Born-digital heritage materials at selected Dutch heritage institutions

an exploratory study

Commissioned by the DEN Foundation
The Hague, February 1, 2010

Maurits van der Graaf
2010-01-29

Pleiade Management and Consultancy BV
Keizersgracht 62
1015 CS Amsterdam
The Netherlands

Tel.: +31-(0)20-488 93 97
Fax: + 31-(0)20-684 77 73
m.vdgraaf@pleiade.nl
www.pleiade.nl

This document is subject to a Creative Commons Licence

CONTENTS

Executive summary 3
1. Introduction 5
2. Methods 6
2.1 Phased approach 6
2.2 Execution 8
3. Status of born-digital heritage material 9
3.1 Introduction 9
3.2 Development stage of born-digital heritage material 11
3.3 Status of the development of digital preservation repositories 15
3.4 Acquisition of born-digital materials 17
3.5 Maintenance and metadata 20
3.6 Problems selecting born-digital material 21
3.7 Digital preservation 23
3.8 Accessibility 25
3.9 Costs of born-digital heritage material 27
4. Need for a cross-institutional approach to born-digital heritage material 28
4.1 Introduction 28
4.2 Collecting born-digital heritage materials 28
4.3 Need for best practices 29
5. Towards a quantitative measurement method for born-digital heritage collections 31
5.1 Possible approaches to a quantitative study 31
5.2 Level of aggregation and granularity 31
5.3 Relevant distinction between born-digital and digitized 31
5.4 Issues regarding the quantitative description of digital archives 32
5.5 Incomparability of different types of collections and/or born-digital objects ... 32
5.6 Designing the questionnaire and the quantitative measurement instrument to be tested 33
5.7 Discussion of the results of the test measurement with a view towards a quantitative measurement instrument for born-digital heritage collections 34
5.8 Proposal for a quantitative measurement instrument for born-digital heritage collections 41
5.9 Size and growth 43
6. Summary and conclusions 44
6.1 Status of born-digital heritage material 44
6.2 Need for cross-institutional approach 46
6.3 Quantitative test measurement 46
Appendix A Overview of born-digital heritage collections at the institutions that were interviewed 48
Appendix B Other survey results 50
Appendix C Survey questionnaire 51

Executive summary

Exploratory study among selected heritage institutions

The goal of this study is to provide a practical contribution to the way in which born-digital collections are managed and maintained by heritage institutions. Forty Dutch heritage institutions were selected because there was evidence that they were dealing with born-digital heritage material. What can the heritage sector learn from these institutions? The exploratory study consisted of three parts: interviews, a number of focus groups and an online survey. A total of 29 heritage institutions participated in at least one of the aforementioned parts of the study. It must be stressed that this selection is not representative, but rather that the institutions are pioneers in the collecting and management of born-digital material.

Results of the status of born-digital heritage material

- *Development stage:* More than half of the respondents already include born-digital heritage material in their collections (execution stage). Of the other organizations, some are at the implementation stage (plans to process born-digital material are currently being made) and some are at the planning or idea formulation stage. (The assumption is that a number of Dutch heritage institutions *outside* the institutions selected for this study are at the idea formulation stage.)
- *Acquisition:* Most of the institutions indicated that they process the born-digital material they get from their usual acquisition sources. For many of these institutions, this meant changing the way they work. More than half of the institutions indicated that born-digital material led to the creation of new collection areas that could be relevant for their collection. Only some of these institutions are doing acquisition in these new collection areas. The interviews revealed two specific aspects of acquiring born-digital material:
    - attempts to reach the producer/archive creator at an early stage of the archive creation
    - new ways of acquiring material due to the emergence of user-generated content
- *Digital preservation*: Permanent access to digital material is forcing heritage institutions to perform new 'technology watch' tasks and actions to guarantee that files remain accessible. The extent of the problems surrounding permanent access is largely determined by the number of different file formats a collection contains: more than 60% of the heritage institutions indicated that their collection contains fewer than 20 different file formats, and 5% indicated that theirs contain more than 50.
- *Accessibility*: More than 65% of the institutions indicated that some of their born-digital material will be accessible via internet. Some of the material, however, cannot be accessed via internet due to copyright and/or confidentiality issues.
- *Costs:* Specific cost items for born-digital material include:
    - the development of software for digital preservation repositories
    - storage costs
    - cost of new tasks related to permanent access to digital material

Wishes regarding a cross-institutional approach

- *Born-digital heritage is being lost:* Interesting Dutch born-digital heritage material is being lost because it is not or is not sufficiently collected by the respective heritage institutions. A clear majority of the heritage institutions confirm this and strongly believes that acquiring born-digital heritage material from these new sources urgently requires a clear allocation of tasks.
- *Best practices:* There is a great need for best practices related to born-digital heritage material in the area of digital preservation (95% of the respondents), acquisition (more than 70%) and selection (some 60%).

Quantitative test measurement

- *Results of the test measurement:* The quantitative test measurement used in this study provides an indication of the size of born-digital collections at Dutch heritage institutions. It is worth noting that new types of born-digital objects, such as websites, games and 3D designs, are currently not collected or are only collected in dribs and drabs.
- *Explosive growth expected:* A large majority of the institutions indicated that they expect their born-digital heritage collections to grow substantially in the next five years.
- *Designing a quantitative measurement instrument*: A quantitative measurement instrument was designed for born-digital heritage collections based on this study. The instrument suggests describing born-digital *archives* differently than heritage collections. Born-digital archives consist of a variety of different objects and file types that, in contrast to other collections, are not seen as separate collection items but as part of an archive or file. For the other types of heritage collections, the suggestion is to distinguish 19 categories of object types.

This research was commissioned by the DEN Foundation and supported by the Dutch Ministry of Education, Culture and Science.

1. Introduction

Exploratory study based on Digitale Feiten

The *Digitale Feiten* (Digital Facts) study carried out by Dutch heritage institutions in 2008 is the Dutch equivalent of the NUMERIC project carried out by the European Commission.[1] The study used a survey to map out the extent of digital heritage collections. The *Digitale Feiten* and NUMERIC studies focused on digitized material: born-digital heritage material was purposely ignored because it was thought to be problematic. This exploratory study was designed to map out the specific problems of born-digital heritage at selected Dutch heritage institutions.

Goal and research questions

The main goal of this study is to provide a practical contribution to the way in which born-digital collections are managed and maintained by heritage institutions such as libraries, archives, museums, archaeological institutions and institutions in the field of architectural history.

The research questions for this study are:

What is the status of born-digital heritage material at selected Dutch heritage institutions?

- *Example collections:* What are examples of born-digital heritage collections in the Netherlands?
- *Acquisition and maintenance:* How are these example collections currently maintained and expanded?
- *Growth:* What are the growth expectations for these collections?
- *Bottlenecks:* What are the major bottlenecks in the maintenance and permanent accessibility of these example collections?
- *Costs:* What knowledge do the involved institutions have about the costs of acquiring, storing, maintaining and making born-digital collections accessible and what can be learned from that?

What is a consistent and repeatable way of quantitatively measuring born-digital heritage collections?

- *Terminology:* Is there a common terminology to describe born-digital heritage material?
- *Classification:* Can a common classification be created to distinguish types of born-digital heritage material that can be reused for quantitative research?
- *Quantification:* How can these born-digital heritage collections be measured?

This report presents the information gathered during the interviews, focus groups and the online survey. The proposed actions and recommendations resulting from this study will be presented separately by the DEN Foundation[2].

[1] Refer to http://www.numeric.ws/

[2] The DEN Foundation (short: DEN) is the Dutch national ICT knowledge center for cultural heritage. DEN supports archives, museums and other heritage institutions to improve their digital strategies and services (see: www.den.nl).

2. Methods

2.1 Phased approach

The survey was carried out in the following phases:

1. **Exploratory round:** An exploratory round consisting of a website and documentation analysis – supplemented by (phone) conversations – enabled 40 Dutch heritage institutions to be identified that were already working with born-digital heritage material or had plans to do so. Twenty of these institutions were approached for an interview. The exploration started in July and finished in August 2009.
2. **Interview round:** Twenty institutions took part in this research phase: nineteen institutions were interviewed once; one institution submitted information in writing. The interviews were conducted between July and the beginning of October 2009. Most of the interviews were conducted before the focus groups, and a few were conducted after the focus groups for planning reasons (see overview on next page).
3. **Discussion report:** A discussion report describing all of issues the respondents raised about born-digital heritage material was written after the interviews. The discussion report was sent to the respondents that were interviewed prior to the focus groups together with an invitation to take part in the focus groups. The report was also sent to a few other people who were not interviewed but who took part in the focus groups (see overview on next page).
4. **Focus groups:** The discussion report was used in two focus groups (30 September, 1 October) to discuss how born-digital heritage material could best be measured quantitatively. A total of thirteen employees from Dutch heritage institutions participated in these focus groups (see overview on next page).
5. **Online survey:** Based on the results of the focus groups and the interviews, an online survey was designed in a first attempt to quantitatively measure born-digital heritage material. The online survey was sent to the 19 interviewees[3] and 21 Dutch heritage institutions for which there was evidence that they were collecting born-digital heritage material or had plans to do so. It is important to note that these results *are not representative for Dutch heritage institutions*. The invitations to participate in the online survey were sent out by email on 4 November. Reminders were sent out on 16 and 25 November. The survey was closed on 1 December. In the end, 22 of the 40 invited institutions completed the survey, representing a response of 55%[4].

[3] The definition of 'heritage material' was refined based on the discussion report and resulted in a previously interviewed organization being discarded. This respondent was not asked to complete the survey (see 3.2).

[4] Of the 11 organizations that did not participate, three responded as follows: one indicated that it did not have any born-digital material, one indicated that it only had digital photo materials and one (with an audiovisual archive of its own events) had been too busy to complete the survey.

Participating heritage institutions	Interviews	Online survey
Aletta (formerly IIAV)		x
Library of the Technical University of Delft	x	x
DANS (Digital Archiving and Networked Services)	x	
DOK Library Concept Center		x
Rotterdam Municipal Archives	x	x
Graphic Design Museum	x	
ICN (Netherlands Institute for Cultural Heritage)		x
IISG (International Institute of Social History)	x	
Netherlands Institute for Sound and Vision	x	x
KITLV (Royal Netherlands Institute of Southeast Asian and Caribbean Studies)		x
National Library of The Netherlands (Royal Library: KB)	x	x
Letterkundig Museum (Netherlands Literature Museum)[5]	x	
Nationaal Archief (National Archives)	x	x
Naturalis (National Museum of Natural History)	x	
Netherlands Architecture Institute	x	x
Netherlands Photo Museum	x	
Netherlands Media Art Institute (NIMk)	x	x
Noord-Hollands Archief	x	x
PARK4DTV		x
Regional Archives Nijmegen		x
RKD (Netherlands Institute for Art History)	x	x
Stadsarchief Amsterdam (Amsterdam City Archives)	x	x
Stedelijk Museum	x	x
Tresoar		x
Leiden University Libraries	x	x
Utrecht University Library (Igitur)	x	
UvA Erfgoed (Heritage Collections at University of Amsterdam		x
V2_, Institute for the Unstable Media	x	x
Zeeuwse Bibliotheek (Zealand Library)		x

Focus group participants
Annet Dekker, Virtual Platform (not a collection-managing organization)
Arie Altena, V2_ Institute for the Unstable Media
Bart Rutten, Stedelijk Museum
Boudewijn Ridder, Netherlands Photo Museum
Corrie-Christine van der Woude, Rotterdam Municipal Archives
Gaby Wijers, Netherlands Media Art Institute (NIMk)
Henk Vanstappen, Netherlands Architecture Institute
Jacob Takema, Rotterdam Municipal Archives
Marcel Ras, National Library of The Netherlands
Mathijs Holwerda, Leiden University Libraries
Milco Wansleeben, DANS
Paul Suijker, Delft University of Technology Library
Remco Verdegem, Nationaal Archief

[5] This interview was conducted by Gerhard Jan Nauta of the DEN Foundation.

2.2 Execution

The study was carried out by Maurits van der Graaf (Pleiade Management and Consultancy) in close collaboration with and under the supervision of Gerhard Jan Nauta (The DEN Foundation: Digital Heritage Netherlands).

The following people were on the project's supervision committee:

- Annet Dekker – Virtual Platform
- Gerhard Jan Nauta – Researcher at The DEN Foundation (Digital Heritage Netherlands)
- Jacqueline Slats – Head of Digital Preservation, Nationaal Archief (National Archives)
- Janneke van Kersen – Directorate for Cultural Heritage, Dutch Ministry of Education, Culture and Science
- Jos de Haan – Professor of ICT, Culture and Knowledge Society, Erasmus University and employed by the Netherlands Institute for Social Research
- Ma Oeh Pe – Project Leader Archives Task Force
- Marco de Niet – Director, The DEN Foundation (chair)

3. Status of born-digital heritage material

3.1 Introduction

Overview of the status of born-digital heritage material

This chapter outlines the status of born-digital heritage material at Dutch heritage institutions based on the interviews and part of the results of the online survey. This introduction defines a number of terms and the significance of born-digital as distinguishing characteristic.

Definition of born-digital heritage

ABC-DE, a dictionary for digital heritage published by The DEN Foundation (2008), defines digital heritage as follows:

Digital heritage: digital form of cultural heritage. There are three types of digital heritage material:

- *Born-digital material*: heritage material that is not available in any other form than digital, such as digital archives, digital art or photos that were taken with a digital camera.
- *Digitized heritage material*: heritage material of which the original form is not digital but that has been reproduced in a digital format.
- *Digital information on heritage*, for example, descriptions, close-ups or digital reconstructions of the heritage object. The information is usually made available in a structured form (for example, a database).

This research uses the above term, whereby we must emphasize that it refers to objects that do not exist in an analogue form.

Difference between born-digital heritage information and digital information on heritage

This study focuses on born-digital heritage. The distinction between born-digital heritage and digital information on heritage is not always easy to make. In many cases, digital information on heritage is born digital and in a number of cases can also be considered as heritage. An example mentioned in the focus groups is the difference between artistic videos (videos intended to be seen as works of art) and videos that document works of art. The respondents in the focus groups attach a different value to artistic videos when it comes to digital preservation: a loss of quality was less acceptable for artistic videos.

Definition of heritage collection

The Netherlands Institute for Heritage (www.erfgoednederland.nl) defines heritage as the tangible (objects) and intangible (stories) expressions of a society's culture that have been handed down from generation to generation. It also states that: 'Heritage makes a society aware of its origins and its culture'.

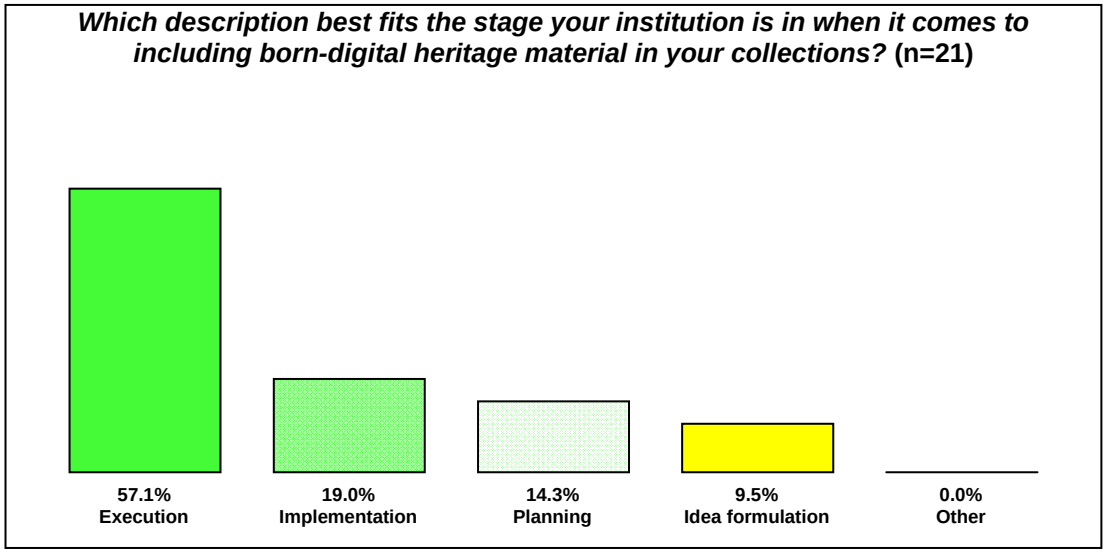
It was important to determine whether this study would include scientific research data: are collections of research data part of heritage or not? We decided to include research data on cultural objects, such as excavations, landscapes and nature in this study.
We also needed to determine how this study would handle the term ‘heritage *collection*’. In the first phase of the study, we interviewed a respondent from Igitur, the department of the University Library of Utrecht that promotes digital publishing. Igitur publishes digital journals, but hands over the individual articles to the National Library of The Netherlands for long-term archiving. Because Igitur's electronic journals are not set up and managed as a collection, they seem to fall outside the notion of heritage collections.

Born-digital as distinguishing characteristic

The interviews revealed that most of the collections contain digitized and born-digital material and that both are managed in the same system. Most of the interviewees stated that it was not common to make a distinction between born-digital material and digitized material: they understood what was meant, but didn't usually make this distinction themselves. Only a few heritage collections contain only born-digital material: for example, the National Library of The Netherlands's collection of archived websites and V2_'s electronic archive.
Although most of the collections contain both digitized and born-digital materials, the distinction between the two is in many cases relevant for the collection manager in terms of acquisition, metadata and digital preservation.

3.2 Development stage of born-digital heritage material

From idea formulation to execution



Which description best fits the stage your institution is in when it comes to including born-digital heritage material in your collections?	n	answers %
Execution: we collect born-digital material and add it to the collection.	12	57.1
Implementation: plans to process born-digital material are currently being executed/implemented.	4	19.0
Planning: we're in the planning stage.	3	14.3
Idea formulation: we're in the idea formulation stage.	2	9.5
Other (please explain)	0	0.0
21 answers		100.0

The above bar chart and table show the results of the question about the stage of development the surveyed institutions' born-digital projects are in.
It is important to repeat that these results are not representative for Dutch heritage institutions: the surveyed institutions were selected because there was evidence that they were dealing with born-digital heritage material.
The results are discussed using information obtained during the interviews.

Idea formulation stage

Almost 10% of the surveyed heritage institutions are formulating ideas on how to handle born-digital heritage material. One of the heritage institutions that is in this stage stood out during the interviews:

- The Netherlands Literature Museum (NLM) collects letters, manuscripts, photos, works of art, cuttings and other objects that belonged to or are about famous Dutch literary personalities after 1750. The organization has a national function in this area. The acquisition and presentation policy is aimed at unique – often

personal – objects that were published in small numbers. The interviewees stated that the museum really has to start thinking about acquiring born-digital material. Kees Fens's archive was mentioned as an example of born-digital material. It is known that Fens actively communicated by email. NLM would like to ingest the email collection, which raises the question of how it should be transferred, described, processed and stored, and made available at a later stage.

- The Leiden University Libraries have always managed the university's archives[6] and collect and preserve the personal archives of important professors and important people in the history of science and national literature. This is partly due to the fact that the university's library is home to the Library of the Association for Dutch Literature (*Bibliotheek van deMaatschappij der Nederlandse Letterkunde*). The interviewees expect born-digital material will soon be an integral part of the university and the personal archives they collect. Paper material in personal archives is already frequently accompanied by digital media. Up to now nothing has been done with this material.

The assumption is that a number of Dutch heritage institutions outside the organizations selected for this study are in this stage: they know born-digital material is increasing and are trying to get a feeling for it.

Planning stage

Almost 15% of the surveyed heritage institutions indicated that they are in the planning stage.
The interviews revealed the Noord-Hollands Archief as an example: a project to ingest and process a private digital archive is about to start. The project is seen as a learning curve for the whole organization. The result of this project is expected to be a set of rules on how born-digital material is to be ingested and processed. The interviewee stated that this preparation is important for the organization because the expectation is that the personal and governmental archives collected will contain more and more born-digital material.

Implementation stage

Almost 20% of the surveyed heritage institutions indicated that they are in the implementation stage.
The interviews revealed the Netherlands Architecture Institute (NAi) as an example: The NAi collects and saves archives created by architects and architectural offices that are winding up their business for a variety of reasons, including retirement. The NAi saves these archives so that researchers can look into the different steps of the design process for a specific building and/or an architect's life work. At the beginning of 2008, the NAi started a project[7] to study how born-digital archives should be treated. A few architectural offices were approached and interviewed about their policy on the use, storage and archiving of born-digital material. Moreover, these offices collected a number of examples of born-digital materials and thus established the test environment. Based on the experience gained from this project, functional

[6] When asked, theLeiden University Libraries turned out to be an exception: other university libraries, with the possible exception of the University of Amsterdam, do not archive their own university materials.
[7] This project is extensively described in 'Connecting the digital with the physical LAM: building a digital repository for the NAi', see conference report *Hybrid Architectural Archives: Creating, Managing and Using Digital Archives* (NAi Rotterdam, June 10-12, 2009).

specifications were written for a production environment that is expected to be up and running soon. During the project, the first major acquisition of born-digital material was made in the form of Dutch architect Carel Weeber's archive.

Execution stage

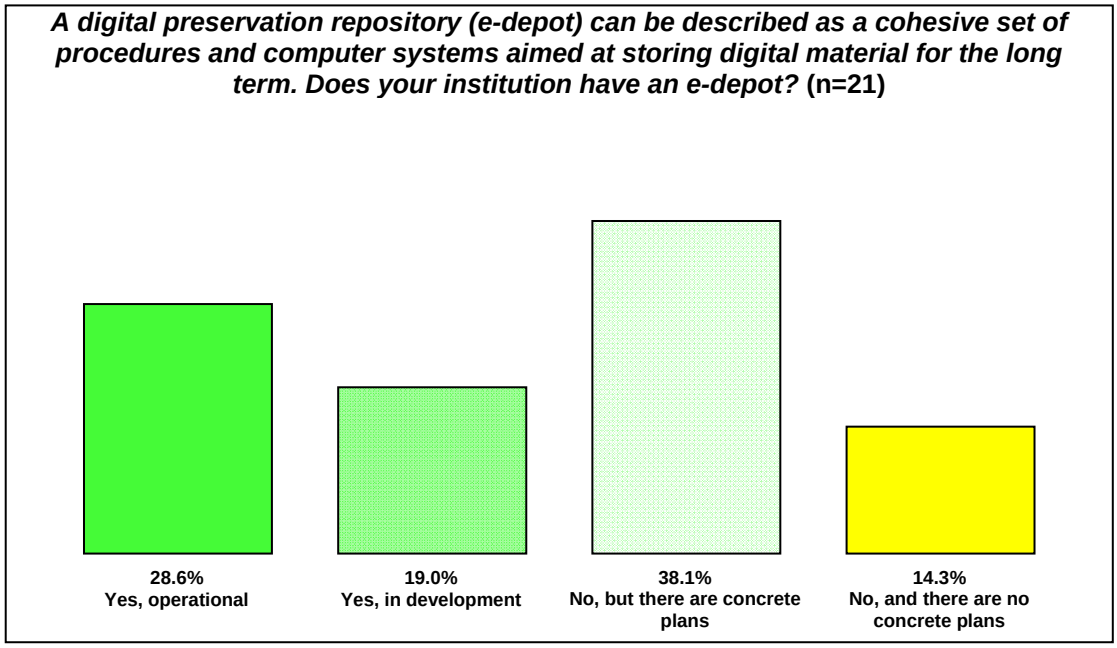
Almost 60% of the surveyed heritage institutions already include born-digital heritage material in their collections (execution stage).
The interviews revealed several institutions as examples. These institutions are described below:

- A few large institutions as pioneers (the National Library of The Netherlands, The Netherlands Institute for Sound and Vision and DANS):
    - The National Library of The Netherlands (KB) currently has two born-digital heritage collections: the so-called *E-depot*, a digital preservation repository containing electronic journals and all other kinds of digital publications that the KB collects as part of its depot task, and the *web archive*, which consists of a selection of important Dutch websites with the goal of ensuring their permanent accessibility.
    - DANS – Data Archiving and Networked Services – collects research data from the humanities and social sciences and ensures the data remain accessible. For this purpose, DANS developed sustainable archiving services in the form of the EASY system (Electronic Archiving System), which is available to all humanities and social sciences researchers as a long-term data storage and retrieval system. An important part of its contents consists of data from archaeological excavation projects.
    - The Netherlands Institute for Sound and Vision archives all of the television and radio programmes produced by the public broadcasting services: the programmes enter the archive in digital format.
- Smaller institutions as pioneers (V2_, NIMk, Netherlands Photo Museum):
    - V2_, the Institute for the Unstable Media, was founded 27 years ago and focused from day one on all kinds of artistic expression that used an electronic signal. The institute has since shifted its focus to digital works of art and organizes events for art and new media. The institute also has an *artist-in-residence* programme and a laboratory that develops technology and helps artists. The digital archive was created as a result of the aforementioned activities and consists mainly of digital photos, event videos and video documentaries of works of art.
    - The Netherlands Photo Museum has two acquisition channels: the purchase of photo series and the ingestion of photographers' archives on retirement. Almost all photographers have switched to digital photography. Considering that the ingestion of archives is lagging by a few years, the archive acquisition channel contains no or very little born-digital material. This is an issue when buying photo series. When the Photo Museum makes a purchase, they acquire the original digital photo file and one or more reference prints. The digital photos can then be distributed for use outside the Photo Museum via its website or stored for the long term.
    - The Netherlands Media Art Institute (NIMk) supports media art in three core areas: presentation, research and preservation, and offers a broad range of services for artists and art institutions through facilities. Born-digital materials play a role in all of NIMk's activities. The institute manages such things as the (artistic) video collections of a number of

institutes and has a preservation strategy for all modern art media in the Netherlands. The master copy of all of the art video collections that have been preserved in the Netherlands are in the institute's safe. It serves as a central depot for the Dutch art video collections. Since 2003, most of the videos were recorded with a digital camera and can hence be called born digital.

- Pioneers in the world of archiving (Nationaal Archief (the National Archives), , Rotterdam Municipal Archives (*Gemeente Archief Rotterdam, GAR* in short) and Amsterdam City Archives:
    - The Nationaal Archief started a project to develop a digital preservation repository (*E-depot*). The project is aimed at supporting the entire archiving chain, from recording to making (born-)digital archive material available. The digital repository that was set up as part of this project went live in 2009.
    - Amsterdam City Archives is very busy developing a digital repository in the form of a searchable repository (completed), a digital preservation repository (completed) and a repository where digital objects can be submitted (work in progress).
    - The people we spoke to from the Rotterdam Municipal Archives stated that the government archives need to make sure the archive producer/creators' digital information is well structured so there are no problems when the information is submitted. The municipality of Rotterdam set up two repositories for this purpose: a digital preservation repository for the archive (*E-depot GAR*) and a repository for the municipal organization (*E-depot CONCERN*). The project consists of much more than just configuring software and hardware, the entire system is considered: employee awareness, procedures, policy and the surrounding organization. The E-depot GAR project is expected to start soon, the E-depot CONCERN may take a bit longer.

3.3 Status of the development of digital preservation repositories



The survey contained a question about the development of a so-called digital preservation repository (*e-depot*), defined as a cohesive set of procedures and computer systems aimed at storing digital material for the long term.[8] The results of the survey are displayed in the above bar chart: less than half of the participating heritage institutions have a digital preservation repository that is operational or in development, almost 40% have concrete plans and almost 15% have no plans.

The interviews revealed the following: All of the institutions that were setting up a digital preservation repository stated that it was intended for both digitized and born-digital materials. A few institutions already need a second generation preservation repository (National Library of The Netherlands, DANS and V2_).

The following issues about the development of software were raised:

- **No standard packages:** No standard (one-fits-all) software packages for digital preservation repositories are currently available on the market. There are also no examples of institutions that have all their ducks in a row, which is why a lot of the interviewees feel they have to reinvent the wheel themselves. This is why the interviewed institutions see the development of a digital preservation repository as a major and expensive undertaking.

[8] Different forms of storage for born-digital objects were mentioned during the interviews:

- **Hard disk; the whole computer:** a few digital art objects are stored by putting the hard disk or the whole computer in the safe.
- **Digital Betacam tapes:** in many cases, videos (both born-digital and digitized) are stored on tape in Digital Betacam format.
- **Servers:** the e-depot systems work with servers on which the data are stored.

- **Collaborative software development:** The reasons mentioned above explain why institutions are seeking to collaborate with similar institutions to develop software for digital preservation repositories. In addition to the example of the Nationaal Archief, which is developing a repository together with a number of other archives and will make the software available to others, it was mentioned in the interviews that V2_ and De Balie are also developing repository software in cooperation.
- **Digital preservation actions in separate module**: A common characteristic of digital preservation repostitories is that the actions to ensure enduring access (emulation - using hardware/software to duplicate the functions of earlier computer systems - or migration of file formats - conversion of data to newer or more common file types) are effectuated in a separate module.
- **Accessibility in separate environment**: In most of the cases, accessibility is also handled in a separate user environment. Nearly all of the interviewees saw opportunities to make parts of the collection accessible to users via internet, provided it does not violate the copyright laws for digital material.
- **Possibilities for shared digital preservation repositories:** An interviewee mentioned that his organization may not need to develop its own digital preservation repository, but could use another (existing) repository instead. Another interviewee explained that conventional institutions always manage their users' physical heritage objects, but that in the case of a digital archive their users would be able to access the material even if it were stored elsewhere.

Work processes and organization

Several interviewees stated that a digital preservation repository is more than just a software environment for digital material: it requires other/new procedures and work processes as well as other/new skills, and according to some, even a different organization. The following comments were made in this context:

- **Other skills, new jobs**: Several interviewees emphasized that collecting and managing born-digital heritage material requires other skills, especially in the area of ICT. The Rotterdam Municipal Archives created a new job: *e-conservator* (digital archivist). This person will enter archival material in the repository and make sure it is permanently accessible. This means that the employee has to have technical skills.
- **Different work processes, different organization:** The work processes for born-digital heritage material can differ from those for traditional heritage material. This is why a number of interviewees believe the way heritage material is acquired, managed and accessed will have to change.
- **Different boundaries between collections:** One interviewee believes that the boundary between the different collections will change: the difference between manuscripts (unpublished material) and 'official' publications changes because information can be published on the web e.g. in the form of blogs and tweets. Another interviewee stated that the websites to be archived were acquired by the information services department instead of by the archives department: websites are actually published materials.

3.4 Acquisition of born-digital materials

Status of the acquisition of born-digital heritage materials

Which of the following statements relate to your institution? [MULTIPLE ANSWERS ARE ALLOWED]	n	answers %
We receive born-digital materials through our usual acquisition channels. Nothing is done with these materials at the moment.	2	9.1
We receive born-digital materials, which we process, through our usual acquisition channels.	15	68.2
Born-digital material has dramatically changed the way our usual acquisition channels work.	8	36.4
New sources/collection areas for born-digital material have been created for our field/domain that could be relevant for our collection.	12	54.5
New sources/collection areas for born-digital material, in which we are doing acquisition for our institution, have been created for our field/domain.	8	36.4
22 answers		100.0

The above table shows the results of the acquisition of born-digital materials by the surveyed heritage institutions.

A few of the preselected heritage institutions (<10%) stated that they receive born-digital materials through the usual acquisition channels, but that they don't do anything with it. Almost 70% stated that they receive born-digital materials through the usual acquisition channels, which they do process. 36% stated that born-digital material has changed the way the usual acquisition channels work.

In addition to the born-digital material that is received through the usual acquisition channels, new sources/collection areas can be created that match the heritage institutions' collection profile. More than half of the surveyed heritage institutions confirm this (54%), and a significant number (36%) of institutions are already collecting born-digital material from these new sources/collection areas.

How has the acquisition of born-digital heritage material changed?

The interviews revealed the following aspects specific to acquiring born-digital heritage material:

- **Contact archive producer/creator in an earlier stage**: The transfer of personal and corporate archives usually takes several decades. This is too long for digital information: it's often not readable anymore and/or is organized in such a way (i.e. so badly) that it would take a lot of manpower to process it. This is why most of the institutions that receive archives from third parties are making plans to approach the archive producer/creator about born-digital materials at an earlier stage. The following types of actions are distinguished:

- ***Information:*** In workshops and meetings, the Netherlands Photo Museum provides photographers with as much information as possible about documenting and storing their photos for the long term.
- ***Guidelines:*** The Amsterdam City Archives have put guidelines in place describing how digital archives are to be delivered to the municipal organization.
- ***Changes in information management:*** One of the municipality of Rotterdam's aims is to implement a digital repository (*E-depot CONCERN*) to improve the way the municipal organization manages information. *E-depot CONCERN* is not only a software environment, but also an organizational entity that will help municipal services to (sustainably) manage digital information. The implementation of the system is an attempt to ensure that digital information is managed such that it can be handed over to the municipal archives without creating any problems.
- ***Information management services:*** Government organizations need to keep different types of information for several years for legal reasons (accountability to citizens, tax authorities and such). A similar rule applies to architectural firms: in addition to their accountability, they are keen on getting inspiration for new buildings from earlier designs. These institutions have to save their born-digital information for several years. This is why the Nationaal Archief is implementing a service that will enable the departmental archives containing born-digital material to be managed. The Netherlands Architecture Institute is thinking about implementing a similar service for architectural offices. This type of service will probably be available for a fee, but is considered to be in the interest of the archival organization because it will make it easier to hand over the born-digital archive at a later stage.

- **Estimate conservation costs when acquiring born-digital art work:** Museums will create checklists so they can estimate the cost of conservation and permanent access as soon as they start considering acquiring born-digital art work. The costs can, of course, differ greatly between digital art works (in contrast to the conservation costs for, say a painting) and can even exceed the purchase value.
- **Acquiring new born-digital heritage domains:** The digitization of our culture creates new domains in which heritage organizations can collect born-digital material. The KB started archiving a selection of Dutch websites and a few other institutions followed suit. This study revealed the following gaps as to collecting new forms of born-digital heritage:
  - ***Websites:*** One of the topics that arose during the interviews is that artists create their own websites and publish a lot of documentation about their own work. None or very few of these websites are being archived. Another topic of discussion was the lack of attention for exhibition websites. Moreover, websites belonging to other people and/or organizations are hardly ever archived.
  - ***New forms of communication via internet***: Blogging, twittering and such were mentioned, but no institutions that include these types of communication in their archives were mentioned during the interviews.
  - ***User environments:*** The Stedelijk Museum started a project that consists of putting (parts of) the collection in an online gaming environment. The intention is to preserve this environment for the long term. Another topic

that arose during the interviews was whether different versions of application files should be kept: there is no policy in this respect.

User-generated content (Web 2.0)

A new way of acquiring born-digital heritage is to let people outside the heritage institutions add information to a database or website. Such applications are denoted as Web 2.0 and the product is user-generated content. The interviews revealed the following interesting examples:

- The Dutch Design Database at the Graphic Design Museum in Breda, where graphic designers can upload their designs. The database currently contains some 10,000 (mainly born-digital) objects. Some of the database's contents are exhibited in the museum, but the intention is to make the content accessible via the web.
- Natural history museum Naturalis manages a website of reports of whales that stranded along the Dutch coast. The database contains a combination of historical and contemporary observations that interested parties can view and add via the website. Another website is the Dutch Species Catalogue, which provides an overview of Dutch biodiversity with photos. Nature photographers can upload their photos of wildlife or plants. The animals or plants are identified and named by experts and published on the website.
-

3.5 Data management and metadata

What are the specific issues concerning metadata for born-digital heritage material? The interviews revealed the following:

- **No physical/analogue counterpart:** Many interviewees emphasized the uniqueness of a born-digital heritage object as opposed to most digitized heritage objects[9] that still have a physical counterpart.
- **New content-related metadata are needed:** Digitized objects are described using the descriptive metadata of the physical/analogue objects. This cannot be done for born-digital objects, which is why (new) descriptive metadata will have to be added.
- **The abundance of digital objects can be a problem:** In a number of cases, the interviewees pointed out that the large amount of digital objects makes it difficult to create descriptive metadata for born-digital objects (see section 3.8).
- **Several versions of one object:** Another issue that was frequently raised was that less-well organized personal archives and archives of small, private institutions frequently contain several versions of one object. This raises the following questions: Does every version have to be saved? How can we assist users in navigating multiple versions?

It was also mentioned in the interviews that technical metadata is usually generated automatically. A number of tools are available or in development. Finally, a few interviewees mentioned the structural metadata of complex digital objects. These structural metadata are used, for example, to define the relationships between the chapters in a book.

[9] A number of interviewees from the field of archiving noted that digitized material replaces physical material in a number of cases. In other words: The physical material is disposed of once it's digitized. In such cases, the digitized material becomes unique.

3.6 Problems selecting born-digital material

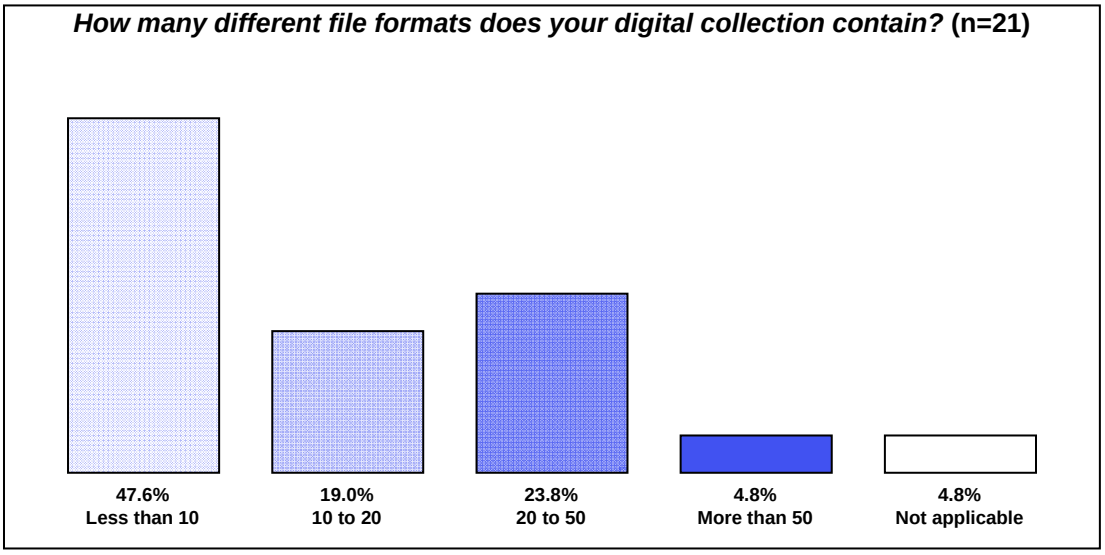
Because there is so much born-digital material, it is particularly important to apply selection criteria at the time of acquisition. The following examples were mentioned during the interviews:

- *Selection at the time of acquisition:* The National Library of The Netherlands (KB) has selected some 2,000 websites for archiving. The goal is to copy each of the selected websites in their entirety two to three times a year. The selection is based on the KB's collection profile. Websites indicated by scholars because they are important to scholarship are also acquired. The goal is to increase the number of archived websites next year to 4,000. Considering the number of websites within the NL domain, this is but a modest number. However, any solution to archive the entire Dutch web would be inpracticable and would at most enable only the presentation layer of websites to be archived. This is why it was decided to archive only a selection (however unsatisfactory this may be).
- *Selecting what should be accessible:* The Netherlands Architecture Institute is gearing up to collect the digital archives of architects. These archives contain files with text, images and audiovisual material, CAD files and such. Each building project has a large number of unorganized files, which will create problems for future users. This is why a few documents will be selected and made accessible as milestone documents. These milestone documents will give users insight into the development of a specific construction project. In other words, a preselection of documents is needed to make it easier to consult the digital archive.
- *Selecting what should be accessible:* The Netherlands Photo Museum notes that digital photography is changing the way photographers work. In the past, a photographer would shoot two roll films of 24 photos for one assignment. Today, a photographer often takes hundreds of digital photos. Photographers work in a very different way. And because they take so many photos, they don’t describe them like they used to. In principle, the Photo Museum is interested in all of these photos: they provide insight into the photographer's workflow and the less good photos can teach just as much as the good photos. The problem lies in making all of these photos accessible.
- *Selecting what should be presented:* V2_ archives visual material of events. The archive's starting point is to make everything accessible. But digital photography creates problems of its own: the photographer currently takes 150 to 200 pictures of an event. How do you present all these photos to users? The idea is tending towards a presentation in several layers.
- *Selecting at the time of dissemination:* The Netherlands Institute for Sound and Vision includes the programmes broadcast by the Dutch TV channels Nederland 1, 2 and 3 in its archive. The institute uses a metadata model that is derived from the Functional Requirements for Bibliographic Records (FRBR), which distinguishes a number of layers, such as work, realization and expression. For example: The work is Pauw & Witteman, the realization is 2009-2010 and the expression is the broadcast on 6 October. Categories are assigned to collected objects: certain expressions get an A: important material of historical value. Category B is assigned to the first game show broadcast in a season. The remaining broadcasts of the same game are then assigned category C. This categorization is a steppingstone for the way the programmes are stored and

accessed (for example, A category programmes are available on line, the other categories are available on request). This distinction in accessibility is not made yet and inflowing digital material is not actively categorized as A, B or C.

3.7 Digital preservation

Digital preservation: Status



The survey[10] contains two questions on digital preservation policies. The first question asks how many different file formats the digital collection contains. This is important for digital preservation. The results are displayed in the above bar chart. Almost half of the surveyed heritage institutions have less than 10 different file formats. 20% of the institutions have between 10 and 20 file formats. More than 25% of the institutions have more than 20 different file formats, 5% even have more than 50.

Which description fits your institution's preservation polices best? [11]	n	%
We save and manage files in the format we receive them in	9	45.0
Files we receive are migrated it to a limited number of file formats	3	15.0
Other	7	35.0
Not applicable	1	5.0
20 answers		100.0

The respondents were also asked to describe their current preservation policy. The results are displayed in the above table. So far, half of the institutions keep and manage the files in the format in which they are received. A relatively small number (15%) migrates incoming files to another file format. An considerable number checked 'Other': the explanations they provided show that some of these institutions apply both strategies: the format of incoming files is migrated, and the file is also saved in its original format.

[10] The study focused on digital preservation issues specific to born-digital heritage materials. More information can also be found in the reports that were written within the framework of the National Digital Preservation Survey.

[11] This report uses the term digital preservation (in Dutch:*digitale duurzaamheid*), but in the survey the term preservation (Dutch: *preservatie*) was used (with an explanation).

Technology watch and actions related to permanent access to digital material
The interviews revealed that the future accessibility of born-digital material is an important issue and that the large number of different file formats could create a bottleneck. A preservation policy needs to be developed for each file format using a so-called technology watch: when could a certain file format become unreadable, which new file formats are being developed for which types of born-digital heritage material. Watching technology and acting accordingly requires a constant effort from the heritage institutions. These are mainly new tasks compared with traditional heritage management, where conservation plays a less important role and a technology watch is pretty much superfluous.

Different demands for permanent access to born-digital works of art and born-digital documentation and/or other materials
Based on the information gathered during the interviews, a distinction can be made between born-digital documentation and art objects. Born-digital works of art that are stored for the long term may not in any way lose quality. Moreover, future viewers must be able to experience the work of art as the artist intended. This puts demands on the technical equipment used to view a work of art. For example, processor speed can affect the way time effects are experienced. For future installations, for instance, this means that the work of art has to be (re-)built on other equipment and that the artist's intentions as to the different technical aspects used during the creation of the work need to be documented.
People interviewed at libraries also stated that the 'look and feel' of a publication is important and that a loss of quality is not acceptable.
A number of interviewees were more pragmatic when it came to documentation about art: some loss of quality caused by digital preservation actions would be regrettable but acceptable. The same seems to apply to a lot of archives, where the authenticity (and the meaning) of the object are not allowed to be affected.

Bottlenecks created by obsolete software and obsolete playback equipment
Examples of special datasets in 'proprietary' file formats and/or on obsolete digital media were mentioned several times in the interviews. The NAi estimates that they will have to buy a number of software licenses so they can continue reading CAD files in proprietary formats. The Rotterdam Municipal Archives put a lot of effort into recovering a three-dimensional scan of an Erasmus statue that was archived on a series of CDs in a proprietary file format. The Netherlands Institute for Art History had tapes that contained a database they couldn't access because they didn't have the appropriate playback equipment.

3.8 Accessibility

Accessibility status

How is the born-digital material in your collection made accessible?	n	answers %
In part online via the internet, in part only onsite (within the institution)	13	61.9
Only onsite within the institution	3	14.3
We haven't looked into the accessibility of born-digital material	3	14.3
Entirely via internet	1	4.8
Not applicable	1	4.8
21 answers		100.0

The above table shows how a collection's born-digital material is made accessible. The majority of the surveyed heritage institutions stated that some of the material is accessible online, via internet. At a number of institutions, born-digital material is only accessible within the walls of the institution, a number of other heritage institutions haven't looked into accessibility yet.

Issues concerning accessibility via internet

Born-digital heritage material is perfectly suited for access via the internet. Many of the interviewed heritage institutions are aware of this and are developing web-based user environments with names such as 'virtual reading room'. The interviews revealed the following issues with respect to born-digital material:

- *Presentation and abundance of digital objects*: The problems associated with presenting large numbers of objects were addressed in section 3.8. The following solutions were suggested:
    - A selection of so-called milestone documents in the architects' archives that are accessible via internet.
    - Presentation in several 'presentation layers'.
- *Presentation in broader contexts*: Several interviewees mentioned interoperability as an important issue when making born-digital heritage material available to users in wider contexts. The relationship between digital collections in different institutions is very important for specific user groups. The user environment Archivista of the Netherlands Architecture Institute (NAi) is OAI compatible: harvesting is used to supplement the institution's own metadata with metadata from other institutions, and the metadata of the NAi can be harvested by others in return.
- Restrictions on internet accessibility due to copyrights or the makers' request to keep the documents confidential (see the next section).

Copyright

The National Library of The Netherlands (KB) also archives e-journals on behalf of publishers. The KB can't make these journals accessible via internet because this right

is reserved to the publisher as copyright holder. The Archiving Agreement with the publishers states that this content can be made available within the walls of the KB and that the material can be included in interlibrary loan traffic within the Netherlands. The KB can make the information accessible via internet in the event of a so-called trigger event, for example, when the publisher's own systems fail.
Other institutions are also frequently bound by agreements with copyright holders. The Netherlands Photo Museum, for example, makes agreements with photographers about making the ingested photos available via internet. These agreements are put in place because the museum believes photos should be accessible from outside its walls. This belief is shared by other heritage institutions. This also requires a separate rights administration (which objects can and cannot be displayed on the web), which in many cases can be stored in the metadata on digital rights management.
Copyrights create problems of their own when archiving and making websites (and other composite digital objects) accessible (at a later stage): the different parts of digital objects often have different copyright holders. Some website owners do not allow the National Library of The Netherlands to make their website accessible (from the web archive) because of the large amount of material their website contains that is owned by other copyright holders.

Confidential material

In a number of cases, access to born-digital materials is restricted to meet the confidentiality wishes of the materials' creators.
A good example is archaeological excavations. The law requires project developers to explore the possible archaeological value of a site. Access to such data can be restricted to the registered archaeologists because demonstrators could use them to stop the project. This confidentiality also applies to research data: the researcher wants to publish his/her findings or specifically grant permission to access his/her data.

3.9 Costs of born-digital heritage material

The new work processes that born-digital heritage material has introduced bring their own cost structure with them. These costs can rocket: the person we spoke to at the National Library of The Netherlands estimates the total annual cost of the digital preservation repository (*E-depot*) somewhere between 3 and 4 million Euros. The Nationaal Archief estimates that the cost of developing a digital preservation repository will amount to some 2.5 million Euros.
The discussions highlighted the difference in cost between traditional and born-digital heritage material. The interviewees mentioned the following cost items:

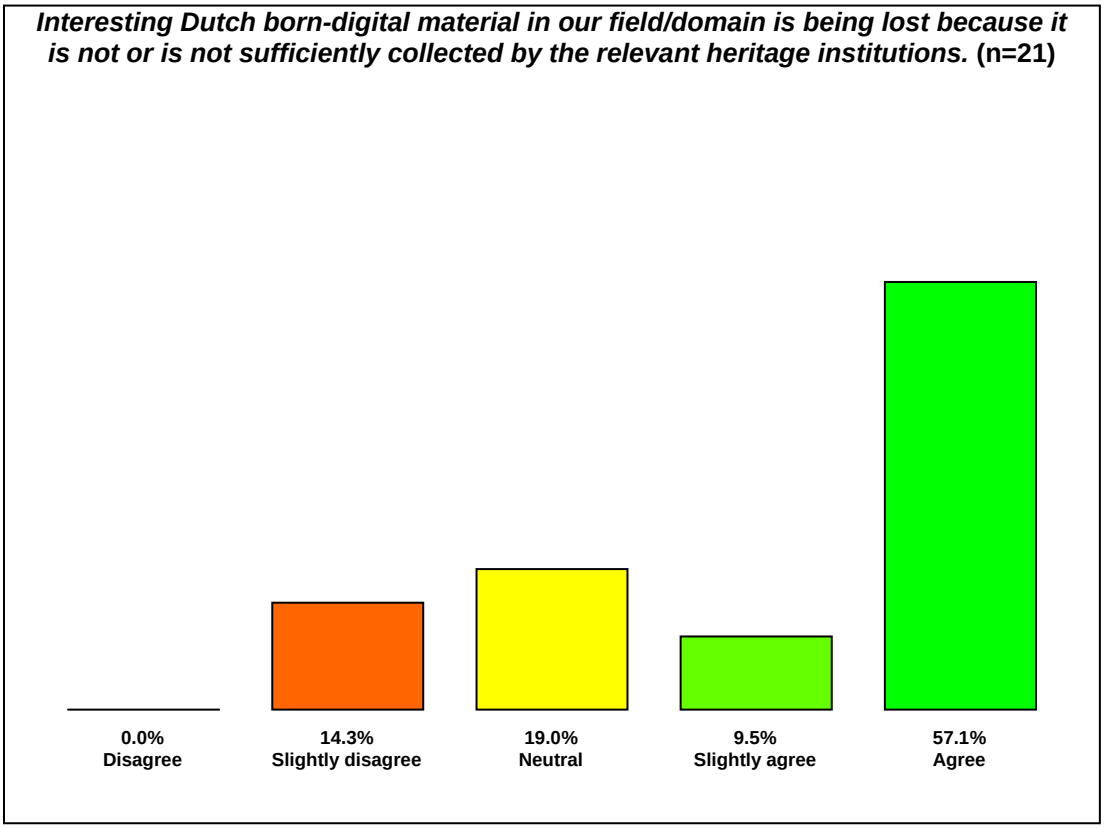
- **Development of digital preservation repository software:** The cost to develop a digital preservation repository system can be very high for a heritage institution. This is mainly due to the fact that standard (off-the-shelf) software is not available, and in many cases, according to the interviewees, “the wheel needs to be invented”.
- **Storage:** The interviewees emphasized that the costs per terabyte of storage space is decreasing, but that these savings are offset by the explosive growth in the number of terabytes needed. Peripherals will also have to be replaced regularly; the period most often mentioned was five years.
- **Technology watch:** A new activity in the area of born-digital heritage material is the previously mentioned ‘technology watch’. The technology used on the internet, in office automation and photo and video equipment is evolving rapidly. New file formats, new forms of communication and such are quick to replace each other. The heritage curator needs to monitor these developments because new file formats will enter the acquisition channel and/or existing file formats will become unreadable when software is upgraded. An interviewee noted in this context: ‘a web archivist is always lagging behind’. He indicated that it doesn't always make sense to periodically archive websites because of the use of new technologies.
- **License fees for ‘old’ software:** A few heritage institutions anticipate that they will have to pay additional license fees to ensure proprietary file formats remain readable.
- **Assigning descriptive metadata/accessibility:** A number of interviewees stated that there is a problem making born-digital heritage materials accessible for their collections. Some objects require more time and effort than traditional objects: this is particularly true for digital works of art. Born-digital heritage material can also be difficult to make accessible because of its large quantity: this may lead to problems with the personal archives of photographers (large number of digital photos) and other personal archives. More work and higher costs are feared.
- **Staffing and organization:** Working with born-digital heritage material requires different skills (more ICT skills). According to a few interviewees, this means higher salary scales and hence higher costs. Another interviewee thinks that a digital library could be cheaper in the long term because less staff is needed to run it.
-

4. Need for a cross-institutional approach to born-digital heritage material

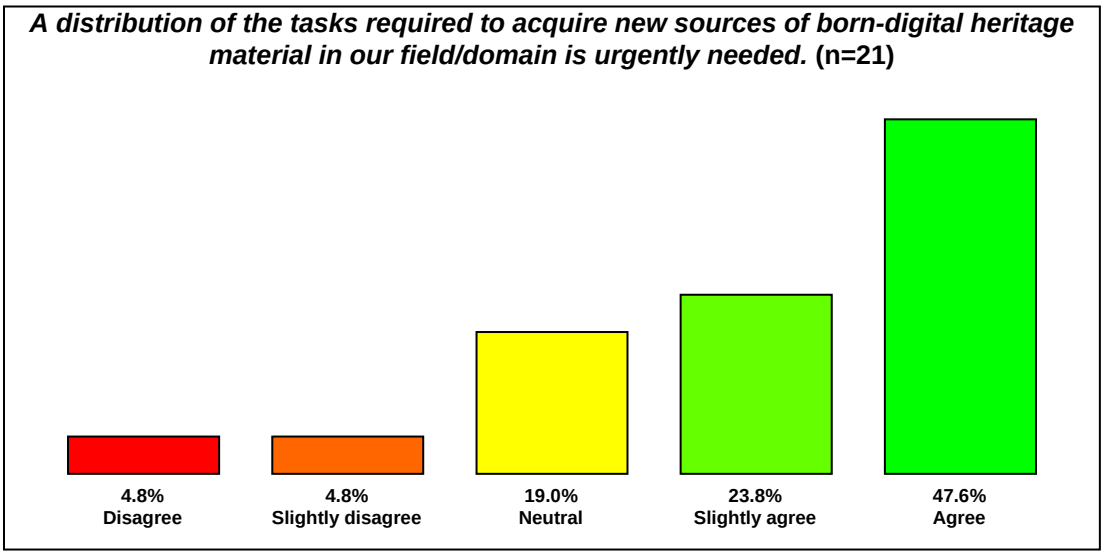
4.1 Introduction

The focus groups revealed a few issues that could benefit from a cross-organizational approach. The options that were best received by the groups were assessed in the online survey. The results are presented in this chapter.

4.2 **Collecting born-digital heritage materials**



The survey contained a number of statements to which the respondents could agree or disagree. The first statement was: Interesting Dutch born-digital material in our field/domain is being lost because it is not or is not sufficiently collected by the relevant heritage institutions.
The results are displayed in the above bar chart. A large majority (more or less) agrees with the statement, almost 20% were neutral, while only a small 15% (more or less) disagree.



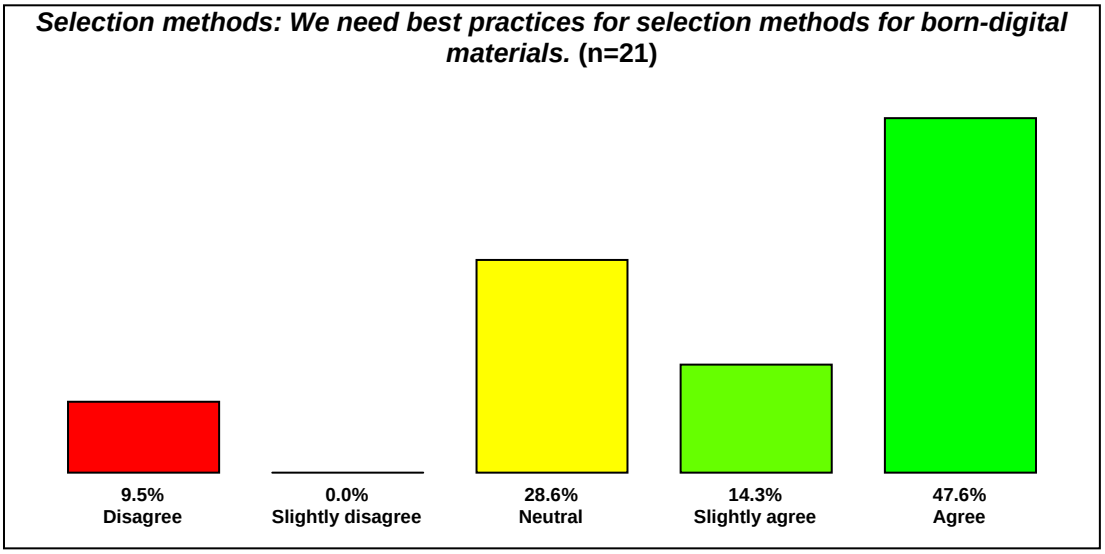
The second statement was: A distribution of the tasks required to acquire new sources of born-digital heritage material in our field/domain is urgently needed.
A large majority of the respondents also agreed (more or less) with this statement. A small 20% were neutral, while less than 10% (more or less) disagreed.

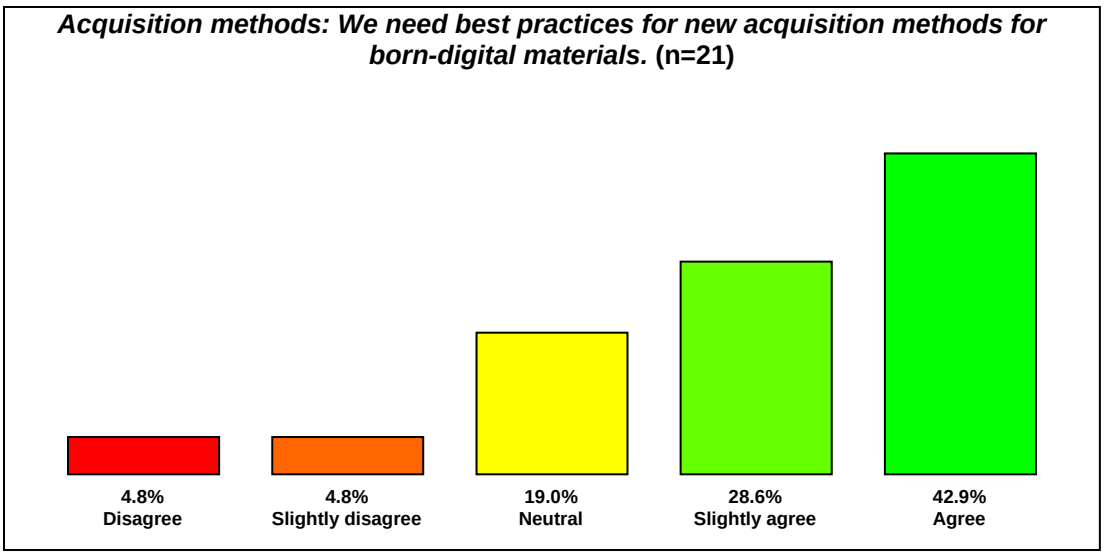
4.3 **Need for best practices**

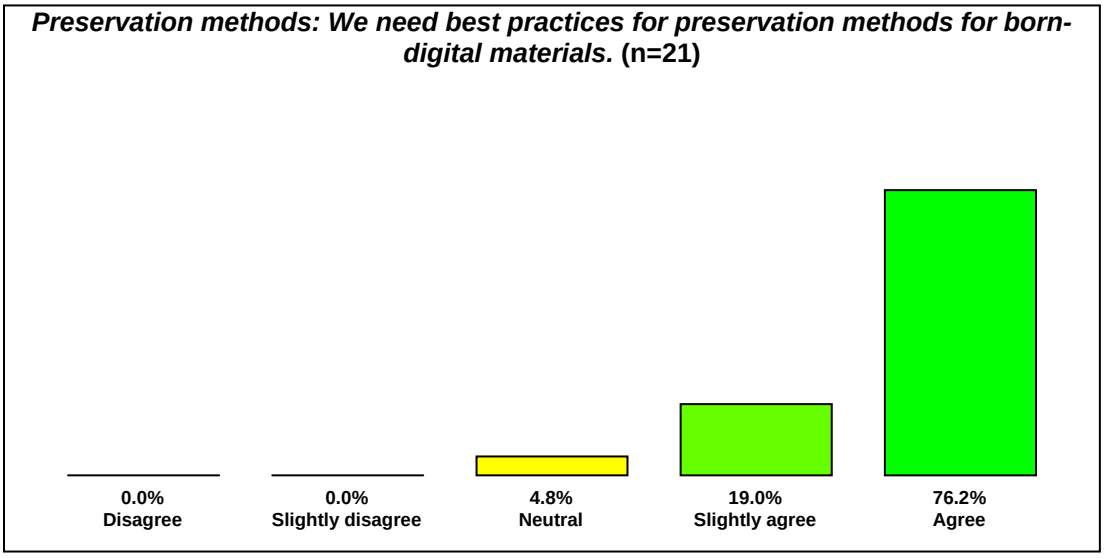
The focus groups also revealed a big need for knowledge and knowledge sharing, the most mentioned form being best practices. The respondents could indicate if they had a need for the development of best practices in the following areas:

- selection methods for born-digital materials;
- new acquisition methods for born-digital materials;
- methods to sustain the accessibility of born-digital materials.

The results are displayed below in three bar charts.







It is clear that the respondents' biggest need is for best practices for the preservation of born-digital materials: 95% of the respondents (more or less) agreed with the statement, 5% were neutral and no respondents disagreed with the statement. Moreover, the need for best practices for new methods of acquiring born-digital material is big: well over 70% of the respondents indicated they were (more or less) necessary. Finally, well over 60% of the respondents need best practices for methods of selecting born-digital heritage material.

5. Towards a quantitative measurement method for born-digital heritage collections

5.1 **Possible approaches to a quantitative study**

The possible approaches to a quantitative study of born-digital heritage collections were discussed extensively during the focus groups. The participants distinguished the following approaches:

- **External orientation:** Users, colleague-institutions, annual reports, policymakers, funders etc. The goals of a quantitative description could be: (1) show what the collection contains and how comprehensive it is, (2) harmonization and allocation of collecting tasks and (3) a national overview and a comparison with other countries.
- **Digital preservation:** Another approach to a quantitative description is to map out the problems of digital preservation. The number of files per file format must be known considering that a preservation strategy needs to be developed for each file format. In most cases, the number of files for each file format is less important: it doesn't really matter if an institution has 100 or 200 files in a specific file type. A special approach is only needed to migrate large quantities of a specific file type.
- **System management/storage:** The third approach that the participants mentioned was the management of the computer systems. This mainly relates to the number of terabytes for the entire collection: the number of terabytes is used to determine the required disk space, storage space, backup processes and such.

5.2 **Level of aggregation and granularity**

At which level of aggregation do the quantitative data of some born-digital objects have to be collected? Several examples were given, whereby the selection of the level of aggregation was less obvious:

- In its archive, V2_ distinguishes *events*. An event usually consists of several files: a video of a lecture, an article, other documentation and such. In other words: the composite object is defined from the user's perspective.
- The Library of the Technical University of Delft has various digital maps of the Netherlands in its collection. Granularity is important here. The top-50 raster map of the Netherlands is one map that consists of 300 sub-maps. Under certain conditions, a section of the digital map can be given to a user: the user is never given the whole map.
- Websites are always clearly delineated. One of the participants in the survey noted that they only archive those parts of websites they find interesting.

5.3 Relevant distinction between born-digital and digitized

The focus groups discussed whether it made sense to distinguish between born-digital and digitized heritage material. It was stated that some collections contain digitized material for which the original had been disposed of. It was also emphasized that there may be different reasons to digitize material (at a certain level of quality): because it increases direct access to the material, or because of permanent access

(online access for the long term). Direct access is the main reason for university libraries to digitize material.

All of the participants agreed that digitized and born-digital are not separate worlds. The discussion revealed that the distinction is mainly made because of preservation (born-digital material does not have an analogue counterpart, whereas digitized material usually does). It makes sense to make the following distinction:

- Traditional/analogue heritage material that can be digitized but does not have a born-digital counterpart (examples: manuscripts, paintings).
- Traditional/analogue heritage material that can be digitized and has a born-digital counterpart (examples: journals and e-journals, books and e-books). It was noted that the design and structure of the born-digital counterpart seems to resemble the traditional/digitized form at first, but continues to evolve. One of the examples mentioned was that journal articles are evolving into enhanced publications: in addition to the traditional text and figures, they also contain video, 3D visualizations, datasets and such.
- Born-digital heritage material that does not have a traditional/analogue counterpart. Examples include websites and databases.

5.4 Issues regarding the quantitative description of digital archives

The focus groups didn't quite understand how to describe digital archives quantitatively: the questionnaire of *Digitale Feiten* distinguishes archives at the macro level, meso level and micro level[12]. The confusion this caused for some of the participating archival specialists could be traced back to the following three issues:

- *Record (micro level) important for digital archives*: Some archives structure their files according to the aforementioned three levels, whereby the micro level 'record' has gained importance for digital archives because the goal is to make the digital archive accessible at this level. It was even argued that the difference between collecting museum objects and digital records is diminishing. A record is increasingly treated as a museum object.
- *Type of record as approach:* An archival organization can also describe its (digital) archive quantitatively by type of 'record': documents, video, CAD/CAM file and such.
- *Process as approach*[13]: Other institutions, such as V2_, structure their archive by (work) processes. In the case of V2_: people, organizations, events and such. A quantitative insight is not only interesting to the collection's users, it is also useful to the manager so he/she can determine the number of items that need to be made accessible.

5.5 Incomparability of different types of collections and/or born-digital objects

Finally, the focus groups emphasized that the different heritage collections cannot be compared because of the different values they represent. The quantitative data for these different collections 'cannot simply be added up'. The difference between an archive and an art collection was discussed, as well as the difference between a

[12] *Digitale Feiten* did not sufficiently explain the meso level, which is why it was not included in this questionnaire.

[13] Wikipedia defines archive as follows: *Archive is process-related information*. (Here, 'process-related' refers to the work process.)

heritage collection and an archive off and documentation about the items in a heritage collection. It is difficult to describe the boundary between these different types of collections. Nevertheless, the focus groups argued that such a distinction is important because it makes the results of a quantitative study more meaningful. One of the examples was videos: art videos have a different value than other videos. For example: the art videos at the Netherlands Media Art Institute cannot be stored in the file format used by the Netherlands Institute for Sound and Vision. The maximum level of quality (in terms of resolution) for the latter file format is not acceptable for art videos. In other words: the different values that are given to the different born-digital objects determine the approach to digital preservation.

5.6 Designing the questionnaire and the quantitative measurement instrument to be tested

Based on the above, the following conclusions were drawn and choices made for the online questionnaire:

- In line with *Digitale Feiten*, the focus was on the external approach, whereby a few questions were also asked about the problems associated with (digital) preservation and computer storage.
- The respondents were asked to characterize their collections so the incomparability of the collections could be addressed.
- An emphasis on born-digital heritage material that does not have a traditional/analogue counterpart. Detailed questions were asked about the possible types of such heritage material.
- A separate question was included for the description of archives, whereby the respondents could choose between the different types of descriptions.

5.7 **Discussion of the results of the test measurement with a view towards a quantitative measurement instrument for born-digital heritage collections**

The question asked in this study is: What is a consistent and repeatable way of quantitatively measuring born-digital heritage collections? The sub-questions were formulated as follows:

- *Terminology:* Is there a common terminology to describe born-digital heritage material?
- *Classification:* Can a common classification be created to distinguish types of born-digital heritage material, a classification that can be reused for quantitative research?
- *Quantification:* How can these born-digital heritage collections be measured?

The results for the different object types are discussed below, using the test measurement that was carried out in this study in the form of an online survey. Twenty-two institutions participated in the survey.

Archives and digital personal files (private papers)

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Archives (macro level)	Quantity	3	Thousands
Archives (micro level)	Quantity	1	Hundreds of thousands
Digital personal files */private papers*	Files	1	Not available

The above table displays the results of the test measurement for archives. Three heritage institutions specified quantities to describe their born-digital archives at the macro level, one heritage institution did the same at the micro level. Someone mentioned that terabytes are sometimes used to expresses the unit of quantification. As previously mentioned, archives can be described in different ways (see page 32): (1) by aggregation level (collection, files, records), (2) by work process and (3) by object type.
All of the eight respondents who answered this question chose the first answer category. The interviews and the focus groups revealed that the macro level and/or micro level are mainly chosen, whereas the latter seems to be gaining importance in the digital environment.
The interviews also revealed that there are considerable differences between government archives and personal archives. Structure, abundance of document versions and so forth are more of an issue in (digital) personal archives, which is why the online survey contained a separate question on digital personal files, which only one heritage institution answered. The suggestion is therefore to include archives only at the micro and macro levels in the quantitative measurement instrument.

Animations

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Animation (e.g. Flash)	Files	1	Some

The above table displays the results of the test measurement for animation. One heritage institution indicated having three animation files in the museum's collection; another institution argues that animation is archived as part of websites and is not counted separately.
Wikipedia defines animation as 'the illusion of motion created by the rapid display of a sequence of stationary images, so-called frames'. One may encounter animations on the internet, in cartoons and in animated films. Graphical formats such as GIF, MNG, SVG and Flash provide animation on computers and the web. One heritage institution argued that they do not record them separately, but include them as part of the website. Another institution, a museum, indicated having three animation files in its collection. The following conclusion can be drawn: when the animation is on the web, it can be considered as a website. If the animation is a stand-alone entity and an art object, it is considered as digital art (digital art objects/installations), if it's an animation file, it's video. This is why it was suggested that the quantitative measurement instrument should not make a separate distinction for animations.

Audio and digital audio books

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Audio files (excl. digital audio books)	Hour	5	Tens of thousands
Digital audio books	Titles	1	Thousands

The above table displays the results of the test measurements for audio files and digital audio books. One institution indicated that it stored audio files on CD ROMs. Information on (digital) audio files (with the measurement unit 'hour') was provided by five of the participating heritage institutions. Separate questions were asked for digital audio books (with the measurement unit 'title'): one heritage institution (the National Library of The Netherlands) had thousands of titles in its collection. This is a meaningful distinction, even though there is something to say for combining both types of audio files.

Websites, blogs, blog entries, tweets, widgets and wikis

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Websites	Sites	8	Thousands
Blogs	Blogs	3	Hundreds
Blog entries	Titles	1	Thousands
Tweets	Micro-blogs	1	Dozens
Widgets	Files	1	Some
Wikis	Quantity	3	Dozens

The above table displays the results of the test measurement for websites, blogs, blog entries, tweets, widgets and wikis.
According to the online survey, eight heritage institutions archive websites and store them in their collections. The other questions related to websites were often left unanswered: three institutions indicated that their collection contains blogs, whereby one institution noted that they consider them as websites. One institution indicated that it stores blog entries separately. Three heritage institutions recorded that their collection contains wikis, while one institution indicated that their collection contains widgets (elements of a graphical interface) and another that their collection contains tweets (also referred to as micro-blogs).
Based on the above, the suggestion is to categorize everything under one term:

- *(parts of) websites with the unit of measurement 'quantity'.*

Databases

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Databases (heritage metadata), created by institutions' professionals	Quantity	4	Dozens
Databases (heritage metadata), created by the public (UGC)	Quantity	1	Dozens

The above table displays the results of the test measurement for databases. The test measurement distinguished between databases (heritage metadata) created by the institutions' professionals and heritage metadata created by the public (*user-generated content*). Although the survey suggested the unit of measurement 'number of records', it turns out that the respondents entered the number of *databases* (four institutions provided information on databases created by professionals, one organization provided information on databases containing user-generated content). Because the interviews revealed a few examples in which historical data, professional data and user-generated content was mixed, the suggestion is to treat them as one group:

- *databases (containing information on heritage; possibly with user-generated content) with the unit of measurement 'number of databases'*

Photos and other born-digital images

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Photos, and other *born-digital* images	Files	10	Hundreds of thousands

The above table displays the results of the test measurement for photos and other born-digital images. Ten heritage institutions entered quantitative data in the online survey. Everything seemed clear and there were no terminology issues.

E-articles (single file and multiple file); e-journals

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
E-journals	Volumes	2	Thousands
E-articles (multiple files; enhanced)	Titles	2	Thousands
E-articles (single file)	Files	4	Millions

The above table displays the results of the test measurement for e-articles and e-journals.
The National Library of The Netherlands (KB) answered with 'millions of single file e-articles, in some cases with supplemental files'. The KB indicated that it does not have any multiple-file e-articles and rightly commented that e-journals have a higher level of aggregation than e-articles. Three other heritage institutions indicated that their collection contains single-file e-articles.
Two archives indicated that their collection contains dozens of multiple-file e-articles (also referred to as *enhanced publications*). This is, however, questionable and could be a misunderstanding.

What are enhanced publications? Woutersen-Windhouwer and Renze Brandsma (2009)[14] define an enhanced publication as a publication that includes research data, additional material for illustration or clarification, post-publication data (such as comments) with an object-based structure. In other words, it's a composite object with elements that consist of different object types. Doorenbosch et al (2009)[15] estimate that the different elements of an enhanced publication will be stored in different 'long-term preservation archives' (for example, the research data at DANS, the text at the KB etc.).

It is clear from the publications that were mentioned that enhanced publications are still evolving and that the term 'enhanced publications' can be interpreted in several ways.

In light of the above-mentioned developments, it has been suggested to discard the aggregation level 'e-journals'. It has also been suggested to discard the distinction between single-file e-articles and multiple-file e-articles. Other types of objects can also be collated objects (for example, e-books, websites etc.) and the distinction is not made there either. This is why we suggest using the following term:

- *e-articles (single file or composite object) with the unit of measurement 'quantity'*

E-books

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
E-books	Titles	3	Tens of thousands

The above table displays the test measurement for e-books. Three heritage institutions indicated that their collection contains e-books. Everything seemed clear and there were no terminology issues.

E-newspapers; online newsletters; word processing documents; e-flyers

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Word processing documents	Files	3	Thousands

[14] Enhanced publications: state of the art, Saskia Wouterse-Windhouwer and Renze Brandsma, in Enhanced Publications, Surffoundation, Amsterdam University Press, 2009
[15] Long-term Preservation of Enhanced Publications, Paul Doorenbosch, Eugene Durr, Barbara Sierman, Jens Ludwig and Birgit Schmidt, in Enhanced Publications, Surffoundation, Amsterdam University Press, 2009

Online newsletters	Issues	4	Thousands
E-flyers	Files	1	Hundreds
E-newspapers	Issues	1	Dozens

The above table displays the results of the test measurement for e-newspapers, online newsletters, word processing documents and e-flyers.
None of the surveyed heritage institutions have e-newspapers in their collections. One heritage institution indicated that they have a license for a number of digital newspapers. Some four heritage institutions do have online newsletters in their collections. However, their explanations revealed that three of the four respondents meant 'titles' as their unit of measurement. Only one heritage institution recognized e-flyers as a collection item. Three heritage institutions provided quantitative data for word processing documents: one heritage institution indicated that these are graduation projects and dissertations/theses.
Based on the above, the suggestion is to use the following terms:

- *e-newspapers with the unit of measurement 'titles';*
- *online newsletters with the unit of measurement 'title';*
- *other digital documents (such as e-flyers, reports and such) with the unit of measurement 'quantity.'*

Digital maps and ground plans

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Digital maps and ground plans	Titles	3	Thousands

The above table displays the results of the test measurement for digital maps and ground plans. Three institutions provided quantitative information. Everything seemed clear and there were no terminology issues.

Digital research files

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Digital research files (for example, archaeological excavation projects)	Quantity (files)	4	Hundreds

The above table displays the results of the test measurement for digital research files. Four heritage institutions provided quantitative information. Everything seemed clear and there were no terminology issues.

Digital art objects/installations; Internet art objects

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Internet art objects	Quantity	1	Dozens
Digital art objects/installations	Quantity	2	Hundreds

The above table displays the results of the test measurement for digital art objects/installations and internet art objects.

Two institutions provided information for digital art objects/installations[16] and one organization provided information for internet art objects. The difference between these two art forms seemed to be clear to the interviewees and did not cause any confusion. Nonetheless both art forms could be combined in a simplified survey.

Games

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Games	Games	1	Some

One heritage institution indicated in the test measurement that their collection contained some online games; another museum indicated that their collection did 'not yet' contain games. Many public libraries do have games in their collection, but probably don't consider them as heritage collections and hence will not store them for the long term.

GIS files

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
GIS files	Quantity	1	Dozens

One heritage institution indicated that they have GIS files (Geographical Information System files): This respondent provided the number of titles because the number of individual files is very big.

3D designs and reconstructions

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Digital (3D) designs of objects and buildings	Quantity	1	Some
Digital (3D) reconstructions of objects and buildings	Quantity	2	Dozens

One heritage institution indicated that they have digital (3D) designs of objects and buildings, two heritage institutions indicated that they have digital (3D) reconstructions of objects and buildings. In light of the relatively small quantities and the small difference, the suggestion is to group this into one term:

- *digital (3D) designs or reconstructions of objects and buildings with the unit of measurement 'quantity'.*

Digital video/film

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Video recordings	Hour	10	Tens of thousands

Ten heritage institutions indicated in the test measurement that their collection contains born-digital video recordings. Everything seemed clear and there were no terminology issues. There is no consensus on the unit of measurement: most of the

[16] One heritage organization indicated that they included video recordings of performances under digital art objects.

heritage institutions used hours as unit of measurement, one respondent entered the number of live stream video files, another used ‘tapes’ as unit of measurement because they're available on DVD. Finally, some respondents indicated that it was difficult to provide quantitative information. It is also worth noting that digital films and digital animated films fall under this category.

Software

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Software (customized)	Programmes	1	1

Only one museum indicated in the test measurement that their collection contains software: software for digital works of art.

Other: spreadsheets, PowerPoint presentations, email folders

Object	Unit of measurement	Number of heritage institutions that have this object in their collection	Total quantity in the collections of the surveyed institutions (completed)
Email folders (attachments incl.)	Files	2	Hundreds
Presentations (e.g. PowerPoint)	Files	1	Some
Spreadsheets	Files	1	Dozens

The above table displays the results of the test measurement for spreadsheets, PowerPoint presentations and email folders.
The results do not provide much to go on: Two institutions indicated that their collection contains spreadsheets, one of these institutions also indicated that their collection contains some PowerPoint presentations and dozens of spreadsheets. It seems that other institutions have grouped similar files under the collective term ‘archive’ and do not see them as separate collection items. This is why the suggestion is to discard this term for the quantitative measurement instrument for born-digital heritage collections.

Conclusions on the quantitative test measurement data

Considering the relatively small number of institutions that participated in the survey, the quantitative results of the test measurement provide no more than an indication of the size of the born-digital heritage collections in the Netherlands. Nonetheless, the following conclusions seem to be justified:

- The surveyed institutions add object types with a traditional and/or digitized counterpart to their heritage collections in large quantities, such as photos, videos, audio files, e-books and e-articles.
- A number of new forms of publication in the born-digital world, meaning objects without a traditional or digitized counterpart, are not collected or are only collected in dribs and drabs by the surveyed institutions. Examples are websites (thousands collected, whereas 3.6 million websites are registered with an NL domain (www.sidn.nl), games (some collected), 3D designs or reconstructions (dozens collected).

5.8 Proposal for a quantitative measurement instrument for born-digital heritage collections

Proposal for a quantitative measurement instrument for born-digital heritage collections		
Term (objects)	These terms also mean:	Suggested unit of measurement
Archives		
Archives (macro level)	Personal archives	Quantity
Records (micro level)		Quantity
Other heritage collections: museum collections, library collections, historical-architectural collections, scientific collections and archaeological collections		
Audio files (excl. digital audio books)		Hours
Databases (containing cultural heritage metadata)	Possibly with user-generated content	Quantity
Digital (3D) designs or reconstructions of objects and buildings		Quantity
Digital audio books		Titles
Digital maps and ground plans		Titles
Digital art objects/installations (incl. Internet art)		Works of art
Digital research files	Such as files from archaeological excavation projects	Files
E-articles	Single-file or composite (enhanced publications)	Quantity
E-books		Titles
E-newspapers		Titles and issues
Photos	Other born-digital images	Files
Games		Games
GIS files		Titles
Online newsletters		Titles (and issues)
Software (customized)		Programmes
Video recordings/Films	Films, animated films	Hours
Websites (and parts of websites)	Blogs, tweets, widgets, wikis	Number of sites
Other digital documents	E-flyers, reports, and such	Titles

The above is a proposal for a quantitative instrument for born-digital heritage collections based on the results of this study. The following should be noted:

- While other heritage collections can be quantitatively studied in a similar way, the suggestion is to study archives in a totally different way because born-digital archives consist of many different types of objects and files that, like for traditional archives, are not seen as separate collection items.
- The proposal has a discrepancy in the level of text aggregation. In the study, electronic articles are studied at the article level, not at the electronic journal level, despite the fact that the broader level was chosen for other forms of text: electronic newspapers and online newsletters. This discrepancy seems to be acceptable because it ties in better with the heritage institutions' current practices.
- Developments in the area of digital material are likely to impact the level of aggregation chosen for a number of objects types in the quantitative measurement instrument for born-digital heritage collections. The aggregation level ‘website’

was chosen based on the data collected during this study, but the level could change if more and more heritage institutions collect only certain parts of websites. It is also likely that the collective term ‘website’ (even an issue of an e-newspaper might be called a website) will be broken down even further by both the users and the heritage institutions.

- Digitization is making some object types more similar to each other: in this proposal, born-digital video and born-digital film are combined under one term, although the corresponding analogue media are perceived as separate objects. Such developments will probably start occurring more frequently.
- The proposed quantitative measurement instrument is a relatively simple instrument that will lower the threshold for heritage institutions participating in a subsequent survey. At the same time, the instrument is expected to provide a good overview of born-the digital heritage collections at (Dutch) heritage institutions.
- Finally, we recommend integrating the quantitative measurement instrument for born-digital heritage collections with the measurement instrument the *Digitale Feiten* project used for digitized collections. The interviews revealed that a lot of heritage institutions are familiar with the distinction digitized-born digital but do not make the distinction themselves. Many heritage institutions perceive their digital collection, which contains both digitized and born-digital items, as an entity.
-

5.9 Size and growth

Size

Twelve institutions answered the question on the size (in terabytes) of their collections. In the explanation, a number of institutions indicated that their collection contains a combination of digitized and born-digital material. The results can be summarized as follows:

- Four institutions indicated that their digital collections consist of up to two terabytes.
- Six institutions indicated that their collections consist of approximately 10 terabytes (8 to 13).
- Finally, two institutions indicated that their collections are bigger: one institution indicated 30 terabytes, and one 2,400 terabytes.

In the explanation, two institutions indicated having difficulty with the way the question - What is the size of your digital collection in terabytes? - was formulated. One institution indicated that they refer more and more to digital material outside the institution. The concern here is whether the material will be available in the long term. This institution wonders if they should harvest the material just to guarantee future access. Another institution also had difficulty with the term 'your collection' and wonders whether it includes licensed materials, such as electronic journals. In a new (subsequent) study, the 'purchase of permanent access' could be included as a criterion.

Explosive growth

The institutions were also asked how much they expect the born-digital part of their digital collection to grow in the next five years. A number of institutions expect explosive growth. The KB expects its collection to grow from its current 13 terabytes to 800 terabytes in 2012. The Rotterdam Municipal Archives also expect strong growth and thinks its archive will reach 500 terabytes in 2012. The Netherlands Institute for Sound and Vision – by far the biggest archive with 2,400 terabytes – did not express its growth in terabytes but indicated that its archive will grow by 32,500 hours of radio and television programmes and by 10,000 music CDs and DVDs[17]. Of the 14 institutions that answered the question on collection growth, 10 institutions indicated that they expected explosive growth in the next five years. Two institutions do not expect substantial growth to take place in the next five years, because (according to one explanation) the handover of born-digital government archives will be well underway and (according to another explanation) the active acquisition policy will have been designed and will have positive outcomes. Two institutions indicated that they expect substantial but not explosive growth.

[17] An indication: Sound and Vision's collection grows every year by about 8,000 hours of television programmes produced by the national public broadcasting corporations. This is about 266 terabytes.

6. Summary and conclusions

6.1 Status of born-digital heritage material

Exploratory study among selected heritage institutions

Forty Dutch heritage institutions were selected because there was evidence that they were dealing with born-digital heritage material. The exploratory study consisted of three parts: interviews, a number of focus groups and an online survey. A total of 29 heritage institutions participated in at least one of the aforementioned parts of the study. It must be stressed that this selection is not representative, but rather that these institutions are pioneers in the collection and management of born-digital material.

Development stage

More than half of the respondents already include born-digital heritage material in their collections (execution stage). Of the other institutions, some are at the implementation stage (plans to process born-digital material are currently being made) and some are at the planning or ideation stage. As far as the status for digital preservation repositories is concerned, some 30% of the institutions have an operational repository, and 20% of the institutions are currently developing one. Almost 40% are drawing up concrete plans, while the rest of the institutions don't have any. A major obstacle in the development of digital preservation repositories is the lack of standard software packages. In some cases, institutions are jointly developing such software, while other respondents see an opportunity to store their material in the repositories of other institutions.

Acquisition

Some of the surveyed heritage institutions indicated that they receive born-digital materials through the usual acquisition channels, but that they haven't done anything with it so far. Most of the institutions indicated that they do process the born-digital materials they receive through the usual acquisition channels. For many of these organizations, this meant changing the way they work. More than half of the institutions indicated that born-digital material led to the creation of new collection areas that could be relevant for their collection. Only some of these institutions are doing acquisition in these new collection areas.

How different are the methods used to acquire born-digital materials from the traditional acquisition methods? The transition period for traditional archives is considered too long for digital archives because of many of the file formats' limited shelf life. The interviews revealed that the main change in the acquisition method was the attempt to reach the archive producer/creator in an early stage of the archive creation. This is done by providing education, guidelines, helping government agencies manage their information and by providing information management services.

User-generated content is another important new way of acquiring or making born-digital heritage material accessible. Web 2.0 applications enable people from outside the institutions to add data and/or heritage material.

Managing born-digital heritage material

This study revealed the following issues with managing born-digital heritage material: Born-digital material is different from digitized material because it is unique (no physical counterpart). This is why content-related metadata have to be added afresh. This may, however, not be easy due to the large number of born-digital objects. Moreover, personal archives in particular often contain several versions of the same born-digital object (think of several versions of a document).

Problems selecting born-digital material

Because there is so much born-digital material, it is particularly important that the right material be selected at the time of acquisition. The interviewees mentioned a number of selection issues when putting the collection together (what and what not to acquire), making the collection available (one of the examples stated was the large number of digital photos a photographer now makes compared with the practice in traditional photography) and when presenting the collection: how can these large volumes of data be presented to the user in a structured way?

Digital preservation

Digital preservation – the actions required to ensure permanent access to the material – is a major issue for born digital materials. Heritage institutions are expected to see the rise of a new activity: a technology watch that flags when a certain file format runs the risk of becoming unreadable, the new file formats that are being developed and the associated activities. The extent of the problem of permanent access is largely determined by the number of file formats a heritage institution's collection contains. One of the questions about this issue was answered as follows: more than 60% of the heritage institutions indicated that their collection contains fewer than 20 different file formats, and 5% indicated that theirs contain more than 50.
The interviews touched on two other issues concerning permanent access. Some respondents saw a difference in the digital preservation requirements for the different collection types: For example, any loss in quality is unacceptable for art videos, although it can be acceptable for video documentaries. Another bottleneck was created by the obsolescence of (licenses for) software used to read proprietary file formats and/or the obsolescence of equipment to read certain digital media.

Accessibility

Born-digital heritage material is perfectly suited for access via the internet from outside the heritage institutions. This is broadly recognized: more than 65% of the institutions indicated that some of their born-digital materials will be available via internet. Another part of these materials cannot be made available via internet due to copyright and/or confidentiality issues. In terms of presenting born-digital material, the respondents mentioned that the abundance of digital objects could cause problems and that when designing the accessibility, the aim should be to present the information in broader contexts (so that the relationship between the different institutions' digital collections is clear to the user).

Costs

Time was also spent in the interviews on the cost structure for born-digital heritage collections. It was obvious that it is different from that of traditional heritage collections: (1) the development of digital preservation repository software is still very expensive because of the lack of off-the-shelf packages, (2) storage costs will

remain high because the explosive growth of collections will more than make up for the low cost of storage media, (3) the costs incurred by new 'technology watch' tasks and actions related to digital preservation are lasting and new for the heritage institutions, (4) the cost of assigning descriptive metadata may increase because of the large amount of born-digital objects and (5) some respondents see additional costs in licenses for 'old' software and potentially in the costs of new – more ICT-focused – tasks.

6.2 Need for cross-institutional approach

The interviews, and the subsequent online surveys, revealed the following issues concerning a cross-institutional approach, which many of the surveyed heritage institutions seem to have a need for:

- Interesting Dutch born-digital heritage material is being lost because it is not or is not sufficiently collected by the respective heritage institutions. A clear majority of the heritage institutions confirm this and strongly believe that acquiring born-digital heritage material from these new sources urgently requires a clear allocation of tasks.
- There is a great need for best practices for born-digital heritage material in the area of digital preservation (95% of the respondents), acquisition (more than 70%) and selection (some 60%).

6.3 Quantitative test measurement

Results of the test measurement

The quantitative test measurement used in this study provides an indication of the size of born-digital collections at Dutch heritage institutions. It is worth noting that new types of born-digital objects, such as websites, games and 3D designs, are currently not collected or are only collected in dribs and drabs.

Explosive growth expected

A large majority of the institutions indicated that they expect their born-digital heritage collections to grow substantially in the next five years. Some of the institutions do not expect the growth to occur for the next five years. The other institutions expect growth, but not explosive.

Designing the quantitative measurement instrument for born-digital heritage material

A quantitative measurement instrument was designed for born-digital heritage collections based on this study. This instrument suggests describing born-digital archives differently than heritage collections. Born-digital archives consist of a variety of different objects and file types that, in contrast to other collections, are not seen as separate collection items but as part of an archive or file.
For the other heritage collections, the suggestion is to distinguish 19 categories of object types. New acquisition methods and/or user behaviour are expected to change the level of aggregation assigned to certain objects over time.

Appendix A Overview of born-digital heritage collections at the institutions that were interviewed

Name of the institution	Born-digital collection components/objects	Comments
Idea formulation and/or planning stage		
Netherlands Literature Museum		Expect born-digital personal archives of literary persons
Netherlands Architecture Institute	Archivalia architects with special component: CAD drawings	Digital preservation repository (e-depot) implementation started; archive per architect, classification by building, selection of milestone documents per building
RKD (Netherlands Institute for Art History)	Some archives have digital files on digital media Inflowing images for image documentation currently born-digital Documentation of technical study of works of art is increasingly born-digital	In consideration: include electronic newsletters; relevant websites
Stedelijk Museum Amsterdam	Digital works of art Videos	Digital works of art stored on separate computer(disks) in safe Videos partly born-digital, partly digitized
Leiden University Libraries	Archivalia	Anticipate future inflow of (1) digital archives from the university (2) digital personal archives
University Library of Utrecht (Igitur)	E-journals	Digital publisher: does not consider this to be heritage
Noord-Hollands Archief	Archivalia	Starting test project digital personal archive
Implementation and/or execution stage		
Library of the Technical University of Delft	Digital maps	Digital maps in special application. Can be searched separately. Expansion planned in the number of maps, applications and user groups.
DANS (Digital Archiving and Networked Services)	Excavation projects; videos oral history projects; various datasets scientific research in the humanities and social sciences	Have a digital preservation repository (EASY). Are developing EASY2 User group consists mainly of researchers
Rotterdam Municipal Archives	Archivalia, websites, data files	Is setting up digital preservation repository GAR for digital archive material. Also involved in repository CONCERN to streamline how municipal organizations manage their data. Started archiving selection of Rotterdam websites. Accessibility still to be developed. Example of a data file: 3D scan of Erasmus statue
National Library of The Netherlands	E-journals; e-books; e-publications; websites	E-depot since 2002. New digital preservation repository completed in 2013. Now mainly digital journals, national and international. Will be expanded to e-books. Also harvest material from university repositories. Selection of Dutch websites (now 2,000, to be expanded to 4,000) is archived two to four times a year. Will be accessible in a few months.
Nationaal Archief (National Archives)	Archivalia	Has a digital preservation repository Is writing guidelines for the delivery of

		digital archives
Naturalis (National Museum of Natural History)	**Enhanced catalogue files** **Natural history observation data files/Web 2.0 data** **Databases with research data**	**Various sources and types of material** **Historical data entered in database; contemporary observations to be added via internet by the public** **Internet site where the public can upload photos of plants and animals for experts to evaluate** **Research data from the institution's own researchers**
Netherlands Photo Museum	**Digital photos (works of art)**	**Digital preservation repository**
Netherlands Media Art Institute	**Videos and installations**	**Collection of born-digital and digitized videos(works of art); single-channel video art is created less often, trend towards installations**
Amsterdam City Archives	**Archivalia**	**Setting up digital preservation repository. Is writing guidelines for the delivery of digital archives for municipal organization.**
V2_ (Institute for the Unstable Media)	**Streaming video, digital photos documents (articles, books, chapters of books)** **video documentaries of works of art, technical documentation of works of art**	**Digital archive available via website with categories: people, works of art, events, organizations, media, articles.** **Develop second generation digital preservation repository together with *De Balie*.**
Graphic Design Museum	**Graphic design**	**Manages Dutch Design database: graphical designers can upload their designs themselves**
IISG (International Institute of Social History)	**Archivalia, meta sources**	**Build up trusted digital repository** **Born-digital collections:** • **Born-digital archives: information is ingested from digital media in the archive** • **Meta sources (large scale data sources for historical research**
The Netherlands Institute for Sound and Vision	**Audiovisual material: radio and television broadcasts, music CDs and DVDs**	**Digital ingestion of radio and TV programmes** **PROARCHIVE: service – digital audiovisual archiving service for third parties** **Research on web-based content (such as films on YouTube)**

Appendix B Other survey results

This appendix contains the other results of the online survey, namely the few questions on how the collections are characterized.

What does your digital collection consist of? (MULTIPLE ANSWERS ARE ALLOWED]	**n**	**answers %**
Born-digital heritage material	**18**	**81.8**
Digitized heritage (not originally digital, but digitized)	**18**	**81.8**
Digital information/documentation on heritage (descriptions, close ups, digital reconstructions etc.)	**15**	**68.2**
22 answers		**100.0**

9. How would you characterize your collection?	**n**	**answers %**
Museum collection	**2**	**10.5**
Library collection	**5**	**26.3**
Scientific collection	**2**	**10.5**
Archive and documentation	**5**	**26.3**
Other	**5**	**26.3**
19 answers (3 blank)		**100.0**

If other, please explain:
New media institution with its own archive
Archive, library and museum collection
Both library and museum collection, archive and documentation, and scientific collection
The collection contains all four of the mentioned elements
We are a merged organization, each of the characteristics applies to us
This particularly refers to our special collections for which we keep receiving graphical collections

Appendix C Survey questionnaire

Exploratory study born-digital heritage materials

The aim of this study is to acquire and build up knowledge of born-digital heritage at Dutch heritage institutions for the benefit of the heritage sector and its policymakers and financiers. Your participation is highly appreciated.
This questionnaire addresses the following topics:
A. Your vision of some issues concerning born-digital heritage
B. Born-digital heritage material and your institution
C. Acquisition of born-digital materials
D. Management, preservation and accessibility of born-digital material
E. Quantitative description of your digital collection
F. Final questions

Can be completed in phases
You can do part of the survey and go back to it later by clicking on the link. The system will let you close the survey once all of the questions have been answered.

Definition of born-digital heritage and collection

There are two types of born-digital heritage material:
- heritage material that is not available in any other form than digital, such as digital archives, digital art or photos that were taken with a digital camera.
- digital information about heritage objects, for example, descriptions, close-ups or digital reconstructions of the heritage object.
The questionnaire also talks about digitized heritage, which is defined as heritage material whose original form is not digital but that has been reproduced in a digital format

Collection: we are asking you for information about the composition of all of your institution's born-digital (sub) collections. If you provide information for only one sub-collection, please make a note in the comments field.

A. Your vision of some issues with born-digital heritage

1. The following issues on born-digital material emerged from a series of interviews and focus groups with your colleagues. What do you think of the following statements?
Indicate your opinion by clicking on the scale (see figure for example).

-2 Disagree
-1: Slightly disagree
0: Neutral/Don't know/Not applicable
1: Slightly agree
2: Agree

Interesting Dutch born-digital material in our field/domain is being lost because it is not or is not sufficiently collected by the relevant heritage institutions.	**← Disagree** -2 -1 0 1 2 **Agree →**
A distribution of the tasks required to acquire new sources of born-digital heritage material in our field/domain is urgently needed.	**← Disagree** -2 -1 0 1 2 **Agree →**
Selection methods: We need best practices for selection methods for born-digital materials.	**← Disagree** -2 -1 0 1 2 **Agree →**

Acquisition methods: We need best practices for new acquisition methods for born-digital materials.	← **Disagree** -2 -1 0 1 2 **Agree** →
Preservation methods: We need best practices for preservation methods for born-digital materials (preservation = permanent access).	← **Disagree** -2 -1 0 1 2 **Agree** →
Please explain:	

B. Born-digital heritage material and your institution

2. Which description best fits the stage your institution is in when it comes to including born-digital heritage material in your collections?	❍ **Execution: We collect born-digital material and add it to the collection.** ❍ **Implementation: Plans to process born-digital material are currently being executed/implemented.** ❍ **Planning: We're in the planning stage.** ❍ **Idea formulation: We're in the idea formulation stage.** ❍ **Other (please explain)**
Please explain the stage your institution is in:	

C. Acquisition of born-digital materials

3. Which of the following statements relate to your institution? [MULTIPLE ANSWERS ARE ALLOWED]	
	❑ ***We receive born-digital materials through our usual acquisition channels. Nothing is done with these materials at the moment.***
	❑ ***We receive born-digital materials, which we process, through our usual acquisition channels.***
	❑ ***Born-digital material has dramatically changed the way our usual acquisition channels work.***
	❑ ***New sources/collection areas for born-digital material have been created for our field/domain that could be relevant for our collection.***
	❑ ***New sources/collection areas for born-digital material, in which we are doing acquisition for our institution, have been created for our field/domain.***
Your explanation of the acquisition of born-digital heritage material:	

D. Management, preservation and accessibility of born-digital material

4. How many different file formats does your digital collection contain?	❍ **Less than 10** ❍ **10 to 20** ❍ **20 to 50** ❍ **More than 50** ❍ **Not applicable**
5. Which description fits your institution's preservation policy best?	❍ **We save and manage files in the format we receive them in.** ❍ **Files we receive are migrated to a limited number of file formats** ❍ **Other** ❍ **Not applicable**
6. How is the born-digital material in your collection made accessible?	❍ **Entirely online, via internet** ❍ **In part online via internet, in part only onsite (within the institution)** ❍ **Only onsite within the institution** ❍ **We haven't looked into the accessibility of born-digital material** ❍ **Not applicable**
7. A digital preservation repository (e-depot) is a cohesive set of procedures and computer systems aimed at storing digital material for the long term. Does your institution have an e-depot?	❍ **Yes, operational** ❍ **Yes, in development** ❍ **No, but there are concrete development plans** ❍ **No, and there are no concrete plans**
Please explain the management, preservation and accessibility of born-digital heritage at your institution:	

E. Quantitative description of your digital collection

We would like you to provide us with a quantitative description of your digital collection insofar as possible. The preliminary study revealed that most of the institutions do not distinguish between born-digital and digitized materials, which is why we combined these here.

8. What does your digital collection consist of? (MULTIPLE ANSWERS ARE ALLOWED]

	❑ ***Born-digital heritage materials***
	❑ ***Digitized heritage (not originally digital, but digitized)***
	❑ ***Digital information/documentation of heritage objects (descriptions, close-ups, digital reconstructions etc.)***

Please explain:

9. How would you characterize your collection?

- ❍ **Museum collection [GO TO QUESTION 11]**
- ❍ **Library collection [GO TO QUESTION 11]**
- ❍ **Scientific collection [GO TO QUESTION 11]**
- ❍ **Archive and documentation [GO TO QUESTION 10]**
- ❍ **Other**

If other, please explain:

ONLY FOR ARCHIVAL COLLECTIONS:
10. Discussions with representatives from the archives revealed that archival institutions can quantitatively describe their digital archives in several ways: by level (i.e. collections, files, records), by the work processes documented in the archiv (i.e. people, organizations, events) and/or by object type (i.e. videos, documents, such as in question 11). How is your archive quantitatively described?

- ❍ **By level (collection, files, records)**
- ❍ **By work process**
- ❍ **By object type**

10A: If your institution describes archives at the macro and/or micro level, please add the description below:

Comments:

	Size	Unit of measurement	Your comments
Macro level collections)			
Micro level (records)			

10B. If you structure your archive according to the work process reflected in the archive, please add your quantitative description below:

Quantitative description by work process:

10C: If you quantitatively describe your archive by object type, please answer question 11.

11. Quantitative description of your collection by type of digital object:
Below is a list of digital objects followed by the unit of measurement in square brackets. There are three columns: the size, your (alternative) term in case you use a different one from the one presented here, and a field for comments (for example, if you use a different unit of measurement). You can skip objects your collection does not contain.

Example: Audio files - Size: 950; Your term if it's different: music; Your comments: the specified quantity is the number of individual files; we do not measure in hours.

	Size	Your term if it's different	Comments
Animation (e.g. Flash) [number of files]			
Audio files (excl. digital audio books) [number of hours]			
Blogs [number of blogs]			
Blog entries [number of titles]			
Databases (heritage metadata), created by institutions' professionals [number of databases]			
Databases (heritage metadata), created by the public [number of databases]			
Digital (3D) designs of objects and buildings [number of files]			
Digital (3D) reconstructions of objects and buildings [number of files]			
Digital audio books [number of titles]			
Digital research files (e.g. archaeological excavation projects) [number of files]			
Digital personal files/private papers [number of files]			
Digital maps and ground plans [number of titles]			
Digital art objects/installation [number of works]			
E-articles (single file) [number of files]			
E-articles (multiple files; enhanced publications) [number of titles]			
E-books [number of titles]			
E-flyers [number of files]			
E-journals [number of titles]			
E-newspapers [number of titles]			
Email folders (attachments incl.) [number of folders]			
Photos and other born-digital images [number of files]			
Games [number of games]			
GIS files [number of files]			
Internet art objects [number of works]			
Online newsletters [number of issues]			
Presentations (e.g. PowerPoint) [number of files]			
Software (customized) [number of programmes]			
Spreadsheets [number of files]			
Word processing documents [number of files]			
Tweets and such [number of micro-blogs]			
Video recordings [number of hours]			
Websites [number of sites]			
Widgets [number of files]			
Wikis [number of wikis]			

Please add comments or additional information (if you are missing digital objects):	
If you cannot provide a quantitative description, please explain why:	
12. What is the size of your digital collection in terabytes?	
13. How much do you expect the born-digital part of your collection to grow in the next five years? How do you express this growth (number of objects, terabytes?)? What determines the expected growth?	
14. Are there any other relevant issues related to born-digital heritage material that were not addressed in this questionnaire?	

F. Final questions

15. Which description matches your institution best?	❍ **Archaeological institute** ❍ **Archive** ❍ **Architectural institute** ❍ **Museum** ❍ **Library** ❍ **University library** ❍ **Other**
If other, please explain:	
16. Name of the institution:	
17. Contact's name and email address:	