Earning trust

A case of digital preservation at two private archives

Robert Gillesse
Digital Archivist
International Institute of Social History
robert.gillesse@iisg.nl

Marcel Ras
Project Manager
Het Nieuwe Instituut
m.ras@nieuweinstituut.nl
marcel.ras@netwerkdigitaalerfgoed.nl

Abstract

In this paper a comparison is made between two private Dutch archives (Het Nieuwe Instituut and International Institute of Social History) and how they collect and preserve born digital collections. In dealing with archival donors, preservation policies and IT implementations they have as much in common as they differ from each other. For instance, gaining the trust of the archival donors in the archival policies of the institute is a top priority for both organizations. Nonetheless, how precisely that trust is shaped is entirely dependent on the type of archives being collected, the context in which the institutes have to operate and the collection and preservation policies of either institute. This paper gives insight in these differences and similarities with the aim to find the areas where the organizations might truly support each other.

Themes of the conference addressed

Taking stock of lessons learned within and across domains. How can we successfully preserve private and confidential material for the long term?

Keywords

Digital Preservation, collaboration, preservation approaches, private archives, trustworthiness

Introduction

Rotterdam's Het Nieuwe Instituut (HNI), which preserves architectural and design collections, and Amsterdam's International Institute of Social History (IISH), focusing on preserving collections about labour history and of socio-economical movements, are two private archives which – looking at digital preservation – have as much in common as they have differences. When looking at these (dis)similarities, there's something essential to be learned: digital preservation cannot be put in a uniform straitjacket and must always be viewed in the context of institutional practices and policies. Notwithstanding there is equally much to learn from these contrasting preservation techniques as from consistencies in practices.

The International Institute of Social History

The IISH is a research institute that examines how work and labour relations develop globally over time. To conduct this historical research, and support other researchers, the IISH collects archives and data from all over the world. The institute was erected in 1935 to save collections from the totalitarian regimes of that era and continues to rescues collections in critical cases. In this respect the IISH has a so called 'safe haven' function. The IISH collections comprise archive, library and audio-visual material with a thematic emphasis on social and emancipatory movements . The Institute holds over 4,000 archives, more than 1,000,000 printed volumes, and an equivalent number of audio-visual items including posters and an extensive photo collection. Examples of better known collections are that of Karl Marx/Friedrich

Engels, Greenpeace, Amnesty International and the CNT-FAI. In the last years the IISH has increasingly collected born digital collections and has also started digitizing a sizeable percentage of the paper collection. Next to the great opportunities these collections offer, they nonetheless pose a new challenge with regards to collecting, preservation and access [1].

Het Nieuwe Instituut

Het Nieuwe Instituut (HNI) is the Dutch agency for architecture, design and digital culture and holds one of the largest architectural collections in the world. It is the State Archive for Architecture and Urban Planning, which includes over 600 archives and collections of Dutch architects, urban planners, professional associations and educational institutions. It comprises a total of some 4,000,000 documents recording 150 years of developments in architecture [2]. Apart from the enormous analog collections, the born-digital collection currently consists of approximately 60,000 files and will continue to grow together with the changing nature of the architectural profession. in 2009, the first digital archive was acquired, the archives of Carel Weeber, containing material from the 60's and 70's. In 2015, the internationally renowned architects MVRDV donated their archives to HNI. This consists of 5TB of data from about 400 projects, designed between 1991 and 2008, as well as backup tapes and a frightening number of old storage media. The archive is of great importance because of MVRDV's pivotal role in Dutch architecture during the last 25 years as it tracks many topical architectural developments from 1990 until today.

Differences, similarities and collaboration

In the paragraphs below the differences, similarities and collaboration between the IISH and HNI are described on the basis of three themes:
1. The acquisition of digital collections
2. A relationship of trust with the archival donor
3. Implementation issues

Within the description of these themes we aim to show that similarities and differences are both a healthy basis for collaboration.

The acquisition of digital collections

Parallels

The fact that HNI and the IISH are both private archives has two main direct consequences for the acquisition of digital collections:

- Both the IISH and HNI are not in the position to impose strict requirements towards archival donors on how archives are transferred to the institute. This is an important difference from governmental archives which have some leeway in demanding the use of - for instance - certain file formats and metadata schemes.
- The archival donors' trust in archival practices is of a very high priority. Donors are under no obligation whatsoever to transfer archives to the institutes, and - especially with collections of a sensitive nature and/or commercial value - archival donors must be completely satisfied that their collections are in safe hands.

Some of the more indirect consequences of the private archive status for the acquisition of born digital are:

- Both institutes receive born digital collections on all kinds of information carriers (hard drives, USB sticks, floppies) or by different digital means (file transfer services, cloud services, ftp). This means a (very) flexible infrastructure is needed for the transfer of archives. This is reflected in

the elaborate pre-ingest processes of both institutes, which are complicated and time consuming.

- Born digital archives can be totally unstructured, with hard to understand file and folder names. In comparison some archives are highly ordered, with logical file folder structures and meaningful filenames. The organisations need, therefore, to be pragmatic when dealing with 'messy' archives. It is not possible for archivists to spend hours on interpreting and deciphering these collections. The expectation is that machine learning and image, sound and entity recognition techniques will help unlock these archives. When dealing with more structured archives the challenge lies in how the institute will be able to represent an original structure in the online presentation. A close relationship with the donors is highly recommended, and in most cases absolutely necessary, to be able to give meaningful access to the archives and understand their complexities.
- Some archives are delivered with descriptive metadata, but in most cases there are no descriptive metadata at all. The organisations therefore need a pragmatic policy for creating archival descriptions for born digital collections. 'Traditional' inventories certainly still have their value but are seen as complementary to machine learning and full text searchability of documents.
- As there is no list of preferred file formats the IISH and HNI both receive a broad variety of file formats and software which are sometimes common but can also be obsolete or obscure. A sophisticated file format policy and normalization strategy is therefore a necessity. There is also a need for an overview of the developments in platforms, software and design process over the years.
- As (born) digital acquisition is on its way to becoming dominant, the IISH and HNI continue to collect analog archives. To make it worse, many digital archives are in fact hybrid archives, and also contain analog materials. There is a need for a sophisticated collection policy in which institutional choices are being made.

Differences
The main difference - when looking at acquisition - lies in the nature of the collections both organisation are acquiring. It can be safely said that the architectural archives of the HNI are in many cases more challenging to ingest, preserve and offer access to, than the collections of the IISH. Architectural objects are in many ways dependent on specific software and, especially when identifying and preserving older objects, can border on digital archeology. Not that the IISH has no problematic objects, but these are individual incidents. Archival policies and practices will therefore differ considerably. For example, for HNI emulation will be a more obvious preservation strategy than for the IISH.

Collaboration
The IISH and HNI are working together in critically reviewing their respective digital preservation policies for the archiving of born digital collections. Especially where collections overlap e.g. with email collections, the organisations are working together to create archival policies. Part of this is done under the umbrella of the Dutch Digital Heritage Network [3].

A Relationship of Trust with the archival donors
Parallels
Gaining the trust of the archival donor with regards to archival practices is paramount for both institutes. Being a private archive the archival donor is under no obligation whatsoever to transfer their (digital) collection to both institutes. So only trust and good relations can persuade archival donors to do this. Besides a good relationship with the archival donor, both the IISH and HNI put a great deal of effort in truly understanding the archives of their donors. We organise conversations with the donors to get a grip on the structure, context and content of their archives. This in-depth effort is reflected in the recent

report of a HNI survey carried out between September 2017 and March 2018 with six architectural firms [4]. The importance attached to certain certification is closely related to this issue of trust. Being a Trusted Digital Repository (TDR) is, for both organizations, seen as particularly important because it reassures the archival donors and funders of our trustworthiness.

Differences
For some donors the IISH really is a safe haven from political and/or societal pressures. Therefore trust for these organizations is mostly described in terms of information security. These donors really want to be sure the archives they hand over will not fall in the wrong hands.

This is different for HNI. Their archival donors also have to trust the institute to protect the -sometimes- confidential nature of the collections, but the priority is different here. HNI is not a safe haven for politically sensitive materials, rather materials that are sensitive in terms of business. The core business of architectural firms are at stake as the archive is packed with innovative designs and approaches, which are central to trade at these firms. The donors want to be reassured that some parts their archives remain closed to the public, and to their competitors.

Collaboration
Both organisations are working together to gain a TDR status in the form of a Core Trust Seal. This is part of a collaborative approach towards Certification developed under the umbrella of the Dutch Digital Heritage Network. This well-balanced method for certification is based on (1) a five step approach; (2) a supporting tool for maturity level checking; (3) translations of the CTS and Nestor standard into Dutch; (4) lessons learnt based on a survey amongst DSA certificate holders; (5) training materials in Dutch [5]. Both the IISH and HNI will follow this approach, learn from each other and provide mutual support where possible.

Implementation issues
Parallels
Both organisations want to use a digital preservation system for the ingest, preservation and management of their digital assets (consisting of born digital and digitized materials). After a thorough comparison between different systems and solutions the IISH decided at the beginning of 2017 to use Archivematica for this purpose Choosing Archivematica. The IISH has been busy implementing Archivematica since early 2017. HNI also made the choice for Archivematica and started implementing it in the beginning of 2018.

Both organisations have begun with the employment of archiving workflows for the most challenging material: the born digital archives. They could have started with the collection of digitized material - of which both organizations have large quantities, and which are much easier to ingest - but the preservation issues of these collections are quite modest compared to the born digital collections. The preservation of born digital collection simply has a higher priority.

Differences
The IISH Archivematica project has been, and in many ways, continues to be challenging. This is mostly caused by the fact that the project proved to be far more than just the application of Archivematica to IISH processes. While implementing Archivematica it soon became clear that it was necessary to reconsider all existing archiving workflows and preservation policies. Looking for instance at the born digital material workflow a new file format policy had to be written and a new way to deal with persistent identifiers was conceived. Pre-ingest and to some extent pre-Archivematica steps (although Archivematica offers some pre-ingest functionality) were also critically reviewed - e.g. the procedure of offloading collections and how the IISH will approach appraisal and selection of born digital collections.

Related projects, such as, a new 'preservation storage' system and developing a new online interface for the digital collections (using IIIF) also had a serious impact on the implementation project.

While the IISH did not start from scratch when implementing Archivematica, HNI did. It could not fall back on earlier steps and had no serious experience with conscious acquisition, ingest and management of born-digital archives. Based on the outcomes of a preservation capability maturity check carried out in 2015 [6], HNI started its first steps in 2017 towards the implementation of a Digital Archive capable to hold the archives from Dutch architecture firms. This has proved to be a challenging path for HNI. Starting with basically nothing, new workflows and infrastructure needed to be developed. Adding to that, HNI also is in the middle of shaping a new collection policy in order to address 'the digital' in the field of architecture [7].

In the context of continuous technological developments, the nature of the architectural practice is changing and so must its archives. HNI research must try to grasp the future form of architectural collections and the risks for long term access to them. The changing nature of architectural practice has far-reaching consequences for the implementation of a digital archive. As things are not yet clear and the objects to be archived are extremely complex, HNI's approach is to set up a digital preservation laboratory. The heart of the laboratory consists of Archivematica and Bit-curator, both installed on three platforms: a Windows, MacOs and Linux environment. Around this core, specific tools and hardware are installed with which different archival materials can be handled. The focus - at the time of writing - is on the pre-ingest and ingest workflows as these are the most challenging components.

Collaboration
As both the IISH and HNI are taking steps towards a professional environment for preserving digital archives, learning from each other is the main objective. Sharing ideas and, where possible, tools is a secondary objective. The IISH already has a great deal of experience in building workflows in Archivematica, HNI can learn from this. Vice versa the IISH can learn from the complexities HNI is facing and the way this is dealt with. Pre-ingest workflows created by HNI will also be shared with IISH.

Collaboration is also evident in the approaches to specific types of material. Currently the IISH and HNI are both working on acquisition and preservation policies for email, social media and websites. While doing this they actively share ideas and approaches. Again, this is linked to the Dutch Digital Heritage Network and its collaborative approach towards appraisal and selection and collection development. Having cooperative tools and approaches for acquisition and archiving collections will encourage and increase efficiency.

Collaboration is also apparent on a broader scale, for example when the IISH hosted the first Dutch Archivematica camp in April 2018. This three day meeting provided a space for anyone interested in or currently using Archivematica to come together, learn about the platform from other users, and share their experiences [8].

Conclusion
Preserving complex born digital objects is a challenging task. It becomes even more challenging when institutes have to work without legal mandates and there is no obligation for donors to follow guidelines and standards. For the IISH and HNI it means that we have to find ways to create incentives for donors to cooperate with us as record keepers. Trustworthiness is by far the most important issue for both the IISH and HNI as this is the starting point for collaboration. It is clear we must build stronger relationships with the donors.

Coalitions between institutions are also becoming even more important to create a stronger common agenda for sustainable access. Together with the Dutch Digital Heritage Network, the IISH founded an Archivematica user group in The Netherlands with HNI as one of its members [9]. Within this group knowledge and experiences can be shared, as well as making a strong case for future developments in Archivematica implementations.

Like many institutes, the IISH and HNI are in the middle of re-shaping their policies and simultaneously processing born digital collections. This requires a pragmatic approach in which small practical steps, based on experiments, can be taken. HNI and the IISH continue to collaborate closely in this project. Not only in using the same preservation system, but especially by sharing knowledge and experience, in having an eye for what we have in common, but also where we differ, so we can truly learn much from each other.

References

[1] For more information on the IISH collection see https://socialhistory.org/en/collections. Accessed 15 April 2018.

[2] For more information refer to the section on collection of the HNI website, available online at https://collectie.hetnieuweinstituut.nl. Accessed 2 June 2018.

[3] The Dutch Digital Heritage Network is a partnership in the Netherlands that focuses on developing a system of national facilities and services for improving the visibility, usability, and sustainability of digital heritage. The network is open to all institutions and organizations in the digital heritage field. http://www.netwerkdigitaalerfgoed.nl/en/. Accessed 2 June 2018.

[4] Ania Molenda, *Between creators and keepers*. HNI internal report on archiving architectural Heritage. May 7, 2018. A public version will be published in August 2018, to be downloaden from the HNI website: https://collectie.hetnieuweinstituut.nl/en/preservation

[5] See for the Dutch Certification roadmap, webwijzer certificering: https://wegwijzercertificering.nl/nl. Accessed on 2 June 2018. This website is in Dutch. A comprehensive description of the Dutch certification approach is described in: Barbara Sierman, Kees Waterman, *How the Dutch prepared for Certification*. Paper presented at iPRES 2017.https://ipres2017.jp/wp-content/uploads/3Barbara-Sierman.pdf. Accessed 2 June 2018.

[6] Michel Emde, Michiel Stornebrink, Henk Wiersema, *Born digital preservation bij Het Nieuwe Instituut*. TNO Report 2015 R10806, June 2015. This is a HNI internal report written in Dutch. Currently not published, but available on demand.

[7] HNI acquisitions policy: https://collectie.hetnieuweinstituut.nl/en/preservation/acquisitions-policy. Accessed 2 June 2018.

[8] A report can be found on the website of the IISH: https://socialhistory.org/nl/digital-preservation-blog/2018/04/account-archivematica-camp-2018-iish-amsterdam. Accessed 5 June 2018.

[9] https://wiki.archivematica.org/Community/Regional_User_Groups#NETHERLANDS_and_FLANDERS. Accessed 2 June 2018.