Getting Persistent Identifiers Implemented By 'Cutting In The Middle-Man'

R. van Veenendaal
National Archives of the Netherlands
The Hague
The Netherlands
remco.van.veenendaal@nationaalarchief.nl

M. Ras
National Coalition for Digital Preservation
The Hague
The Netherlands
marcel.ras@ncdd.nl

M.C. Dangerfield
Rotterdam City Archives
Rotterdam
The Netherlands
mc.dangerfield@rotterdam.nl

ABSTRACT

The Persistent Identifier (PID) project of the Dutch Digital Heritage Network (DHN) promotes sustainable access in the cultural heritage domain by stimulating the use of PIDs as references to (digital) cultural heritage objects. PIDs are long-lasting references to objects and offer a solution for link rot, which often results in 404 or Not Found error messages.

Cultural heritage organisations, regardless of size, are often hesitant to implement PIDs in their systems. They lack knowledge of PIDs, are unaware of the capabilities and benefits of PIDs, and fear possible complex and costly implementation processes as well as the maintenance costs for sustained services.

To address these issues, the DHN PID project focussed on:

- Raising awareness of (the importance of) PIDs in cultural heritage organisations

- Increasing knowledge regarding the use of PIDs within cultural heritage

- Supporting the technical implementation of PIDs in cultural heritage collection management systems

At the time of writing, four vendors have implemented PIDs in the collection management systems (CMS) they provide[1], and a dozen cultural heritage institutions are taking part in this pilot implementation phase. More importantly however, is that through these CMS vendors, PIDs have become more available and affordable as a sustainable long term access solution for hundreds of institutional digital collections and their users. To ensure that the information, training and education the project provided about PIDs can be disseminated and used as widely as possible, the project created (best practice) documentation in Dutch and English, along with a unique PID Guide for learning about PIDs and taking the first steps towards selecting a suitable PID solution, along with a PID implementation Roadmap. This was an effort to encourage the adoption and use of PIDs in the cultural heritage domain, by utilising their existing collection management system vendors.

The concept of 'cutting in the middle-man' – cooperating with collection management system vendors and supporting the implementation of PIDs in their products – has proven extremely successful, and the outcomes of this project may help cultural heritage initiatives in other countries to get PIDs implemented in their organisations too.

1 INTRODUCTION

Link rot, links resulting in a 404 or Not Found HTTP error message, is a nuisance to internet users and a threat to the sustainable availability of online cultural heritage and scientific information. Especially when links have been harvested and are (re)used in cultural heritage portals and aggregators like Europeana, broken links can have a significant impact.

Museums, libraries, archives and scientific organisations in the Netherlands are joining forces to fight link rot by implementing Persistent Identifiers (PIDs) in their (collection management) systems. PIDs are unique and permanent links to (digital) objects.

The Digital Heritage Network (DHN) established a PID project to raise awareness of PIDs, increase knowledge regarding the use of PIDs, and support PID implementation projects.

2 DIGITAL HERITAGE NETWORK

The DHN is a partnership focused on developing a network of national faculties and services for improving the visibility, usability, and sustainability of digital heritage in the Netherlands. The network was established as an initiative by the Dutch Ministry of Education, Culture and Science and is undertaken by several key institutions and experts relating to the area of digital heritage [1].

The DHN has developed a three-pronged strategy covering Visible, Usable and Sustainable aspects of Digital Heritage. For each of these aspects a work package has been established outlining projects necessary to achieve their core goals. The Sustainable work package is coordinated by the Netherlands Coalition for Digital Preservation (Nationale Coalitie Digitale

---

[1] These vendors, and the collection management systems they implemented PIDs in, are: Picturae (Memorix), DeventIT (Atlantis), De Ree (MAIS-MDWS), Cit (The Museum System).

Duurzaamheid, NCDD). The PID project[2] is one of the projects in the Sustainable work package, and is managed by the National Archives of the Netherlands. The first phase of this DHN program ran from September 2015 to June 2017.

3 PERSISTENT IDENTIFIERS

A Persistent Identifier is a unique and permanent link to a (digital) object, e.g. scan, audio file, metadata record, web page, or book. In the context of the internet, PIDs are posed as a solution to the problem of link rot: where web pages become inaccessible when a link or site address is changed. This is especially common when websites link to each other. When site addresses change (because the contents have been moved to a new page, for instance, or because the organisation maintaining the site has changed its domain name), the links no longer point to a valid address, resulting in a 404 or Not Found HTTP status code. With the rise of aggregators like Europeana3, and the increasing availability of (linked) open data, link rot has become a serious threat to the sustainable findability of digital cultural heritage. The metadata of digital objects are harvested on a large scale and reused on various websites, and if the original address is then changed, thousands or even millions of links can suddenly break.

4 THE PERSISTENT IDENTIFIER PROJECT

The Persistent Identifier project promotes sustainable access within the cultural heritage domain. Cultural heritage organisations, regardless of size, are often hesitant to implement PIDs. They lack knowledge of what PIDs are, don't know about the capabilities and benefits of PIDs, and fear a possible complex and costly implementation process as well as the maintenance costs for a sustained service.

To address these issues, the project worked on:

- Raising awareness of (the importance of) PIDs in cultural heritage organisations
- Increasing knowledge regarding the use of PIDs within cultural heritage
- Supporting the technical implementation of PIDs in cultural heritage collection management

The project ran from September 2015 to June 2017, and will continue in a follow-up project in the second phase of the DHN program.

4.1 Raising Awareness

To raise awareness, the PID project first created an inventory of cultural heritage organisations currently using PIDs, and those that did not. Organisations using PIDs were asked to share their (best) practices with us. Organisations not yet using PIDs were invited to workshops and presentations, for which training and education material was created. A PID Roadmap was also written,

explaining the most important steps in a PID implementation project: preparation, implementation and publication, and the subsequent management of PIDs.

The PID project made extended efforts to educate a wider audience about the importance of PIDs, by publishing articles in the magazines Archievenblad and Informatieprofessional [2], also resulting in an online PID dossier[4], and regular communication through the DHN's social media channels. Three videos were recorded and published in Dutch and English: 'What are Persistent Identifiers', 'Choosing the most suitable Persistent Identifier', and 'Implementing Persistent Identifiers'[5].

4.2 Increasing Knowledge

Once a cultural heritage organisation is aware of the importance of PIDs, they often notice that they need to learn more about PIDs and the various PID options, and eventually select one (or more) to implement. There are multiple PID options, including Archival Resource Keys (ARKs[6]), Digital Object Identifiers (DOIs[7]), the Handle System[8], OpenURL[9], Persistent Uniform Resource Locators (PURLs[10]) or Uniform Resource Names (URNs[11]). Each system has its own particular properties, communities, strengths and weaknesses. But which one is best suited to a particular organisation's needs? Choosing a PID solution is an important decision with long-term implications. To help cultural heritage organisations learn and think about important PID aspects, and to guide them through their first steps towards selecting a PID solution, the project developed a PID Guide. The Guide was developed in collaboration with PID experts from Delft University of Technology (for DataCite DOI), SURFsara (the Handle System) and the National Library of the Netherlands (for the URN:NBN namespace). DataCite DOI, the Handle System and URN:NBN are the three most used PID solutions in the Netherlands.

4.3 The PID Guide

The PID Guide reuses the idea of Voting Advice Applications[12] and guides participants through 25 statements about PIDs. The participant is presented with 5 options to choose from per statement. These options reflect a 5-point Likert scale [3]. The outcome of this guide points towards a preference for one

---

[2] See http://www.ncdd.nl/en/ncdd-projects/digital-heritage-network/project-persistent-identifiers/, visited 2017-03-29.

[3] See http://www.europeana.eu/, visited 2017-03-29.

[4] See http://www.informatieprofessional.nl/bijdragen/2016/06/dossier-oplossingen-voor-/, visited 2017-03-29.

[5] See the NCDD YouTube channel
https://www.youtube.com/channel/UCvHDarEQeREkRVLdzJSbKWw, visited 2017-03-29.

[6] See https://confluence.ucop.edu/display/Curation/ARK, visited 2017-03-29.

[7] See https://www.doi.org/, visited 2017-03-29.

[8] See http://www.handle.net/, visited 2017-03-29.

[9] See http://www.oclc.org/research/activities/openurl.html, visited 2017-03-29.

[10] See https://archive.org/services/purl/, visited 2017-03-29.

[11] See https://en.wikipedia.org/wiki/Uniform_Resource_Name, visited 2017-03-29.

[12] See https://en.wikipedia.org/wiki/Voting_advice_application, visited 2017-03-29.

or more PID solutions. More importantly, the participant has learned and thought about 25 important PID issues in the process.

The preference(s) calculated by the Guide should not be seen as a definitive choice for a particular PID solution. Other factors not represented in the guide may affect the final decision. An organisation may also be part of a community that has a strong preference for a particular PID solution. Where the Guide does not rule out any PID solution, an organisation could have showstopper criteria for selecting a PID solution, such as cost, level of support or formal standardisation.

PID solutions are rarely restricted to national borders, and while the PID Guide was written with Dutch requirements in mind, the 25 statements and technical setup are transferable internationally. We therefore welcomed reviews of the PID Guide by the International DOI Foundation[13] and nestor[14]. Their feedback helped improve the PID Guide before publication. Additionally, the PID Tutorial held at iPres 2016 [4], resulted in interest from other countries (to translate the PID Guide) and communities to create a more comprehensive PID resource (including, e.g., more PID solutions). The PID Guide has a CC0 license[15] and is freely reusable[16]: www.ncdd.nl/en/pid-wijzer.

At the time of writing, 91 participants have used the PID Guide. Participants are asked to provide information about their type of organisation, job title and collection management system. This gives the project valuable information regarding the types of organisations using the PID Guide, and their (resulting) preferences. After filtering out incomplete entries, the data from 69 participants, including 23 international participants, was used to calculate preferences for PID solutions in the organisation types archive, library, museum, science and other.
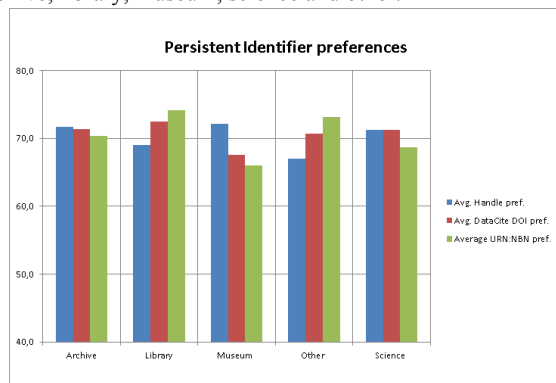

Figure 1: preliminary PID Guide preferences for organisation types.

The bar chart from figure 1 is based on the average preferences resulting from the PID Guide. For each participant, the resulting preference scores for all (3) PID solutions are used as input. The chart shows a preference for handles in museums (based on 18 participants), a preference for URN:NBN in libraries (20 participants), and a preference for DataCite DOI in research institutions (10 participants). Archives (16 participants) have a less clear preference for a particular PID solution. In the Other category, URN:NBN is preferred. This is however based on a limited set of 5 participants, mostly from commercial organisations. Given the limited number of participants for this survey and the fact that no immediate follow-up research has yet been undertaken, figure 1 is a first sketch of the kind of information that can be gathered from the PID Guide. In time, we hope to receive more data from PID Guide participants, and analyse the PID Guide's database in greater detail. The results could help focus future PID advocacy and implementation efforts.

4.4 Cost

The cost of (implementing) PIDs is an important issue. While some PID solutions are free, i.e. have no annual fee, the combined cost of implementation, maintenance and use is never zero. In the past, the cost of PID solutions has been a major inhibiting factor in PID implementation in cultural heritage organisations. For example during the Continuous Access to Cultural Heritage (CATCH, 2004-2014) and CATCHPLUS (2009-2012) projects, important steps were taken towards implementing PIDs in cultural heritage organisations, but the cost of using PIDs remained an obstacle for widespread uptake[17]. The timing of this PID project was more fortunate. The establishment of EPIDC[18], the foundation of DONA[19] and changes in SURFsara's organisational policies led to affordable Handle System based PID services becoming available for cultural heritage organisations. With URN:NBNs provided for free by the National Library of the Netherlands, and clients of Delft University of Technology's DataCite DOI service required only to pay an annual fee of €750 per prefix, the basic annual cost for using the major PID solutions in the Netherlands are all below €1,000 and within financial reach of most cultural heritage organisations regardless of size[20].

One important cost issue remained: the cost of implementing PIDs within collection management systems. This is where the cutting in the middle-man strategy of the PID project came into play. Instead of requiring individual cultural heritage organisation

---

[13] See https://www.doi.org/, visited 2017-03-29.

[14] See http://www.langzeitarchivierung.de/Subsites/nestor/EN/Home/home_node.html, visited 2017-03-29.

[15] See https://creativecommons.org/publicdomain/zero/1.0/, visited 2017-03-29.

[16] We are discussing the transfer of the stewardship of the PID Guide to the Open Preservation Foundation, thus further increasing its international visibility and sustainability.

[17] See for CATCHPLUS http://www.catchplus.nl/en/pids/index.html, and, for information about the cost obstacle https://www.edata.nl/0903_010615/pdf/Het_bemiddelen_tussen_vraag_en_aanbod_zit_in_mijn_genen.pdf (in Dutch), both visited 2017-03-29.

[18] See http://www.pidconsortium.eu/, visited 2017-03-29.

[19] See https://www.dona.net/, visited 2017-03-29.

[20] Free DOI services, like Zenodo, figshare and Dataverse were not taken into consideration. Their modus operandi include uploading objects, and most cultural heritage organisations have published or want to publish and manage their collections in their collection management systems.

to finance a tailor-made PID implementation project, the PID project offered to pay for the initial implementation costs to the vendors, under the proviso that a generic and reusable solution was developed. In theory, the implemented PID solution should become available to all users of that (collection management) system.

4.5 Cutting in the middle-man

Via the projects' various communication channels, workshops and presentations, word spread that support for PID implementation was available. This led to talks with vendors of collection management systems working with Dutch cultural heritage institutions. As a result, at the time of writing, 4 of the 5 key software vendors have started implementing PIDs in their systems, together with a variety of pilot institutions:

- Picturae[21], together with the Noord-Hollands Archief, Archief Eemland, Regionaal archief Zutphen, Groninger archieven and Regionaal archief Tilburg, have made handles available in Memorix software,
- DeventIT[22], together with the Centraal Bureau voor de Statistiek, Regionaal Archief Nijmegen and Gemeentearchief Ede, are implementing handles in Atlantis,
- De Ree Archiefsystemen[23], together with Regionaal Archief Rivierenland, are implementing handles in MAIS-MDWS,
- Cit[24], distributor of The Museum System, are implementing handles with the Wereldculturen group, which counts the Tropenmuseum, the Afrika Museum and Museum Volkenkunde as user institutions.

These vendors and their clients discussed possibilities and their institutional preferences, and in all cases, a preference for the Handle System based PID service from SURFsara was chosen. One of the benefits to this solution seems to be that PID registration and maintenance can be handled automatically, via an APID. The PID project has also started discussion with the Dutch representatives of the Axiell Group[25] regarding the implementation of PIDs in the Adlib software.

Each vendor has a specific business model and annual fee schedule for the provision of PIDs in their systems. Most see PIDs as a service that should be available in modern collection management systems, and keep their fees relatively close to the minimum tariff set out by SURFsara's PID service. The aim of the project was to make the provision of PIDs advantageous for all parties, which has been achieved by this cutting in of the middle-man.

4.6 Best Practices

The collection management system vendors that implement PIDs and the cultural heritage organisations that use them, were asked to share their experiences from the implementation projects. Their shared experiences will help encourage other organisations in the future. Their experiences have been gathered in a best practice document. As the template for these documents has the same headings as the PID implementation Roadmap, a consistent feedback loop was created: organisations starting a PID implementation project can use the PID Roadmap, and their best practice documentation helps fill a library of PID implementation project experiences, which helps improve the Roadmap.

5 CONCLUSION

The DHN PID project set out to raise awareness of PIDs, increase knowledge regarding the use of PIDs, and support PID implementation projects. The project's results show that the project's approach was successful.

The PID information, training and education material used in workshops and presentations contributed to raising awareness, together with PID Roadmap and best practice documents. The unique PID Guide is still actively used by organisations for learning about PIDs and taking the first steps towards selecting a suitable PID solution. SURFsara's Handle System-based PID service has become available to cultural heritage organisations at an affordable price, ensuring that all major PID solutions in the Netherlands are now available for less than €1,000 per year. Most importantly, by cutting in the middle-man, PIDs have been implemented in (4) leading collection management systems in the Netherlands, and are available to all their (inter)national customers.

Although the DHN is a national program, the results of the PID project can have an international impact. DataCite DOI, URN:NBN and the Handle System are available internationally. Other countries or communities can also choose to reuse the materials, PID Guide and the concept of cutting in the middle-man.

REFERENCES

[1] Digital Heritage Network, National Digital Heritage Strategy, retrieved March 29, 2017, from DEN Foundation: http://www.den.nl/art/uploads/files/Publicaties/20150608_Nationale_strategie_digitaal_erfgoed_Engels.pdf.
[2] M. Rusken. 2016. Persistent Identifiers. *Informatieprofessional 2016*, 4 (2016), 24-27.
[3] R. Likert. 1932. *A technique for the Measurement of Attitudes.* Archives of Psychology, 140, 1-55.
[4] J. Clark, R. van Veenendaal, M. Lunghi and J. Hakala 2016. Persistent Identifiers for Cultural Heritage. In *Proceeding of the 13th International Conference on Digital Preservation* (iPres 2016). Swiss National Library, Bern, 309-310.

---

21 See https://picturae.com/, visited 2017-03-29.

22 See https://www.deventit.nl/, visited 2017-03-29.

23 See https://www.de-ree.nl/, visited 2017-03-29.

24 See http://www.go2cit.nl/en/, visited 2017-03-29.

25 See http://alm.axiell.com/, visited 2017-03-29.