PersID: BUILDING A PERSISTENT IDENTIFIER INFRASTRUCTURE
Adrian Price

Introduction:

Over the last year 6 European countries – Denmark, Germany, Italy, Netherlands and Sweden – through 10 organisations, have been engaged in building a persistent identifier infrastructure for the scientific community: PersID. The PersID[1] initiative started under the auspices of Knowledge Exchange[2], but has now expanded to embrace other partners as well[3]. The project lead is being undertaken by the SURF foundation in the Netherlands. The purpose of the project is to enable institutions engaged in providing search and access services for any type of information resource (publications and parts of publications, photographs, video, maps, recordings etc) to name, reference and locate these objects with an emphasis on permanence, i.e. through policies and mechanisms which aim at persistent identification.

PersID organisations in participating countries will utilise existing services or establish new services, which will create the means whereby information resources of all types can be named and located in standardised ways. Most of the participating countries already have mechanisms in place on which they can build organisations which will be able to participate in PersID.

Although a persistent identifier does not in itself by any means ensure persistent access, it is a necessary step along the way to ensuring a "shelf mark" for distributed internet resources, to aid in long-term preservation, and a mechanism by which resources can be related to one another in a stable way. In this way, resources can enrich each other for the benefit of users who wish to utilise these connections.

This article will not describe the technical or administrative details behind persistent identifiers. For an overview of the field and the challenges involved, see for example: http://nbn-resolving.de/urn:nbn:de:gbv:7-isbn-90-6984-508-3-8.

What is the basis of PersID?

PersID builds largely on already existing persistent identifier services offered by several of the participating countries. These services have all chosen to use the URN (Uniform Resource Name) as a means of naming resources[4], and hence PersID has also chosen to embrace the URN standard. Put briefly: each member organisation will use the URN format as a means of naming (thus identifying) a resource, which is independent of where it is located. PersID and the local services will then be used to locate the whereabouts of the identified resources, just as one today locates a resource through a URL (Uniform Resource Locator) and the DNS. The difference being, that a URL solely locates a resource while a URN identifies a resource and PersID will be used to locate them. The next step will be to identify/locate on a timescale which resembles permanence, and ensuring access through activities of preservation. That though will be the next step for individual partners in the initiative.

PersID has chosen the URN system. Those partners who already have identifier mechanisms in place, all use the URN out of several identifier systems available: DOI's[5] are well known in the publishing sector and the Handle[6] system (which DOI utilises) is well known in, for example, the repository sector. Other identifier systems include ARK[7] and PURL[8].

As the project partners are centred around national libraries, the focus will be on using the NBN (National Bibliography Number)[9] URN namespace, which is under the control of national libraries.

The URN is by no means a new invention. Collectively, millions of URNs have already been issued to information resources by the PersID organisations and other URN services, and the time has arrived to further develop these services to become an important part of the information infrastructure.

---

[1] PersID: http://www.persid.org/
[2] Knowledge Exchange: http://www.knowledge-exchange.info/
[3] For a list of PersID partners and contact people see: http://www.persid.org/partners.html

[4] The basis of a standardised URN is: http://www.ietf.org/rfc/rfc1737.txt and syntax: http://www.ietf.org/rfc/rfc2141.txt
[5] DOI: http://www.doi.org/
[6] Handle: http://www.handle.net/
[7] ARK: http://www.cdlib.org/inside/diglib/ark/
[8] PURL: http://purl.oclc.org/
[9] NBN: http://www.ietf.org/rfc/rfc3188.txt

The PersID "meta-resolver" and the PersID local resolvers:

At the core of PersID will be a central so-called "resolving service", a server which will allow the input of a "URN name" and which will thereafter enable users to locate resources on the internet. This is the PersID "meta-resolver", which will be a service established and maintained by the PersID partners.

The "meta-resolver" will enable a user, for example a researcher at a university in search of a referenced article, to enter into the "meta-resolver" search field, the URN of the required resource and be linked directly to the pdf version of the article, if it is known by the "meta-resolver". "Is known" by the "meta-resolver" means here, that the "meta-resolver" is able to redirect the URN of the resource via a URL to the server which can provide access to the pdf, which is being sought. Alternatively, the "meta-resolver" will be able to return metadata associated with the resource (if available). It is being decided at the moment what level of metadata should be presented by local services, but the main focus is on providing access to the resource itself.

These resources can be offered by one of three types of services, as they are seen by the PersID "meta-resolver":

1. Trusted URN identifier services, where the services are known and registered by PersID and which comply with a set of policies regarding permanence.

2. Other trusted identifier services using other identifier services and which do explicitly comply with policies regarding permanence.

3. Other services which do not explicitly comply with PersID policies and thereby it is not certain what will be returned by these services.

In case 1, the "meta-resolver" will delegate the resolving of URNs to the local servers of PersID partners and others who comply with policies regarding permanence. In case 2, the "meta-resolver" will delegate the resolving of other identifier systems to the (known and trusted) resolver services of these identifier systems. And in case 3, the "meta-resolver" will redirect identifiers to the servers the "meta-resolver" deems to be best suited at the time - and with the necessary reservations.

Case 2, where the "meta-resolver" will be able to accommodate the resolving of other identifiers than URNs, illustrates the explicit policy of PersID not to be "URN centric". PersID realises that there will NEVER be one identifier system in use. The element of overriding importance is that the custodians of information resources introduce mechanisms whereby these resources can be permanently identified, be given access to and can be used to relate resources to one another in a secure manner. These "mechanisms" will be many and infrastructures built to accommodate them should be built to also accommodate the others in existence.

How far is PersID now?

Early in 2011 the "meta-resolver" should be an active service and those partners who comply with PersID policies and services, should be accessible through the PersID "meta-resolver". The details of PersID policies are at the moment being discussed and formulated.

It is also an important element that the PersID service should be able to continue after the project period has ended. This means that some kind of organisation must be built around the service which will enable the present and new partners to continue developing the service. We expect that the details of this organisation will be in place during early 2011.

The near future for proper identification and connection:

Identifiers hold the key which will enable information resources to be related to one another in a stable and (thus) meaningful way. This is what the basis of the "semantic web" is all about, or what has also been described as "Web 3.0" – which, judging by the example of Web 2.0, it might be a good thing to be hip about! Of course, Google is God, but there is very little intelligence in what it finds, it is not able to find and connect together resources which are directly (semantically) related to one another. These relationships have to be encoded by people or advanced machines, and at their basis will have to be persistent and unambiguous resource identifiers, be they documents (of all types), people or organisations. Thus, "Web 3.0" will be well served by identifiers!

Contact PersID:

PersID is by no means isolated to the institutions already mentioned. The partners are interested in expanding to include other countries and institutions, as a way of enabling better user access to information resources through standardised means for a far wider group of users. Go to the PersID website and contact either a PersID contact person in an individual country or the PersID initiative as such. We think that resource naming and identification is an important issue and welcome comments and questions - and new partners. PersID the project can be reached at: info [at] persid.org

Adrian Price, ap@life.ku.dk. Information Specialist, Co-ordinator, Research Documentation Section, The Faculty of Life Sciences Library, University of Copenhagen