Choosing a Persistent Identifier Type for Your Digital Objects

Remco van Veenendaal

National Archives of the Netherlands

DPC Technology Watch Guidance Note

August 2023

DigitalPreservationCoalition

This report is published by the Digital Preservation Coalition (DPC). The DPC is an international charitable foundation which supports digital preservation and helps its members around the world to deliver resilient long-term access to digital content and services. In addition to the publication of reports on a range of themes which cover the state of the art in digital preservation, the DPC also supports its members through community engagement, targeted advocacy work, training and workforce development, identification of good practice and standards, and through good management and governance. Its vision is a secure digital legacy.

Discover the DPC's publications, including the latest updates and revisions, at: https://www.dpconline.org/digipres/discover-good-practice/tech-watch-reports

Find out more about the DPC, the support it offers and how to become a member at: https://www.dpconline.org/about/join-us

1   Introduction

This report is intended to help you get started using persistent identifiers (PIDs) for digital objects. Its intended audience is people who are involved in digital preservation in heritage and research organizations. The report answers questions such as: 'What are persistent identifiers?', 'Why are they important?', 'Which type should you choose?', 'Are you ready for them?', and 'How should you implement them?'. The report does not specifically cover persistent identifiers for people, organizations, grants, workflows, and so on, but some of the same general concepts would also apply.

2   What are PIDs?

According to the online Cambridge Dictionary, 'persistence' can be defined as the fact that someone or something persists. One of its synonyms is tenacity: the determination to continue what you are doing. An 'identifier' is defined as a set of numbers, letters, or symbols that is used to represent a piece of data.

A persistent identifier could therefore be defined as a set of numbers, letters or symbols that has the determination to continue to represent (or identify) a piece of data. However, numbers, letters, or symbols do not have the determination to do something, so this persistence must be implemented by people, through policy, procedures, systems and services.

In the DPC's Digital Preservation Handbook, a **persistent identifier** is defined as 'a long-lasting reference to a digital resource' (Digital Preservation Coalition, 2015). Additionally, it typically '(…) has two components: a unique identifier; and a service that locates the resource over time even when its location changes. The first helps to ensure the provenance of a digital resource (that it is what it purports to be), whilst the second will ensure that the identifier resolves to the correct current location'.

Persistent identifiers can refer to both digital and physical objects. Note that in some cases, a persistent identifier will not link directly to an object or resource, but to a 'landing page' with information (or metadata) about the object.

3   Why are PIDs important?

Persistent identifiers are an important tool in the digital preservation practitioner's toolbox and have been in existence for more than 20 years. Use of persistent identifiers is considered good practice for those involved in providing long-term access to digital resources as illustrated by the examples below.

- The first of the FAIR Guiding Principles defined by Wilkinson, Dumontier, Aalbersberg *et al.*, (2016) is: 'F1. (meta)data are assigned a globally unique and persistent identifier'.
- Requirement 12, Discovery and Identification, of the CoreTrustSeal Standards and Certification Board (CoreTrustSeal, 2022) asks for persistent identifiers: 'Once discovered, digital objects should be referenceable through full citations, including persistent identifiers (PIDs) to help ensure that they can be accessed into the future'.
- The managed level of the Metadata Management section of the DPC's Rapid Assessment Model (Digital Preservation Coalition, 2021) states that 'Persistent unique identifiers are assigned and maintained for digital content'.
- The nestor seal (Bähr, Gerdes, Harmsen *et al.*, 2019) and ISO 16363 (ISO, 2012) also expect you to use persistent identifiers.

Those who access digital content (whether human or machine) expect that the same reference will return the same digital object at a (much) later time. What they do not want to receive is an 'HTTP 404 Not Found' status code (also referred to as 'link rot') (IANA, 2022), or a different object entirely (known as 'content drift'). These two issues combined are sometimes referred to as 'reference rot' (Jones, Klein and Van de Sompel, 2021).

Demonstrating that reference rot is an existing problem, Zittrain, Bowers and Stanton (2021) found that of *The New York Times* articles between 1996 and 2019, 25 per cent of external links were inaccessible (link rot). In addition, from a sample of 4,500 links still accessible, 13 per cent did not lead to the original content (content drift). For journalistic fact-checking or reproduction of research these findings are very problematic.

The *Digital Preservation Handbook* notes that reference rot is caused not only through technological issues (such as server failure) but more commonly through human and organizational issues – 'Organizations transfer journals to new publishers, reorganize their websites, or lose interest in older content, leading to broken links when you try to access a resource. This is frustrating for users, but the consequences can be serious if the linked resource is essential for legal, medical, or scientific reasons' (Digital Preservation Coalition, 2015). The use of Persistent Identifiers can help to solve this problem.

4   How do PIDs work?

Persistent identifiers are not complex. There are two things to keep track of: the persistent identifier, and what it resolves into. You can compare persistent identifiers to national identification numbers used on ID cards or passports. Those numbers are used by governments as a means of tracking their citizens (Wikipedia, 2023). The identification number does not change but is associated with the current name and address of a citizen. If that citizen moves house, the number's associated address needs to be updated.

A persistent identifier works similarly: an identifier is associated with an object's current location. If that object's location changes, the identifier's associated address – often a URL – needs to be updated.

If you want to get a glimpse of what goes on behind the scenes, take a look at, for example, a record of a persistent identifier of the Handle System. Visit the Handle System resolver (CNRI, 2023). Enter the handle for Rembrandt van Rijn's *The Night Watch*: 10934/RM0001.COLLECT.5216. Check 'Don't Redirect to URLs', click 'Submit' and take a look at the resulting handle record. What you should see is a 'Handle Values for 10934/RM0001.COLLECT.5216' table with two rows. The first row contains the URL that the handle resolves into. The second row contains administrative information: who has which create, read, update and delete rights.

Persistent identifiers usually consist of a *prefix* and a *suffix*. The prefix uniquely identifies the organization while the suffix uniquely identifies the object; the combination should be globally unique. The prefix is assigned to an organization by the persistent identifier type provider, and the suffix can be defined by the organization itself (for example, representing the object's collection code, or a unique identifier).

A *resolver* is used to make persistent identifiers actionable on the internet. In practice, a resolver is an online service that resolves persistent identifiers into locations (URLs). Persistent identifiers are usually written as a combination of resolver, prefix and suffix, the prefix and the suffix forming the persistent identifier.

Applying the concepts of resolvers, prefixes and suffixes to various persistent identifier types, you will see that they are very similar in form (with an additional label for ARKs and URNs). The following list, inspired by ARK Alliance (2022) shows five example web addresses that consist of a resolver service and a persistent identifier:

- https://n2t.net/ark:/**99999**/*12345*   (ARK)
- https://dx.doi.org/**10.99999**/*12345*   (DOI)
- https://hdl.handle.net/**99.99999**/*12345*   (Handle System)
- https://purl.org/**99999**/*12345*   (PURL)
- https://<various>/urn:**99999**:*12345*   (URN(:NBN))

5   Which PID types are there?

For digital objects, there are a range of different identifiers to choose from, including those listed below:

- Archival Resource Key - ARK: https://arks.org/
- Digital Object Identifier - DOI: https://www.doi.org/
- Handle System: https://www.dona.net/handle-system
- Persistent URLs - PURL: https://purl.archive.org/
- Uniform Resource Name (National Bibliography Number) - URN(:NBN): https://en.wikipedia.org/wiki/National_Bibliography_Number

This list is not exhaustive and perhaps oversimplifies a more complex landscape. The DOI Foundation is a 'community of communities', with, for example, the Crossref (2023) and DataCite (2023a) communities providing slightly different services (Crossref, 2019). URN(:NBN) is an example of how a URN namespace is used by a community; it is not a persistent identifier type in itself. Also, DOI uses the Handle System technology for the resolution part of the DOI system.

6   How do you choose a PID type?

Persistent identifiers are all meant to solve the same problem and there are many to choose from. One of the challenges can be simply deciding which persistent identifier type is right for your digital content and organizational context.

One of the readily available tools to help with this is the *Guides to Choosing Persistent Identifiers* (Madden, van Horik, van de Sandt *et al*., 2020) compiled by the European FREYA project. Note that these guides have a broader scope than this report, including persistent identifiers for publications, datasets, people, organizations, and software as well as objects.

Another helpful resource is *Developing Identifiers for Heritage Collections* (Kotarski, Kirby, Madden *et al*., 2021). It was created in the context of the Towards a National Collection programme (Towards A National Collection, 2023) and 'describes the different aspects of developing persistent identifiers (PIDs) in the context of Heritage Collections'. The section 'Which persistent identifier does what?' helpfully compares the functionalities of a long list of persistent identifier types.

The *Persistent Identifier Guide* (Dutch Digital Heritage Network, 2022a) 'guides you through 25 statements, helps you learn and think about important persistent identifier subjects, and guides your first steps towards selecting a system'. Note that this resource focuses on the Dutch digital heritage sector but is also available in English. It compares Archival Resource Keys (ARK), DataCite DOIs, the

Handle System and URN:NBNs. A Persistent Identifier Guide Toolkit is also available to enable others in the community to create guides for different languages, sectors and contexts.

In addition to these recommended tools and resources, the websites of persistent identifier type providers are also helpful to collect and compare information, and talking to others is also an invaluable source of information. Ask your collection management system provider which persistent identifier type they implemented in their system or ask those working at similar organizations which persistent identifier type they chose and why.

7 Are you ready for PIDs?

How do you know if your organization is ready for persistent identifiers? Asking yourself these questions may help:

1. Do you have digital **objects** or digital metadata about your objects?
2. Do your objects have unique **identifiers**?
3. Do you have a **website**? (Note, this may be a third-party platform such as Dataverse or Zenodo)
4. Do you **publish** your objects, using your unique identifiers, on your website?
5. Do you manage your objects in a **systematic** way, for example in a collection management system?
6. Do you have to do **long-term preservation** of your objects?
7. Do you have to provide **long-term access** to your objects?
8. Do you have the option and means to **implement** persistent identifiers in your systems?
9. Do you have the means to **manage** and **maintain** your persistent identifiers?
10. Do you have a plan to let your users **make use of** your persistent identifiers?

If you answer 'yes' to these questions, read on to find out how to implement persistent identifiers. If you answered 'no' to one or more of these questions, you may benefit from carrying out additional work before moving forward with implementation.

8 Implementing PIDs step-by-step

This section is a summary of the *Stappenplan voor implementatie van Persistent Identifiers* (*Step-by-step plan for the implementation of persistent identifiers)* ([Dutch Digital Heritage Network](#), 2022b).

8.1 Preparation
1. Decide which **objects** you want to assign persistent identifiers to.
2. Decide **where** you will create and maintain persistent identifiers, for example, where in your processes and in which systems.
3. Look for **good practice** of how others implemented persistent identifiers, use your network and speak to others.
4. Choose a persistent identifier **type** that is suitable for your situation.
5. Choose a **provider** of that persistent identifier type and read implementation guidance from that provider.
6. Read **example persistent identifier policies** from others, for example, from the Bibliothéque nationale de France ([National Library of France](#), 2019), the British Library ([British Library](#), 2021) or the Nationaal Archief ([National Archives of the Netherlands](#), 2023). Another example is the persistent identifier policy for the European Open Science Cloud ([European Commission, Directorate-General for Research and Innovation](#), 2020).

8.2   Implementation

1. Establish a **policy, and a plan** for create, read, update and delete operations on persistent identifiers (more on this in section 8.4).
2. Decide how you will **form** your persistent identifier suffixes, for example, with a number, the object's collection code or a Universally Unique Identifier (UUID).
3. **Choose the system** in which you will implement the persistent identifier operations.
4. **Implement the operations** in that system. You may have the capabilities to do this in house or outsource the job to, for example, the system vendor.

8.3   Publication

1. **Publish your persistent identifiers** so that they are visible to your users (human and machine).
2. **Publish your persistent identifier policy**, both internally and externally.
3. **Communicate** the existence of your persistent identifiers, where to find them, and how to use them.

9   Maintenance

Remember that persistent identifiers will not persist over time without active management and maintenance. Having implemented persistent identifiers, you will have to maintain them. Also from the *Stappenplan* are these things to consider (Dutch Digital Heritage Network, 2022b).

1. Ensure all stakeholders in your organization are aware of the importance of having, managing and maintaining your persistent identifiers. Ask yourself, 'If everyone who was involved in our persistent identifier implementation were to leave the organization, how would staff know how to manage and maintain them correctly?'
2. Be prepared to take action when – not if – the locations of your objects change. Know whose responsibility it will be to perform which action when that time comes.
3. If an object that has a persistent identifier is deleted, create a 'tombstone page' that informs the user that the object is no longer available. Ideally it should also state when the object was removed and why. Your persistent identifier provider may have its own guidance on this (see for example DataCite 2023b).
4. If possible, create an agreed plan for what will happen if the department or organization responsible for maintaining your persistent identifiers should cease to exist.

10 Conclusion

The use of persistent identifiers is an important part of digital preservation good practice in supporting access to digital objects over the longer term. Using persistent identifiers increases your trustworthiness as an organization.

There are different types of persistent identifier, and you should choose one to meet your own requirements and context. The good news is that there are tools and communities that can help with this and some of these have been outlined in this guidance note.

Once implemented, persistent identifiers should be maintained and supported over time to ensure they provide a reliable way of referencing digital objects. Even if your digital objects persist, they are far less likely to be discoverable, usable, and useful without well-maintained persistent identifiers. Be persistent about your persistent identifiers!

11 Acknowledgements

I want to thank Jenny Mitcham of the Digital Preservation Coalition for inviting me to write this report. It is a great opportunity to raise awareness of the importance of persistent identifiers. I also want to thank Frances Madden (Technological University Dublin), Tamara van Zwol (Dutch Digital Heritage Network) and René van Horik (DANS-KNAW) for their comments on an earlier draft.

12 References

ARK Alliance (2022) 'What ARK, DOI, Handle, PURL, and URN have in common'. Available at: https://web.archive.org/web/20230814121517/https://arks.org/about/comparing-arks-and-other-identifiers/

Bähr, T., Gerdes, T., Harmsen, J. H. *et al*. (2019) 'Erläuterungen zum nestor-Siegel für vertrauenswürdige digitale Langzeitarchive. Version 2.1.'. Nestor-Materialien 17. Available at: https://web.archive.org/web/20230815101700/https://d-nb.info/1189191830/34 [German only]

British Library (2021) 'British Library Persistent Identifier Policy'. British Library. Available at: https://web.archive.org/web/20230815101752/https://bl.iro.bl.uk/concern/reports/4a68b806-8c2f-4d89-a0dd-c30e6ff98e91

CNRI (2023) 'Resolve a Handle and View the Values'. Available at: https://web.archive.org/web/20230815000849/http://hdl.handle.net/

CoreTrustSeal Standards and Certification Board (2022) 'CoreTrustSeal Requirements 2023–2025 (V01.00)'. Zenodo. Available at: https://web.archive.org/web/20230815102004/https://zenodo.org/record/7051012

Crossref (2019) 'DataCite – Crossref'. Available at: https://web.archive.org/web/20230529191157/https://www.crossref.org/community/datacite/

Crossref (2023) 'You are Crossref'. Available at: https://web.archive.org/web/20230813232625/https://www.crossref.org/

DataCite (2023a) 'Welcome to DataCite'. Available at: https://web.archive.org/web/20230814050941/https://datacite.org/

DataCite (2023b) 'Best Practices for Tombstone Pages'. Available at: https://web.archive.org/web/20230530001745/https://support.datacite.org/docs/tombstone-pages

Digital Preservation Coalition (2015) *Digital Preservation Handbook, 2nd Edition*. Available at: https://web.archive.org/web/20230809220313/https://www.dpconline.org/handbook

Digital Preservation Coalition (2021) Digital Preservation Coalition Rapid Assessment Model (DPC RAM) (version 2). Available at: https://web.archive.org/web/20230815102602/https://www.dpconline.org/docs/miscellaneous/our-work/dpc-ram/2433-digital-preservation-coalition-rapid-assessment-model-v2/file

Dutch Digital Heritage Network (2022a) 'Persistent Identifier Guide'. Available at: https://web.archive.org/web/20230608192923/https://www.pidwijzer.nl/

Dutch Digital Heritage Network (2022b) 'Stappenplan voor implementatie van Persistent Identifiers'. Available at: https://web.archive.org/web/20230815102722/https://www.pidwijzer.nl/en/implementation [Dutch only]

European Commission, Directorate-General for Research and Innovation (2020) 'A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC)'. Publications Office. Available at: https://web.archive.org/web/20230815102834/https://op.europa.eu/en/publication-detail/-/publication/35c5ca10-1417-11eb-b57e-01aa75ed71a1/language-en

IANA (2022) 'Hypertext Transfer Protocol (HTTP) Status Code Registry'. Available at: https://web.archive.org/web/20230801205330/https://www.iana.org/assignments/http-status-codes/http-status-codes.xhtml

ISO (2012) ISO 16363:2012. 'Space data and information transfer systems — Audit and certification of trustworthy digital repositories'. Available at: https://www.iso.org/standard/56510.html [accessed 15 August 2023]

Jones, S., Klein, M. and van de Sompel, H. (2021) 'Robustifying Links to Combat Reference Rot' Code4Lib, 50 edn. Available at: https://web.archive.org/web/20230604005954/https://journal.code4lib.org/articles/15509

Kotarski, R., Kirby, J., Madden, F. *et al*. (2021) 'Developing Identifiers for Heritage Collections'. Zenodo. Available at: https://web.archive.org/web/20230815103218/https://zenodo.org/record/5879089

Madden, F., van Horik, R., van de Sandt, S. *et al.* (2020) 'Guides to Choosing Persistent Identifiers – Version 3'. Zenodo. Available at: https://web.archive.org/web/20230815103340/https://zenodo.org/record/4192174

National Archives of the Netherlands (2023) 'Persistent identifier policy'. Available at: https://web.archive.org/web/20230509174513/https://www.nationaalarchief.nl/archiveren/nieuws/nieuw-pid-beleid-nationaal-archief-online-beschikbaar [Dutch only]

National Library of France (2019) 'The ARK identifier (Archival Resource Key)'. Available at: https://web.archive.org/web/20230815103705/https://www.bnf.fr/fr/lidentifiant-ark-archival-resource-key [French only]

Towards a National Collection (2023) 'Towards a National Collection | Collections United'. Available at: https://web.archive.org/web/20230629214521/https://www.nationalcollection.org.uk/

Wikipedia (2023) 'National identification number'. Available at: https://web.archive.org/web/20230806181324/https://en.wikipedia.org/wiki/National_identification_number

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al*. (2016) 'The FAIR Guiding Principles for scientific data management and stewardship'. *Sci Data* 3, 160018. Available at: https://web.archive.org/web/20230815105007/https://www.nature.com/articles/sdata201618

Zittrain, J., Bowers, J. and Stanton, C. (2021) 'The Paper of Record Meets an Ephemeral Web: An Examination of Linkrot and Content Drift within *The New York Times*'. Library Innovation Lab, Harvard Law School. Available at: https://web.archive.org/web/20230815105105/https://dash.harvard.edu/handle/1/37367405