

Overview of PID Systems for objects, December 2022



About this overview

A Persistent Identifier (PID) is a unique, persistent identification code for a digital object.

The information in this document was initially based on the PID Overview presentation by Jonathan Clark of the International DOI Foundation during the iPRES 2016 tutorial Persistent Identifiers for Cultural Heritage [1]. This overview was last updated in December 2022.

The focus of this overview is on PID Systems for object identifiers, such as identifiers for cultural heritage objects, scientific datasets and publications. PID Systems for this purpose are (in alphabetic order): ARK, DOI, Handle System, PURL, URN and URN:NBN. We included URIs because of their use in Linked (Open) Data. For PID Systems for other purposes, see e.g. the Guides to Choosing Persistent Identifiers of the FREYA Project: <https://zenodo.org/record/4192174>.

This document is provided as part of the Persistent Identifier Guide pidwijzer.nl, a Dutch Digital Heritage Network product maintained by the National Archives of the Netherlands. The PID Guide helps you learn and think about important PID subjects, and guides your first steps towards selecting a PID system.

What the PID Systems in this overview have in common

- Choosing between PID Systems is not about right or wrong, but about the best solution for your situation
- All aim to solve the problem of references failing to take you to the referenced object, also known as broken links and often resulting in 404 error messages
- All require maintenance: you must update your PIDs when the location of the referenced object changes, e.g. because your (domain) name changes or your systems (and the object ids therein) change
- All have been in existence for more than 20 years
- All can be used for identification and citation of digital, physical or abstract objects
- All are similar in form, with a PID consisting of a **prefix** and **suffix**, and the addition of a **resolver** to make the PID actionable. The prefix identifies your organization, the suffix your object. The suffixes can have any form you like (UUID, collection code, number, etc.) and a rule of thumb is: the less semantics, the more persistent. ARKs and URNs have an additional **label** "ark:" or "urn:". Parts are separated by a slash, or colon with URNs.
- (Hypothetical) examples¹ are:
 - <https://n2t.net/ark:/99999/12345> (ARK)
 - <https://dx.doi.org/10.99999/12345> (DOI)
 - <https://hdl.handle.net/99.99999/12345> or <https://hdl.handle.net/99999/12345> (Handle System)
 - <https://purl.org/99999/12345> (PURL)
 - <https://<various>/urn:99999:12345> (URN and URN:NBN)

ARK

- Archival Resource Key (ARKs)
- Introduced in 2001
- Developed by California Digital Library (CDL), managed by the ARK Alliance, an international community around ARKs
- Documented in informational RFCs



¹ Inspiration: <https://arks.org/about/comparing-arks-and-other-identifiers/>.

- “Archival Resource Keys (ARKs) serve as persistent identifiers, or stable, trusted references for information objects. Among other things, they aim to be web addresses (URLs) that don’t return 404 Page Not Found errors. The ARK Alliance is an open global community supporting the ARK infrastructure on behalf of research and scholarship.” <https://arks.org>
- See also: https://en.wikipedia.org/wiki/Archival_Resource_Key

DOI



- Digital Object Identifier (DOIs)
- Established in 1996
- DOI is an implementation of the Handle System, using a reserved range of prefixes (10.*). It is governed and managed by International DOI Foundation (IDF), a not-for-profit member organization
- DOI is an international standard: ISO 26324, 1 May 2012
- “We are an international community of communities bound by a common interest in persistent infrastructure. So far, we have welcomed agencies that manage communities spanning entertainment, standards, the built environment, natural history collections, scholarly communications, and research data. (...) A DOI name is a digital identifier of an object, any object — physical, digital, or abstract. DOIs solve a common problem: keeping track of things. Things can be matter, material, content, or activities.” <https://www.doi.org/>
- See also https://en.wikipedia.org/wiki/Digital_object_identifier

Handle System



- Handle System (handles)
- Established in 1993
- Developed by CNRI, and governed and managed by DONA Foundation, a not-for-profit member organization
- Documented in informational RFCs
- “The Handle System is a rapid-resolution, globally distributed system run by multiple groups that the public can use for resolving identifiers (handles). (...) The Handle System is a trademarked service of the DONA Foundation and is the Identifier Resolution component of the Digital Object Architecture. (...) A key part of the Handle System is the GHR [Global Handle Registry], which contains records of prefixes allotted to local handle service providers. A client that queries the GHR, will typically learn the network address(es) and certain relevant security information of the local handle services to query for the corresponding handle record.” <https://www.dona.net/handle-system>
- See also https://en.wikipedia.org/wiki/Handle_System

PURL



- Persistent Uniform Resource Locator (PURLs)
- Introduced in 1995
- Developed and implemented by OCLC, but currently the Internet Archive holds the resolver service and its administration interface
- “A PURL is a persistent URL, it provides a permanent address to access a resource on the web. When a user retrieves a PURL they will be redirected to the current location of the resource. When an author needs to move a page they can update the PURL to point to the new location. (...) The PURL service is an initiative of the Internet Archive, a 501(c)(3) non-profit, building a digital library of Internet sites and other cultural artifacts in digital form.” <https://purl.archive.org/>
- See also https://en.wikipedia.org/wiki/Persistent_uniform_resource_locator

URN

- Universal Resource Name (URNs)
- Introduced in 1994, formalized in 1997

- There is no central governance for URN and no central resolving infrastructure. National libraries in Europe have established their own subgroup of URN (URN:NBN) and operate a joint resolving infrastructure
- Documented in Internet standards
- *“A Uniform Resource Name (URN) is a Uniform Resource Identifier (URI) that uses the urn scheme. URNs are globally unique persistent identifiers assigned within defined namespaces so they will be available for a long period of time, even after the resource which they identify ceases to exist or becomes unavailable. URNs cannot be used to directly locate an item and need not be resolvable, as they are simply templates that another parser may use to find an item.”* https://en.wikipedia.org/wiki/Uniform_Resource_Name

URN:NBN

- National Bibliography Number (URN:NBNs)
- Introduced in 2001
- URN:NBNs are publication identifier systems used by national libraries in countries such as Germany, Italy, Norway, The Netherlands and Sweden. Each national library uses its own NBN strings; there is no global authority which controls them. Thus, NBNs are unique only on national level. In combination with the national URN country code prefix, they are globally unique. Resolvers also operate on a national level, some for more than one country.
- The URN namespace for NBNs has been assigned and is described in IETF RFC 8458
- *“National Bibliography Number (NBN) is a group of publication identifier systems used by national libraries in countries such as Germany, Italy, Finland, Norway, The Netherlands and Sweden. There is no global standard for the contents of NBNs; instead, they have a country-specific format. NBNs are typically used for documents which do not have a publisher-assigned identifier such as an ISBN. They can be used to identify media archived in national libraries, such as Ph.D. theses.”* https://en.wikipedia.org/wiki/National_Bibliography_Number

Note that this is an example of how a URN namespace is used by a community, not a separate PID system.

(Linked Data) URI



- Uniform Resource Identifier (URIs)
- The publication in 2005 of IETF RFC 3986's content as the full standard STD 66 reflected the establishment of the URI generic syntax as an official internet protocol. The Semantic Web uses the HTTP URI scheme to identify both documents and concepts in the real world. See also https://en.wikipedia.org/wiki/Uniform_Resource_Identifier.
- In Tim Berners-Lee's 5-star Linked Open Data scheme, URIs are a requirement for 4 or 5 stars. See also https://en.wikipedia.org/wiki/Linked_data.
- Each URI begins with a scheme name that refers to a specification for assigning identifiers within that scheme e.g. http:, ftp:, mailto: file:, etc.
- See for more information on (Cool) URIs <https://www.w3.org/TR/cooluris/> and <https://www.w3.org/Provider/Style/URI>.

Note that URI is not a PID system. But especially with Linked (Open) Data, providing persistent or Cool URIs is important for sustainable accessibility. As with PID systems, persistence is a promise that requires attention and maintenance.

References

[1] Clark, J., van Veenendaal, R., Ras, M., Lunghi, M. & Hakala, J. (2016). Persistent Identifiers for Digital Cultural Heritage. *Proceedings of the 13th International Conference on Digital Curation. Tutorial held at iPRES 2016, Bern (309-310). Bern, Switzerland: Swiss National Library.*