

Checklist: Persistent identifiers

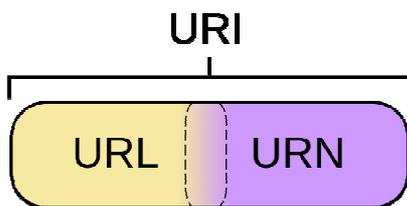
Persistent Identifiers (PID) are unique character strings¹ attached to various items. Attached to digital items, they are a prerequisite for linked data. Persistent identifiers can be applied to both objects and concepts.

Each cultural institution should have a general policy for identifiers. It is important that identifiers are designed in such a way that they are globally unique and persistent.²

If identifiers are only used internally, they should be designed in such a way that they can be integrated into formal standards and systems when used externally.

Different kinds of identifiers

There are different kinds of identifiers. A URI (Uniform Resource Identifier)³ is an identifier of a resource available online or in a data network. A URI can be classified as a Uniform Resource Locator (URL), a Uniform Resource Name (URN), or both.



Picture 1. A URI can be a URL (Uniform Resource Locator)⁴, a URN (Uniform Resource Name) or both.

(Source: "URI Euler Diagram no lone URIs" by David Torres original author derivative work: Qwerty0 (talk) - URI_Venn_Diagram.svg. Licensed under CC BY-SA 3.0, through Wikimedia Commons.⁵)

PIDs are a combination based on the identifier at the institution, the resource type and information on how to reach the resource. There are several services allowing links to be made between the identifiers used within the management systems of the institutions and external identifiers which are created when the information is, for example, made available online. With a resolution service, PIDs can be created for identification of digital files. However, a resolution service can also enable reference to several files or digital objects, and even enable storage of metadata about a resource.

¹ Siffer code or human readable

² For more information on this, see "Persistent Identifiers (PIDs): Recommendations for Institutions", ATHENA <http://www.athenaeurope.org/getFile.php?id=779>, part 4, p. 29.

³ http://en.wikipedia.org/wiki/Uniform_resource_identifier

⁴ http://en.wikipedia.org/wiki/Uniform_resource_name

⁵ http://commons.wikimedia.org/wiki/File:URI_Euler_Diagram_no_lone_URIs.svg#/media/File:URI_Euler_Diagram_no_lone_URIs.svg and http://en.wikipedia.org/wiki/Uniform_resource_name#/media/File:URI_Euler_Diagram_no_lone_URIs.svg

Today, most libraries use URNs (Uniform Resource Name) for publications, which through a resolution service can be combined with URLs (Uniform Resource Locator). In this way it is possible to make a distinction between a unique name and a location, when a named resource or its copies are located at multiple locations, or when you need to separate the concepts when a specific analogue identifier is implemented in the URI. Sometimes it could also be relevant with other sorts of identifiers, as for example LSID (Life-Science Identifier).⁶

Systems for persistent identifiers

There are both internal and external systems that can handle persistent identifiers. If your institution is using an internal system, make sure that the system is designed according to recommended standards, for example regarding the design of URIs. If URIs are created at the institution, it is important that they are as stable and persistent as possible, as it is more specifically explained in the European Commission's recommendations⁷ and W3C recommendations, for example on "Cool URI".⁸ PID systems are used increasingly today for digital resources as URIs are generally considered to be inadequate in the long run (because domain names changes often over time). However, domain names are persistent as long as you retain registration of your domain name. In practice, this means that PIDs created by a PID system could be less persistent as they are technically dependent on expertise and custom-build software.

When selecting an external system, it is important to make sure that the system is also technically reliable, authoritative, flexible in the presentation of metadata and interoperable.⁹ In relevant cases, even a resolution service that redirects the browser to a current copy of the item could be used.

With the help of a so-called content negotiation a server could be configured to give different representations of a resource, based on which representation of a URI is requested; for example, a browser can get an HTML document or a PDF file, while an RDF reader gets an RDF file.¹⁰

Persistent identifiers at your institution

- Is there a policy for persistent identifiers at your institution?
- Are the identifiers unique at your institution?

⁶ http://www.tdwg.org/fileadmin/subgroups/guid/LSIDs_for_Biologists.pdf

⁷ https://joinup.ec.europa.eu/catalogue/asset_release/study-persistent-uris-identification-best-practices-and-recommendations-topi

⁸ <http://www.w3.org/TR/cooluris/>

⁹ For more information on this, see "Persistent Identifiers (PIDs): Recommendations for Institutions", ATHENA <http://www.athenaeurope.org/getFile.php?id=779>, part 5.

¹⁰ https://joinup.ec.europa.eu/catalogue/asset_release/study-persistent-uris-identification-best-practices-and-recommendations-topi

- Are the identifiers unique globally?
- Are the identifiers persistent at your institution?
- Are the identifiers persistent globally?
- If identifiers are only used internally, are they designed in such a way that they can be integrated for external use in a unique and permanent way?
- Are identifiers already implemented in the institution's existing systems (e.g. databases used for registration/cataloguing processes)?
- Do you re-use unique identifiers that already exist (not creating a new identifier?)
- Is there a search function that creates new pointers, for example when updating the links?
- If persistent identifiers are created by the institution as URIs, do they follow the structure recommended by the European Commission: "http: // {domain} / {type} / {concept} / {reference} /"?¹¹
- Is an external service for identifiers used (such as for example a Handle resolution service <http://www.handle.net/>)?

System for persistent identifiers at your institution

- Is there a system for persistent identifiers at your institution?
- Is the system used internally?
- Is the system used externally?
- Is the system integrated with other systems at the institution, for example a collection management system?
- Is there a plan for the management and long-term preservation of the system?
- Is there an introduction to the system for everyone who uses identifiers for registration of cultural heritage material?
- Is the system technically reliable, in terms of:
 - security?
 - backup with redundant technology?
 - automatic update of the register of identifiers?
- If it is an external system which is used to assign identifiers/resolvers to resources, is it managed by an authoritative and credible organisation?
- Is the system adaptable with master data and able to expose data qualitatively and with high granularity?
- Is the system interoperable and using generally accepted international standards/open standards?

¹¹ "Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC", European Commission, 2012: <https://joinup.ec.europa.eu/sites/default/files/D7.1.3%20-%20Study%20on%20persistent%20URIs.pdf>

Examples of standards and systems for persistent identifiers

Uniform Resource Name (URN):

<http://tools.ietf.org/html/rfc3406>

URN: NBN (National Bibliography Numbers):

<http://www.kb.se/isbn-centralen/urnnbn/>

Persistent URL (PURL):

<https://purl.oclc.org/docs/index.html>

Digital Object Identifier (DOI):

<http://www.doi.org/>

Archival Resource Key (ARK):

<http://www.ifla.org/best-practice-for-national-bibliographic-agencies-in-a-digital-age/node/8793>

Open the URL:

<http://www.oclc.org/research/activities/openurl.html>

Handle system:

http://en.wikipedia.org/wiki/Handle_System

Persistent identifier component, a part of ECK, (Europeana Connection Kit) developed by the Europeana Inside-project:

http://www.europeana-inside.eu/documents/communication_material/eck-guide.html

Some interesting links on persistent identifiers

EPIC (European Persistent Identifier Consortium):

<http://www.pidconsortium.eu/>

"Persistent Identifiers (PIDs): Recommendations for Institutions", Athena:

<http://www.athenaeurope.org/getFile.php?id=779>

"State of the art report on persistent identifier standards and management tools", Linked Heritage, 2013:

<http://www.linkedheritage.eu/index.php/en/215/persistent-identifiers>

Learning Object: "Persistent identifiers - what if?" Linked Heritage:
<http://linkedheritage.cab.unipd.it/training/LO-01/en/00-about.html>

Learning Object: "Persistent identifiers - Commercial and Heritage views", Linked Heritage:
<http://linkedheritage.cab.unipd.it/training/LO-05/en/00.html>

"Persistent Identifiers for Cultural Heritage", the Digital Preservation Europe, Briefing Paper:
http://www.digitalpreservationeurope.eu/publications/briefs/persistent_identifiers.pdf

"Implementing Persistent Identifiers: Overview of concepts, Guidelines and Recommendations", Consortium of European Research Libraries
European Commission on Preservation and Access, 2006:
<http://www.ica.org/5694/paag-resources/implementing-persistent-identifiers-overview-of-concepts-guidelines-and-recommendations.html>

IFLA Persistent Identifiers:
<http://www.ifla.org/best-practice-for-national-bibliographic-agencies-in-a-digital-age/node/8789>

"Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC", European Commission, 2012:
https://joinup.ec.europa.eu/catalogue/asset_release/study-persistent-uris-identification-best-practices-and-recommendations-topi
and an extended web version <http://philarcher.org/diary/2013/uripersistence/>

"Persistent and unique identifiers", CLARIN, 2008:
<http://www.clarin.eu/sites/default/files/wg2-2-pid-doc-v4.pdf>

The CIDOC Statement on LOD identifiers:
http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/PDF/StatementOnLinkedDataIdentifiersForMuseumObjects.pdf

MuseumID - Identifying museums and museum objects on the Internet:
<http://museumid.net/documentation>

Best practices from the W3C-community:
<http://www.w3.org/TR/2015/WD-dwbp-20150224/#dataIdentification>
<http://www.w3.org/TR/ld-bp/#HTTP-URIS> (in particular "URI construction")

This checklist was developed by Digisam together with experts from Sweden and abroad. We would like to give our special thanks to Leif Johansson at SUNET, Gordon McKenna at Collections Trust and Antoine Isaac at Europeana.

Version 1.0. License CC-0. A more up to date version could be available at www.digisam.se

Digisam is a secretariat for National coordination of digitisation, digital preservation and digital access to cultural heritage. The secretariat is a government-established activity at the Swedish National Archives.