assessment of Citizens' Benefits was used in analysing the benefits. It has been used in other evaluations of governmental IT projects in Finland, too. The model identifies qualitative and cost benefits not only for the organizations but also for the citizens, taking into account the number of transactions per citizen per annum, and the time and other resources consumed by a transaction. In this model, the emphasis of the benefit analysis was put on monetary benefits. The model identifies benefit factors and estimates their importance as well as the speed and probability of their realization. [For an introduction in English, see e.g. (van den Hurk, 2008).]

International Comparisons

LTP is a subject of avid interest and a number of projects around the world. National and other large-scale approaches are not uncommon. However, as far as costs and cost-benefit analyses were concerned, specific information was hard to come by: data were often scattered, preliminary, or they could not be disclosed. The case studies of the LIFE project were important, as was information obtained from the National Digital Heritage (NDHA) programme of New Zealand.¹

Basic Assumptions

The Concept of a Shared LTP System

The aim of the NDL project is to create a centralized system for digital LTP. This system should be able to provide preservation services to different kinds of organizations, curating different types of materials and having different customers.

The 2007 working group studied several options quite thoroughly and ended up recommending a centralized model. This model has also earned wide support among organizations participating in the NDL Project.

The cost-benefit analysis used a hypothetical model of two geographically separate places of operation that both have facilities for ingest, preservation, and dissemination of digital materials.

In the analysis, LTP is always understood to cover three aspects of preservation, namely

- bit-level preservation,
- preservation of interpretability, and
- preservation of original experience.

More information about the concepts used in the analysis is available in the Final Report of the LTP Section (Long-Term, 2010).

Users of the LTP System

The NDL is intended for archives, libraries, and museums supervised by the Ministry of Education and Culture. There are about two hundred such organizations. On the national level they include the Finnish National Archives Service, the National Library of Finland, the Research Institute for the Languages of Finland, the National Board of Antiquities, the Finnish National Gallery, the National Audiovisual Archive, and the Finnish Museum of Natural History.

Amount of Data

In 2009, two surveys were carried out to take stock of the size of the digital collections that might potentially be ingested into the LTP system.

The estimated amount of data for the present and near future is shown in Table 1. Typically, a small number of organizations provide most of the material in the library and archive sectors respectively, whereas in the museum sector digital materials are more evenly distributed over the organizations.

A question about the expected growth of digital collections from 2008 to 2025 yielded the following, very tentative, figures:

- archives, from 250 Tb to 3,000 Tb;
- libraries, from 65 Tb to 600 Tb; and
- museums, from 14 Tb to 270 Tb

All respondents felt very unsure about these figures. It must also be noted that data from scientific research are under-represented in the figures. It appears that at least the amount of scientific data in need of preservation is underestimated. Also, the survey omitted all potential partners not supervised by the Ministry of Education and Culture. Their inclusion could change the numbers significantly.

Parts of the surveys were repeated in late Spring 2011. According to the preliminary analysis some organizations have raised their estimates substantially. The study also showed that services related to *emulation* as a preservation strategy are not expected from the shared LTP system by potential participating organizations. The results will be published later in 2011.

Table 1: Materials in the NDL, estimated in 2008

Type	Amount of data (Terabytes)			
	2008	2009	2010	2011
Text documents	64	164	190	216
Still images	6	10	17	25
Moving image	11	25	31	37
Audio	17	23	28	33
Reference records	<10	<10	<10	<10
Web Archive	8	17	26	35
Radio and TV Archive	2	57	112	167
Total	108	296	404	513

Table 2: Assumed Building Phases of the shared LTP System

Year	Phase	Organizations	Materials
1	Functional requirements completed; developing tools and supporting services; planning acquisitions		
2	Developing tools and supporting services, putting out tenders, start piloting		
3	First back-end system connected; integration project; going into production	5 organizations 5 systems	300 Tb
4	1st production phase: first preservation location in production. — Increasing capacity; use and maintenance.	20 organizations 10 systems	700 Tb
5	1st production phase. — Preparing geographical expansion. 2nd phase tenders.	80 organizations 20 systems	1,000 Tb
6	2nd production phase: two locations in production. — Increasing capacity.	140 organization 30 systems	1,400 Tb
7–12	2nd production phase. — Capacity increases 15–25% annually; updating hardware and software; at year 11, replacing LTP software.	209 organizations 40 systems	4,000 Tb

Construction Stage of an LTP System

For the purposes of the cost analysis, it was necessary to make assumptions as to how an LTP system will take shape. We assumed that the system would take four years to go into production, and that it would be completed by year seven. Year-by-year assumptions as to the number of participating organizations and the amount of data were also necessary. These are summarized in Table 2.

Costs of the Shared LTP System: Methods

The LIFE model was the starting point for analyzing the cost of the shared LTP System. It is a life-cycle model for assessing present and future costs of LTP. The model identifies six stages of the life cycle — Creation/Purchase, Acquisition, Ingest, Bit-Stream Preservation, Content Preservation, and

Access. These stages help to position costs on a time scale and identify cost peaks. The LIFE model required slight modifications for the purposes of the present analysis. The modifications mainly concerned the question of repeated stages: in the transfer of materials to the shared LTP systems some components of the Ingest stage are repeated.²

Costs of Digital Preservation

General Assumptions Concerning Costs

The starting point of the analysis was that a shared LTP system would be built. The general architecture of the system was fairly well planned and agreed upon by the time the analysis was done.

Cycles of Hardware and Software Replacement

- Increase of disk space, 1 year
- Disk arrays, 3 years
- Servers, 3–5 years
- Tape robot, 5–8 years
- Tape drives, 3–5 years
- Network, 3–5 years
- System administration and control software, 3–8 years
- LTP software, 5–10 years
- Format-dependent software (for accessing preserved materials), 3–5 years

Amount of Materials to be Preserved

Building on the basic assumptions (above) it was further postulated that:

- All materials will be transferred to the LTP, although gradually.
- By end of 2011 the amount of materials is 700 Terabytes.
- The amount of materials will increase by 15% annually.
- File format migrations will happen every ten years for all kinds of materials, i.e. 10% of the total collection will be migrated every year. Both versions are supposed to be of equal file size, and both will be

preserved. Ten percent more storage space will be needed every year because of migrations.

- Updates (e.g. revisions of metadata) usually do not necessitate copying the objects and thus do not increase the need for storage space.

Cost of Human Resources

General statistics of 2009 were used as the source for HR costs and their rise over time. Further estimations were made about the costs of the workspace, employers' additional costs, costs of outsourced services, etc. The exact figures are very much dependent on local conditions, so they are not presented here.

Number of Organizations Using the Shared LTP System

It was assumed that all potential libraries, archives and museums would join the shared LTP System. Therefore, the following estimates were made of the number of participating organizations:

- Libraries: 30 organizations, 10 systems
- Archives: 15 organizations, 15 systems
- Museums: 164 organizations, 15 systems
- Total: 209 organizations, 40 systems.

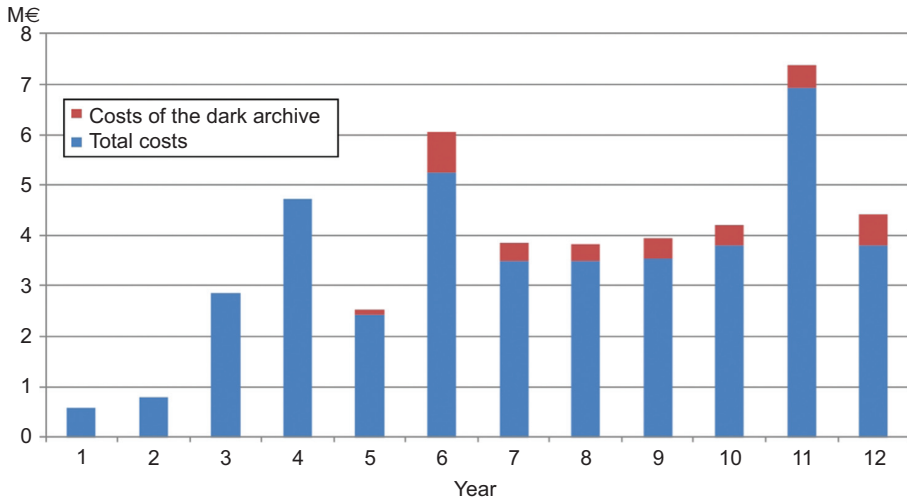
Costs of the Ingest Stage

Work done in the participating organizations was excluded from the calculations. The role of the LTP System is considered to be in support and consultation. After the initial phase, one person-year was thought to be sufficient.

Costs of the Bit-Level Preservation Stage

Materials should be stored in several copies, on different media and in different locations. In the hypothetical model, there were (from year six on) full LTP services available at two separate locations, each keeping digital objects at least in three copies on at least two types of media. The costs of one additional dark archive were calculated, too. The dark archive was understood to be a geographically separate place of storage only to be used internally by the LTP system.

Fig. 2: Total Costs of the Shared LTP System.



The access time to any materials was allowed to be 'a few seconds' at most.

Capacity, Purchase Costs, and Maintenance Costs of Storage Media

These costs were predicted to go down:

- The price of storage tape (relative to capacity) will drop by half every three years.
- The price of hard disks (relative to capacity) will drop by half every two years.
- However, same amount of money will be invested in new models of storage systems, thus increasing the capacity.
- From year seven on, the annual costs of the storage systems will remain stable. The demand for more space will be covered by decreasing prices.
- Consumption of electricity will decrease by 10% per byte per annum.

Other Assumptions

Maintenance of hardware, operating systems etc. will take five person-years, the dark archive adding one person-year.

There will be an 'LTP application' that provides tools for integrity control, digital signatures management and preservation procedures (such as file format migrations). Various ways of obtaining such a software system (licencing, building one internally, or any combination of these) were assumed to cost the same.

Costs of the Content Preservation Stage

It is important to note that an exact division of labour greatly influences the costs, as those costs incurred by the participating organizations are not shown.

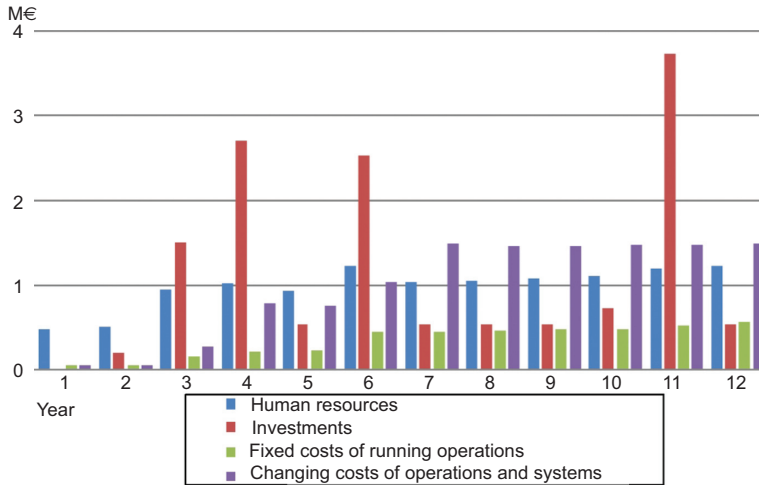
- It is assumed that materials will be ingested in the shared LTP system in an up-to-date format. File format migrations will become necessary by year eleven.
- Ten percent of the materials will be migrated to a new file format every year.
- There are 20 major types of file formats. Assessing their status and migrations will take about six person-years from year eleven on.
- Manual checking and assessing of materials is understood to be carried out by the participating organizations.

Costs of the Access Stage

These are thought to be relatively low.

The annual total costs of the shared LTP system are shown in Figure 2. The main breakdown of the costs is shown in Figure 3.

Fig. 3: Main Breakdown of Costs of the Shared LTP System.



Qualitative Benefits of Digital Preservation and the Shared LTP System

In choosing the methods, the NDL Project was informed by the eServices and eDemocracy Acceleration Programme (SADE), run by the Ministry of Finance. The SADE Programme has created a procedure for assessment of benefits, based among others on the Standard Cost Model for Citizens. According to the SADE procedure, the project focussed was on benefits the costs and benefits of which could be measured in money.

In the SADE procedure, a benefit index is calculated. On an a scale from 0 to 100, it reflects

- the importance of a beneficial factor,
- the time span of the realization of the benefit, and
- the probability of realization.

None of the benefits of a shared LTP system achieved the highest scores. It must be noted that the model is designed to strongly favour fast implementation and quickly realized benefits.

The analysis of the concept of a shared LTP system identified eight qualitative benefits, listed in Table 3 with their benefit index numbers.

Table 3: Qualitative Benefits of a Shared LTP System

Benefit	Importance	Time scale	Probability	Benefit index
(1) It supports organizations in fulfilling their legal obligations.	3	2	3	81
(2) It helps organizations to focus on core functions.	3	2	3	81
(3) It reduces overlapping operations.	2	2	3	63
(4) It makes it easier to bring skills and knowledge together so that they are not dependent on one person.	2	2	3	63
(5) It reduces risks of failure in digital preservation	3	1	3	63
(6) It creates better processes and services.	2	2	3	63
(7) It makes it possible for future generations to use and reuse the materials preserved.	3	1	3	63
(8) It enables seamless cooperation and share of resources across organizational borders.	3	1	2	38

Quantitative Benefits

Central to the SADE model is the translation of qualitative benefits into amounts of money. These monetary benefits may be divided into

- savings in performance;
- savings in comparison to alternative modes of operation;
- increased profits; and
- other savings.

In the case of the shared LTP system, the most visible monetary benefits are the following:

Savings in Using Shared Solutions (benefits 3 and 4)

The analysis indicates that savings of circa €30 million during the planning and implementation stages might be obtained by adopting the shared LTP system. When in production, the savings would be around €8 million per annum.

These calculations assume an alternative situation with five separate LTP systems. Savings are obtained in both human resources and in investments. For example, as software costs do not increase with the amount of materials preserved in a system, they are multiplied in the alternative scenario.

Savings in Costs of Preservation Management (benefits 2 and 6)

Savings in performance during the production phase are estimated at €2.5 million per annum. It is assumed that in the alternative scenario every organization designs processes for management, distribution and quality assurance, and that every organization allocates to these functions 60% of the hours of 2 persons.

Savings Created by the Value of the Digital Materials Being Preserved (benefit 5)

The savings created by the fact that digital preservation is indeed taken care of are shown to be €0.5 million per annum. Digital preservation produces benefits through reuse of materials in education, creative activities and research, as well as in digital products and services based on the materials preserved.

For purposes of the analysis, it was estimated that two percent of *digitized* materials would be destroyed or damaged annually without an LTP solution and thus should have to be redigitized. The cost of digitization was put at €1.3 per object. [The figure was based on data in (Numeric, 2009), and infor-

mation gathered from the organizations participating in the NDL Project.] The benefits of digitization per se (rather than preservation of the results), such as reduced costs of stacks and easier access, were not taken into consideration.

As far as born-digital materials were concerned, the savings were believed to be 'several million' euros per annum. As above, the rate of deterioration was assumed to be two percent annually. For many born-digital materials it is impossible to recreate them once they are damaged. The savings are the result of a lessening burden of curating damaged materials, and of the productive use of materials preserved. Monetary losses caused by damage to born-digital materials were, of course, hard to calculate. It was postulated that the costs are significantly higher than those of redigitization.

The final analysis shows a benefit of €30 million in the development and implementation phases when using a shared LTP system, compared to a model with many independent systems. During the production, the savings are estimated at €10 million annually.

Conclusions

Cost-benefit analysis of LTP proved difficult. That hardly surprised anyone, for the time-span of analysis is very long; future developments in information technology are relatively hard to predict; and comparable data from other projects were scarce. Lastly, the organizations that provided information for the analysis felt very insecure about their prognoses.

Throughout the analysis, it was obvious that mainstream tools of cost-benefit analysis are not particularly suited to an analysis of LTP, as they showed a tendency to favour quick implementation and quick returns on investment.

Nevertheless, even allowing for a wide margin of error, the results of the analysis are compelling. They certainly seem to confirm the widely held conviction that extensive cooperation of archives, libraries, and museums makes sense — at least in the field of LTP and at least from a financial point of view.

The main lesson from the NDL Project was that cost and benefit analysis of LTP could, in fact, be done. The particular figures from this exercise very probably become obsolete in the near future — some are obsolete already.

A more important and more permanent result is the framework for thinking about costs and benefits.

The next round will be easier.

References

Enterprise, 2010: The National Digital Library project/Ministry of Education and Culture: The National Digital Library — Enterprise Architecture. V1.0. 2010. <http://www.kdk.fi/en/enterprisearchitecture>.

Hormia, 2010: Hormia, Kristiina: Libraries, archives and museums working together! Learning by doing! Making the collections and services of libraries, archives and museums digitally available. Presentation slides, LIBER 2010. <http://www.statsbiblioteket.dk/liber2010/presentations/Hormia.pdf>.

Life, 2008: Ayris, P., Davies, R., McLeod, R., Miao, R., Shenton, H., Wheatley, P. : The LIFE2 final project report. The LIFE2 Project. 2008. <http://www.life.ac.uk/2/documentation.shtml>.

Long-Term, 2010: The National Digital Library project/Ministry of Education and Culture: The National Digital Library Initiative — Long-Term Preservation Project. Final Report. V. 1.0. 2010. http://www.kdk.fi/images/stories/LTP_Final_Report_v_1_1.pdf.

Numeric, 2009: Chartered Institute of Public Finance and Accountancy (CIPFA): NUMERIC — Developing a statistical framework for measuring the progress made in the digitisation of cultural materials and content. Study deliverable No 8: Study Report. Study findings and proposals for sustaining the framework. May 2009. http://cordis.europa.eu/fp7/ict/telearn-digicult/numeric-study_en.pdf.

van den Hurk, 2008: Joey van den Hurk, Peter Rem, Milan Jansen: Standard Cost Model for Citizens. User's guide for measuring administrative burdens for Citizens. Ministry of the Interior and Kingdom Relations, The Hague 2008. <http://www.whatarelief.eu/publications/standard-cost-model>.

Notes

¹ <http://ndha-wiki.natlib.govt.nz/ndha/>

² Unlike in the original model, in the analysis described here the Acquisition stage is understood to cover all steps before the transfer, including cataloguing. Ingest, on the other hand, only refers to the ingest into the LTP system.