digital preservation europe

Persistent Identifiers systems in the Public Administration sector

An identifier is any label that allows us to find a resource. Identity card numbers; fingerprints; phone numbers; street addresses; proper names; geographical coordinates; international standard book numbers (ISBN) - all these are identifiers. Here we are mainly concerned with identifiers for digital objects created, modified and made accessible in the course of public administrations activities.

On the Internet the most widely known identifier is the Uniform Resource Locator (URL), which allows users to find a resource by listing a protocol, domain name and, in many cases, file location.

A persistent identifier is, as the name suggests, an identifier that exists for a very long time. It should at the very least be globally unique and be used as a reference to the resource beyond the resource's lifetime. URLs, although useful, are not very persistent. They only provide a link to the resource's location at the moment in time they are cited, if the resource moves they no longer apply.

Persistent identification of a resource is critical for assuring the reliability of its retrieval; the validity of an information system and its trust depend on having its resources reliably retrieved and therefore from their persistent identification.

Persistent Identifiers in Public Administrations

eGovernment has the potential to save European taxpayers 15 billion Euros a year (1). eGovernment-powered services are becoming faster and more personalized and may also strengthen democracy by improving two-way communication between the citizens and their government. The main dimensions by which eGovernment systems are evaluated include trust, transparency, cost, delivery and uptake (2).

Public sector information (PSI) is an extraordinary resource for the EU's digital content industry - the overall market size for public sector information in the EU is estimated at € 27 billion. Despite its economic value, much of Europe's PSI is not fully re-used. The European Commission completed in November 2008 a consultation in view of its upcoming review of the 2003 PSI EU directive to overcome barriers that limit the re-use of PSI. The stakeholders expressed the need for technical measures introducing metadata and identifier standards for facilitating improved discovery of services and developing asset registries (3) (4). The top recommendation in the consultation of the WG on Information Management Standards and Data Quality of the ePSIplus network is that "PSI assets should be identified with an international standard persistent identifier" (5).

The US federal administration CENDI organisation posit that persistent identification is essential for providing e-government services and information. Without such persistence citizens who try to access government services from outdated references will receive annoying messages rather than valuable government information and services; this will frustrate citizens, increase maintenance costs and potentially result in the withholding of services or information from the public (9).

The CEN/ISSS eGovernment Focus Group identify three classes of standards (Semantic, Technical and Process standards). It classify identifier scheme among the Semantic standards that are concerned with the meaning of entities; they provide mechanisms to identify people, organizations, documents and other types of objects. The report was presented at the inaugural meeting (2/2008) of the CEN/ISSS workshop on Sharing eGovernment Resources whose agenda includes common agreed approaches to facilitate the discovery of, query of, access to, and federation of eGovernment resources (10).

Persistent Identifiers as part of the Trust infrastructure of eGovernment

Persistent identifiers are a key component that must be addressed in any strategy for long term preservation of digital resources. (7). Placement of trust allows actions that otherwise are not possible, but there is an element of risk until "proof" by experience provides the parties a more solid basis. Initial establishment of trust therefore utilizes "icons of conviction" or symbols that induce risk taking. Trust is an action that involves the placement of resources (financial, intellectual, informational, etc) under control of a trusted party while there is no "provable" commitment. In the domain of eGovernment, trust is not wholly defined by the security of the communication technology or by the individual conception of trust engendered through education and socialization. However in the "information society" development of trust can combine these two factors to assist in the making of educated decisions in situations typically characterized by uncertainty (6). We may conclude that the use of persistent identifiers in a public administration environment will have both instrumental and iconic roles in developing trust in the system.

Further information and resources

1. Culture and Society:: eGovernment > Overview - **Putting citizens first**
http://ec.europa.eu/information_society/tl/soccul/egov/index_en.htm

2. **eGovernment 'all about trust and transparency'**
http://www.euractiv.com/en/infosociety/egovernment-trust-transparency/article-166456

3. **Organisational change for citizen-centric eGovernment: Issues, Policy and Strategy** May 2008.Includes A Handbook for Citizen-Centric eGovernment.
http://www.ccegov.eu/Downloads/cceGov%20Think%20Papers%20and%20Handbook%20-%202008.pdf

4. Results of the Online Consultation of stakeholders "**Review of the PSI Directive**"
http://ec.europa.eu/information_society/policy/psi/docs/pdfs/online_consultation/report_psi_online_consultaion_stakeholders.pdf

5. ePSIplus Conference: Conference Proceedings and **Recommendations to the Commission Consultation on the Review of the PSI Directive**
http://www.epsiplus.net/events/epsiplus_conference_psi_re_use_who_takes_action_next/conference_proceedings_report

6. Frank Wilson, **Trust and Identity in Interactive Services**: Technical and Societal Challenges. cc:eGov Organisational change for citizen-centric eGovernment – Think Paper 11. ECOTEC.
http://www.ccegov.eu/Downloads/Paper%2011%20-%20Trust%20and%20identity%20in%20interactive.pdf

7. Jens Ludwig and Markus Enders. **Trustworthiness and Interoperability of Persistent Identifiers and Resolvers**. Nestor Workshop (Network of Expertise in Digital Long-Term Preservation), November 2007.
http://nestor.sub.uni-goettingen.de/pi_2007/index.php?show=programm

8. PADI: **Preserving Access to Digital Information, Persistent Identifiers.** National Library of Australia.
http://www.nla.gov.au/padi/topics/36.html

9. **Persistent Identification: A Key Component of an e-Government Infrastructure**. CENDI Persistent Identification Task Group (US federal government)

10. Draft Final Report of the Project Team of the **CEN/ISSS eGovernment Focus Group**, CEN, 2008 http://www.egovpt.org/fg/Report

Sources on Persistent Identification

1. Emma Tonkin. **Persistent Identifiers: Considering the Options**. Ariadne Issue 56, 30-July-2008
http://www.ariadne.ac.uk/issue56/tonkin/

2. **PILIN** Guideline Identifier Association Guidelines.
http://resolver.net.au/hdl/102.100.272/WBNMH9DQH

3. Examples of a Persistent Identifiers Resolver at the German National Library http://www.persistent-identifier.de/english/204-examples.php

4. Persistent Identifier Scheme for digital collections at the National Library of Australia
http://www.nla.gov.au/initiatives/nlapi.html

Benefit of Persistent Identifiers

Globally unique identification can help to reduce confusion over multiple versions of a given resource. They can help to improve the ease of locatability of distributed resources thereby facilitating access and re-use of resources for new research. Identifier strategies can integrate legacy naming systems and promote interoperability.

Long-term value: The application of identifiers may indicate a level of commitment on the part of the creating organisation. This can have a positive impact on the levels of trust towards that institution. Identifiers may help to provide provenance information which can positively impact the authenticity of a resource over time.

Persistent Identifier Approaches

There are several commonly used persistent identifier approaches that are described in another DPE briefing paper:
http://www.digitalpreservationeurope.eu/publications/briefs/persistent_identifiers.pdf

Resolver Service

A persistent identifier system separates the resources name from the resource location. A resolver system translates names (identifiers) into locations. A resolver service redirects a request for a digital object to the current storage location of the object. The resolver parses the persistent identifier and, based on a set of rules held in a configuration file, determines a destination URI to redirect the user to. The destination URI may be a file system location or an address of an appropriate application based on the type of object being delivered.

Persistent identification systems can associate extra informatin with the persistent identifer of the digital object (PID): the URLs of copies of the object, metadata, authentication system information and more.

Uniform Resource Name (URN) (8)

A Uniform Resource Name (URN) is a standard, persistent and unique identifier for digital resources on the Internet.

All URNs will include a Namespace Identifier (NID) code and a Namespace Specific String (NSS). The NID indicates the identification system being used for the URN and facilitates the interpretation of the NSS. The NSS is the local code that identifies the individual document.

The international ISBN and ISSN agencies are registering URNs using 'ISBN' and 'ISSN' as the NIDs.

Other persistent identifier approaches include **the Handle System; DOI** – Digital Object Identifier**; PURL** – Persistent URL**; ARK** - Archival Resource Key; **XRI** – OASIS eXtensible Resource Identifier; **N2T** – Name to thing.

Dov Winer, *Israel National Library*
dov.winer@gmail.com