



## Persistent Identifiers in brief

### State of art of PI technologies and initiatives

Persistent identification of Internet resources is an important issue within the life cycle approach to cultural and scientific digital library applications. The **instability and uncertified credibility of these identifiers reduce the extent to which the Internet can be used** as a medium for research and dissemination of scientific and cultural contents. There are many systems that address these issues, but **no single technology is likely to prevail at a global level.**

#### *Persistent URL (PURL)*

PURLs were developed by the Online Computer Library Center, Inc.(OCLC). Functionally, a PURL is a URL that, instead of pointing directly to the location of an Internet resource, points to an intermediate resolution service. The latter redirects the request to the actual location of the resource by using the standard web servers capabilities.

<http://purl.org>

#### *Uniform Resource Name (URN)*

The purpose of a Uniform Resource Name is to provide a globally unique, persistent, location-independent resource identifier which can be used for identification, for access to resource characteristics or for access to the resource itself. The URN specification is part of the IETF family of specifications encompassed by the Uniform Resource Identifier (URI) framework. This framework also includes URLs, which specify both a protocol and a location in order to give access to resources on the Web. IANA (Internet Assigned Numbers Authority) is the registration authority for URN namespaces. URNs are designed to enable heterogeneous namespaces mapping (which share URN the properties) onto a URN-space, and therefore enable the reuse of well-known identifiers.

Unlike URLs, URNs are not directly actionable (browsers generally do not know what to do with a URN) because they have no associated global infrastructure that enables resolution (such as the DNS supporting URL). Although several implementations have been made, each proposing its own means for resolution through the use of plug-ins or proxy servers, an infrastructure that enables large scale resolution



has not been implemented.

As an example in the library domain, the NBN PI schema follows the URN standard. The NBN namespace, as a Namespace Identifier (NID), has been registered and adopted by the Nordic Metadata Projects but is being separately implemented by individual systems with no reference implementation which enable the coordination of information sources. In fact, several national libraries have developed their own NBN systems for national and international research projects; several implementations are currently in use, each with different metadata descriptions or granularity levels.

<http://tools.ietf.org/html/rfc2141>

### ***Handle system***

The Handle System was developed by the Corporation for National Research Initiatives (CNRI). The Handle System is a general-purpose global name service that allows secured name resolution and administration over networks such as the Internet. The Handle System manages “handles”, which are unique names for digital objects and other Internet resources. A naming authority is authorised to create and maintain handles; the identifiers must be unique to that authority but have no prescribed syntax.

<http://www.handle.net>

### ***Digital Object Identifier (DOI)***

DOIs are names assigned to objects with intellectual properties such as electronic journal articles, images, learning objects, eboSKO, and any kind of content. They are used to provide current information, including where resources (or information about them) can be found on the Internet. The DOI is therefore an actionable identifier and its use will immediately enable resolution over the Internet. The DOI identifier itself is part of the DOI system that provides a framework for managing intellectual content, for linking customers to content providers, facilitating electronic commerce, and enabling automated copyright management for all types of media. The DOI system is an application of the Handle System; it also adds to the latter an approach based on additional metadata formats and interoperability standards, such as the Online Information Exchange (ONIX) metadata format, policies, procedures, business models, and application tools.

<http://www.doi.org>



### ***Archival Resource Key (ARK)***

The ARK schema is another proposal intended to facilitate the persistent naming and retrieval of information objects. A founding principle of the ARK is that persistence is purely a matter of service and is neither inherent in an object nor conferred on it by a particular naming syntax. ARK aims at gaining acceptance by providing peculiar functionalities, such as the capability of separating the univocal identifier assigned to a resource from the (potentially multiple) addresses that may act as a proxy to the final resource. An ARK, however, provides extra services above and beyond that of an ordinary URL. Instead of connecting to one thing, an ARK should connect to three things: the object itself, metadata, and a maintenance commitment from the current server. The resolution service returns a brief machine- and human-readable metadata record in an Electronic Resource Citations (ERCs) format.

<http://www.cdlib.org/inside/diglib/ark>

### ***Info URI***

The Info URI schema was developed within the library and publishing communities to expedite the referencing of information assets that have identifiers in public namespaces but have no representation within the URI allocation. The "info" URI scheme exists primarily for identification purposes. Implementations must not assume that an "info" URI can be dereferenced for obtaining a digital representation of the resource identified by the URI.

It is simply an identifier with no associated resolution or metadata, although some of the registered schemes may provide this functionality.

<http://info-uri.info>

### ***OASIS Extensible Resource Identifier (XRI)***

The purpose of XRI is to define a URI scheme and a corresponding URN namespace for distributed directory services that enable the identification of resources (including people and organizations) and the sharing of data across domains, enterprises and applications.

<http://www.oasis-open.org/committees/xri>



Fondazione

RINASCIMENTO  
digitale

nuove  
tecnologie  
per i beni  
culturali

Fondazione promossa da Ente Cassa di Risparmio di Firenze

### ***Library of Congress Control Number (LCCN) Permalink***

The Library of Congress has developed the “LCCN Permalink” a new persistent URL service for creating links to bibliographic records in the Library of Congress Online Catalogue using the LCCN. Permalinks can be used to cite items from the Library's collection in bibliographies, reference guides, emails, blogs, databases, Web pages, etc. This category of permalinks is referenced by bibliographic records in the LC Online Catalogue in MACHine-Readable Cataloguing (MARCXML), Metadata Object Description Schema (MODS), and Dublin Core formats.

<http://lcn.loc.gov>

### ***Enhancement of Persistent Identifier Services - Comprehensive Method for Unequivocal Resource Identification (EPICUR)***

A notable experience that tries to address this challenge is the EPICUR Project at the German National Library. It is linked to other national registries as well as to the commercial DOI platform. The resolution service is able to resolve the NBNs generated by National Libraries linked to the DOI namespace.

<http://www.persistent-identifier.de>

### ***NBN Italy***

In Italy, a consortium led by Fondazione Rinascimento Digitale is developing a novel persistent identifier architecture, based on the NBN namespace, with additional features (e.g., for resolution and robustness) and solutions that recall the DNS architecture. They have designed a distributed system, according to which registered institutions can generate and resolve the NBN identifiers associated with their sub-namespace. Institutions can also resolve NBNs generated by others linked to the infrastructure, as well as the DOI namespace, through communication with the central node. The open source code is going to be reused by other countries. An Internet Draft proposal has come out from this project and the research will continue in PIF.

<http://www.rinascimento-digitale.it/indexEN.php?SEZ=531>

### ***Name to Thing (N2T)***

Via Bufalini, 6  
50122 Firenze  
Tel 055 2613904  
Fax 055 2613906



Fondazione

RINASCIMENTO  
digitale

nuove  
tecnologie  
per i beni  
culturali

Fondazione promossa da Ente Cassa di Risparmio di Firenze

N2T proposed by John A. Kunze at the California Digital Library (CDL), tries to handle the integration of different naming techniques and tackle the challenge of PI longevity and interoperability through the definition of a single consortium. This consortium manages a global access point redirecting requests to individual PI resolution services.

<http://www.n2t.info>

### ***Life Science Identifiers (LSIDs)***

A LSID is a persistent, location-independent resource identifier for uniquely naming biologically significant resources on the Web. The LSID concept proposes an approach to naming and identifying data resources stored in multiple, distributed data stores that tries to overcome the limitations of naming schemata used today. By defining a simple, common way to identify and access biologically significant data, independently of whether data is stored in files, relational databases, in applications, or public data sources, LSID provides a naming standard underpinning for wide-area science and interoperability.

<http://lsids.sourceforge.net/>

### ***Network of Expertise in Long-Term Storage of Digital Resources (NESTOR)***

This is a cooperative project (funded by the BMBF, German Ministry of Education and Research) involving a consortium of libraries, archives and museums as well as leading experts, who form a network of expertise in long-term preservation and long-term availability of digital resources. In particular, the standards working group of the German NESTOR initiative has undertaken a project for the DIN (German Institute for Standardization) in order to explore the real long term validity of identifiers, what technical and organizational conditions are necessary for persistent identifiers to become interoperable with each other, and how to extend the notion of trustworthiness in the area of repositories and in the persistent identifiers domain.

[www.langzeitarchivierung.de/index.php?newlang=eng](http://www.langzeitarchivierung.de/index.php?newlang=eng)

### ***Persistent Identifier Linking Infrastructure (PILIN)***

Via Bufalini, 6  
50122 Firenze  
Tel 055 2613904  
Fax 055 2613906



Emphasis in the PILIN Project is put on building an identifier management infrastructure based on a technology (the Handle System) that is now under development through the auspices of CNRI in order to underpin sustainable global identifier infrastructure. The project aims to take advantage of existing governance and consultative mechanisms within the ARROW (The Australian Research Repositories Online to the World) environment for ensuring relevant and sustainable outcomes and an optimal return on investment. The project will be run in partnership between ARROW and the University of Southern Queensland to develop a technology-independent model for persistent identifiers and services. This model will extend the lifetime of the infrastructure by allowing services to be mapped to new technologies over time. It will also contribute to the international understanding of the requirements of PI infrastructures.

<http://www.arrow.edu.au/PILIN.php>

Different application scenarios and knowledge domains are expected to introduce independent systems for the purpose of customizing the metadata formats associated with PIs. In fact, existing PI techniques compose a fragmentary landscape because each of them addresses only a particular set of domain requirements.

From an accurate analysis of these systems we have realised that they vary in terms of architecture, scalability, actionability, resolution mechanisms, metadata associated with resources, data item granularity, and security.

Some of these technologies (e.g., PURL, Permalink) focus only on stability of the URL without providing additional info about resources, like type, authenticity, rights. Most of them work only in one system, with unique archive and unique user interface, they have no relationship with other PIs.

Final users must learn all the specific syntaxes and retrieval interfaces. Some technologies (e.g., DOI, NBN, ARK, info URI) bring metadata but at the moment there is not consensus on the schema to be adopted. Some PI systems are tailored on specific communities or service requirements and are not adaptable to other contexts. The handle system is a strong technology suitable in particular for environments where security is a priority. One application of the handle, DOI is surely a 'success story' on this field, thanks to the strong commitment from its community, and clear economic value/model for the applications, but we must envisage also some systems more open and free. The N2T is a new proposal facing two main issues: first, co-existence of different technologies through a central server 'translating' the different syntaxes; second, longevity of the archive through a consortium of 200 big institutions but this approach is difficult in Europe due to the different contexts and cultural background. Finally, the NBN up to now has not developed a common approach/interface for different environments and countries.

In conclusion, this scenario shows that:



Fondazione **RINASCIMENTO**  
**digitale**  
nuove  
tecnologie  
per i beni  
culturali

Fondazione promossa da Ente Cassa di Risparmio di Firenze

- **We cannot expect/impose a global unique technology;**
- **trustability/longevity of the institution maintaining the service is crucial;**
- **granularity of what a PI can point to is widely different in each application;**
- **user interface must be friendly and possibly unique for different PI systems.**

Via Bufalini, 6  
50122 Firenze  
Tel 055 2613904  
Fax 055 2613906