BRIEFING PAPER October 2006

DCC

Persistent Identifiers

The persistent identification of digital resources can play a vital role in enabling their accessibility and re-usability over time. However, progress in defining the nature and functional requirements for identifier systems is hindered by a lack of shared agreement on what identifiers should actually do. To some, an identifier system is strictly a means of providing a unique name to a digital or analogue resource — either globally or locally. To others, identifier systems must also incorporate associated services such as resolution and metadata binding. Specific requirements will differ, but it is vital that institutions recognise that the application and maintenance of identifiers forms just one part of an overall digital preservation strategy. Without adequate institutional commitment and clearly defined roles and responsibilities, identifiers cannot offer any guarantees of persistence, locatability, or actionability in the long or short-term.

Short-term Benefits and Long-term Value

Short-term benefits:

- Globally unique identification can help to reduce confusion over multiple versions of a given resource
- Persistent identifiers 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

HE/FE Perspective

"At the end of the day, the only guarantee of the usefulness and persistence of identifier systems is the commitment of the organisation which assign, manage and resolve identifiers."
***Stuart Weibel**, Senior Research Scientist, OCLC*

e-Science Perspective

"Shortcomings in scientific data provision and inter-disciplinary use have been identified. Data sources are widely unknown and data are often archived without context. The suggested solution is publication of primary data together with long-term archiving and context documentation at scientific community level. The use of persistent identifiers (DOI/URN) together with meta information for citing electronic publications (ISO 690-2) allow for persistent access within the Internet and for integration of scientific primary data into standard library catalogues. Primary data are then searchable together with scientific articles and can be cited within literature like other publications."
***Michael Lautenschlager,** Director, World Data Center for Climate at Model and Data/Max-Planck-Institute for Meteorology*

Issues to be Considered

Functional requirements will differ depending on the individual institution, but the following questions should be carefully considered prior to the implementation of any identification strategy:

- What should the identifier be identifying — the resource, the location, the metadata, or all of the above?
- Does the identifier need to be globally or locally unique?
- What basic functionality is required of the identifier scheme (identification vs retrievability)?
- What level of granularity is required?
- Are there legacy naming systems that need to be incorporated? If so, how will interoperability between naming systems be handled?
- Will opaque or semantic identifiers be used?
- Versioning can be problematic. When does a resource change significantly enough to warrant the application of a new identifier?
- How will metadata be stored and bound to the identified resource?
- Can the identification strategy scale to meet future needs?
- At what stage in the workflow will identifiers be applied to a resource?
- Who will be responsible for managing the identifiers over time?
- How will the assignment and long-term management of identifiers be financed?

Additional Resources

- NISO Identifiers Roundtable [http://www.niso.org/news/events_workshops/ID-06-wkshp.html] and report [http://www.niso.org/news/events_workshops/ID-workshop-Report2006725.pdf]
- Proceedings from the DCC workshop on Persistent Identifiers. [http://www.dcc.ac.uk/events/pi-2005/]
- Proceedings from the ERPANET workshop on Persistent Identifiers [http://www.erpanet.org/events/2004/cork/index.php] and final report [http://www.erpanet.org/events/2004/cork/Cork%20Report.pdf]
- eBank UK Data Citation Links [http://www.ukoln.ac.uk/projects/ebank-uk/data-citation/]
- PADI [http://www.nla.gov.au/padi/topics/36.html]
- NOIDs [http://www.cdlib.org/inside/diglib/ark/noid.pdf]
- InfoURI [http://info-uri.info/registry/OAIHandler?verb=ListRecords&metadataPrefix=oai_dc]
- Digital Object Identifier (DOI) [http://www.doi.org/]
- Persistent Uniform Resource Locator (PURL) [http://purl.org/]
- Archival Resource Key (ARK) [http://www.cdlib.org/inside/diglib/ark/]
- Handle System [http://www.handle.net/]