

Appendix 6. Persistent Identifier, Digital Author Identifier

Persistent Identifier

In the era of paper the International Standard Book Number (ISBN), a unique, numerical commercial book identifier, has been developed. Each edition and variation (except reprinting) of a book is given an ISBN. In the digital age, there is a growing need for such a unique, numerical, identifier for digital publications as well. Moreover, not just for publications, but for all kinds of digital objects.

On the Internet, we consider the URL as the identifier of a digital object. However, we are all familiar with broken or dead links link that points to web pages that are permanently unavailable.

An URL might change overtime, due to server migrations and other technical reasons, with undesired consequences for links and citations within scholarly communication.

Therefore a 'persistent identifier' is needed with which a digital object is permanently associated. This persistent identification number always refers to the digital object to which it has been assigned, regardless of the underlying locator technology (at the moment these are web addresses; in the future, however, an object's location may be completely different).

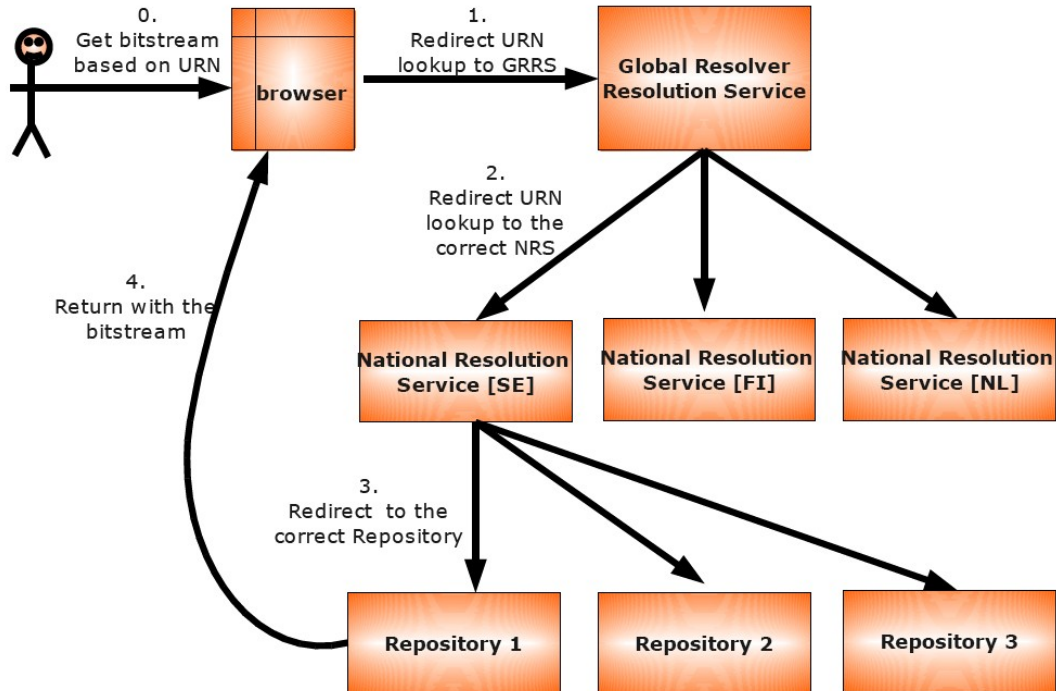
In several countries, a system for such a persistent identifier has been developed and 'national resolvers' have been set up. A resolver is a look-up mechanism that negotiates with the client application and transforms an identifier into the current location of a web resource. Common identifiers in the case of scholarly communication are DOI, Handle and URN:NBN. In case of DOI and Handle the resolution mechanism is located in the US at [CNRI](#). In case of URN:NBN resolution mechanisms are hosted by a national organisation, often this is done by the National Library.

Every digital object is assigned a number that represents that object or information for all time. Even if technology moves on, the national organisation will make sure the current location is up to date. A stable information infrastructure makes research citations a lot more reliable.

The differences between persistent identifiers are described by Hans-Werner Hilse and Jochen Kothe in [Implementing Persistent Identifiers](#). Also an article [Persistent Identifiers: Considering the Options](#) in Ariadne issue 56 by [Emma Tonkin](#)

The Scandinavian countries, Germany, Czech Republic and the Netherlands are using URN:NBN. The main reason for choosing urn's is that they are a future-proof internet standard. The only drawback now is that urn is not actionable without using a http resolution address as a prefix. Work still needs to be done on [integrating URN in the DNS](#) system by using [NAPTR records](#) that is also used for VOIP phone calls.

Recently Norway, Sweden, Finland, and the Netherlands have come to a promising proposal for a Global Resolver of Persistent Identifiers (URN:NBN). In cooperation with representatives of the Hopkins and Berkley Universities (US) a working [proof of concept](#) of a global resolver (GRRS) has been developed. This GRRS integrates four different national resolvers into one global resolver. The GSRS ([n2t.info](#)) receives the identifier from a browser plug-in and redirects the browser to the appropriate national resolver where the browser again is redirected to the current location of the web resource. The architecture of this multi-system process is depicted below.



Implementation plan on using URN:NBN persistent Identifiers

First of all we would like to say that the persistency of identifiers and web resources is not about the technology one uses, but about organisation and sustainable business models. For more information about Persistent Identifier policies take a look at the successful [Persistent Identifier Linking \(PILIN\) project](#) in Australia that is part of the [ARROW project](#).

To setup a persistent identifier program based on [National Bibliographic Numbers \(NBN\)](#) URN identifiers and a resolver one needs to take the following steps:

1. **Working group:** Create a working group that manages all the technical and organisational details of such project. Also think about the syntax that is going to be used. For example `urn:nbn:{country}:{sub-namespace}:{repositoryid}-{localid}`. Country is the short name of the country, sub-namespace represents web resources that come from the repositories, repositoryid is a two digit representation of the repository, localid is the identifier generated at the repository. This can for example result in the following identifier for one publication `urn:nbn:ie:ui:21-1234/5678`.

2. **Formalities:** Since the urn:nbn:ie namespace is by default claimed by the National Library, one has to arrange an agreement with the National Library to use a sub-namespace for scientific material. This name should be short and have no semantic meaning. For example urn:nbn:ie:ui, or urn:nbn:ie:oa, or urn:nbn:ie:sp.
3. **Registration Agency:** Create a registry in which repositories are given a short random number of two digits. This will create a sub-namespace in which a repository autonomously can distribute persistent identifiers for their publications. For example Trinity College Dublin (TCD) is registered as 21. The namespace for TCD to operate in will be urn:nbn:ie:ui:21.
4. **Implementation at local level:** Each repository must generate persistent identifiers for each publication within their namespace that is provided and store this identifier in the database record. For example TCD can use existing identifiers to add after their namespace followed by a dash. In case TCD uses handle, the identifier for one publication could look like the following urn:nbn:ie:ui:21-1234/5678. In case TCD uses database numbers urn:nbn:ie:ui:21-15874. (Make sure to store the identifier and not generate them on-the-fly. In case of database migrations might change these numbers and persistency is lost.)
5. **Transport of identifiers and URL's:** Each repository must generate a DIDL package in which the URN and URL are included. See the **Fout! Verwijzingsbron niet gevonden.** section in the main report.
6. **National Resolution Service:** A national resolver can be made by harvesting the DIDL packages from each repository where the URL and URL bindings are extracted and stored. A web location must be created where the user or machine can go to for resolution of the identifier. For example <http://resolver.ie> where the user can insert an identifier and receive the current location of the web resource. E.g. <http://resolver.ie/urn:nbn:ie:ui:21-1234/5678> resolved to <http://repository.tcd.ie/1234/5678>

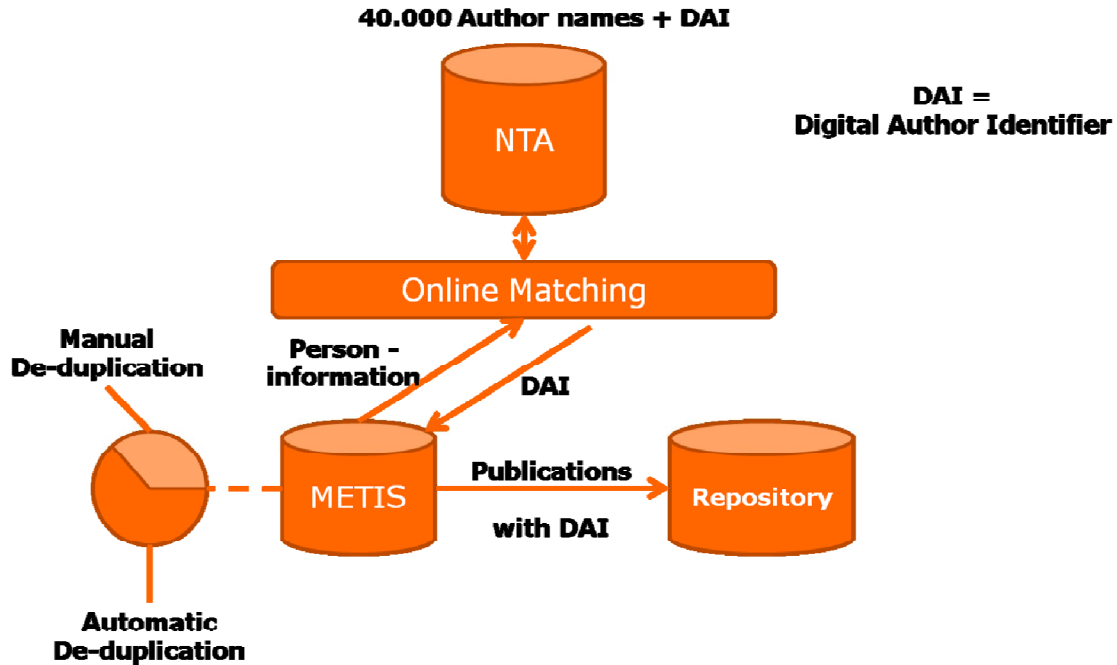
Digital Author Identifier

In a way the Digital Author Identifier (DAI) is also a persistent Identifier for authors. The purpose of the DAI is to get rid of the naming confusions of authors. Authors might have different name variations in journals where they tend to use different writing styles, or name variations appear during important changes in the author's lives, e.g. marriage. Using the DAI an end-user or system might trace the article back to the same person entity.

To put this in the perspective of the IReL open project one is able to make a publication list of an author who has his/her publication scattered throughout several repositories on an aggregated level based on the DAI, and not on unreliability of name variations. Of course this is only possible when the DAI is somehow provided in the metadata of the publication.

When implementing the DAI one could think about integrating the DAI in other systems related to the academic information domain in which the researcher appears. For example the Finance System, the Current Research Information System (CRIS) and the (Federated) Identity Management System.

The Dutch implementation of a national DAI is involved with several legacy systems. Personal information about the author such as their name and its variations, date of birth, institutions and the period they have worked at an institution are stored in a secure database which is called the National Thesaurus for Author names (NTA), currently maintained by OCLC. The NTA is part of an older shared bibliographic cataloguing project started a few decades ago. In the NTA all authors are given a DAI. Then in the Netherlands there is a CRIS called METIS, which contains all research projects and author names from a few decades ago. However the METIS systems are installed separately at each university and have no connection between each other. The authors in the METIS systems are given a DAI by matching the personal information with the information in the NTA, when there is a match a DAI is given back to the METIS system. Publications that came out of a research project can be fed into the repositories with a DAI. The illustration below depicts the process described above.



ISNI

The DAI used in the Netherlands is not just a random database number. It is a string of characters that fits in the long-anticipated International Standard Name Identifier ([ISNI](#)), due to OCLC's involvement in the ISNI project. The International Standard Organisation started this project 27729 in 2006 and expects to finish in 2009. The former name of the project was "International Standard Party Identifier", and the secretary [moved](#) from the Library and Archives Canada to the NISO organisation. To read about the developments see this [presentation](#).