

A future for our digital memory: born-digital cultural heritage in the Netherlands

Inge Angevaare

This article provides an overview of the status quo with regard to collecting and preserving born-digital cultural heritage objects in the Netherlands. Since there, as elsewhere, the cultural heritage sector is still grappling with the realities of web 2.0, the article also offers some more speculative thoughts on where web 2.0 may take us, as well as some practical suggestions for the next steps cultural heritage organisations could take.

Introduction

Throughout the ages, the introduction of new technologies has caused substantial insecurities. What is this? How can I use it? Can I trust it? What are the drawbacks? The first reflex is often to fall back on the well-known world, either by refusing to use the 'novelties' or by trying to model their use to well-known practices. At the beginning of World War I early radio transmissions were distrusted by senior commanders in the English Navy in battlefield conditions and as a consequence many lives were lost. The first automobiles were made to look as much as possible like horse-drawn carriages and the first index of the world wide web, published around 1994, was a sturdy printed book. Such is human nature.

At present we are some 15 years into the digital and internet revolution and in this article I wish to explore how well libraries and museums have adapted to the new reality, focusing on born-digital cultural heritage material. I will largely base my analysis on the situation in the Netherlands, which was described in two recent reports: the Netherlands Coalition for Digital Preservation (NCDD)¹ carried out a national survey of digital preservation activities (and gaps) in 2009, which was published under the title *A future for our digital memory*² (henceforth referred to as the 'NCDD report'), and Digital Heritage Netherlands (DEN), the national knowledge centre for ICT applications

in the cultural heritage sector,³ carried out more detailed research into born-digital collections in a selected number of Dutch cultural heritage institutions in late 2009/early 2010 (henceforth referred to as the 'DEN survey').⁴

Take-up of digital materials

The NCDD report identified three public (sub)sectors, each with their own dynamics: government and archives, the scholarly community, and the cultural heritage sector. Marked differences were noted with respect to the degree to which digital objects have become an integral part of life in these communities. Science, technology and medicine lead the way; they have become 95% born-digital. The humanities follow at a somewhat slower pace, as the object of research is still often a physical object. Government and archives lag behind, but their use of digital objects is on the rise. With the exception of audiovisual media, however, the cultural heritage sector still pretty much revolves around the physical object. This is to be expected, as the benefits of digital information are quite immediate for science and technology, whereas it has always taken time for 'culture' to become 'heritage'. This almost 'natural' time lag enabled cultural heritage institutions to postpone their acquisition and/or conservation decisions until a certain amount of hindsight had brought to light

what was worth acquiring/conserving at all. But, as we shall see, this relatively comfortable position is about to change radically.

The technical challenges of digital objects

Digital objects have a number of features which fundamentally affect the traditional workflow of libraries and museums. They have a particular form of fragility which needs to be addressed if we are to make the cultural heritage available to future generations. This fragility is primarily a consequence of the very characteristic that makes them so fast, user-friendly and ubiquitous: digital objects consist of binary codes which are machine-readable and machine-readable only. Humans need the intervention of media, hardware and software to be able to access the information. As these evolve rapidly, digital objects run the risk of becoming illegible with time. Whereas printed objects can usually survive years and years of 'benign neglect' in our stacks, digital objects may lose their usefulness within a 5- or 10-year time span.

Two main strategies have been developed to address those risks:

- Migration: transferring digital objects from obsolete media and file formats to newer/more robust media and file formats;
- Emulation: software can be developed to enable modern computers to 'emulate' the old software/hardware combination on which the digital object was created.

Both preservation strategies require any digital collection to be monitored *continually* in a technically advanced storage environment by a staff trained to diagnose possible risks and address them in time. As the information world expands and develops rapidly, this is a very research-intensive – and thus resource-consuming – task for custodians of digital collections, a task of such proportions that it can only be carried out by large-scale organisations or collaborations between such organisations.

As if this were not enough, the word *continually* implies yet another complication: caring for digital objects starts at the moment of creation, when important technical choices are made which affect the longevity of the object for all time to come. Some file formats are more robust than others; key metadata are essential for recreating the

environment in which the object was created or for migrating the object to new environments without severe loss of data. This means that in the digital era acquisition and preservation decisions have to be made *before* a certain lapse of time has allowed us to determine which objects are worth acquiring and preserving at all.

Born-digital objects are even more special

Digital objects come in many shapes and sizes. In the cultural heritage sector we generally distinguish between:⁵

- Digital information about (physical or digital) objects, such as descriptions, detail photos or digital reconstructions. In the Dutch national survey cultural heritage institutions often listed these as their only digital 'collections'. They are mostly to be found in structured databases such as collection management systems.
- Digitised reproductions of physical objects, mostly produced to facilitate internet access to (representations of) physical collections.
- Born-digital objects.

When questioned about this distinction, participants in the DEN survey indicated that they understood the differences, but that in practice such differences played no role in their dealings with digital objects. This is a clear indication that policy making with regard to digital objects is relatively underdeveloped in the cultural heritage sector, because if we take a closer look at these categories, we can discern a number of properties that may have substantial impact on the level of resources we are willing to allocate to their acquisition and preservation:

- Born-digital objects do not have a physical counterpart, which means that loss of the digital object is irreparable. The loss of digitised objects may be a financial and organisational setback, but it is not irreparable.
- Born-digital objects are often more complicated in a technical sense than digitised objects. Whereas the latter are often relatively simple pdfs, tiffs, or jpegs of erstwhile printed documents, born-digital objects increasingly make use of the full potential of digital information processing by combining various techniques and file formats – which complicates long-term preservation.
- Whereas the provenance of digitised objects can

usually be traced, born-digital objects can be produced and uploaded by anyone anywhere on the internet. This makes it very difficult to develop any kind of systematic collection policy. In addition, the type of cradle-to-grave care which digital objects require is very difficult to organise when one does not know who the producer is or where he/she is to be found.

- Digitisation is mostly carried out by organisations with relevant expertise, which ensures that proper metadata are added to the content: born-digital objects often lack the structured metadata necessary to properly render the binary code.
- Many born-digital objects never reach a definitive, 'archival' state as they are created, used and reused, mixed and mashed to new digital objects by an (inter)active web 2.0 community.

It is obvious from the above that born-digital objects run far greater risks than digitised objects – not only in a technical sense, but also, and perhaps even more importantly, in organisational terms. What goes on on the world-wide web to a large extent eludes the organisational structures we have inherited from the printed era. Anyone with a personal computer and an online connection can now 'publish' content anywhere on the web and backbones of the printed library era, such as national bibliographies and national legal deposit systems, are now failing us – especially where born-digital material is concerned.

Born-digital content in the Netherlands

To start on a positive note, there are some categories of born-digital content that have relatively easily found their way into Dutch (and I would imagine British) cultural heritage institutions. This does not always mean that preservation and access in the long term are as yet guaranteed, but it does mean that at least the *responsibility* has been assigned and accepted for securing long-term access to these materials:

- publications: Early in the 1990s the Koninklijke Bibliotheek (KB) became aware of the potential of the digital revolution and decided to develop a digital deposit library. The e-Depot became operational in 2003 and provides permanent access to all digital publications of Dutch origin and those which are deposited in the universities' digital repositories. In line with the internationalisation of the publishing industry, the e-Depot also offers archiving facilities to international scholarly publishers; so far it is mostly scientific (STM) publishers who have availed themselves of this opportunity.
- output of Dutch public broadcasting companies: the Dutch public broadcasting companies had already joined forces in the analogue era to archive their output collectively in the Institute for Sound and Vision and this practice was continued with respect to digital broadcasts.
- archaeological data: The Valletta Treaty of 1992 included a legal obligation to preserve data from archaeological sites and in the Netherlands all parties involved successfully joined forces to develop the EDNA repository which is hosted by Data Archiving and Networked Services (DANS), the national archive for the humanities and social sciences.
- archive collections: although for a long time digital workflows still resulted in printed records, the National Archives (NA) and the municipal archives of Rotterdam and Amsterdam have opened up digital repositories in 2009 and are expected to extend their services to collections not subject to the Public Records Act in due course (although implementation of such services may yet take some time).

When we take a closer look at these initiatives, it becomes clear that born-digital materials are at present collected, preserved and made available successfully in situations where the number of producers is limited and roles and responsibilities with regard to collection and access remain more or less the same as in the analogue era (publications, public broadcasting companies). Where society has a very direct interest in preserving and accessing records, at least the legal responsibility has been assigned (although records management *practices* still leave much to be desired).

Most of the cultural world hardly meets these requirements. There is an unlimited number of potential producers and the urgency to preserve cultural heritage is not at all as self-evident as, for instance, in the scholarly community or with respect to public records. Even in the analogue era this was clearly demonstrated by the relatively weak position of the cultural heritage sector when it came to securing funding in the Netherlands, and this situation only got worse with the onset of the credit crunch. The very ubiquitousness of digital media and digital content may even have contributed to a reduced regard for digital cultural content and thus for the urgency of preserving it. In the Netherlands institutions seem to have a harder time making the case for initiatives such as web archiving than they

do, for example, for digitisation of reputable physical collections.

Be this as it may, a number of Dutch cultural heritage institutions have begun collecting born-digital objects, although it is not always clear whether they intend to provide long-term access:

- The Netherlands Architecture Institute (NAi) aims to preserve and provide long-term access to selected architects' archives – whether analogue or digital. The plans for a digital repository are quite concrete and dedicated staff have been recruited, but sufficient funding for setting up the repository is as yet unavailable.⁶
- NIMk, the Netherlands Media Art Institute,⁷ conserves a number of media art collections for Dutch museums. Increasingly these collections are digital, but NIMk has not yet developed plans for long-term preservation.
- V2, the Institute for the Unstable Media,⁸ maintains an on-line archive of events and objects, but its long-term prospects are unclear.
- The Koninklijke Bibliotheek has begun harvesting websites which are deemed to be representative of Dutch society, language and culture. In a few years some 10,000 websites will be harvested on a regular basis and these will be preserved in the KB's e-Depot.
- The Netherlands Institute for Sound and Vision intends to harvest websites of public broadcasting companies and is extending services to regional and local broadcasting companies.
- The Netherlands Photo Museum in Rotterdam reports that it is increasingly acquiring digital photographs, but it has not yet developed long-term preservation facilities.
- The first true web 2.0 collection initiative is reported by the Graphic Design Museum in Breda, where graphic designers can upload their designs into the Dutch Design Database.⁹ Some 15,000 designs have been uploaded so far. At the moment, these can only be viewed in the museum, but there are plans for online access.

The above list may not be exhaustive, but even if it were twice as long as it is, it would not hide the fact that collection and preservation of digital-born cultural expressions has yet to take off in any real terms, in the Netherlands as elsewhere.

Organisational challenges

The technical challenges of collecting and preserving born-digital objects were described

earlier in this article. In the mid-1990s, when the impact of digitisation first came to be an area of study, these technical challenges took pride of place. Fifteen years on, the focus has shifted to the organisational issues. This does not mean that the technical work is done – that is an ongoing task with the introduction of every new piece of hardware and software – but rather that technical risks have come to be seen as a surmountable barrier. As the BBC's Richard Wright stated in 2008: 'Risk can always be reduced by adding money: more copies, more devices, more reliable devices, less data per data manager'.¹⁰

Money is always an issue in the cultural heritage sector and born-digital objects only add to the problems because of the substantial expense involved in developing trustworthy repository facilities *and* because of the enormous scale of digital production in the world today.

But money is not the only problem. Cradle-to-grave care for born-digital objects is especially difficult to organise in the cultural heritage sector which is, of all the sectors surveyed in the Netherlands, the most diverse. In scholarly communications and public records management it may be difficult to organise efficient co-operation between producers and guardians of born-digital materials, but in the cultural heritage sector it seems all but impossible.

The Dutch strategy for securing long-term access

The NCDD is, at the time of writing, developing a strategic approach to ensuring permanent access to digital objects in the Netherlands.¹¹ The coalition has come to the conclusion that any infrastructure for permanent access has to be characterised by a high level of co-operation, especially between those involved in the same information chains *within* sectors, and that any approach must have a scale that matches the scale of the challenges. Thus it is being proposed that within each (sub)sector a national organisation, with the proper network, expertise, experience and funding, will take the initiative in bringing all parties in their sector together to develop the level of co-operation and scale needed, and that their work will be supported by the Ministry of Education, Culture and Science, both in terms of a mandate and in terms of funding. Both for public records and scholarly communications, such prime movers can easily be identified (the KB, DANS (Data Archiving and Networked Services)

and the National Archives respectively) and in the cultural heritage domain Sound and Vision will take the lead for digital audiovisual collections. The remainder of the cultural heritage sector, however, presents a problem. It is so diverse and so little experience has as yet been developed with regard to digital collections, that a prime mover for digital collections cannot, as yet, be designated. All I can say for now is that the NCDD will, of course, continue to further the debate on an infrastructure for the cultural heritage sector and, meanwhile, concentrate its efforts on providing as much practical guidance as possible to help digital collections survive the present transition times.

Future prospects

Speculating on what the future will bring is always a hazardous undertaking, but perhaps you will allow me to make a few observations of a more general kind which may help us think more clearly about what is happening in the world today.

First, let us not underestimate the impact of the internet, and especially the web 2.0, revolution. We are in the middle of a fundamental paradigm shift that affects the very fibre of our societies, and it is only logical that dealing with this will take us some time – a generation or so might very well be required.

At the same time, it is imperative that we stop allowing the scale of what is happening from paralysing us – ‘us’ being the ‘traditional’ cultural heritage institutions. For however chaotic web 2.0 may seem, it is also immensely creative and to some degree it will develop its own solutions within its own logic. Although these may not always be ‘our’ solutions, we must keep an open mind about them – in order to prevent the new automobiles from looking like horse-drawn carriages.

Libraries may well find that they need not worry about preserving born-digital collections, as publishers increasingly offer content on an online subscription basis, long-term access to the originals being secured by national and international deposit libraries. Museums may find that dedicated institutions will spring up to deal with born-digital art and that a new division of labour emerges whereby some organisations concentrate on physical objects and others on digital objects. Institutions may come to deal with the cradle-to-grave issue by developing a tiered preservation strategy: simply preserving the bits for the first 10 years, and only deciding on a more intensive preservation strategy when the material has proven its cultural heritage

value. If we take our cue from developments in scholarly communications, new modes of co-operation may arise, less based on geographical criteria and more on substantive criteria.

Lastly, it may be reassuring to realise that over 90 per cent of the web contains information that is temporary or not worth collecting and preserving. The fact that we could, theoretically perhaps, preserve all of it should not guide our decisions. As our resources to manage collections are limited, selection is still the key, now perhaps more than ever.

Next steps for cultural heritage institutions

Can any practical guidance be distilled from the above for cultural heritage institutions? The first suggestion must be not to undertake collection and preservation of digital objects as individual organisations, but to seek co-operation and pool available resources. Fortunately, the digital era allows institutions to co-operate in the back office, while at the same time maintaining their individual identities in the front office, so this need not affect institutions’ positions *per se*. Rather, collaborative efforts should help bring about a sound division of labour between institutions.

Secondly, in this seemingly increasingly chaotic world, it is more important than ever for each organisation to take the time to draft the clearest possible mission statement and apply it scrupulously to every executive decision. This may seem self-evident, but in the many discussions I have taken part in during the past year, it has become clear that this is where many smaller cultural heritage institutions still have not developed, I would say, a professional or business-like attitude. For without a clear image of the institution’s identity and mission in life, it will be impossible to make any, let alone the right, choices.

Thirdly: do not overrate the importance of money or allow the absence of adequate funding to paralyse the institution. Very small practical measures in the realm of sound IT management can already have a significant effect on the longevity of digital collections, at least for the next 5 or 10 years. To mention just a few: when acquiring a digital object, find out as much as you can about the technical environment in which it was created, and document this information; store your content on servers rather than cd’s or dvd’s; make regular back-ups and send these to a colleague; and so on.

This brings me to my fourth and final point: invest in *knowledge* before anything else. Most cultural heritage institutions will not have the resources to do their own research and development, but there are enough prime mover organisations out there which are more than willing to share their expertise. Organisations such as the Dutch NCDD and the British Digital Preservation Coalition¹² will do their utmost to provide the platforms for acquiring the knowledge and expertise, but it is up to each individual organisation to put it to work.

Revised version of a paper given at the OKBN*ARLIS/NL study day, Themadag Born-digital, held in Amersfoort 30 November 2009.

References

1. Nationale Coalitie Digitale Duurzaamheid, <http://www.ncdd.nl/en/index.php>.
2. 'Toekomst voor ons digitaal geheugen' (Nationale Coalitie Digitale Duurzaamheid, 2009), <http://www.ncdd.nl/activiteiten-natverkenning.php>; a 20-page English-language summary, 'A future for our digital memory,' is available at <http://www.ncdd.nl/en/ddnl.php>.
3. Digital Heritage Netherlands: 'let's build together,' <http://www.den.nl/english/>.
4. Maurits van der Graaf, 'Born-digital materialen bij een selectie van Nederlandse erfgoedinstellingen, een verkennend onderzoek' [Born-digital materials in a selected number of Dutch cultural heritage institutions, an exploratory report] (Digital Heritage Netherlands, 2010), http://www.den.nl/getasset.aspx?id=Rapporten/Born-digitalErfgoedmaterialenBijNIErfgoedinstellingen_versie20100205.pdf&assettype=attachments.
5. Ibid.
6. An extensive English-language summary of the NAI's proposed digital repository can be found in Henk Verstappen, 'Connecting the digital with the physical LAM: building a digital repository for the NAI' (paper presented at the Hybrid Architectural Archives Conference, Nederlands Architectuurinstituut, Rotterdam, June 10-12, 2009), http://conference.nai.nl/mmbase/attachments/525859/2.2_Henk_Vanstappen.pdf.
7. Netherlands Media Art Institute, <http://nimk.nl/eng/>.
8. Institute for the Unstable Media, <http://www.v2.nl/home/>.
9. Graphic Design Museum, <http://www.graphicdesignmuseum.nl/en/exhibitions/current/dutch-design-database/309>.
10. Chris Rusbridge, quoting on the Digital curation blog, entry posted September 29, 2008, <http://digitalcuration.blogspot.com/2008/09/ipres-2008-risks-and-costs.html>.
11. At the time of writing, draft texts were still being discussed among the NCDD members and with the Ministry of Education, Culture and Science. The final text will be translated into English and made available at <http://www.ncdd.nl/en/over-beleid.php> later in 2010.
12. Digital Preservation Coalition, <http://www.dpconline.org/>.

Inge Angevaare

Co-ordinator

NCDD, Netherlands Coalition for Digital Preservation

PO Box 90407

2509 LK The Hague

The Netherlands

Email: inge.angevaare@kb.nl