INVESTIGATING THE COST OF DIGITAL PRESERVATION.
A SUMMARY

What is the cost of maintaining the long-term availability of digital materials from the domains of archives, libraries, culture, science, and media? In 2016 and 2017, the Digital Heritage Network researched the matter. This document summarises their approach as well as their findings.

Objective of the study
The volume of digital information is increasing rapidly and hence the cost of sustainable management and preservation of digital heritage. This increases the need to know and control costs. In many cases, at least until recently, digital preservation has been project-based. Funding flows are often determined by these projects, which means that resources for long-term management are not always structurally available.

In 2016, the above gave rise to the start of the project ‘Insight into the cost of sustainable access’. It is part of the Sustainable Digital Heritage program of the Digital Heritage Network (NDE), which is coordinated by the Netherlands Coalition for Digital Preservation. Its objective is to visualize and control the cost of digital preservation, both at the institutional and domain-wide levels.

That objective is divided into four parts:
1. Acquiring more insight into the cost of sustainable access to increase grip and control;
2. Being able to compare the cost of digital preservation between institutions. This increases the knowledge about cost factors as well as grip and control of costs;
3. Being able to make statistical forecasts;
4. Being able to make strategic choices based on the three above-mentioned aspects, at the institutional and domain-wide levels.

Developing a cost pricing model
The first stage of the project was dedicated to developing a cost pricing model (with BMC Consulting) that could provide the institutions concerned with meaningful insight into the cost structures, the composition of those structures and the cost drivers. The model provides insight into the hard financial figures of management, retention and access of digital heritage. By linking the cost of digital preservation to the cost drivers, the model can perform statistical forecasting to substantiate strategic choices. The model builds on the international CCEx model developed within the framework of the 4C project.

The cost pricing model can now be used by heritage institutions to map out the cost of sustainable access to digital collections and compare it to other institutions.

First group of data is in
In 2016, the data of nine Dutch heritage institutions were fed into the cost pricing model. These figures provide a good first impression of the current cost structure of digital preservation. A cost comparison was also carried out. Since only a limited number of institutions have been entered, the results of the comparison are indicative only and not representative of the heritage sector. However, the experience of these institutions has shown that the cost pricing model certainly has potential and can be a useful tool for comparing the cost of digital preservation with other institutions, in addition to insight into an institution’s own cost structure, and for calculating correlations and regression to explain the outcomes for digital collections and compare it to other institutions.

www.den.nl/nde • www.ncdd.nl
Research method

The following diagram shows the various stages of cost research:

Participants in the research were the National Archives, the National Library, EYE Film Museum, Rotterdam City Archives, the Historical Centre of Overijssel, the International Institute of Social History, SURFsara, DANS and The New Institute.

Potential of the cost pricing model

Applying the cost pricing model for digital preservation has a number of benefits, both for individual institutions and for collaboration within and among domains:
- The model provides insight into an institution’s cost of digital preservation and reinforces its grip on it.
- It provides insight into the cost drivers. If the model is fed data for several years, it is possible to discern the impact of the various drivers and their correlations, if any. With sufficient data present, the model can also provide statistical forecasts.
- Using the model leads to an increased awareness of the relationship between the primary process (collections and digital preservation) and the financial function, and vice versa, and to prioritization and awareness of the importance of finance within the organisation.
- It offers the ability to compare costs to other institutions that have been using the model (within and between domains) and to search for collaboration based on the insights gained.
- It provides insight into the differences among cost drivers, contexts and objectives of the institutions. This enables institutions to understand choices they have made, revise them if necessary, and base their strategy upon accordingly.
- Finally, it is possible to link the data to the CCEx model database and thus make international comparisons. The figure below shows the structure of the cost pricing model for digital preservation:

The Dutch model versus CCE

The Dutch cost pricing model for digital preservation builds on the CCEx model developed by 4C. However, the Dutch model goes one step further. First, the cost is described in more detail based on Activity Based Costing. This gives organisations more insight into costs and more control of them. Secondly, the model links costs to cost drivers (for example personnel costs when checking metadata or the cost of storage per TB of data), which makes it possible to draw conclusions about interrelationships. In other words, how does a change in the qualitative factors affect the costs? This helps to determine the financial consequences of certain developments, which may be interesting in strategy formulation. This is also a first step towards determining the social efficiency of digital preservation, for example by making social cost-benefit analyses.
Findings and conclusions

1 Determining the cost of digital preservation is still in its infancy
The research shows that most institutions are still beginning to develop their digital preservation, particularly in terms of financial engineering. Only the larger institutions with more digital preservation experience have a relatively good picture of the cost, and the cost drivers, of the process. For medium-sized and small participating institutions, painting the full financial picture was a considerable effort or even proved impossible.

2 The cost pricing model for digital preservation provides insight into the costs and promotes management and control
Participating institutions indicate that the model has provided more insight into the cost of digital preservation for their own organisation. Feeding the model with data has made the participants more aware of the layout of their primary process and their financial function. The model also provides management information. This allows for more grip on processes and associated costs. For multiple institutions, the model’s Activity Based Costing required a new way of accounting, which is now mostly focused on financial accounting principles rather than substantive governance. In case of substantive control, the costs are derivative.

3 There is as yet insufficient critical mass
At the time of writing this report, there were not enough participating institutions to enable the comparison of the cost of digital preservation between institutions (Objective 2) and to calculate correlations and regression and thus explain the outcome (Objective 3). As a result, the potential of the model cannot be fully utilised. A Roadmap has been developed to create the critical mass that is required to gain insight into the cost of digital preservation at a domain-wide level.

4 Costs related to digital preservation are at the forefront of the process and in coordinating process activities
A large part of the costs are incurred in the Selection/Pre-Ingest and Ingest stages. This means that if the data provided are selected, ordered and described better by the providers or caretakers, the institutions can reduce costs. In addition, we see that the costs largely relate to the coordinating process activities (Metadata, Preservation Management, Infrastructure and ICT). This is because many institutions are still investing in digital preservation and the associated infrastructure. This involves depreciation costs. We observe that most institutions act on their own in this regard. Collaboration within domains is beginning to take place, but inter-domain collaboration hardly occurs at all.

5 Costs related to preservation actions are not yet visible
Costs incurred for performing preservation actions, such as emulation and migration, are not – yet – visible. The reason is simple: these costs have not been made by the institutions concerned at this time. These costs are expected to show up in the model as soon as practice in this area changes.

6 Personnel costs are a significant component
The work to be performed in processing digital collections proves to be quite labour-intensive. Personnel costs make up over 50% of total costs in most of the institutions surveyed. The relatively high proportion of personnel costs makes it important to share knowledge with other institutions. This may allow for better management of personnel costs and optimization of efforts.

Follow-up

Based on these findings, a Roadmap for subsequent steps has been made. It describes the steps to be taken in various phases. Its aims are to further develop the cost pricing model itself and to strengthen collaboration among institutions and domains by increased understanding of the cost of sustainable preservation.

The first phase was completed in 2016 and resulted in the cost pricing model, which has been filled out by nine heritage institutions. The second phase, starting in 2017 and running until the end of 2018, is aimed at increasing the number of participants in the model and setting up recurrent data monitoring in order to highlight long-term developments.
This will increase the volume of data so that cost effects can actually be calculated from substantive choices (‘turning the knobs’) and the model’s potential can fully be exploited. The third phase, from mid-2018, will be aimed at better matching demand and supply of sustainable preservation facilities by analysing the figures collected in the model. This will lead to greater insight into the effects of strategic choices (which knobs can be turned). Finally, the fourth phase should make it possible to make social cost-benefit analyses with which to further strengthen strategy formation at heritage institutions.

More information

If you would like to know more about this project or if you want to offer your data to the cost pricing model, please contact Marcel Ras, Program Manager NCDD/ Digital Heritage Network, by email: marcel.ras@ncdd.nl.

You can also check out the page Costs of sustained access at the NCDD website and the factsheet Tools and tricks for sustained access to your collections (PDF), listing all results from the Sustainable Digital Heritage work package (2015 – 2017) in two pages.