TOGETHER WHERE POSSIBLE, INDIVIDUALLY WHERE NECESSARY

Constructing a network of nationwide facilities together

NCDD, 2015
TOGETHER WHERE POSSIBLE, INDIVIDUALLY WHERE NECESSARY

Constructing a network of nationwide facilities together

Summary and conclusions of an investigation into a national infrastructure for sustained access to digital information

Marcel Ras, Marco de Niet, Joost van der Nat. March 2015
SOGETHER WHERE POSSIBLE, INDIVIDUALLY WHERE NECESSARY

Constructing a network of nationwide facilities together

In 2014, the National Coalition for Digital Preservation (NCDD) carried out a study into the possibilities for developing a network of nationwide facilities at public organisations for ensuring sustained access to digital information in the Netherlands. Based on concrete and attainable scenarios, these organisations and their clients can make well-founded policy choices, rearrange resources, and better designate responsibilities on a national scale. At the same time, the Cultural Coalition for Digital Preservation (CCDD) investigated the state of affairs concerning sustained access to “born digital” heritage in the cultural sphere. This research produced an inventory and assessment of the different approaches to preservation and accessibility of born digital cultural heritage. The realms of art, film, photography and architecture were examined in particular detail. Both research projects were carried out with financial support from the Dutch Ministry of Education, Culture and Science.
Sustained digital access

It is up to collection managing institutions to ensure that meaningful objects and information concerning the world we live in remain accessible for posterity. For years, clear demarcations separated different types of objects and their collection: you visited an archive for source research, a library to borrow books, and a museum to admire historical objects. But thanks to rapid computerisation and the development of the Internet, we have now started to produce entirely new objects in digital form. Over a relatively short period of time, huge amounts of digital sources have been created in all sectors, in part through the large-scale digitisation of authentic sources, and in part through the accretion of born digital objects. Using computers, we produce documents, photographs, films, music, art, databases, games, websites, blogs, tweets, multimedia applications, etcetera. Digitisation also means that other forms of access have become possible. Collections are now accessible through large, overarching portals or linked data, and are easily connected to each other. This has opened up unprecedented possibilities for research and the use of these collections. The use of digital techniques and advanced search capabilities can greatly enhance a collection’s societal value.

The documentation of modern life occurs in various digital forms, is often complex, and comprises an enormous amount of data, ranging from official government documents to artists’ personal email correspondence, and from tweets on important events to news websites. If we do nothing, we risk losing a large proportion of today’s history. Where previously paper, for instance, was literally patient, today’s bits and bytes require constant attention. We will need to
ensure that the digital objects created today are preserved and remain accessible in the future. This task raises new questions on the topics of authenticity and reliability, intellectual property, selection, privacy, and internationalisation.

The collection managing institutions of the Netherlands acknowledge their responsibility to guarantee the long-term accessibility of valuable digital sources. The challenges, however, are huge: continuous technological development, rapidly increasing volumes, the high costs of preservation, demanding user groups, and the growing dependence on technology necessitate making certain choices. Dialogue with data producers is also of increasing importance.

Achieving digital sustainability may justifiably be called one of the greatest challenges of our generation. We are the first to have had to develop methods and systems on a large scale to keep digital objects accessible in an authentic form for the future. After a few years of learning and pioneering work, we have now entered a new era. Poor digital preservation will cause irreparable damage to the public domain and the missions of heritage and knowledge institutions. In this era, the loss of digital objects means a loss of history, and lax handling of research data spells disaster for science and may lead to data fraud. Data value increases as accessibility is enhanced, and scientific transparency is served by the professional data management and archiving.

**Collaboration**

The NCDD was founded in 2008 by a group of public organisations which consider the long-term preservation of digital data collections to be one of their core tasks. The coalition aims to construct a shared organisational and technical infrastructure to ensure the long-term usability of digital files. This cannot be done all at once. A realistic approach would involve a gradual build-up, taking into consideration the differing responsibilities, roles, velocities, and means of all parties involved.

The Cultural Coalition for Digital Preservation (CCDD) is a network of professional cultural institutes working on sustained access in the cultural sector. In this sector, more and more collections are now digitised, and increasing numbers of digital objects are being produced and archived. Consider digitally created art, for instance, or digitised paintings, photographic collections, documents, and large databases of digital information on heritage collections. This requires heritage institutes to accumulate knowledge and expertise on sustained access to these digital collections and information. Preferably, this should occur internally. Some institutes have been working on this issue for some time, while others have only just started.

Over the past few years, important steps have been taken. Methods and systems have been developed to keep digital objects accessible in an authentic form for the future. At the start of the 21st century, the National Library of the Netherlands constructed an operational e-depot for digital publications. The Netherlands Institute for Sound and Vision is a leading organisation at
home and abroad in the sphere of management and preservation of digital audio-visual heritage. The National Archives are expanding their current e-depot facilities into a shared infrastructure for all digital archives in the Netherlands. DANS possesses a digital archive for research data going back to the 1960s.

The facilities developed so far have been created on a project basis and in large part with external financial support. Most of these facilities are now operational or even in need of a second generation. At this moment, it is essential that the operationalisation of management and preservation is stimulated, expanded, and fortified at a national level, with an eye to the international framework. This is a joint task for the government, NCDD, and other concerned parties. Further operationalisation and expansion is an important step forward in the public mission of the institutes involved, and certainly also for the thousands of other organisations in the spheres these leading institutes operate in.

This step forward will require more than just internal policy choices or rearrangement of resources by the leading institutes. The creation of a stable organisational and technical infrastructure which ensures the preservation and sustained accessibility of digital information is of crucial importance to science, culture, and society. This is beyond the individual capacities of the institutes involved. Collaboration increases efficiency, offers access to knowledge present elsewhere, ensures greater profits are made from achieved results, and allows everyone involved to better connect to large-scale digital developments. It also makes it easier to determine how to deal with the interweaving of different types of material. The boundaries between data, documents, and publications are blurring, and old definitions are becoming obsolete. Developments in the field of open data only accelerate this process. On the one hand, we must prevent activities from being duplicated. On the other hand, gaps must not be allowed to appear in collection policies at the national level.

An investigation into a national infrastructure for sustained access

Cross-domain collaboration, with an eye to the differences between the domains (“Together where possible, individually where necessary”), lies at the heart of the NCDD’s approach. Without collaboration, individual institutes will keep reinventing the wheel and expanding their digital archives in inefficient ways. This would mean a duplication of work and a waste of time, money, and energy. Economies of scale make it easier for the many smaller Dutch institutes to profit from available facilities, services, and knowledge as well.

In the past few years, major steps forward have been taken: operational e-depots have been developed and put to use, much technical knowledge has been obtained, and we’ve seen some good examples of services to third parties: services provided by one collection managing institute to another, but also public-private collaboration. On the other hand, much is still poorly understood. How large is the Netherlands’ digital collection now? What part of it has been properly archived – and what exactly can be considered “properly archived”? What are the
needs of Dutch collection managing institutes in the sphere of sustained digital archiving? What are the costs to be considered? And what is the scalability of our facilities?

With this research, the NCDD has attempted to develop a clearer picture of the necessary stable organisational and technical infrastructure ensuring preservation and sustained accessibility of digital information. What would such a nationwide infrastructure look like? How do we ensure its scalability? What is the best approach to realising this infrastructure? If we assume this is a growth model, what would this model look like? And which steps will we need to take in both the short and long terms?

The research in three phases

In the desk-based first phase of the research project, a working definition of infrastructure was established. What do we mean when we refer to a “stable organisational and technical infrastructure for sustained access to digital information”? We can be sure that an infrastructure is more than a collection of “cords and sockets”. It does not just comprise the technology, but also, crucially, the organisational facilities surrounding that technology. Naturally, it is important to make qualitative technical arrangements when it comes to matters like storage, networks, hardware, software, and applications. On top of that, however, we are concerned with “digital information and archive management”, the primary process by
which digital objects are processed, stored, and managed. To properly implement this process, development and exchange of knowledge is necessary, as are research and training. These aspects are very closely linked with the preservation policies of the individual institutes. A digital archive needs to be trustworthy for it to last: grant providers, data producers, depot donors, and users must be able to trust the managers of a digital archive. They must be able to assume that their digital collections are safely managed and kept accessible by collection managing institutes. As such, the quality and reliability of work processes and management systems will have to be subjected to standardised regulations. To this end, certification trajectories have been created, which are determinative in the construction of an infrastructure. Finally, legislation, tasks and responsibilities also influence the nature of the infrastructure.

These considerations have resulted in the following analytical framework for an infrastructure, as used in this research.

Figure 1
Elements of an infrastructure
The second phase of research involved fieldwork. Extensive interviews were conducted with representatives of a number of large institutes, among them NCDD coalition partners. The goal was to obtain insight into the state of the currently existing infrastructure and to sketch out a common vision for a desirable future situation.

All interviewees were asked about their positions on various infrastructural elements, both in the present state of affairs and in an ideal future. Which elements of the infrastructure, as defined by this research, are currently present, and how are they executed? The current facilities, often controlled by individual institutes, were discussed in detail, as well as the possibilities for having parts of these facilities provide services to third parties. Which infrastructural elements can be jointly developed and managed?

These interviews have provided us with a detailed image of the possibilities for collaboration. A number of facilities should obviously be handled individually by each institute. Ingesting digital files, for example, is part of the core task of collection managing institutes, and as such is handled internally. This will not change. Supporting processes like preservation planning and preservation watch, however, are highly suitable areas for collaboration. The ICT elements, too, offer great opportunities for working together and making use of shared facilities. Hardware, storage, and technical management especially are facilities that can be administered in a network, instead of being constructed or purchased separately by each institute. The same goes for knowledge development and training, certification, and the creation and management of persistent identifiers.

A concrete picture emerges of which infrastructural elements could be shared, which is a key factor in creating an organisational and technical infrastructure for sustained access. It might be better to speak of a network of nationwide facilities, as large parts of this national infrastructure have already been realised. We do not, therefore, need to develop an entirely new infrastructure from scratch. Existing facilities must instead be shared and opened up to third parties. In this way, a network of facilities will gradually materialise.

**Growth scenario**

Finally, in the third phase of our research, based on the analytical framework and the field results from the previous two phases, a number of possible scenarios for developing a nationwide infrastructure were explored.

The central tenet of this research project (and of the NCDD as a whole) is that collaboration will increase effectiveness and efficiency. The aim is to allow every organisation in the Netherlands using digital archives to render their digital objects accessible, and to keep them that way, at the lowest possible cost. Through economies of scale the cost efficiency of work can be greatly enhanced.

These scenarios were developed based on the assumption that the more we collaborate, the greater the increase in effectiveness and efficiency will be. It must be said that in practice,
this assumption is not always valid. Above all, it is important to differentiate between administrative and operational responsibilities. The first are fixed by tasks and legislation, while the second may be outsourced where possible. A cross-domain approach is not always feasible, for various reasons:

- The nature of the digital material involved requires different preservation strategies.
- Standards for publication of material differ between domains.
- Different demands, including legal requirements, apply to the selection, supply, and storage conditions of the material.
- There are domain-specific service needs.
- A domain is heavily fragmented.
- Specific agreements were made on the funding of digital archiving within a domain.

In each case, then, the frameworks for collaboration will need to be closely examined. This brings us to the approach advocated by the NCDD: Together where possible, individually where necessary.

From the various scenarios explored, one preferred scenario for a network of distributed national facilities has emerged. A distinctive feature of this scenario is that institutes retain their own facilities where necessary. Wherever possible, facilities are shared.

This scenario provides for three trajectories:

- The development of a network of technical facilities. Storage facilities and hardware, among other things, belong to this bottom layer.
- Assistance and consultation. This involves services like training, research, persistent identifiers, and certification.
- Distributed applications for digital preservation.

This scenario is a growth model in which the three trajectories can operate in parallel. Ultimately, it is aimed at creating a network of facilities capable of substantiating collaboration to fulfil existing needs. As a caveat, the existence of a technical infrastructure (A) is a prerequisite for a distributed network of facilities at the application level (C). The first steps towards achieving (A) have already been taken: the Netherlands Institute for Sound and Vision offers sustained storage and access services to third parties through their e-depot. The government’s Consolidation of Data Centres program will create a nationwide data provision service of four or five data centres by 2020. DANS is collaborating with 3TU and SURFsara (Research Data Netherlands) to realise a federative data infrastructure for sustained access to research data. This model includes both technical and organisational facilities. Along the same lines, the Archief2020 program is developing shared facilities for the archiving domain.

Figure 2 represents a distributed network of nationwide facilities, showing all infrastructural elements and the collaborating institutes’ level of involvement in each. Green boxes represent elements which may be shared. Red boxes, on the other hand, refer to elements that each organisation must realise individually. Within this model, every institute tasked with assuring digital preservation has the opportunity to clarify which of these elements are specific to itself or
to the domain in which it operates. In addition, the institutes may indicate if an element should be a national shared service or that international facilities can be used instead. This presupposes that (1) each institute knows its needs and (2) the services offered are clearly defined and certified where necessary.

The NCDD does not consider a “one size fits all” solution feasible. Inevitably, a balance will need to be struck between the individual approach, domain-specific collaboration, and cross-domain collaboration. It is not yet clear what this balance will look like. This is due, among other things, to a lack of insight into the domain-specific aspects of digital preservation.

Figure 2
The distributed landscape of facilities for sustained access
In between the legally defined tasks of individual institutes, which must be respected, and the looming perspective of large-scale cross-sector collaboration lies a large grey area, which has as yet been insufficiently mapped. To what extent, for instance, are archiving institutes bound by sector-specific standards for sustained archiving? How far will the centralisation of the storage of immensely diverse digital research material in the scientific domain go? And is it a good idea for the museum sector as a whole to have its own, recognisable e-depot? Over the coming period, the NCDD will continue to chart this grey area by inviting various institutes to apply their own perspective to the distributed model and share their vision of a common approach.

At the outset of this investigation, the intent was not only to develop a preferred scenario as described above, but also to substantiate it financially based on existing costs of and investments in the management, preservation, and exploitation of the digital archives of NCDD partners. This financial underpinning is expedient for three reasons:

1 - to obtain insight into the differences and similarities between the cost structures in the various domains represented by the NCDD;
2 - to enable a calculated estimate of the necessary investments over the coming five to ten years, to further develop these national facilities;
3 - to develop a business case in order to determine how much every institute which will make use of these national facilities can contribute to their funding.

Over the course of the research, however, it proved impossible to realise this goal within the available time, due to a lack of clear and comparable financial data. The NCDD coalition partners are individually working towards a financial model based on total cost of ownership, but this has not yet been fully put into practice. For this reason, financial data capable of being compared within a single model have been unavailable. The NCDD will work on this matter by creating a Dutch research trajectory based on the instruments developed in the European 4C (Collaboration to Clarify the Cost of Curation) project.

A two-pronged approach

The CCDD’s investigation into the state of sustained accessibility of born digital heritage in the cultural sector has produced a clear picture of the state of affairs on the preservation and management of this heritage. The collection and sustained storage of born digital heritage by heritage institutes in the Netherlands is still in the early stages. Both the amount of material collected and the ways in which it is managed and preserved differ markedly between domains, and a uniform approach spanning all domains is, for the moment, impossible. Within domains, too, there are large differences between individual institutes on the approach taken to the problems associated with born digital heritage. There is a need for active domain-based knowledge centres and networks. Solutions by domain are currently easy to implement, and relatively limited efforts can yield great results.

For the moment, the most realistic approach for the cultural heritage sector is a domain-based one, before the next step towards a nationwide network can be made. There is a
need for basic facilities, specialised e-depots, and knowledge centres. Conversely, the library and museum sectors are at a much higher stage of organisation and already have a partially operational network in place. In these domains, overarching collaboration can now commence.

The NCDD’s research has taken a top-down approach, sketching out a model for a broad nationwide network of facilities. This is, however, not the only possible approach. A parallel bottom-up trajectory is needed to allow smaller parties to contribute to the discussion and help determine solutions based on their impact on them.

**The next steps**

Now that a preferred scenario has been developed, plans can be made for the next steps in strengthening the shared practice of digital preservation in the Netherlands.

The multi-year plan established by the NCDD and the Network for Digital Heritage (NDE) offers a backdrop for these steps. The NDE is a collaboration aimed at developing a system of nationwide facilities and services for improving the visibility, usability, and maintainability of digital heritage. It was initiated by the Ministry of Education, Culture and Science, and among its participants are several large national institutes working on the professional preservation and management of digital data, as well as a growing number of parties and persons from inside and outside the heritage sector. Digital preservation is a key part of NDE’s agenda, under the banner of Sustained Digital Heritage.

Over the coming years (2015-2016), the NCDD will coordinate the activities around this theme.

Specifically, activities will be centred along three lines:

1. Scalability of facilities
2. Cost management
3. Roles and responsibilities in collection development

Along these lines, projects will be carried out which contribute to the realisation of the distributed model and the necessary facilities, but also projects which meet the direct needs of smaller institutes: a combined top-down and bottom-up approach.

Furthermore, the NCDD will perform several flanking activities, aimed at broad knowledge development and communication. For instance, the NCDD is investing in the development of practical digital preservation training aimed at professionals, as well as bringing together existing knowledge bases on this theme.

This research has yielded a good conceptual framework for the NCDD coalition partners and other institutes to start taking steps in their own fields towards a joint implementation.
The investigation into a national infrastructure for sustained access to digital information was carried out between April 2014 and February 2015 by Joost van der Nat at the request of the NCDD and with financial support from the Ministry for Education, Culture and Science. This research is part of the National Coalition for Digital Preservation’s multi-year plan 2013-2018.

The conclusions and recommendations of the NCDD research will be published in the report Onderzoek naar een landelijke infrastructuur voor duurzame toegang tot digitale informatie. Samen bouwen aan een netwerk van landelijke voorzieningen (in Dutch).

The conclusions and recommendations of the CCDD research have been published in the brochure Born digital cultureel erfgoed is bedreigd erfgoed. Op weg naar een generieke workflow voor born digital erfgoed binnen de domeinen kunst, film, fotografie en architectuur (in Dutch).
The Netherlands Coalition for Digital Preservation (NCDD) was founded in 2008 by organisations whose core business includes long-term preservation of digital information in the public domain. Members of the NCDD are the National Library, the National Archives, Data Archiving and Networked Services (DANS), the Netherlands Institute for Sound and Vision, and the Cultural Coalition for Digital Preservation, represented by EYE Film Museum and the Digital Heritage Foundation. The NCDD acts as a platform for sharing knowledge and expertise. It coordinates the development of a nationwide network to guarantee sustained access to digital information in the public sector.

www.ncdd.nl