ERPANET Training Seminar on Persistent Identifiers

June 17-18, 2004
ERPANET Training Seminar on Persistent Identifiers

BRIEFING PAPER
Introduction

Efficient and permanent access to digital information is crucial for government bodies, businesses, cultural heritage professionals, scholars and researchers to make key decisions and to generate new knowledge. While simple uniform resource locators (URLS) have largely been used to identify and access these digital resources to date, they are by no means reliable or persistent. URLS can be altered for any number of reasons such as changes to organisations and staff, to file structures and to web site content or navigation. As digital resources are cited more and more frequently in scholarly research papers, business and government reports it becomes even more pertinent that continued access to these resource can be guaranteed. Any loss of access can result in diminished trust and confidence in the organisation that created or cited these digital resources. While persistent identifiers can help to ensure that any organisations’ digital resources are permanently locatable, unless their application forms part of an overall digital preservation strategy there is no guarantee that once located the digital resource will be actionable.

The Seminar

This two-day training seminar will provide participants with an overview of four major persistent identification strategies: the Digital Object Identifier (DOI), the Persistent Uniform Resource Locator (PURL), the Archival Resource Key (ARK) and the Handle System. In addition, the seminar will present information on related initiatives including the OpenURL, InfoURI and Uniform Resource Name (URN). Speakers will share their experiences with the application of persistent identifiers from a range of perspectives including the cultural heritage, publishing, government and scientific sectors. Two breakout sessions will enable the participants to explore the issues raised during the seminar in a more in-depth manner and to share their own experiences, interests and concerns.

Seminar Topics

- Introduction to the topic and related issues
  The first day will focus on providing some background information on the development of persistent identifiers and will explore some of the issues related to the persistent identification of digital resources. An overview of the Uniform Resource Name (URN), OpenURL and InfoURI will be presented and will be followed by an update from the Dublin Core Metadata Initiative working group on persistent identifiers.
  
- Overview of current strategies
  Overviews of four major persistent identification strategies will be presented. The overviews will cover issues such as how the strategy works; its main
features; the costs associated with implementation and on-going maintenance; and how the system helps to ensure the long-term access to and digital resources.

- **Breakout session 1**
  This session will allow the participants to explore and discuss the features of the persistent identification strategies and related issues more implicitly.

- **Reports from a range of discipline on experiences with persistent identifiers**
  Day two will focus on sharing experiences on the implementation and research into persistent identifiers from the cultural heritage, publishing, government and scientific sectors.

- **Breakout session 2**
  This session will focus on some key questions regarding the implementation of persistent identifiers and their overall role in a digital preservation strategy. Participants will be asked to consider specific roles and responsibilities with regards to the application and maintenance of persistent identifiers.

**Seminar Challenges**

During the course of the seminar, participants will be able to explore the following issues.

- What are persistent identifiers and why are they important for my digital resources?
- How do the URN, OpenURL and InfoURI work with persistent identifier strategies?
- How do the major strategies preserve access to digital resources?
- What are the cultural heritage, publishing, government and scientific sectors doing with regards to persistent identifiers?
- How can persistent identifiers fit in with a digital preservation strategy?

This seminar has been designed to benefit participants from a wide range of backgrounds and will be of interest to anyone considering the long-term accessibility and reusability of their digital resources.
Programme

Thursday, June 18

09:00 – 09:15  **Opening Remarks**  
Seamus Ross (ERPANET)  
Carol Quinn (University College Cork)

09:15 – 09:45  **Keynote Address**  
Norman Paskin, International DOI Foundation (IDF)

09:45 – 11:15  **Panel Discussions: Related Issues**  
- **Overview of URN**, Kathrin Schroeder, (Die Deutsche Bibliothek)  
- **Overview of OpenURL**, Tony Hammond, (Nature Publishing Group)  
- **Overview of InfoURI**, Stu Weibel, (OCLC)  
- **Update from the DCMI working group on persistent identifiers**,  
  Robin Wilson (TSO) and Stu Weibel (OCLC)

11:45 – 12:15  **Overview of ARK**, Mary Heath (California Digital Library)

12:15 – 12:45  **Overview of PURL**, Stuart Weibel, (OCLC)

14:00 – 14:30  **Overview of Handle System**, Larry Lannom, (CNRI)

14:30 – 15:00  **Overview of Digital Object Identifiers (DOI)**, Norman Paskin, (IDF)

15:30 – 17:00  **Breakout session**

17:00 – 17:30  **Report back**

Friday, June 19

9:00 – 10:00  **EPICUR Project**, Kathrin Schroeder, (Die Deutsche Bibliothek)  
**National Digital Data Archive of Hungary**, Andras Micsik, Hungarian Academy of Sciences

10:00 - 10:30  **DiVA Project**, Eva Muller (Uppsala University)

10:30-11:00  **Publishing perspective**, Cliff Morgan  
(John Wiley and Sons, Ltd.)

11:30 – 12:00  **JISC Report: Digital Object Identifiers for Publishers and the e-Learning Community**, Robin Wilson, (TSO)

13:30 - 14:00  **Scientific perspective**, Michael Lautenschlager (WDCC)

14:00 – 14:30  **Government perspective – Canada**, Cecil Somerton, (Treasury Board of Canada)

15:00 – 16:30  **Breakout session**

16:30 – 17:00  **Wrap up and conclusions**
Requirements for digital object identifiers as drafted by The Stationary Office (TSO)

http://www.tsoid.co.uk/Information.aspx?PageID=Information

Content is generally not actionable i.e. it remains a static resource which resides in a location (or often multiple locations) and is viewed as a collection of standalone entities. In order to become useful, content should be a living, breathing, evolving collection of resources, each of which is associated with the additional information necessary to invoke actions and reactions in users, systems and other resources.

Resources, Locators and Identifiers
Identification is a fundamental concept in computing. Operating systems and programs have to keep track of data and objects which reside in memory or file systems. They achieve this by "pointing" at content, allowing it to be located and displayed on demand.

Digital identifiers need to fulfill a more sophisticated role than just pointing to a location. They must tell us more about the resource we're looking for than where it resides (or used to reside).

The Four Pillars of Information Assurance
While the concept of information assurance may hint at security, this is only one element of the overall mix. Information assurance is more about maintaining a holistic approach to information management, ensuring the integrity, authentication and confidentiality of the content that an organisation creates, holds or disseminates. It has major implications for a number of areas, including corporate governance and risk management – but its wider role is to support the information management strategy as a whole. Four key factors underpin these ideas and are crucial if organisations are to have full confidence in their information assurance.

Global
Nowadays, information and content must be frequently shared outside the organisation that created it. Digital communities can be regional or global in nature. The management of content therefore needs to reflect this. Assets must be registered globally, with a worldwide guarantee that asset X is exactly what it is supposed to be, and always will be.

Uniqueness
One of the major problems with information is that it's difficult to find exactly what you want. Too often there are multiple iterations of documents, information is repurposed or amended or its location is changed. A unique number assigned to a piece of content guarantees that the information found is exactly what is sought.

Resolution
For content to be useful it is advantageous that it is a) connected to something and b) actionable. Only then does it become a useful resource. Resources should not be viewed in isolation, but as parts of an overall picture.

Persistence
The persistence of content is crucial in that it moves around and can therefore be difficult to locate. Any service facility for information management should offer guarantees that required content is kept and that – wherever it may reside – users can find exactly what they want and can access it easily.