

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
www.nationallibrary.fi

Abstract

This paper is a presentation of the cost-benefit analysis of digital long-term preservation (LTP) that was conducted as a part 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 a 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 a LTP system will incur, over 12 first years, cumulative costs of 42 million euros, i.e. in average 3.5 million euros per annum. Human resources and investments on information technology are the major cost factors. – After the initial stages, the analysis predicts annual costs of circa 4 million euros.

The analysis compared scenarios with and without a shared LTP system. The results point to remarkable benefits of a shared system. In the costs of developing and implementation stages, a shared system shows an advantage of 30 million euros, against the alternative scenario consisting of 5 independent LTP solutions. During the later stages, the advantage is estimated at 10 million euros per annum. The cumulative cost benefit over first 12 years would be circa 100 million euros.

Introduction

Securing LTP requires planning of sustainable models of 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 the present understanding, digital materials cannot survive without constant assessment and prevention of risks. Resources,

from electricity and cooling to highly skilled IT specialists will be needed permanently.

Better understanding of 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 it to the decision-makers.

Background

The National Digital Library Project

The NDL Project was launched in summer 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 a solution of LTP of these materials. The Project is a part of wider efforts of improving 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 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 in the 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 in the LIBER Annual Conference in 2010 [Hormia 2010].) The enterprise architecture of the NDL is has been published [Enterprise 2010].

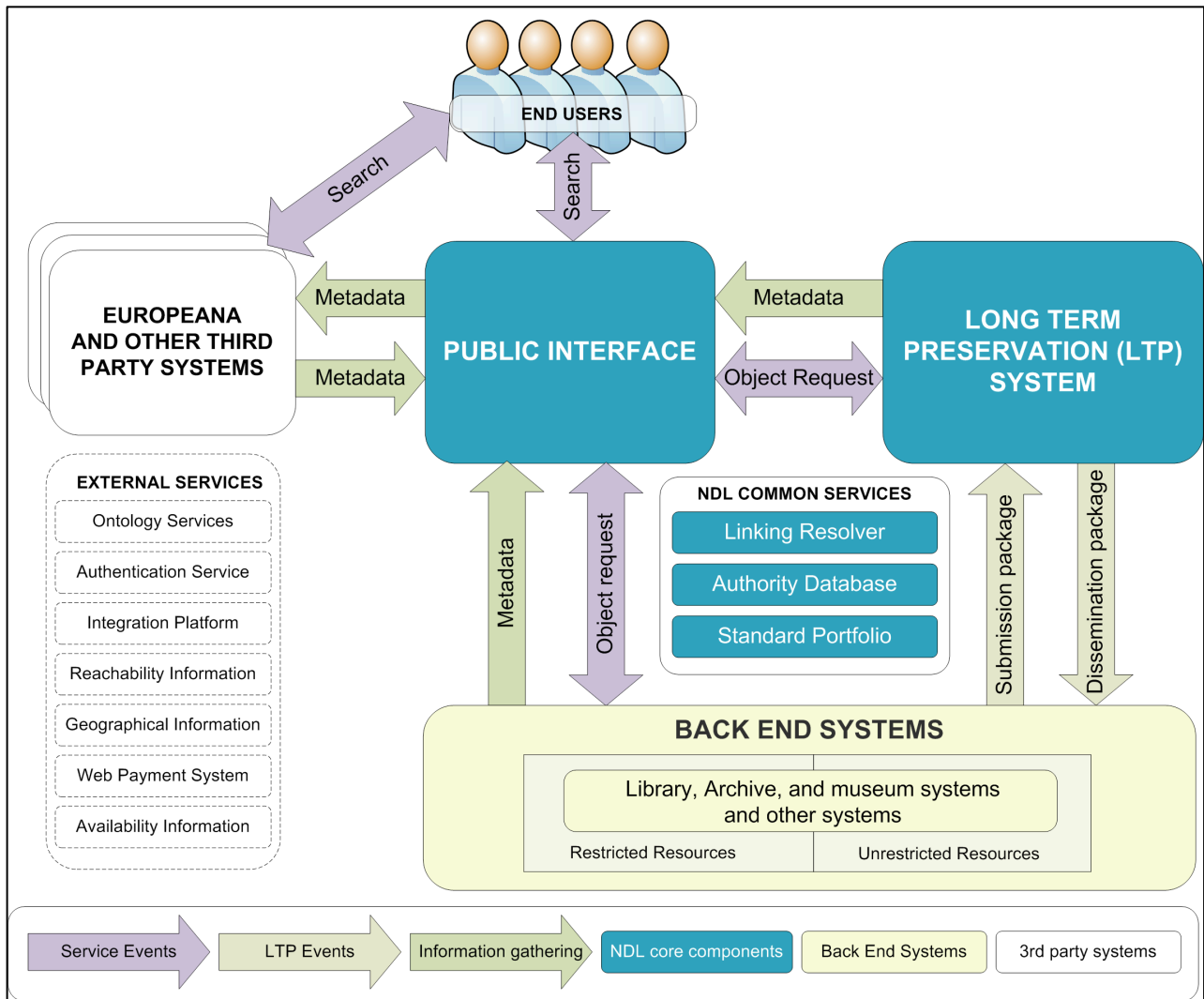


Figure 1. The National Digital Library

Long-Term Preservation in the NDL Project

The idea of extensive cooperation between 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 a task to draw up national principles of LTP, and to prepare a first draft for organizing LTP, paying special attention to benefits of 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 a shared LTP systems and services are both feasible and desirable.

The LTP Section of the NDL Project continued on 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 writer 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 of the library sector in preparing the cost-benefit analysis and supporting surveys.

Cost Analysis and Benefit Analysis

Economics 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 cycles.

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 policy-making, economic and environmental benefits are converted into money values for easier comparison.

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

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 that identifies qualitative and cost benefits not only for the organizations but also for the citizens, taking into account number of transactions per citizen per annum; and time and other resources consumed by a transaction. In this model, emphasis of the benefit analysis was put on monetary benefits. The model identifies benefit factors and estimates their importance as well as quickness 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 cost and cost-benefit analyses were concerned, data was often scattered, preliminary, or it could not be disclosed. Case studies of the LIFE project were important, as was information obtained from the National Digital Heritage (NDHA) programme of New Zealand.¹

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

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 options quite thoroughly and ended up to recommend 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 readiness of 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 concept used in the analysis is available in the Final Report of LTP Section [Long-Term 2010].

Users of the LTP system

The NDL is intended for archives, libraries, and museums in the field of 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, 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 made to the size of digital collections that might potentially be ingested to the LTP system.

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

Question about growth of digital collections from 2008 to 2025 yielded 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

Scientific research data is badly represented in the figures.

All respondents felt very unsure about these figures. It seems 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 estimations materially. This study also showed services related to *emulation* as a preservation method are not expected from the shared LTP system by potential participating organizations. The results will be published later in 2011.

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 1. Materials in the NDL, estimated in 2008

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	1 st production phase: first preservation location in production. - Increasing capacity; use and maintenance.	20 organizations 10 systems	700 Tb
5	1 st production phase. - Preparing geographical expansion. 2 nd phase tenders.	80 organizations 20 systems	1,000 Tb
6	2 nd production phase: two locations in production. - Increasing capacity.	140 organization 30 systems	1,400 Tb
7-12	2 nd production phase. - Capacity increases 15-25% annually; updating hardware and software; at year 11, replacing LTP software.	209 organizations 40 systems	4,000 Tb

Table 2. Assumed Building Phases of the shared LTP System

Building stages of a LTP system

For the purposes of the cost analysis, it was necessary to make assumptions on how a LTP system will take form. In these assumptions, the system would take 4 years to go into production, and it

would be complete by year 7. Year-by-year assumptions of the number of participating organizations and amount of data were also necessary. These are summarized in Table 2.

Costs of the Shared LTP System: Methods

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 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. LIFE model needed some slight modification for the purposes of the present analysis. Mostly it the question of the repeated stages: transfer of materials to the shared LTP systems repeats some features of the Ingest stage².

Costs of Digital Preservation

General Assumptions Concerning Costs

The starting point of 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 (see page 3) it was further postulated that:

- All materials will be transferred to LTP, although gradually.
- By end of 2011 the amount of materials is 700 Terabytes
- The amount of materials will increase 15 % annually.

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 hands means only the ingest to the LTP system.

- File format migrations will happen every 10 years for all kinds of materials, i.e. 10 % of the total will be migrated every year. Both versions are supposed to be of equal file size, and both will be preserved. 10 % more storage space will be needed every year because of migrations.
- Updates (e.g. revisions of metadata) usually do not necessitate recopying the objects and thus do not increase the need of 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 costs of workspace, employer's additional costs, costs of outsourced services, etc. The exact figures are very dependent of the 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, following estimations about the number of participating organizations were made:

Libraries	30 organizations, 10 systems
Archives	15 organizations, 15 systems
Museums.....	164 organizations, 15 systems
<u>In total.....</u>	<u>209 organizations, 40 systems</u>

Costs of the Ingest Stage

Work done in the participating organizations was excluded from calculations. The role of the LTP System is considered to be in support and consultation. After the initial phase, 1 person-year was thought 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 6 on) full LTP services available at two separate locations, each keeping digital objects at least in three copies at least on two types of media. The costs of one additional dark archive were calculated, too. The dark archive was understood to be geographically separate place of storage only to be used internally by the LTP system.

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

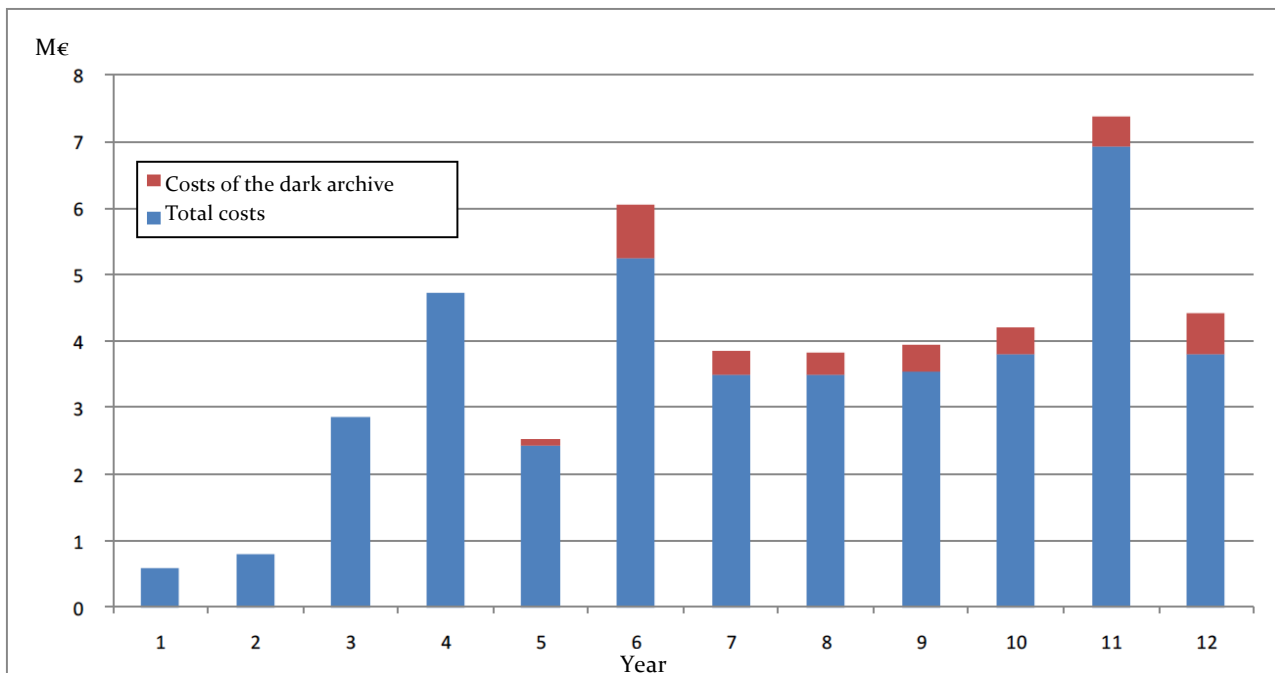


Figure 2. Total Costs of the Shared LTP System

Capacity, Purchase Costs, and Maintenance Costs of Storage Media

These costs were predicted to go down:

- Price of storage tape (relative to capacity) will half every three years.
- Price of hard disks (relative to capacity) will half every two years.
- However, same amount of money will be invested in new models of storage systems, thus increasing the capacity.
- From year 7 on, the annual costs of storage systems will remain stable. The demand for more space will be covered by cheapening prices.
- Consumption of electricity will decrease by 10 % per byte per annum.

Other Assumptions

Maintenance of hardware, operating systems etc. will take 5 person-years, the dark archive adding one person-year.

There will be a "LTP application" that provides tools for integrity control, digital signatures management, and preservation procedures (such as file format migrations). Different 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 notice that the exact division of labour influences greatly the costs, as those incurring in 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 11.
- 10 % 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 6 person years from year 11 on.
- Manual checking and assessing of materials is supposed to happen in participating organizations.

Costs of the Access Stage are thought to be relatively small.

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

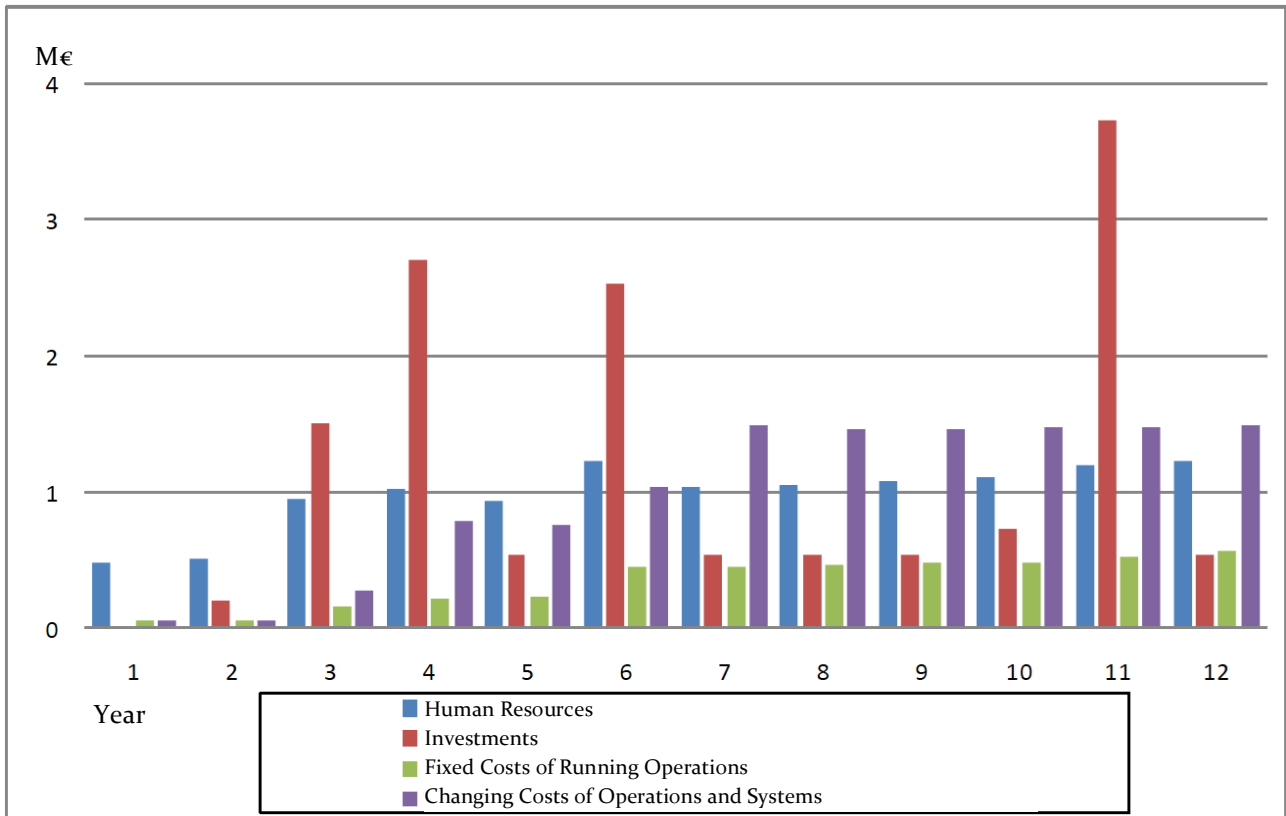


Figure 3. Main division of Costs of the Shared LTP System

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, focus was on benefits whose costs and benefits 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 the realization.

None of the benefits of a shared LTP system achieved highest scores. The model is designed to heavily reward 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.

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 gather skills and knowledge so that they are not dependent on	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 to the future generations to use and reuse the materials pre-	3	1	3	63
(8) It enables seamless cooperation and share of resources over organizational borders.	3	1	2	38

Table 3 Qualitative Benefits of a shared LTP system

Quantitative Benefits

In the SADE model, it is central to translate 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, 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 euros 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 euros per annum.

The calculation assumes an alternative case of 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 euros per an-

num. It is assumed that in the alternative scenario every organization creates the processes of management, distribution and quality assurance; and that every organization allocates to these functions 60 % of person hours of 2 persons.

Savings created by value of digital materials being preserved (benefit 5)

The savings of the fact that digital preservation is indeed taken care of are shown to be 0.5 million euros per annum.

Digital preservation produces benefits through reuse of materials in tuition, creative activities and research, as well as in production of digital products and services based on materials preserved.

For purposes of the analysis, it was estimated that 2 % of *digitized* materials would be destroyed or damaged annually without a LTP solution and should be redigitized. The cost of digitization was put at 1,3 € per object. (The figure was based on data in [Numeric 2009], and information gathered from the organizations participating in the NDL Project.) Such benefits of digitization (rather than preservation of the results) as reduced costs of premises and easier access, were not taken into consideration. It should be noted that benefits of digitization, such as reduced costs of premises and easier access, were not taken into consideration, either.

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 2 % annually. Some materials may not be recreated. The savings are caused by lessening burden of curating damaged materials, and of the productive use of materials preserved. Monetary losses caused by damages to born-digital materials were naturally hard to calculate. It was postulated that the costs are significantly greater than those of redigitization.

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

Conclusions

Cost-benefit analysis of LTP proved difficult. It hardly surprised anyone. The time-span of analysis long; the future development of information technology is relatively hard to predict; and comparable data from other projects were scarce. 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 do not always lend themselves easily to analysis of LTP. The models used were tuned to reward quick implementation and quick returns on investment.

Nevertheless, even allowing 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 for the NDL Project was that cost and benefit analysis of LTP could be done. The figures very probably become obsolete in near future – some are obsolete already. More important and more permanent result is the framework for thinking about costs and benefits.

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 N° 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>