

Preserving digital heritage: the UNESCO Charter and developments in the Netherlands

Marco de Niet, Titia van der Werf and Vincent Wintermans

Abstract

The e-Depot of the Netherlands National Library and Images for the Future, the mass-digitisation project for audiovisual heritage from the Netherlands carried out by a partnership led by the Netherlands Institute for Sound and Vision, are two large scale projects for long-term preservation and digitisation in the Netherlands. Comparing the results achieved and the challenges met in these projects, this study gives insight in how assumptions and problems in the field of digital preservation have evolved in the past decennium. This will fuel the discussion on the revision of the Charter for the Preservation of the Digital Heritage.

Biographies of authors

Marco de Niet is the director of the DEN foundation, the Dutch knowledge centre for digital heritage. He studied History of the Book at Leiden University. He is actively involved in the Europeana Network and core partner in the ENUMERATE project that measures the progress of digitisation in Europe. He is a founder member of the Dutch Memory of the World Committee, a board member of the Dutch Museum Register and a member of the Council for Dutch Language and Literature. Before DEN he worked at the National Library of the Netherlands as head of Innovative Projects and Digital Preservation.

Titia van der Werf is senior programme officer at OCLC Research, based at OCLC's office in Leiden. Between 1993 and 2002 she worked for the R&D department of the National Library of the Netherlands and contributed significantly to early research into digital preservation and design of the first e-Depot system. She was actively involved in the development of the Open Archival Information System (OAIS) Reference Model and was on the team that drafted the

UNESCO Charter on the preservation of digital heritage (2003). She is a member of the Netherlands Memory of the World Committee.

Vincent Wintermans studied Slavic languages at Leiden University. He is policy officer at the Netherlands National Commission for UNESCO and an observer attached to the Netherlands National Committee for the Memory of the World Programme. His current professional interests are open access publishing, documentary heritage, endangered languages and communication & development. He carries out research on the history of the work of UNESCO and the IIC in the field of linguistics and bibliography.

1. Introduction

UNESCO's General Conference adopted the Charter for the Preservation of the Digital Heritage¹ in 2003. This non-binding standard-setting instrument, targeted mainly at governments, explains the particular vulnerability of digital heritage, asks for immediate action to protect it and recommends the development of strategies and international cooperation. In the same year, UNESCO published Guidelines for the Preservation of the Digital Heritage², written by Mr Colin Webb of the National Library of Australia. These Guidelines elaborate the issues raised by the Charter in great detail, both from a management perspective and from a technical and practical perspective.

Article 12 of the Charter requests UNESCO to 'determine, on the basis of the experience gained over the next six years in implementing the present Charter and the Guidelines, the

¹ Charter for the Preservation of the Digital Heritage: http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html

² Guidelines for the Preservation of the Digital Heritage: <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>

need for further standard-setting instruments for the promotion and preservation of the digital heritage'. The results of the survey conducted in 2009 were unsatisfactory: only a small number of Member States responded and the outcomes were not entirely clear. Some countries argued for a new standard-setting instrument, or for the 'upgrading' of the Charter to a higher level instrument. Some asked for efforts to make the Charter better known.

In 2009, the Unesco Netherlands National Commission started a project with the aim of answering on a more empirical basis the question of whether a revision of the Charter and the Guidelines is possible and necessary. In a workshop organised in the framework of the IV International Conference on the Memory of the World Programme *Culture-Memory-Identities* (Warsaw, 18-21 May 2011), it presented and analysed examples of how Charter and Guidelines have been used in South Africa, Brazil and Poland. The results of this preliminary step, to be published in the proceedings of the conference³, show that the Charter has been a useful instrument in the three countries considered. Some themes that might be included in an addendum to the Guidelines were tentatively proposed: technical innovations like cloud computing, the appearance of commercial actors in the field of digital preservation, the experiences of non-Western countries in digitising their heritage, and the responsibility for digital heritage that cannot be attributed to a country.

When UNESCO announced that the revision of the Charter would be one of the subjects to be treated at the Conference *The Memory of the World in the Digital Age*, the Netherlands National Commission decided to analyse experiences gained in the Netherlands with long-term mass preservation of digital heritage in order to detect major developments in the field that could be the subject of further standard-setting by the Organisation.

³ V. Wintermans 'An Addendum to the Charter and Guidelines for the Preservation of the Digital Heritage', <http://www.unesco.nl/documents/documenten-natcom/Addendum%20Preservation%20Digital%20Heritage.pdf>

2. Digital preservation in the Netherlands

Dutch heritage institutions participated in the very first European projects to develop methods and technologies for digital preservation in the 1990s. Soon after, the Dutch government was actively involved in the realisation of the UNESCO Charter on the Preservation of Digital Heritage. The National Library of the Netherlands (Koninklijke Bibliotheek, KB⁴) was the first national library in the world to present an operational longterm archives for electronic journals, in part thanks to a substantial grant from the Dutch government. With a similar governmental commitment, the Netherlands Institute for Sound and Vision (Nederlands Instituut voor Beeld en Geluid⁵) became one of the international front-runners of long-term preservation and management of digital audiovisual heritage.

In 2007, the KB and DANS, the national scientific data archive⁶, took the initiative towards a national and more structured approach to digital preservation of information in the public domain. Following the examples in the United Kingdom (Digital Preservation Coalition⁷) and Germany (NESTOR⁸), both organisations invited other public institutions with digital preservation as a core task to join forces. Together they founded the National Coalition on Digital Preservation (NCDD⁹). The mission of this foundation is “to ensure that there will be a stable organisational and technical infrastructure in the Netherlands for the preservation of

⁴ www.kb.nl

⁵ www.beeldengeluid.nl

⁶ www.dans.knaw.nl/

⁷ www.dpconline.org

⁸ <http://www.langzeitarchivierung.de>

⁹ www.ncdd.nl

and permanent access to digital information that is relevant to science, culture and society at large."

The first assignment for the NCDD was a national survey of the digital preservation situation in the sciences, the public media, the cultural heritage sector and the government. The results were published in the report *A Future for our Digital Memory* (July 2009)¹⁰. From the survey, it was obvious that the role of some front-runners was crucial for the progress made in the various sectors: the e-deposits of DANS and the KB took care of scientific data and publications, respectively. The National Archives and some innovative archives, including the city archives of Amsterdam and Rotterdam, laid the foundation for a networked archiving system for governmental documents and data. Sound and Vision built the largest public digital archive in the country for audiovisual material. The weakest sector was the cultural heritage sector, notably the museums. In this sector, there was not one obvious front-runner to take up the responsibility for developing and managing a digital archive for cultural data and objects. Following on from this conclusion, various cultural institutions took the initiative to install a Cultural Coalition on Digital Preservation, which main tasks are raising awareness, developing policy and sharing knowledge and research. Due to the lack of resources and the lack of an organisational infrastructure it cannot set up such a cultural digital archive in this point in time.

The time when digital preservation was mainly a research topic is well behind us. Digital information is omnipresent and the lack of professional digital preservation strategies and solutions will inevitably result in loss of public data. In the past decade, large-scale systems for the preservation and management of digital data have become operational in the

¹⁰ www.ncdd.nl/en/publicaties.php

Netherlands. However, these infrastructures are not yet fully mature. Archiving systems are constantly under development and are facing a rapidly growing flood of digital information that seems to be in a permanent state of flux. Institutions that manage these systems are still trying to find reliable financing and design business models to sustain the high costs of continuous development and maintenance into the future. Also, the digital collections of many institutions are not yet included in digital preservation solutions. With the current economic downturn and a government that is cutting its budgets for science, culture and the public sector, it is a big challenge to improve and expand digital preservation practices. This article is intended to provide some inspiration for future developments by presenting the lessons learned from two Dutch pioneers that have built large scale systems for the preservation of digital objects that fall within the scope of UNESCO's Charter on the Preservation of Digital Heritage.

3. Introduction to *Beeld en Geluid (Sound and Vision)* and the *Koninklijke Bibliotheek (KB)*

The Nederlands Instituut voor Beeld en Geluid (Netherlands Institute for Sound and Vision) is a cultural-historical organisation that collects, preserves and opens up the Dutch audiovisual heritage for various user groups, such as media professionals, education, science and the general public. It is one of the largest audiovisual archives in Europe. It is located in the city of Hilversum and is housed in one of the most eye-catching buildings in the country. The collection contains more than 800,000 hours of television, radio, music and film from the beginning in 1898 until the present day, including the complete radio and television archives of the Dutch public broadcasters. The institute currently manages over 70 percent of the

Dutch audiovisual heritage. Digitisation is a core task of the institution, both for conservation and for enhancing access.

The Koninklijke Bibliotheek, National Library of the Netherlands (KB), was founded in 1798 and is based in The Hague. The KB is the one of the largest libraries in the Netherlands together with the university libraries of Amsterdam and Leiden. The KB is both a research library and, since 1974, a deposit library. The Netherlands is one of the few countries in Europe where there is no deposit law for books and publications. The deposit library is based on mutual agreements with publishers and their representatives. The KB became an independent body in 1982. In the early 1990's, the KB set up its own research department, which helped the library to modernise its services and processes and acquire a strong international profile. The current policy plan of the KB puts digital information at the heart of its mission.

4. Mass digitisation as a catalyst

In the past decades, major infrastructural developments and projects in the Netherlands were financed from the profits made from exploiting the country's gas reserves. The Dutch government is closely involved in this industrial enterprise, and the Dutch treasury benefits each year from several billions of euros generated from monetising these natural reserves. Until recently, special governmental programmes were set up to invest these extensive tax revenues in all kinds of public services, including digital infrastructures for science and culture. Science and culture are, together with education, the main areas of the Dutch Ministry of Education, Culture and Science (Onderwijs, Cultuur & Wetenschap, OCW¹¹). Both institutions described in this paper are supported and financed by this Ministry. Sound and

¹¹ www.rijksoverheid.nl

Vision is operated under the directorate for Media, Literature and Libraries, while the KB is positioned under the directorate of Research & Science. Both directorates have invested extensive funds in digitisation programmes and digital preservation systems to support the digital transition in these institutions.

In 2000, more or less simultaneously with their project to develop an e-Depot for electronic publications, the KB initiated the Memory of the Netherlands programme for the digitisation and online accessibility of cultural heritage. This programme, inspired by the American Memory project of the Library of Congress¹², consisted of three components: 1) a funding scheme for smaller institutions to have visually interesting collections digitised, 2) an expert centre to set up quality guidelines for digitisation of cultural heritage, and 3) a website through which all the digitised collections could be accessed¹³. Currently, the website contains close to 800,000 cultural objects from 100 institutions. It was intended to have the digital masters created in the Memory of the Netherlands programme stored in the e-Depot of the KB. This plan has not yet materialised due to technical constraints of the e-Depot system, which was primarily designed for the preservation of scientific journals, but also due to the lack of a proper business model to sustain and preserve these digitised collections over a longer period of time (see also below).

In July 2007, an even larger digitisation project was launched. This project, Beelden voor de Toekomst (Images for the Future), aimed at the restoration, preservation and digitisation of Dutch audiovisual collections.¹⁴ The project was initiated by the Ministry of OCW and the digitisation includes the collections from Sound and Vision, the Dutch Film Museum EYE

¹² <http://memory.loc.gov/>

¹³ www.geheugenvannederland.nl

¹⁴ <http://beeldenvoortoeekomst.nl/>

and the photographic collections of the National Archives. The total budget was originally 154 million euro, of which 64 million had to be paid back to the government before the year 2025, e.g. from revenues earned through cultural entrepreneurship with the digitised collections. In 2010, a mid-term evaluation by the Dutch independent research organisation TNO made clear that the expectations regarding the revenue to be generated from the market could not be met, and the overall budget was reduced to 115 million euro. This means that less material can be preserved and digitised, but the obligation to repay some of the funds was dropped altogether. Currently (Summer 2012), almost 85,000 hours of video, 84,000 hours of audio, 16,000 hours of film and close to 2.5 million photos have been digitised. As for Sound and Vision, their budget within the project is 85 million euro and a total of 200,000 hours of material will be preserved and digitised.

For both the KB and Sound and Vision, these large-scale digitisation projects influenced the setup of the digital preservation systems, as these systems had to support the ingest of both new (born-digital) content from external suppliers as part of their ongoing mission (publishers and broadcasters respectively) and of digitised counterparts of historic collections of various origins. In terms of the UNESCO Charter on the Preservation of Digital Heritage, both institutions were in a good starting position to address the major challenge of digital continuity in all its complexity (Charter, Article 5).

5. System development and technology partners

In 2000, the KB and IBM started building the Digital Information Archiving System (DIAS) that was to become the technical core of the infrastructure for KB's e-Depot for electronic publications. It was clear from the start that this project could not rely on out-of-the-box solutions alone, because at that time no solution readily addressed both the aspects of large

volume and durable storage as well as the long-term preservation requirements. As well as implementing the e-Depot system, IBM, KB and the British Library carried out a “Long-Term Preservation Study”¹⁵ to define the requirements of the OAIS Preservation Planning function as an extension of the basic content management and storage functionality provided by the initial DIAS version. By 2012, the KB had decided that DIAS was no longer a viable solution and terminated its contract with IBM. What had happened?

There were two main reasons for KB’s decision. As the KB needed to ingest more and more different types of content, coming from other sources than e-publishing (specifically digitisation and web archiving), the operational DIAS system did not evolve sufficiently, and the ingest procedure became a bottleneck. The system did not support important standards or interoperable interfaces to connect with preservation tools and methods that were coming out of research projects. The negotiation process with IBM for requesting changes and new functionality and getting rapid software upgrades was not without its frictions. KB experienced this as a vendor lock-in constraint.

It was also a setback that DIAS did not evolve into a widely used market product, which was the original intent and hope of both parties. With only the KB in the Netherlands and the Deutsche Nationalbibliothek (DNB) in Germany as DIAS users, the product could never become profitable for IBM and the costs remained high for the customer libraries. As a result of KB’s decision to terminate its relationship with IBM, the e-Depot is currently undergoing the most fundamental migration in its history.

¹⁵ http://www.kb.nl/hrd/dd/dd_onderzoek/reports/1-overview.pdf

DIAS is being replaced by a totally new system that is designed, built and implemented in-house. The KB has come to the conclusion that there are no adequate systems available in the market for what they want to achieve with its e-Depot and that they do not want to repeat the vendor lock-in experience. Open-source products are integrated in the new system, if fitting the requirements – for example, software to support some of the workflows. The digital preservation approach implemented in DIAS with the “Reference Platform” infrastructure and view paths for specific digital object types is done away with. The new system is specified on the basis of 15 years of R&D into digital preservation and more than 10 years of practical experience with the e-Depot service. The whole operation includes a medium migration exercise – migrating from optical disks to hard disks – and a complete redesign of the metadata management – migrating the storage, retrieval, discovery and preservation metadata into a single system. The archival information packages (AIPs) will be repackaged from a tar file into a PREMIS-XML file. The KB expects that adopting the PREMIS standard will allow for a better administration of the ‘events’ that objects in the repository undergo – which is crucial information for the long-term preservation, as migrations and other transformations will continue into the future. As issues of scalability and rapid change will not disappear, the KB is preparing itself for recurring system migrations in cycles of 10 to 15 years.

The digital archives of Sound and Vision have completely different beginnings, but there are interesting similarities to note in the further developments. The creation of the digital archives in Hilversum was not a conscious decision of a heritage institution to start expanding its usual business into the digital domain, but the inescapable result of the decision taken by the Dutch public broadcasters to digitise the production process. By 2006, the whole broadcasting processes for production, distribution and playout had been digitised.

For the technical infrastructure of its archives, Sound and Vision had always relied on the Nederlandse Omroepproductie Bedrijf (NOB), an audiovisual production company that was part of the Dutch public broadcasting landscape, and a close neighbour of Sound and Vision in more than just the geographical sense. In 2006, NOB Cross Media Facilities was taken over by Technicolor, a division of the French multinational Thomson. Both changes (digitisation and privatisation) evidently had major effects on operations at Sound and Vision. Thomson set up the archives as a storage-as-a-service facility, meaning that Sound and Vision paid per unit of storage capacity used. With the increasing volumes of born-digital content coming out of the production process and the digitised historical collections from the Images for the Future project, costs rose at an alarming rate from 2007 onwards. Sound and Vision took the decision to disentangle itself from Thomson. This proved to be a difficult operation because the storage-as-a-service model meant that Thomson was providing services to other clients using the same infrastructure. At the time of its conception, this “shared platform” philosophy was thought to be a cost-efficient solution. Over time, Sound and Vision found itself to be the largest customer by far, but being constrained by obligations from Thomson to other, smaller clients. Sound and Vision experienced this as a vendor lock-in constraint. Because of the many dependencies, it concluded that tendering for a new archiving solution was impossible and decided to build a completely new archiving facility from scratch.

This was a major decision because it entailed that a vast body of technical knowledge had to be acquired by its staff in order to catch up on archiving expertise and skills that had previously been outsourced to Thomson. Disentanglement of the processes, requirement specifications, design, development and implementation of a new archiving environment all had to be carried out in parallel with the continued and uninterrupted service delivery to the broadcasters and other customers. According to Sound and Vision, its new environment operates at far lower costs.

The in-house system of Sound and Vision is built to support critical stages in the heavy-weight broadcasting production process: each production is immediately archived and available for reuse after broadcast in the controlled environment for the Hilversum broadcasting companies. Storage requirements are evidently geared to high-volume storage, as the archives have now reached close to 6 Pb of capacity usage. Preservation functionality was not a high-priority for Sound and Vision when they started their in-house development project. Medium migration has been the obvious preservation action in 2011 and 2012 and is not considered to be very complex. Formats seem to have a longer lifespan than expected a few years ago, so format migration is not an immediate concern.

When asked about the lessons learned, both KB and Sound and Vision give the same answer: do not buy solutions or outsource tasks that you do not fully understand. Make sure you have the expertise in house to specify everything you need and to enter into dialogue with vendors and technology partners at a professional level.

Although the term ‘vendor lock-in’ does not occur in the Charter or Guidelines for the Preservation of the Digital Heritage, both documents have something to say about co-operation with industries. The Charter stresses the need for reliable systems (Article 5) and the need to share technical knowledge and experience between industries and heritage institutions (Article 11), but it does not warn against overdependence on parties that, by their nature, are primarily driven by profit. The Guidelines (p. 113) point to the benefits of purchasing off-the-shelf solutions, such as storage, which is core to digital archiving, but do not provide guidance in forging sustained business relationships with parties in the IT industry. Such relationships require an understanding of and respect for each other’s interests, in particular in new areas where market opportunities still need exploring and requirements

are still developing. Digital preservation is clearly such an area: both the industry and the community of practice have not yet fully matured. Defining generic requirements and designing robust systems for the market requires close collaboration between the community of practice and the industry.

Interestingly, both the KB and Sound and Vision have “disentangled” themselves from their industry partner and they have opted for in-house development and as much use of open standards as possible.

Open-source software is incorporated where useful. Development of open-source solutions is not a priority for KB and Sound and Vision, considering the specialised nature of some of the requirements and the limited number of potential members of that community of practice.

Being neither open source nor vendor solutions, the development and maintenance of the in-house solutions of Sound and Vision and KB will rely heavily on the institutions' expertise and budgets.

6. Geared to growth

The KB and Sound and Vision are both transitioning to large-scale archiving systems that can accommodate all their digital collections (both born-digital and digitised). The KB refers to the new system as their Digital Stacks ('Digitaal Magazijn') to underscore the multi-purpose aspect of the archive. In the past 15 years, both institutions have not only witnessed a remarkable growth in scale, but they have also had to deal with a growing variety of content. This leads us to the question: how did their collection policies and selection criteria evolve?

Sound and Vision is the result of the merger in 1997 between the business archives of the public broadcasting service, the film archive of the Netherlands Government Information

Service, the Film and Science Foundation (Stichting Film en Wetenschap), and the Dutch Broadcasting Museum (Nederlands Omroepmuseum). The core and largest collection is the material produced by the Dutch public broadcasting service. No initial selection of born-digital broadcast materials takes place; all are automatically ingested in the system on a daily basis. This has been done at the request of the broadcasters. In the longer run, however, the Institute expects to implement select criteria, just as it did in analogue times. It will then, for example, store the entire series of news bulletins, but only a few instalments of popular daily game shows.

Sound and Vision's policies are geared to growing its mass archiving facility and allowing AV collections from other institutions to make use of it, in search for economies of scale. Likewise, the objective of the Images for the Future digitisation programme is to digitise as much as possible without applying selection criteria at the item level. Selection is very time-consuming, it argues, and too much of the project budget would be lost on this activity. Moreover, it can be argued, selection already took place while building the analogue and does not need to be repeated. However, priority is given to collections on obsolete or fast deteriorating carriers and as the funds are not sufficient to digitize all items, some selection at the collection level does take place. In the bewildering mass of different file formats, and confronted with legacy metadata of hugely varying quality, the institute looks for 'sweet spots', an optimal balance between production volume, available budget, time constraints, quality and present and future archival, preservation, access and repurposing requirements.

To be more specific, Sound and Vision currently has approximately 6 PB MXF D10 (30 of 50) material, plus WAV (broadcast wave format). In addition, it has 2 PB of DPX files, which grow in numbers a fast rate. The so-called 'digital masters' are stored on two mirrored library systems, on separate locations. All tapes have duplicates, one duplicate is shelved (i.e.

the data tape is not in an actual library). Browse files are online in MPEG1 format + keyframes in JPG en MP3 format. These amount to some 170 TB. Same storage practice applies; stored online twice + backup on tape. An important additional feature is delivering of parts of videos.

At the end of the 1990s, the KB had extended its voluntary deposit agreement with the Dutch Publishers Association (Nederlands Uitgeversverbond) to cover electronic publications (offline and online). In 2002, a bilateral e-archiving agreement was signed with Elsevier and included all Elsevier journal titles. After these major achievements and the successful implementation of its e-Depot system at the end of 2002, the KB felt confident enough to open up its facilities to other major international journal publishers¹⁶. Currently it has a storage capacity of 12Tb and contains over 18 million digital publications (mostly scientific articles). Its objective is to collect 80% of the output of the STM publishers.

Alongside the international e-Depot, the KB maintains another digital archive for the digitised collections that are the result of KB's mass digitisation programmes: Metamorfoze¹⁷, Memory of the Netherlands¹⁸, Newspapers Online¹⁹ and more recently, the Early Dutch Books Online programme, which aims at digitising all Dutch titles published between 1750 and 1940 held at the KB and the University Libraries of Leiden, Amsterdam and Utrecht²⁰. The books digitised under the Google contract are also part of this programme. The digitised collections currently total a storage capacity of 470 Tb. The KB has also been archiving a

¹⁶ <http://liber.library.uu.nl/index.php/lq/article/view/7866/8062>

¹⁷ www.metamorfoze.nl

¹⁸ www.geheugenvannederland.nl/?en/homepage

¹⁹ <http://kranten.kb.nl/>

²⁰ www.earlydutchbooksonline.nl/en/edbo

growing selection of Dutch websites since 2005 and it keeps back-up copies of the Dutch university institutional repositories. All these digital collections have not been ingested into the current e-Depot, but will be incorporated in the new Digital Stacks.

The collection and acquisition policies of both institutions are geared to growth. Both are investing in large-scale digital back-end infrastructures and facilities in order to be able to offer these to other parties (e.g. publishers and memory organisations) with a view to achieving economies of scale and recovering costs. Interestingly, the KB has started to develop a valuation model which can help in judging the value of a collection; different preservation levels can then be offered at different costs based on this. The Charter declares that ‘the main criteria for deciding what digital materials to keep would be their significance and lasting cultural, scientific, evidential and other value’ (Article 7). The KB example shows that these criteria can also play a role in deciding on the preservation regime to be applied to a specific collection.

7. Publishing and archiving formats

A selection criterion that has been discussed in the digital preservation debate since the late 1990s is the format of the content. It has been argued that memory institutions with a mission to preserve digital materials should promote best practices to standardise the processes of content creation, and discourage the use of formats that are more prone to obsolescence. Some archives have actually followed this principle by limiting the variety of formats in which content files can be submitted. What is the experience of Sound and Vision and the KB?

The e-Depot of the KB generally receives PDF files, but there is a clear trend of diversification of incoming formats, as scientific publications are increasingly accompanied

with data sets and other materials ('enhanced publications'). Moreover, as mentioned earlier, the KB itself is increasingly using its digital stacks for other materials than publications, which also leads to a larger variety of formats to be ingested in the e-Depot.

The Sound and Vision archives receive formats that are generally created by trusted parties, like the MXF format supplied by the public broadcasters. In general, Sound and Vision has control over the formats, in particular the digitised products created in-house. It has just started the largest film digitisation project in the world with the objective to preserve deteriorating analogue film materials. This project delivers files in the DPX standard, which is excellent for long-term storage but less suitable for direct online access. Professional users will get previews and the possibility to order a copy in the nearby future.

Both organisations promote best practices and publish guidelines about the best formats for archiving, but they do not restrict submissions of suppliers on the basis of formats. An important argument for adhering to these guidelines is that this enables the archiving institution to guarantee long-term access.

The Charter states in Article 5 that 'long-term preservation of digital heritage begins with the design of reliable systems and procedures which will produce authentic and stable digital objects', and in Article 11 that 'industries, publishers and mass communication media are urged to promote and share knowledge and technical expertise'.

The Guidelines stress that long term preservation is only possible when preservation issues are taken into account from the very beginning, i.e. when a digital item is created. It is critical about the role of the industry: 'currently, preservation efforts have to work against the prevailing trend of digital technology, and how it is developed and used'. (5.2.5) Again, in

chapter 13: *'digital materials are created by producers who are not necessarily concerned with long-term availability. [...] Without some kind of intervention, it is unlikely that digital heritage materials will automatically be made in ways that will minimise costs and remove barriers to preservation'*. (13.4) More specifically about file formats, the Guidelines remark that creators *'should be encouraged to use very widely adopted, well-standardised file formats that fit their purposes. Generally speaking, data in simpler formats using open source, non-proprietary software are easier to preserve (although some proprietary applications achieve such widespread use that they may be accepted as an industry standard, especially if their specifications are openly published). [...] If access or copying barriers are considered necessary to protect intellectual property, they may well make preservation impossible. Arrangements will be needed to allow preservation processes such as copying to take place'*. (13.13)

The ten years of experience at the KB and at Sound and Vision do not seem to support the worries of both Charter and Guidelines concerning the file creation format. The two archiving services have built relationships of trust with the publishing and broadcasting industry, respectively, and they are confident that these parties share an interest in the long-term availability of their content.

8. Finances and business models

"The costs of preservation programmes are hard to estimate because they encompass so much uncertainty, including evolving techniques, changing technologies and very long timeframes. [...] Total costs are also likely to remain high, including set-up costs and significant recurrent costs." (Guidelines, p. 23).

Based on the experiences in the Netherlands, it is safe to say that the costs of development and operational management of large-scale digital repositories, such as those at Sound and Vision and the KB, can easily run into millions of euro per year.

Both systems started as project-based preservation programmes, with support from the national government in an international (mostly European) research context. From the beginning it was clear that the goal had to be the establishment of large-scale operational systems. Both institutions were confronted with a fast growing influx of digital materials that had to be dealt with as part of the institutions' missions. This probably made it easier for them to present compelling cases of urgency to the Dutch government to invest in their e-deposits. The KB received a fixed addition to their annual budget for running the e-deposit of international publications. It was calculated that the current annual cost of the international e-deposit is around 1.3 million euro per year. It is obvious that costs will rise considerably once the digitised collections described above have been added.

Sound and Vision receives a budget of 2 million euro per year to maintain its digital archive. However, with a storage volume of an impressive 6 Pb (growing to 10 Pb by the end of 2014), this will not prove to be sufficient. It is estimated that an annual budget of over 3 million euro would be needed to maintain the system and services that come with it.

Besides governmental funding, both organisations have committed a lot of their own resources as well for the development of their archives, for instance by organising temporary matching budgets to (research) grants and by permanently reallocating budgets to the maintenance of the digital archives, as there are recurrent costs "associated with staff, accommodation, energy supplies, network use, telecommunications costs, storage media such as disks and tapes, and consumables" (Guidelines, p. 56). With linear or maybe even

exponential growth of the content of digital archives, these cost factors can easily run out of control. "Reliable preservation programmes must be sustained over long periods, so they require business models that guarantee adequate resources will continue to be available. Unfortunately, such guarantees are rare in the real world. Most programmes have to survive with less certainty." (Guidelines, p. 54).

The European Commission and the Dutch government support the emergence of more public-private partnerships (PPP) in the area of digitisation and digital infrastructures in order to enlarge private investments. As described, Sound and Vision and the KB both engaged in public-private partnerships. In general, it can be stated that these PPPs were quite successful during the research and development stages, as both parties were willing to invest money and knowledge. For the operational management, this relationship turned into a more traditional supplier-client-model. In that respect, it should be remembered that "while suitable service providers may be found to carry out some functions, ultimately responsibility for achieving preservation objectives rests with preservation programmes, and with those who oversee and resource them" (Guidelines, p. 24). The Guidelines clearly state here that the ultimate responsibility for digital preservation cannot (or should not) be privatised. Sound and Vision and the KB have both secured their responsibilities and they have tightened, or in the case of the KB completely taken over, control of the development and maintenance process. The Dutch cases seem to indicate that it is probably easier to set up PPPs around research and development than identify business opportunities with the operational management of digital archives. In this respect, the concern should be noted that the European Commission does not address digital preservation as a separate research topic anymore in their Horizon 2020 strategy. This may seriously reduce the opportunities for continued PPPs in digital preservation research.

Besides public-private partnerships as an additional source of funding, there is also another opportunity: exploiting the facilities of the large-scale digital archives by providing digital preservation services to smaller affiliated institutions. Both Sound and Vision and the KB have experimented with this. The KB tried to offer its repository as a facility for storing the master files of digitised cultural heritage resources from other institutions. The KB did a feasibility study to set up such a 'TIFF archive' for smaller archives, libraries and museums, and the conclusion was that (for now) a business model that was acceptable to all parties involved could not be achieved. It turned out that most institutions were not willing to pay (yet) for storing the master files in a central facility. Sound and Vision ran a similar project called ProArchive. One of the lessons learned from the ProArchive project was that services for permanent storage should be separated from services for metadata management. Sound and Vision offered both as part of a package deal, but it turned out that many institutions were not interested in participating, as they required different metadata management options than those that Sound and Vision provided. Although some customers use the full archiving infrastructure, Sound and Vision decided to make its infrastructure more flexible in order to provide more versatile services.

Another possibility for creating additional revenues is charging end users for content delivery. This is complicated for a variety of reasons. KB and Sound and Vision are both non-profit organisations, and cannot participate in economic competition as private partners can. In addition, in many cases they do not own the rights attached to the content in their archives. It would take a lot of resources to clear those rights. As a result, it is not clear from the outset what the economic benefit for the institution itself would be.

A final possibility for sustaining the digital archives is to implement a mechanism of payment upfront, rather than charging the user: the party that delivers the content pays for the cost of

digital preservation. The main motivation for such an approach would be to strengthen the joint responsibility for continuity of the preservation value chain. Publishers, scholars and archiving institutions, all have their own interests in achieving that continuity. Publishers can use the public digital archives for keeping their content alive; archiving institutions can get rid of complex licensing agreements, which may jeopardise long-term access once an agreement has expired. Scholars have an interest in safeguarding the integrity of the scientific record.

Services like Portico²¹ show that collaboration between libraries and publishers can be beneficial to both. The KB is exploring options to provide services in Europe that are similar to those of Portico. These services should enable the KB to run the international e-deposit of scientific journals without being dependent on additional government funding anymore.

9. Organisational impact

Although some of the traditional tasks of cultural heritage institutions provide a good foundation to accept digital preservation responsibility (Guidelines p. 49), it is obvious that digital preservation programmes also require new skills and expertise and bring new ways of working, maybe even new organisational structures to cultural heritage institutions.

In recent years, there have been various initiatives in the Netherlands to better understand the digital transition that is currently taking place in cultural heritage institutions. Both quantitative and qualitative research among archives, libraries and museums have shown that there is hardly any activity left that has not been affected by the emergence of ICT.²² The need to share knowledge, experiences and best practices has been a strong driver to organise

²¹ www.portico.org

²² www.den.nl/art/uploads/files/Publicaties/BusModIn_eng_final.pdf

various national and international conferences, workshops and other meetings on digital preservation issues, including file formats, standardisation, business models, system development, acquisition policies, emulation etc.

Because of their long-standing involvement with digital preservation, both the KB and Sound and Vision are living proof of the impact of ICT on the mission and main tasks of cultural heritage institutions. In the past decade, both institutions have become international expert centres on digital preservation. The KB is a founding member of both the Alliance for Permanent Access (APA²³) and the Open Planets Foundation (OPF²⁴), a community hub for digital preservation whose main goal is to jointly manage and improve tools and research outcomes. Sound and Vision houses the European PrestoCentre, a membership-driven organisation that brings together a global community of stakeholders in audiovisual digitisation and digital preservation to share, work and learn²⁵. Sound and Vision also maintains an extensive knowledge base in Dutch on preservation and digitisation of audiovisual collections²⁶.

How fundamental some of the changes in the ways of working in both institutions have been, can be illustrated by the Images for the Future project at Sound and Vision and by the Metamorfoze programme at the KB. Metamorfoze is the national programme for the preservation of paper collections (in libraries and archives) in the Netherlands. Until only a few years ago, both programmes considered preservation on film to be the most reliable preservation strategy for fragile physical originals. More or less simultaneously, they radically

²³ www.alliancepermanentaccess.org/

²⁴ www.openplanetsfoundation.org/about

²⁵ <http://www.prestocentre.org/about-us>

²⁶ <http://www.avarchivering.nl/>

changed the preservation strategies, abandoned the use of analogue copies and made the switch to preservation digitisation and preservation imaging. As a consequence, all guidelines had to be redefined and expertise in this area had to be safeguarded. As Hans Westerhof, Vice-Director of Sound and Vision puts it: “The new archivist is someone who understands networking and storage, he or she is a computer scientist or an information analyst.”

As for their internal structures, both organisations have strong research departments which operate independently from the daily operations of the digital archives. Looking at the various organisational models described in the Guidelines (p. 59), KB’s model can probably best be described as "a series of specialist units looking after different aspects", while Sound and Vision’s model is "a matrix of people working in different areas, responsible to an overall programme manager". Both institutions made the deliberate decision to invest in their own staff to increase their knowledge of digital preservation to become less dependent on external companies and be more in control. As Westerhof puts it: "Don't buy what you don't understand".

As rather large cultural institutions, they are able to make these kinds of investments in hiring experts as part of their regular staff. They both agree that smaller institutions should be able to tap into this new kind of knowledge and get their support. This is in accordance with the Guidelines, which states that not all institutions with a traditional heritage role "should try to become digital heritage managers: in some cases the resources and expertise required are just not available" (Guidelines, p. 49). However, this does not mean that the KB and Sound and Vision do not acknowledge that there is no additional expertise available at smaller institutions. There are several institutions who have specialised in the preservation of specific collections, such as born-digital art. The KB and Sound and Vision also recognise that other institutions have their own responsibilities regarding the preservation of their collections.

Building an inclusive community of practice, however, proves difficult to realise; there are many factors that can get in the way of closer collaboration, such as lack of trust, absence of appropriate business models, lack of common standards, lack of proper agreements on selection and acquisition and the absence of open technology. The challenge for the entire heritage community in the Netherlands and beyond is to build a strong professional network that can effectively and efficiently bring together the various expertise and solutions necessary to safeguard all the digital heritage collections that are covered by UNESCO's Charter.

10. Conclusions

We do not claim to draw general conclusions about the state of digital preservation in the Netherlands, based on the two examples discussed here²⁷. However, the two digital archives discussed in this article are important players in the field of long-term, mass preservation of digital heritage. Their development shows that the field has made progress, beyond the phase of pilot projects that was the typical situation when the Charter was written. These examples do not allow us to conclude that we have arrived in the safe haven of 'comprehensive and reliable preservation programmes' that are the ideal of the Guidelines (p. 23); for this, the development of digital technology is far too rapid and volatile, the preservation research community far too loose and exclusive and funding far too insufficient and unsecured. Yet, ten years of practice in developing and managing permanent large scale and complex digital archives have given the organizations involved a lot of precious knowledge and experience that could not have been achieved otherwise.

²⁷ Information on experiences gained in 20 European countries can be found in the proceedings of the Conference 'Aligning National Approaches to Digital Preservation' (<http://educopia.org/publications/ANADP>) held in Estonia in May 2012.

(1) Archiving the products of large digitisation projects has significantly contributed to the level of expertise that the KB and Sound and Vision now possess, forcing them to get a firm grip on the problem of scalability. The need to tackle the challenges of scalability forced them in turn to redesign their archiving systems and introduced the opportunity to achieve economies of scale and to devise new business models.

(2) The high costs of mass storage and maintenance of digital content, often much underestimated by funders, are pressing the heritage institutions to devise new cost-recovery services to make ends meet. Both KB and Sound and Vision have experimented with providing paid services. There are no fully developed business models yet, but progress is expected. The KB strives to set up an international e-Depot of scientific journals completely independent from government funding by the year 2014. Sound and Vision aims to recover some of its costs by providing archiving-as-a-service to other institutions. For instance, it provides an archiving service to the ECV, storing all major football league matches and it stores some smaller additional collections for some public broadcasters at their request. These are paid services.

(3) We saw that both institutions struggled to free themselves from depending too much on their technology partner. What the partnership experiences with IT vendors have shown is that successful collaboration in funded research projects does not transition easily into a more business-like supplier-customer relationship. What stands out is the awareness of the institutions involved that they need to fully understand their requirements and the technology involved in order to be able to negotiate as peers. It has also become clear that the market for archiving solutions that support digital preservation is not sufficiently mature yet. It remains to be seen if the alternative choice for in-house solutions makes the costs more controllable.

(4) Conclusions (2) and (3) have important consequences for the organisation of heritage institutions and the skills required of librarians and archivists. Heritage institutions need to hire and retain highly skilled IT specialists in order to be able to carry out their tasks in a digital environment. Developing new cost-recovery services and exploiting economies of scale will require more trust-based partnerships with the suppliers of content. Trust is based on good relations and clear agreements. Cooperation between heritage institutions and building inclusive communities of practice are even more necessary in 2012 than they were ten years ago when the Charter stressed the need to ‘democratise access to digital preservation techniques’ (Article 11).

The writers wish to express their gratitude to Mr Hans Westerhof and Mr. Johan Oomen (Netherlands Institute for Sound and Vision), Mr Marcel Ras and Ms Astrid van Wesenbeeck (National Library of the Netherlands) for generously providing all information needed to write this article. Responsibility for all mistakes in fact or judgement rests, of course, with the authors.