Overview

Persistent identifiers are a prerequisite for sustained and reliable resource discovery. A number of persistent identification systems have been developed that use different technical approaches and operate at different levels. This briefing paper introduces some of the systems currently available to support persistent identification of digital resources, and points to resources to help deployment within institutional repositories.

What are persistent identifiers?

The use of location-based identifiers such as the Uniform Resource Locator (URL) can lead to problems accessing resources over time. When accessing a resource via a hyperlink users may receive a "404 - page not found" error, which is caused by the resource being moved, relocated or renamed while external links to the resource remain unchanged.

Persistent identifiers attempt to solve the problems of resource identification and long-term access to online digital materials. A persistent identifier allows the resource to be uniquely identified in a way that will not change if the resource is renamed or relocated, and will persist regardless of the protocol used to access it. This means that a resource can be reliably referenced for future access by humans and software.

An important part of persistence is organisational policy, not just the adoption of technical solutions. In essence, the persistence of an identifier is only effective if a Naming Authority maintains and manages the persistence of the association between the identifier and the object identified and, at the same time, the content provider has to guarantee the stability and continuous access to the object identified. It is clear that the persistence cannot be self claimed and it needs a formal commitment between the institution that works as a Naming Authority and the community served.

Different systems in use

Examples of persistent identifiers in use include ARK, DOI, Handle, ISBN, ISSN, PURL, URI, URL and URN. Within the repository environment a common subset of identifiers are often used, the most relevant of which are:

- **URI** (Uniform Resource Identifier). Provides a standard, unique method for identifying a resource. It consists of a scheme name (http), a domain name (www.example.org) and a path (/example.htm). URLs (which specify a resource location) and URNs (which specify the resource name independent of location) are both forms of URIs. Many persistent identification systems are based on or extend the URI framework.

- **URN** (Uniform Resource Names) are intended to serve as persistent, location-independent resource identifiers and are designed to make it easy to map other namespaces (that share the properties of URNs) into URN-space. Therefore, the URN syntax provides a means to encode character data in a form that can be sent in existing protocols, transcribed on most keyboards, etc.

- **PURL** (Persistent Uniform Resource Locator). Takes the concept of the URL and adds a resolution service context. Instead of the URL pointing directly to the resource in question, it references an intermediate PURL resolution service. This service is used to look up the actual address of the URL before redirecting to the appropriate resource.

- **Handle**. Identifies the resource address by a unique handle assigned by a common registration service. When the browser receives a handle, it is sent to the global registration service for resolution and redirection to a local handle server which in turn

can resolve the local part of the identifier to the resource in the repository.

- **DOI** (Digital Object Identifier). Makes use of the Handle system as a resolution service but gives additional services such as descriptive metadata about the digital object that is being represented. DOIs are mainly used by publishers to identify journal articles, but have also been used by repositories to identify documents or data.

Using PIs with digital repositories

Persistent identifiers are widely used in digital repositories, providing valued-added services to a user: guaranteeing globally unique identifiers through a standard namespace adoption, a persistent relationship between an identifier and the same resource over time, and reliable services for managing and using these identifiers over time. Making use of this infrastructure will provide a foundation for utilities such as: Access to resources over time. Identifier infrastructure can support access to resources as they are moved from one repository or data store to another.

- **Discovery:** Persistent identifiers provide the link between the discovery of resources from federated repositories and access to those resources. The metadata returned from a discovery service can contain an identifier that is resolved to the location of the resource at the time of discovery.

- **Enhanced resource management**: Identifiers can be used to model complex relationships between resources and make it possible to manage compound resources and distributed copies of a resource, to distinguish between different versions of a resource, and to help

decide which is the most appropriate copy or version for a user.

- **Rights management:** Identifiers can be used to associate rights information with resources. They can also be used to track the relationship between a resource and derived works.

How can a repository administrator make use of a global identifier system?

Most repository platforms have been designed for persistence, in several cases allowing implementation of these services with minimal additional resource (an example is the Handle System included in the DSpace repository software). This might involve registering with a resolution service to receive a sub-namespace. Once this sub-namespace has been received, the repository software can be configured appropriately by the system administrator. The identifiers vary in particular for their scope, trustworthiness, authority and business model. Each institution must define its identification requirements before choosing a solution.

Conclusion

Persistent identification plays a key role in ensuring long-term access to objects within repositories, but it doesn't happen automatically. Persistent access needs to be backed by institutional policy, and can be assisted by a number of technical solutions. Remember, don't lose your audience to 404 errors, by ensuring works in the repository have identifiers that persist through time.

References & Further information

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

Catalogue of criteria for assessing the trustworthiness of PI systems (nestor)
http://nbn-resolving.de/urn:nbn:de:0008-20080710227

Briefing Paper updated by Emanuel Bellini, Foundazione Rinascimento Digitale

Repositories Support Project http://www.rsp.ac.uk/

The Repositories Support Project (RSP) aims to co-ordinate and deliver good practice and practical advice to HEIs to enable the implementation, management and development of digital institutional repositories