A Dutch approach in constructing a network of nationwide facilities for digital preservation together

Joost VAN DER NAT
National Coalition Digital Preservation
P.O. Box 17 097, 1001 JB Amsterdam
The Netherlands
+316 5204 8788
jvdnat@ncdd.nl

Marcel RAS
National Coalition Digital Preservation
P.O. Box 90 407, 2509 LK Den Haag
The Netherlands
+316 1477 7671
marcel.ras@kb.nl

ABSTRACT

Cross-domain collaboration lies at the heart of approach of the Dutch Coalition for Digital Preservation (NCDD) [1]. Sharing responsibilities, knowledge and infrastructures is an absolute necessity for the realisation of high-quality, effective and efficient digital archive- and information management.

In November 2016, the NCDD research on the construction of a cross-domain network of facilities for long-term access to digital Cultural Heritage in the Netherlands was rewarded the Digital Preservation Award 2016 in the category Research and Innovation [2]. According to the judges the research report presents an outstanding model to help memory institutes to share facilities and create a distributed, nationwide infrastructure network for Digital Preservation.

This model describes a network of collaborative facilities as a back bone for sharing facilities within the Netherlands. It describes all the elements of an infrastructure for digital preservation. Not only the technical parts of it, but also its organisational, legal and financial parts. Describing them as elements, building-blocks, helps us to understand them and to get a better picture of what is already available on the one hand, and the demand on the other hand. One should notice that this model is based on the most desirable infrastructure. It describes the broad picture and enables us to create a roadmap towards the desirable situation. As always, the proof of the pudding is in the eating. This paper describes the theory behind the model and the steps to put the model into practice in an operational sense.

KEYWORDS

Digital Preservation, Collaborative Digital Preservation, National Infrastructure, Services and Facilities.

1 INTRODUCTION

Cross-domain collaboration lies at the heart of approach of the Dutch National Coalition for Digital Preservation (NCDD) [1] for the construction of a national shared infrastructure for Digital Preservation. It is key to realizing high-quality, effective and efficient digital information management.

The NCDD partners are advancing this collaborative approach by searching for the best solutions across the board of the public domain. This explicitly includes the interests of smaller organizations which, due to a lack of technical facilities, organization and knowledge, are not capable of ensuring reliable digital management on their own. In 2013 NCDD made it part of her strategy to

work on this collaborative model that should result in a distributed national infrastructure.

Following on a national survey [3], the NCDD in 2010 formulated a strategic agenda [4]. This agenda consisted of a description of the major steps to be taken on a national level in the Netherlands in order to address the issues described in the survey. The strategy is centred on four themes:

1) knowledge-sharing;
2) development of a scalable and usable infrastructure for long-term management of digital information;
3) cost management; and
4) development of co-ordination in collection development policies.

It was also thought necessary to create a sense of urgency towards policy makers on all levels, with the message that we had to act, and act on a national level, to ensure long-term access of digital information. Within the sense of urgency, the focal point was the development of a national infrastructure. Therefore, NCDD and especially the partners within the NCDD took the lead in addressing the problem on a policy level, but also on a practical level. It was decided that under the umbrella of the NCDD coalition, the large heritage institutes in the Netherlands would work out a "collaborative model", setting up collaborative facilities or share facilities where possible. Which in reality would not always be the case.

2 NATIONAL COLLABORATION

Under the motto "Joining forces for our digital memory", a research project was started in 2014 which was commissioned and financed by the Ministry of Education, Culture and Science. This project had the purpose to find out what level of differentiation between the domains offers the best balance for efficiency. Without collaboration, inefficiencies loom, while individual institutes continue to expand their digital archives and may be reinventing the same wheel over and over again. The project's objective was and is to avoid duplication of work, and to avoid wasting time, money, and energy. Economies of scale make it easier for the many smaller Dutch institutes to profit from available facilities, services, and expertise as well. Policy makers can now ponder the question "The same for less money, or more for the same money?".

The result of the project [5] was a model, which will be described in the next paragraphs. Next steps were taken within the national collaborative framework of the Dutch Digital Heritage Network (Netwerk Digitaal Erfgoed, NDE) [6]. This partnership has the intention of developing a network of nationwide facilities and services for improving the visibility, usability, and sustainability of digital heritage. The network was established on the initiative of the Ministry of Education, Culture and Science and consists of a

number of large organizations occupying key positions in the field of digital heritage.

The NDE has developed a three-pronged strategy covering Visible, Usable and Sustainable Digital Heritage, respectively. A work package has been established for each of these aspects, outlining the projects necessary to achieve its central goals. The work package on Sustainable Digital Heritage is led by the NCDD.

Following on the results of the NCDD research project, a project within the sustainable digital heritage work package was constructed. This project built on the constructed model and developed a catalogue of existing service of all elements of the infrastructure for digital preservation. Also, case studies were carried out in order to get a better picture of the demand for services. Case studies should bring Supply and Demand together.

3 THE RESEARCH

With this research, the NCDD has presented a practical and down-to-earth framework for all parties involved, providing a clear picture of what such nationwide infrastructure might look like, including its scalability, and various growth models, culminating in the steps we need to take to achieve this, in both the short and long term.

3.1 Desk Research

Based upon extensive desk research [7 – 20] a model was developed for the catch-all term "infrastructure for digital preservation".
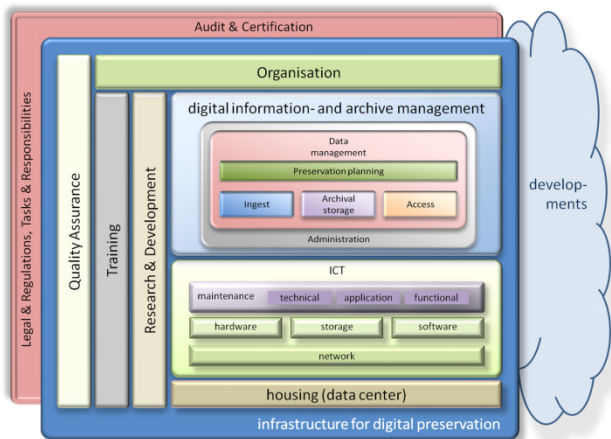

Fig 1 Infrastructure for digital preservation

In Figure 1 the constituent elements allow for an overall look at all the required elements and its relations. This shows that the necessary elements are not just IT-elements, but also involve knowledge, policies, quality control, training and many more organisational aspects.

The central part of the model consists of the OAIS functional entities. These elements are the heart of every preservation system and are the core of the processes of organisations, when concerning themselves with digital preservation. As shown in the figure above, these OAIS-functions are part of the core of the model.

Based on this high-level model we developed a "Business – IT stack" for digital preservation, with Building Blocks representing all elements we consider to be the part and parcel of the digital preservation environment.
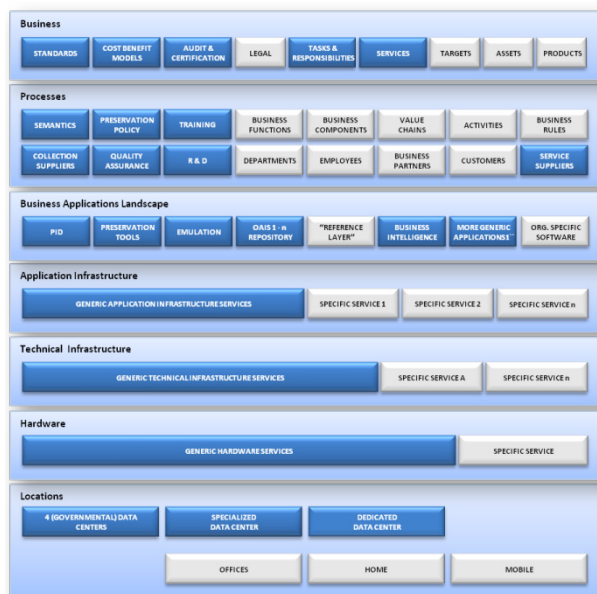

Fig 2 Building Blocks for digital preservation

In figure 2 all these elements are depicted as Building Blocks. From IT-elements as storage facilities, to standards, training, R&D, and semantics. Any organisation, or collaborative effort, obligated to ensure long-term access to digital information should have (most of) these elements organised. The model then focuses on those elements which could be shared within a network or a collaborative effort. The *dark blue* Building Blocks are the ones that are potentially shareable. The *grey* ones seem by definition organisation specific.

Each building Block can be seen as a Service. Services can be grouped together in Facilities [20].

3.2 Fieldwork

Extensive interviews with representatives of a number of large institutes [21], among them NCDD coalition partners, provided insight in the state of the current infrastructure and helped to sketch out a common vision for a desirable future situation.

Current facilities, often controlled by individual institutes, were discussed in detail, as well as the possibilities to share facilities and services with third parties. These interviews gave us a detailed picture of the possibilities for collaboration as well as domain specific services needing an individual approach.

As an *example*: for one group of elements, the "Support of the primary process of digital preservation", the results of the interviews look like this:
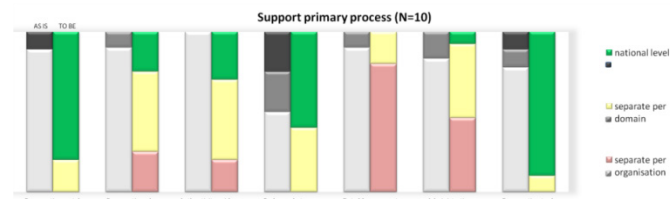

Fig 3 Diagram cluster Support of Primary Process

The *black – grey* bars refer to the current situation, *coloured* bars refer to the preferable future situation.

E.g.: the dataset to the left: "Preservation Watch" is presently done mainly at the organisation level, whereas it is seen as a service that should be done at national level.

For all groups of elements of the defined infrastructure a concrete picture emerges of which infrastructural elements could be shared, which is a key factor in creating an organisational and technical infrastructure at the national level for sustained access without unnecessary overlaps. It might be better to speak of a network of nationwide facilities, as large parts of this national infrastructure have already been realised. An entirely new infrastructure from scratch is not necessary. Instead, existing facilities can be shared and opened up to third parties. In this way, a network of facilities will gradually evolve and materialise.

3.3 Scenario's

Applying the outcomes of the fieldwork to the Building Block model, the following representation was developed:

Fig 4 Building Blocks as entities to be shared or organisation / domain specific

Having the shareable elements defined, scenarios were developed in order to set out the lines for collaboration. The Building Blocks are grouped in three sets. The preferred scenario is depicted as follows:

Fig 5 Preferred scenario for developing a National Infrastructure for LTDP

Technical infrastructure (A) and Preservation Services (B) can run parallel.

Technical infrastructure (A) is conditional for Distributed applications for Digital Preservation (C).

3.4 As is and To be

Developments for implementation are already in progress. For example: the first steps for the Technical Infrastructure are being made with the development of four Dutch "Governmental Data Centres". For Dutch Archives, the National Archive is deploying its Facility the e-Depot. And the NCDD is developing Services for Digital preservation in its present projects.

Many (semi)commercial solutions already exist (e.g. LOCKSS, Preservica, Archivematica to name just a few). They can be considered to offer a Facility with several embedded Services.

Our model would be instrumental to describe a desired (To be) situation of Services. It would then be possible to analyse the gap between existing and desired Services. This would be the foundation for a policy to develop new, and adapt existing Services.

4 MATCHING SUPPLY AND DEMAND

The project carried out within the framework of the Digital Heritage Network has its focus on bringing the model into practical use. Therefore, it was important to have a clear view on supply and demand. Having a model and having services in place is by no means a certainty for usage of the model and the existing services. Three things are necessary for this:

- an overview of tools and services in place, to be used as collaborative elements in a network: a catalogue of services;
- a common articulation of the demand for services; what do we expect from collaborative services?
- a specific articulation of the demand on an operational level; what types of objects need to be preserved under which mandates.

As these three elements where defined, case studies were done to create the connections between supply and demand based on the building blocks of the model. The case studies are the "proof-of-concepts" for the model. Based on the outcomes of the case studies the model could be adapted. The main goal is to bring together the Demand for Services for long term digital preservation with the Supply.

Fig 6 Supply and Demand for Facilities

4.1 Defining Supply: Catalogue of services

A first version of a Catalogue of services has been built. In this Catalogue, a list of Services is made available to small(er) organisations who are in need of Services for long term digital preservation. These services are mainly the digital archiving services in place at the large scale Cultural Heritage Institutes like the National Library, The National Archives and the Dutch Institute for Sound and Vision. Examples are:

a) the Dutch National Archive is deploying its "e-Depot" in a two-tier solution whereby 11 regional "Historical Centres" make use of a certified repository. They, in turn, can service archives of provinces, water authorities, local municipalities and even private collections;
b) the Dutch Institute for Sound and Vision can serve as the dedicated repository for audio-visual objects in the Netherlands;
c) the NCDD recently launched a Wiki (Dutch only) [22]. Its first topic is "Preservation Policy".

4.2 Articulating Demand: definition

As explained above, the model for a network of distributed facilities for long-term preservation has different objectives. It helps to map the Dutch digital preservation infrastructure in order to create a nationwide infrastructure. It also is a theoretical model which helps us to define the elements of a digital preservation infrastructure. Even more, it models all elements of an infrastructure, not only the technical parts. Alongside this model organisations will be able to overview their needs and their own organisation. With the model, we where able to create a Catalogue of existing services in the Netherlands. At the flip-side of the coin, the model can help defining the demand for services. Cultural Heritage organisations having digital collections and, being aware of the fact that they need to be preserved for the long-term, can use the model to find out what they should organise. They can use the Catalogue to find out whether these services are available within their own domain and serving their needs. The model will help them define their demand.

4.3 Articulating Demand: operational level

Next step is within the operational level. Making demand more practical in knowing what exactly an organisation needs. Is there an urgent need for preservation storage, or development of a preservation policy? Should digital objects be migrated to another file-format or is there a need for an emulation-service?

Handing over digital collections to an e-depot service brings many questions, beginning with organising the mandate, knowing what types of digital objects you will hand-over and what are the volumes of it? It is therefore crucial to assist small(er) organisations in the articulation of their operational needs. To help organisations in this, a knowledge-driven Decision Support System is being developed.

5. THREE CASE STUDIES

As mentioned above (see paragraph 4) case studies will bridge the gap between the theoretical model and the operational Supply and Demand. Three examples of case studies are:

5.1 Frisian museums

The Dutch province of Friesland has some 40 museums, with Frisian cultural heritage collections. Within the project of Virtual Fryslân, financed by the Provincial government of Friesland, a large deal of these collections is being digitized. The project runs from 2015 until 2018 and it should digitize about 55% of the collections of the Frisian museums, which is about 450.000 objects. However, within the project no obligations for long term preservation of these digitised objects were described.

This turned out to be a perfect case study to find out how to map the demand of the Frisian museums to existing supply of facilities. Therefore, we first needed to have facts about the digitised collections (types of objects, file formats, storage facilities, volumes, etc). A survey was held under 30 participating museums. Some of the outcomes:

a) total volume is 1,25 Terabyte;
b) 2/3 of the museums preserve their collection in a sustainable way;
c) 57 % stores the collection within their premises;
d) 19 % do not have a back up facility.
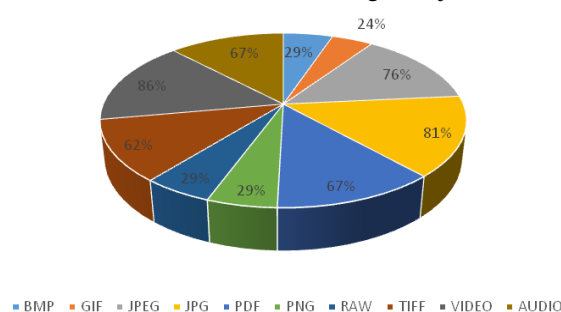e) 10 different file formats of stored digital objects:

Fig 7 Storage formats used by Frisian museums

Very interesting outcome of the survey is that 70% of the participating museums does have a policy or agreements for long-term preservation of the digitised collections. However, the financial means to preserve the collections for the long-term are low and will diminish in the forthcoming years. So there is an urgent need for a collaborative facility to be used by the Frisian museums as a group. As we now have the operational demand mapped out, we can start finding solutions within the network. Finding a supplier will be much easier now, as there is a clear view on what is needed.

The report of this study is to be published soon.

5.2 "Vidceedings" of the Dutch Parliament

The Proceedings of the Dutch Parliament are by law to be archived with the Dutch National Archive (NA). These were, for centuries, paper documents. After the years stated in the Archival Law they are to be transferred to the NA. But with digitisation comes innovation, also in these long-standing traditions.

Nowadays video recordings are made as well. These recordings need to be archived by the NA, conforming the law. However, the NA currently neither has the expertise, nor the (technical) infrastructure to do so for long term digital preservation. There are two ways to deal with this issue. Either the NA has to build up expertise and redevelop its infrastructure in order to meet the obligations it has. Or the approach could be to start collaboration with an institute which does have expertise and infrastructure in place already. In the Netherlands this is the Dutch Institute for Sound and Vision. The NA and Sound and Vision started a pilot project in which they clarified all aspects of this possible collaboration. Main point is that the NA does have the legal obligation and will always remain responsible for long term availability of the video notes. But the task

to carry out this responsibility can be sourced out. The pilot consists of the clarification of the organisational-, legal-, financial- and technical issues involved. It will result in a business model and firm specifications of service level agreements. This is an example of how the theoretic model can be executed in real life.

5.3 "De Digitale Stad"

De Digitale Stad (DDS; The Digital City) is the oldest Dutch virtual community and played an important role in the internet history of Amsterdam and the Netherlands. For the first-time internet was (free) accessible to general public in the Netherlands. DDS is an important historical source for the early years of the internet culture in the Netherlands. The virtual city and its inhabitants produced objects, ideas and traditions in new digital forms such as web pages, newsgroups, chat, audio and video. This social network proved to be very successful and had 140.000 inhabitants in 2000. But in 2001 it was taken offline.

Ten years later, in 2011, the Amsterdam Museum started the project "re:DDS", the reconstruction of DDS [23]. Not only to tell and show the story of this unique internet historical monument, but also –and more important- to raise awareness about the risk of the loss of our digital heritage. This was the beginning of a case study in web archaeology: how to excavate, reconstruct, preserve and sustainably store born-digital data to make it accessible to the future generations.

The excavation has taken place, which resulted in reconstructions of two of the three phases of DDS [24]. Most parts of the environments are uncovered, but the next step is to revive the environments and put them in display at the Museum. And then the final step has to be taken, bringing the excavated data into a digital archive for long-term preservation. Here is where the collaborative services model comes in. As DDS is on the demand-side of it, a project has started in which different scenarios are worked-out for bringing the data from the excavation into the e-depot. Alongside three scenario's all aspects of preservation of the DDS-data are described (technical-, legal-, organisational- and financial) using the model for a distributed network of facilities.

6. ONGOING AND FUTURE WORK

The presentation of a national strategy and the establishment of three Work Programs are an important development, which brings many existing initiatives and plans together. This is a start of an integrated approach for access to and preservation of Dutch digital heritage. The timing is perfect as there is a growing community of professionals involved in digital preservation. The level of knowledge exchange and the willingness to collaborate is growing too.

The model of Building Blocks has served as the basis for defining projects, and will continue to do so for future developments.

Our challenge is now to bridge the gap between Theory and Practice even further. Supply and Demand have to be brought together. We are now developing a model to formally describe Services. This would enable us to find the gap between existing Facilities & Services, and the desired set thereof.

The ultimate goal is to develop a toolset that will assist in matching Demand and Supply for Facilities & Services for Digital Preservation.

7. ACKNOWLEDGEMENTS

8 REFERENCES

[1] Link to the Dutch National Coalition for Digital Preservation (NCDD) (English): http://www.ncdd.nl/en/about-the-ncdd/ (English)

[2] Link to DPC, winners of Digital Preservation Award 2016: http://www.dpconline.org/events/digital-preservation-awards/the-winners

[3] http://www.ncdd.nl/documents/NCDDToekomstDEF2009.pdf (Dutch only)

[4] http://www.ncdd.nl/documents/NCDDToekomst_2_Strategischeagenda.pdf (Dutch only)

[5] Link to English Summary of report: http://ncdd.nl/site/wp-content/uploads/2014/06/summary_NCDD_research_DEFWEB.pdf

[6] The Network Digital Cultural Heritage in the Netherlands: http://www.ncdd.nl/en/ncdd-projects/digital-heritage-network/ (English) The participants in this network are national organizations with large digital collections and a mandate to preserve them, like the National Library (Koninklijke Bibliotheek or KB), the Institute of Sound and Vision (BenG), the Royal Netherlands Academy of Arts and Sciences (KNAW), the National Archives (NA), the Cultural Heritage Agency of the Netherlands (RCE) together with other partners like for example the knowledge centre Digitaal Erfgoed Nederland (DEN) and the National Coalition for Digital Preservation (NCDD).

[7] The Final Report of the Blue-Ribbon Task Force; 2010; Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information

[8] Consultative Committee for Space Data Systems; January 2002; Reference Model for an Open Archival Information System (OAIS)

[9] Barbara Sierman; december 2012; Het OAIS-model, een leidraad voor duurzame toegankelijkheid

[10] UNESCO conference; december 2013; A Digital Roadmap for Long-Term Access to Digital Heritage

[11] Erfgoedinspectie; december 2012; Beperkt houdbaar? Duurzame toegankelijkheid in een digitale omgeving bij de rijksoverheid

[12] Software Engineering Institute; november 2010; CMMI® for Services, Version 1.3, Improving processes for providing better services

[13] 4C Project; February 2015; Investing in Curation, A Shared Path to Sustainability

[14] Projectbureau I-strategie; juli 2012; Doelarchitectuur Digitale Duurzaamheid Rijksoverheid

[15] Landelijk Overleg Provinciale Archief Inspecteurs; mei 2008; ED3: Eisen Duurzaam Digitaal Depot. Toetsingskader voor de beheersomgeving van blijvend te bewaren digitale informatie

[16] APARSEN (Alliance for Permanent Access to the Records of Science Network); January 2014; D21.1 Overview of Preservation Services

[17] DANS / Data Seal of Approval Board; July 2013; Data Seal of Approval, Guidelines version 2

[18] DCH-RP (Digital Cultural Heritage Roadmap for Preservation); April 2013; D3.1 Study on a Roadmap for Preservation (revision 3.1)

[19] PARSE.Insight consortium; June 2010; Deliverable D2.2, Science Data Infrastructure Roadmap

[20] http://www.noraonline.nl/images/noraonline/1/12/Doelarchitectuur_Digitale_Duurzaamheid_%28v1.0%29.pdf (Dutch only, page 11)

[21] Instituut voor Beeld en Geluid, Data Archiving and Networked Services (DANS), International Institute for Social History (IISG), Koninklijke Bibliotheek, LIMA (platform for sustainable access to media art), Nationaal Archief, Nationaal Museum voor Wereldculturen, Het Nieuwe Instituut, Rijksdienst voor het Cultureel Erfgoed (RCE), Rijksmuseum, Stadsarchief Amsterdam.

[22] http://wiki.ncdd.nl (Dutch only)

[23] https://hart.amsterdam/nl/page/521/re-dds (Dutch only)

[24] Tjarda de Haan, "The Digital City Revives" A Case Study of Web Archaeology: Paper - iPRES 2016 https://phaidra.univie.ac.at/detail_object/o:503160?SID=10011580&actPage=1&type=listview&sortfield=uw.general.title,SCORE&sortreverse=0