

*erpa*Seminar

Persistent Identifiers

University College Cork, Ireland

June 17-18, 2004

FINAL REPORT

Persistent Identifiers

*Cork, Ireland
17-18 June 2004*



Table of Contents

Table of Contents	3
Executive Summary	4
Introduction.....	5
Seminar Setting.....	5
Aims and objectives	5
Keynote Address: The Development of Persistent Identifiers	7
Overview of the Uniform Resource Name (URN)	8
OpenURL – The Rough Guide.....	8
An Overview of InfoURI.....	9
DCMI Persistent Identifier Working Group Session Abstract.....	10
Update from the CENDI working group on persistent identifiers	11
Overview of the Archival Resource Key (ARK): Establishing Persistence for Digital Collections.....	12
Overview of Persistent Uniform Resource Locators (PURLs)	13
Overview of the Digital Object Identifiers (DOI) System	15
The Implementation of Persistent Identifiers: Case studies	17
The Library Perspective – Die Deutsch Bibliothek, Hungarian Academy of Sciences (MTA), the National Széchényi Library and Uppsala University Library	17
The Publishing Perspective – John Wiley & Sons Ltd.	19
The Higher Education Perspective – JISC Report on Digital Object identifiers for the Publishing and e-Learning Community, The Stationary Office (TSO)	20
The Scientific Perspective: World Data Centre for Climate (WDCC).....	21
The Government Perspective: Persistence and Access to Information in the Government of Canada.....	21
Conclusions	23
Appendix One: Seminar Programme.....	25
Appendix Three – Additional Seminar Reports.....	28
Appendix Four – Sponsoring Organisations.....	29

Executive Summary

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 web-based 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 web-based 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 resources can be guaranteed. Any loss of access could result in diminished trust and confidence in the organisation that created or cited the digital resources. While persistent identifiers can help to ensure that an organisation's digital resources are permanently locatable in theory, unless their application forms part of an overall digital preservation strategy with the necessary levels of commitment, there can be no guarantee that once located the digital resource will be actionable.

This two day training seminar provided 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 presented information on related naming schemes including OpenURL, InfoURI and Uniform Resource Name (URN). Speakers shared 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 enabled 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.

Many of the experts were surprised to learn that the Uniform Resource Name: National Bibliographic Number (URN:NBN) is being widely used among several European libraries. Improved communication between R&D and other stakeholders is something that all of the participants feel would be beneficial for ensuring interoperability with naming schemes across disciplines and nations.

Throughout the two days, it became clear that the publishing sector has progressed the use and management of persistent identifiers more than any other discipline. As such, the speakers and participants agreed that more work is needed to identify user-requirements for a wider range of disciplines including government, scientific research, cultural heritage and education. It was suggested that ERPANET and the new JISC-funded Digital Curation Centre might be best placed to coordinate further research in this area. Encouragingly, communications via the DC-Persistent Identifiers listserv have been very active since the workshop and many of the workshop participants have joined the debate.

Introduction

This international seminar examined current strategies and practical uses of persistent identifiers (PIs) among the cultural heritage, scientific, publishing and government communities and discussed how they can best be applied to the long-term management of and access to digital data.

Seminar Setting

The seminar was held at the University College Cork (UCC)¹ and attracted thirty-four participants from the United Kingdom, Austria, Canada, Ireland, Sweden, Hungary, Germany, Denmark and the United States. ERPANET is extremely grateful to the UCC and its President, Professor Gerard Wrixon for providing the venue for the seminar and for hosting an evening reception for the participants. Special thanks go to Michael Holland and Carol Quinn for their invaluable assistance in ensuring that this seminar was a success.

ERPANET would also like to thank MINERVA for funding the participation of some of the workshop participants from Hungary. Specific thanks go to David Dawson (MLA², MINERVA).

Finally, ERPANET would like to thank the Digital Curation Centre (DCC)³ for its assistance in promoting this event.

Aims and objectives

One of the major aims for this seminar was to provide an international forum to exchange information about the persistent identification of digital resources across different disciplinary, institutional, and national contexts. The objectives were to:

- Provide an overview of the major current strategies for persistent identification of electronic resources.
- Identify and discuss the key technical, management, and policy considerations for the successful implementation of persistent identifiers.
- Provide a networking opportunity for seminar participants to meet with other researchers, data managers, information specialists, archivists, and policy experts across disciplinary and national boundaries.

Day one of the seminar opened with words of welcome from ERPANET Coordinator Peter McKinney and University College Cork Archivist Carol Quinn. Following this, Norman Paskin, Director of the International DOI Foundation (IDF) delivered the keynote address to outline the history of persistent identifier development. The keynote address was followed by an overview of several naming schemes related to

¹ University College Cork (UCC) <http://www.ucc.ie>.

² Museums, Libraries and Archives Council (MLA) <http://www.mla.gov.uk/index.asp>;
MINERVA http://europa.eu.int/comm/education/programmes/socrates/minerva/ind1a_en.html.

³ Digital Curation Centre (DCC) <http://www.dcc.ac.uk>.

persistent identification including InfoURI, OpenURL and URN. Updates from the Dublin Core Metadata Initiative (DCMI) Working Group on Persistent Identifiers and CENDI's Working Group on Persistent Identifiers showed some of the current activity into PIs. These talks provided the participants with an understanding of the history of persistent identifiers, their relation to naming schemes and gave a flavour of some of the current research into their application. The seminar then moved on to examine some of the major naming strategies developed to facilitate the persistent identification of electronic resources. Detailed overviews were presented on Persistent Uniform Resource Locators (PURLs), Archival Resource Keys (ARKs), Handles, and Digital Object Identifiers (DOIs).

Day two focused on the practical use of PIs from a range of perspectives including the cultural heritage, higher and further education, publishing, scientific and government sectors. Breakout discussion groups and reporting sessions on both days allowed participants and speakers to delve more deeply into some of the issues raised during the presentations. The dialogue was lively and produced some very interesting ideas. This report will follow the structure of the seminar and provide a brief synopsis of the presentations and discussions that took place.

Keynote Address: The Development of Persistent Identifiers

The naming of unique electronic resources on the web has been confusing to say the least. Norman Paskin, Director of the International DOI Foundation⁴, warns that we must avoid falling into a word-trap. We must question what we mean by persistent and what we mean by identifier. The meanings for both will vary from person to person and from discipline to discipline. For instance, in some disciplines, persistent refers to a span of time well beyond the life of the actual resource itself. In other areas, ten to fifteen years is considered adequately persistent. As such, each organisation must be explicit about what level of persistency they require from the outset.

Equally vital is ensuring that we know what we are identifying. The identifier must do more than simply identify a given object. It must also describe the resource precisely to ensure that the specific object requested is always the item that is returned. This is problematic when there are a multitude of formats and versions in existence. The extent to which an object is described will depend largely on the needs of the assigner and their user community. Some may find that returning a 'bag of bits' is ample while others will require more intellectual content information. Granularity requirements will impact the choice of identifier that an organisation will implement.

Since the late 1980's, several naming schemes have been developed. Systems such as the Uniform Resource Name (URN), Persistent Uniform Resource Locators (PURLS), Uniform Resource Identifiers (URI) and Handles emerged during this time. The release of the Armati Report in 1995 provided a framework for the development of persistent identification naming schemes, including the Publisher Item Identifier (now known as the ISTC) and the Digital Object Identifier (DOI). The report listed requirements such as interoperability, persistence, as well as global scalability, uniqueness and meaning.

Paskin pointed to the automation of ISBN number application in the 1960's and the automation of resource sharing through the World Wide Web in the 1990's as major factors impacting interoperability. Indeed, the majority of participants agreed that any viable persistent identification system must enable interoperability. However, Paskin urged participants to remember that only a small number of all electronic resources are made available via the Internet. As such, PI naming schemes must be flexible enough to cover non-networked materials and work across a variety of standards and disciplines. Metadata will be vital to describe relationships between electronic resources – to describe versions, various formats and context – but Paskin emphasised that no single metadata scheme will fit all purposes. As such, the development of ontologies will be extremely important to enable existing metadata to be mapped to various schemas using data dictionaries.

Paskin concluded by reiterating some fundamental concepts that must be addressed before implementing any persistent identification system. We must understand what it is that we wish to identify and what level of granularity will be required. We must know what we want the persistent identifier to resolve to. We must consider what level of metadata we will make available with the persistent identifier and how this metadata will be interoperable with other schemas. And we must address how the costs of implementing the necessary infrastructure for persistent identification and

⁴ International DOI Foundation <http://www.doi.org>.

resolving will be met and sustained for the long-term. Are the costs associated with persistent identification the responsibility of the content creator, the registry agent or the end-user? As would become clear throughout the course of the event, the above concepts and other basic user requirements remain a key area to be addressed by a wide range of user communities. The majority of the speakers urged participants to work together to establish frameworks of basic requirements that could then push forward development within specific persistent identification strategies.

Overview of the Uniform Resource Name (URN)

Following the keynote address, Kathrin Schroeder of Die Deutsche Bibliothek⁵ offered a brief overview of the Uniform Resource Name (URN). The URN is an official Uniform Resource Identifier (URI) scheme and can be used to identify both abstract and physical resources. In addition, the URN can apply to whole objects or individual parts of an object (for example, an aggregate learning object). Based on the DNS system, the URN requires that a globally unique namespace be registered with Internet Assigned Numbers Authority (IANA)⁶. This action in itself implies a certain level of commitment on the part of the assigner. The URN is hierarchical in nature and the model allows for existing (legacy) naming schemes to be incorporated. The URN syntax involves three sections. The first is the URN itself which specifies the naming authority. Following the URN, the Namespace Identifier (NID) refers to an established naming scheme such as the ISBN. Completing the syntax is the Namespace Specific String (NISS) which identifies the specific item. The system requires a plug-in to resolve from the PI to the item and, as such, is not directly actionable. However, the system is widely used by the European Library community and to date, the Internet Engineering Task Force (IETF)⁷ has approved over twenty-one name space schemes.

OpenURL – The Rough Guide

Tony Hammond of Nature Publishing Group⁸ then provided participants with an overview of the OpenURL⁹. The OpenURL was developed as an extension of research carried out by the University of Ghent in 1999 to link heterogeneous collections within the institution. While not a persistent identification strategy in itself, the OpenURL works with PIs to provide users with access to extended services.

Unlike opaque URIs, OpenURLs are designed to be non-opaque and instantly recognisable. The OpenURL is an actionable URL that transports metadata about a requested object. As the OpenURL describes the resource rather than merely pointing to it, it offers the end user more control in determining exactly what is returned to them. In addition, the OpenURL can return information on both print and electronic resources. As mentioned previously, the ability to return information about

⁵ Die Deutsche Bibliothek <http://www.ddb.de>.

⁶ IANA <http://www.iana.org/>.

⁷ IETF <http://www.ietf.org/>.

⁸ Nature Publishing Group <http://www.nature.com>.

⁹ OpenURL <http://library.caltech.edu/openurl/>.

non-networked resources is considered to be an important requirement for persistent identifier strategies.

In February this year, JISC¹⁰ funded EDINA¹¹ and UKOLN¹² to research the creation of an OpenURL Router. This Router provides a central registry of institutions' OpenURL resolvers and aims to enable institutional resolvers to be more persistently available. There are currently more than one hundred installations of the OpenURL and an OpenURL Framework has been submitted as a draft standard¹³ to ANSI/NISO. The framework has two main deliverables: the first being a data model and the second being a registration mechanism to publish identifiers and metadata formats. This new framework will make the OpenURL more useable beyond the bibliographic context for which it was originally created. The standard is expected to be signed-off by ANSI/NISO in late 2004.

An Overview of InfoURI

Stuart Weibel from the Online Computer Library Centre (OCLC)¹⁴ further explored some of the issues highlighted during Norman Paskin's keynote address and presented information on the InfoURI scheme. Like Paskin, Weibel agreed that we need persistent identifiers for a multitude of reasons. Accordingly, it is highly unlikely that any one system could ever meet the needs of all users. Again, participants were urged to consider specific requirements that they might want from persistent identifiers. Weibel introduced a layered model of the various factors that could impact user requirements for persistent identifiers. The layers of the model represent social, business, policy, technology and functionality implications that must be considered prior to implementing any strategy. This 'layer cake' model could be a very useful starting point for any organisations looking into the implementation of a persistent identification system. Weibel also stressed the need to know what we want to identify and how long we need to identify it for. Here he emphasised that regardless of technological innovations and persistent identification systems, commitment on the part of the assigning organisation is by far the single greatest factor affecting of any resource's persistence.

Weibel then went on to describe the InfoURI¹⁵ scheme. InfoURI is being developed by the library and publishing communities to incorporate legacy identification systems within the web context regardless of the object's format. Therefore, digital and physical objects can be identified with InfoURIs. InfoURI only specifies the namespace authority. For example, a digital object identifier would appear as `info:doi/` under the InfoURI scheme. The rest of the identifier is strictly up to the naming organisation. As such, semantics are beginning to creep into InfoURIs. This could become problematic in the future.

Unlike URN, InfoURI is a naming scheme that focuses exclusively on identification. It does not offer resolution to the object directly or via a plug-in. Nor do they make any

¹⁰ The Joint Information Systems Committee (JISC) <http://www.jisc.ac.uk>.

¹¹ EDINA <http://www.edina.ac.uk>.

¹² UKOLN <http://www.ukoln.ac.uk>.

¹³ The OpenURL Framework for Context-Sensitive Services
http://library.caltech.edu/openurl/StandardDocuments/Z39_88_Pt1_ballot%20final.pdf.

¹⁴ Online Computer Library Centre (OCLC) <http://www.oclc.org>.

¹⁵ InfoURI <http://info-uri.info/registry/docs/misc/faq.html>.

guarantees regarding the persistence of the item itself. However, these are services that could evolve over time. At this point, Weibel urged participants to separate the concepts of identification and resolution. While an identifier can be actionable, it must be clearly understood that persistent identification does not necessarily imply resolution. Again, with more research into user requirements, tailored persistent identification systems can be developed to meet a wide range of user needs. InfoURI is currently in its formative stage and is the focus of much debate on the DCMI Persistent Identifier listserv.

DCMI Persistent Identifier Working Group Session Abstract

Stuart Weibel (OCLC¹⁶) and Robin Wilson (The Stationery Office¹⁷) provided an update from the DCMI Persistent Identifier Working Group (DCMI-P-ID)¹⁸. The group was formed to examine current PI strategies and to make recommendations for their use. The speakers emphasized that while the research being carried out by the DCMI working group looks at how PIs can enhance metadata, efforts must look beyond the Dublin Core. Indeed, the working group will collaborate with other international standards bodies to ensure that maximum benefits can be achieved. The group has been involved in the e-Government Envoy research being carried out in the UK¹⁹. While the DCMI working group has no objections to making recommendations regarding the use one or many PI strategies, the identification of a single strategy is not their aim. Rather, they are looking to gather use case studies and information on implementation so that they may offer guidance on a range of strategies. As such, the focus of the working group is of an informational rather than prescriptive nature.

In January 2004, the group distributed a survey to identify current use, research and implementation plans for persistent identification strategies. In addition to revealing that persistent identifiers have not yet earned the trust of most of the respondents, results also show that much more work must be done to identify user requirements across various domains. The survey will run through to the end of September and the results will be presented in a report at the DC-2004 conference to be held in Shanghai²⁰. The DCMI Working Group urged participants to become involved in making their requirements known – both from the decision-maker and the end-user point of view. Without this vital information, it will be virtually impossible to make appropriate recommendations regarding strategies for specific uses. To start with, we must at least be able to define what we want to identify with the application of PIs. For instance, we may want to identify a particular resource or perhaps just the metadata for the resource. Additionally, we may want to identify an aggregate of resources.

Specifically, the research being carried out by the DCMI Working Group aims to establish a broader understanding of PI functionality and to enable institutions to

¹⁶ Online Computer Library Centre (OCLC) <http://www.oclc.org>.

¹⁷ The Stationary Office (TSO) <http://www.tso.co.uk>.

¹⁸ DCMI Persistent Identifier Working Group <http://dublincore.org/groups/pid/index.shtml>.

¹⁹ E-Government Envoy <http://www.govtalk.gov.uk/egif/contentmment.asp#table5>.

²⁰ International Conference on Dublin Core and Metadata applications 2004
<http://dc2004.library.sh.cn/>.

make informed choices regarding the available strategies. The group also hopes to provide more definitive information regarding the potential costs involved with the application and maintenance of PIs as well as their potential benefits.

Weibel and Wilson concluded by stating that this seminar provided a great opportunity for experts and end-users to discuss specific requirements for persistent identifier systems. A working group listserv has been set up to enable interested parties to provide information on use cases or simply to engage in the ongoing discussions²¹.

Update from the CENDI working group on persistent identifiers

Larry Lannom, CNRI followed the report from the DCMI working group with a short description of the work being done on persistent identifiers by CENDI²². CENDI is an interagency cooperative organisation composed of scientific and technology information managers from several departments of the United States Government. Their mission is to develop more effective scientific and technological information support systems for the US Government. A task group was created to investigate how persistent identifiers might be used to improve the effectiveness of information support systems under the E-Government Act of 2002. The task group have found that persistent identifiers can help to ensure that the public have consistent, reliable and accurate access to government information, but only if their application forms part of an overall digital preservation framework. At the moment, the US Government uses PIs most commonly to manage long-term access to document-like objects but they will likely be used to help manage complex digital objects over time.

As was demonstrated throughout the course of this seminar, persistent identifiers alone are not enough to guarantee the long-term access to digital resources. Rather, they are one component of an overall digital preservation strategy. As with many other digital preservation initiatives, the main challenges hindering the large-scale uptake of persistent identifiers are social and economic rather than technical. As such, CENDI recommends that the application of PIs be incorporated into existing policies and guidelines. The development of a core set of persistent identification metadata to including elements for discovery, digital rights management, preservation and associated services for a range of disciplines will be necessary for the long-term management of and access to digital resources. CENDI believe that greater cooperation on national and international levels will be required for any successful mass implementation of a persistent identification strategy to be effective. As such, they have engaged in discussions on the potential government uses of persistent identifiers with the Stationary Office (TSO) and the E-Government Envoy in the UK. Like Weibel and Wilson, Larry Lannom also firmly believes that more work on the gathering of user requirements is urgently needed.

²¹ Archives of DC-PERSISTENT-IDENTIFIERS@JISCMail.AC.UK
<http://www.jiscmail.ac.uk/lists/DC-PERSISTENT-IDENTIFIERS.html>.

²² CENDI <http://www.dtic.mil/cendi/>.

Overview of the Archival Resource Key (ARK): Establishing Persistence for Digital Collections

Mary Heath then presented an overview of the Archival Resource Key (ARK)²³. She began by explaining that one of the major challenges she sees is guaranteeing the same longevity and persistence of access to digital collections as has been achieved with print collections. At California Digital Library (CDL)²⁴, persistence of access to digital resources for the University of California is seen as a key service issue. In addition, establishing a high level of trust and credibility among their user community is also of major importance. CDL developed the Archival Resource Key (ARK) as a way to both build credibility and ensure persistence. During her talk, Heath explored the development of ARKs, the services associated with them, and the ways in which ARKs can help establish credible persistent access. She also looked at related metadata issues particularly as they pertain to the establishment of a preservation repository for digital objects.

While acknowledging that the ARK is the new kid on the block, Heath believes that the ARK addresses some needs that are not met by other strategies. For example, in addition to being actionable and globally unique, ARKs provide end-users with a statement of organisational commitment on the part of the assigner. This is the biggest difference between the ARK and other strategies.

The basic concept for the ARK is as follows:

- The Naming Assigning Authority (NAA) assigns an ARK to a digital resource. The NAA could be library, archive, publisher or other organisation accepting stewardship of the resource. Each NAA is identified as a unique number by the registration agency. Object names assigned by NAAs can be complex and hierarchical in nature and can reveal variants.
- Then the Name Mapping Authority (NMA) resolves the ARK through a simple file lookup. NMAs are meant to be temporary, disposable and replaceable. However, NAAs and NMAs can be the same organisation.

ARK naming is based on a strict set of rules. The naming structure facilitates hierarchical naming, so that one ARK could point to a book, a chapter or a specific page. The structure also enables various formats to be identified in the same manner. ARKs provide three main services. The first is the access to the digital resource, the second is access to the metadata associated with the digital resource and the third is an organisational policy statement on the level of commitment that they accept for the long-term stewardship of the resource.

The costs associated with the implementation of this system are mainly centred on the creation of metadata and the development of a persistence policy. Costs are also incurred through the maintenance of necessary mapping tables. The exercise of drafting a persistence policy could be a valuable starting point for many organisations. By conducting an audit of what electronic resources they have, the organisation can decide what level of commitment they can guarantee to preserve access to these resources. This system is quite unique as it clearly requires

²³ Archival Resource Key (ARK) <http://www.cdlib.org/inside/diglib/ark/>.

²⁴ California Digital Library (CLD) <http://www.cdlib.org>.

philosophical consideration on the part of the organisation for the creation of a persistence strategy and commitment statements for each item being named. This combination of human commitment and technology could be exemplar for all existing and emerging PI strategies.

Overview of Persistent Uniform Resource Locators (PURLs)

Stuart Weibel again took the stage to give the participants an overview of the PURL²⁵. PURLs resulted from Uniform Resource Name (URN) working group meetings held in 1996. They were designed to be an interim solution to some unanswered questions. Basically, PURLs are URLs and they function using the basic redirect facility. As such, they work in all of the browsers currently being used and require no special plug-ins or tools.

To date, URLs have performed two major functions on the Internet. The first is naming the resource and the second is locating the resource. Working with this underlying premise, PURLs map a symbolic identifier to a network location. Basically, a simple table with two columns is used to record both the symbolic identifier and the network location of the actual resource. The table is behind the scenes, so the end-user is unaware of this redirect function. Organisations are not required to run their own PURL server and the open-source PURL software is available for download free of charge. Additionally, the software can be freely modified by the organisation to meet their specific needs.

Ultimately, PURLs are managed URLs. Their implementation does not require any adherence to policies. In fact, there is no requirement to use the term PURL in the actual identifier, however Weibel recommends this practice to provide transparency. This flexibility could be seen either as a positive or a negative depending on the organisation in question. One major drawback of the system is that it offers strictly a one-to-one rather than one-to-many resolution. As such, it offers no granularity. Again, whether this is a positive or a negative aspect depends completely on each organisation's requirements.

While PURLs can guarantee persistency of the name, they cannot guarantee perpetual access to the resource. However, it is important to note here that no naming scheme can actually guarantee perpetual access unless there is sufficient commitment on the part of the naming organisation.

PURLs offer tools for the management of digital networked resources and namespaces. However, Weibel insists that while technology can assist in the persistent identification and resolution of digital resources, the problem is fundamentally a social one. Many of the workshop participants found the flexible and customisable aspect of PURL software to be of interest. In fact, the level of interest in PURLs was surprising to Weibel and reinforced his belief that better communication was necessary between user communities and persistent identifier developers.

²⁵ <http://www.purl.org>.

Overview of the Handle System

Larry Lannom, from CNRI²⁶ then went on to provide an overview of the Handle System²⁷. The Handle System is a 'general-purpose global name service that allows secured name resolution and administration over networks such as the Internet'²⁸. Handles are unique names for digital objects and networked resources. The Handle System protocol defines its own resolution and administration protocol directly on top of the network layer. It can be used within the context of the web but is not dependent on it nor is it dependent on DNS.

The Handle System consists of the Global Handle Registry (GHR), which operates as the root of the Handle System, and many Local Handle Services (LHS's). A resolution request for a specific handle, with no caches involved, would start at the GHR, which responds with LHS information identifying the appropriate LHS to go to for final resolution of the given handle. Each Handle Service, including Global, consists of one or more Sites and each Site consists of one or more servers. Resolution authority is divided among the LHS's according to the handle prefix, which is generally numeric. Each organization responsible for minting and maintaining handles is given one or more unique numeric prefixes simply handed out in numeric order. As such, Handles avoid semantic problems that could occur if the organisation is taken over or changes its name. For instance, institution 'X' might be represented by the prefix 123. However, there is no reason that Handles can not be used with semantic naming and a few older prefixes are semantically meaningful, e.g. loc for the Library of Congress. The resolution request for a Handle using a given prefix, or naming authority, would be passed to the GHR which then redirects the client to the LHS that administrates the specific institution's Handles. This is often the institution that created the resource, but a Public Local Handle Service is available for those who simply want to try out the Handle System or who would rather not run their own local server. This distributed architecture is a key feature of the Handle System.

The Handle System reference implementation, both client and server, are openly available for free downloading for research purposes. Handles offer granularity as one Handle can refer to several versions of a given digital resource or indeed associated services. The Handle protocol and reference implementation is based on Unicode (UTF-8) and allows for international characters to be incorporated.

Organizations can choose to run the Handle System on their own or they can join a group effort, the most prominent and successful of which is the DOI. Collaborating on policy, including the very important issue of interoperable metadata, and infrastructure can reduce the resources required of any individual organization in utilizing handles to their full potential. The number of such groups is growing, which adds a certain level of credibility that can be lacking in some less established identifier systems.

²⁶ <http://www.cnri.reston.va.us>.

²⁷ <http://www.handle.net>.

²⁸ Handle System Overview, November 2003 <http://www.handle.net/rfc/rfc3650.html>.

Overview of the Digital Object Identifiers (DOI) System

Norman Paskin of the International Digital Object Identifier Foundation (IDF)²⁹ then described the Digital Object Identifier (DOI). The DOI is based on the Handle System infrastructure and enables granular, scalable and secure identification and resolution. 'The DOI is a system for interoperably identifying and exchanging intellectual property in the digital environment'.³⁰ Not just an identifier, the DOI also enables resolution to a collection of value-added services.

DOIs are based on open standards and provide unique naming and directly actionable resolution. Each DOI has the ability to return several options to a variety of versions, formats or services. Paskin emphasized that metadata is vital to identify and contextualise content. Without it, you cannot describe the often-complex relationships that exist between versions, formats and related resources. Metadata is also essential to enable the re-use of digital content. By re-using existing metadata, DOIs help to make the time consuming process of assigning metadata far more efficient and cost effective. Associated metadata fields can vary greatly from one schema to the next. To accommodate this the DOI data model consists of a data dictionary³¹ and a framework for applying it. Together these provide tools for defining what a DOI specifies (through use of a data dictionary), and how DOIs relate to each other, (through a grouping mechanism, Application Profiles, which associate DOIs with defined common properties). This provides semantic interoperability, enabling information that originates in one context to be used in another in ways that are as highly automated as possible, and so re-use existing metadata. The DOI system uses an interoperable data dictionary built from an underlying ontology. This is the Indecs data dictionary, incorporating both the ISO/IEC MPEG-21 Rights Data Dictionary (for which IDF is the formal agency), and other terms such as ONIX as used in the publishing industry. DOIs need not make use of this data model, but it is envisaged that many will.

DOI is implemented through a federation of Registration Agencies (RAs) which use policies and tools developed through a parent body, the International DOI Foundation (IDF). The IDF is the governance body of the DOI system, which works with RAs and with the underlying technical standards of the DOI components to provide a technical and social infrastructure (including creation, maintenance, registration, resolution and policymaking of DOIs) available to any DOI registrant, and requires that no third party licenses might reasonably be required to practice the DOI standard. DOI resolution is freely available to any user encountering a DOI. Registration agencies pay the IDF an annual membership (\$35,000) and a per DOI charge of 0.04 cents for each DOI registered, and are free to establish their own added-value services. The IDF provides implementation through agreed standards of governance and scope, policy, to define "rules of the road". It also provides a technical infrastructure (resolution mechanism, proxy servers, mirrors, back-up, central dictionary) and a social infrastructure - persistence commitments, fall-back procedures, cost-recovery (on a self-sustaining model), and shared use of the system. The IDF is not a standards body, but a central authority and maintenance agency. IDF delegates and licenses authority to use the system through Registration

²⁹ IDF <http://www.doi.org>.

³⁰ DOI News - October 2004

³¹ DOI Factsheet: DOI and Data

Dictionaries <http://www.doi.org/factsheets/DOIDataDictionaries.html>.

Agencies, each of which can develop its own applications and use DOI in "own brand" ways appropriate for their community.

Commitment to persistence is stated through a formal legal agreement with the not-for-profit IDF. Through this agreement, if an RA becomes no longer viable, the DOIs revert back to the stewardship of the IDF. In the event that the IDF becomes defunct, an arrangement has been made to ensure that the DOIs are transferred to an external archive. This model offers a solution to long-term sustainability that has been lacking in other PI systems. The IDF also serves as a governance body and ensures that any improvements made to the DOI system are implemented by RAs. The IDF currently has thirty-four members³². Members may be Registration Agencies or supporters of the development work. To date, ten registration agencies have been approved by the IDF. These RAs manage over 15 million DOIs, of 1200 individual naming authorities (users). The DOI System has the flexibility to deliver identification and resolution services that fulfil the requirements of any application domain. However, these don't come "in a box" since someone needs to build the specific social and technical structures to support the particular requirements of a community (such as scientific data). The rules about what is identified, and whether two things being identified are (or are not) "the same thing", are made at a lower level: in a specific application of the DOI. This is a role of DOI Registration Agencies. This provides an identification system of enormous flexibility and power while hugely increasing the importance of an explicit structured metadata layer, since without this the identifier essentially can have no meaning at all outside a specific application. At the moment, the majority of DOIs resolve to text-like documents but the IDF is actively researching the application of DOIs to other formats including images, sound and scientific data.

³² For a list of IDF members and Board members, see <http://www.doi.org/idf-member-list.html>.

The Implementation of Persistent Identifiers: Case studies

Day two examined the use of persistent identifiers from the library, higher education, publishing, scientific and government perspectives.

The Library Perspective – Die Deutsche Bibliothek, Hungarian Academy of Sciences (MTA), the National Széchényi Library and Uppsala University Library

Kathrin Schroeder explained the use of the URN:National Bibliographic Number (NBN) within Die Deutsche Bibliothek (DDB). The DDB decided to adopt the URN:NBN as a strategy to provide more persistence with regards to the access of their digital resources. This strategy follows the URN specifications as outlined above but incorporates the NBN as a legacy-naming scheme. The URN:NBN was chosen by the DDB due to its simple, flexible approach. The DDB is the main naming authority for the resources they hold, but they can also accommodate sub-name spaces to allow for the sharing of resource administration. THE DDB split metadata management from bibliographic and URN management to allow for greater flexibility. Based on earlier work on the CARMEN-AP4³³ project to facilitate persistent access to on-line dissertations, the EPICUR project aims to replace URLs with managed, resolvable URNs for all of the objects currently archived by the DDB. The DDB server can perform external URN requests from European colleagues such as the University of Uppsala in Sweden. A free download of the DDB plug-in is available from the EPICUR web site³⁴. Additional research is being carried out to incorporate digital library tools into the EPICUR project. The system seems to be working well for the DDB and other European libraries. Schroeder emphasized the importance of implementing organisational policies to ensure that the URN:NBNs or indeed any other PIs are managed effectively over the long-term. She also stressed the need to ensure that the application of persistent identifiers forms part of an overall digital preservation strategy to ensure that long-term access can be guaranteed.

The library perspective was also represented through presentations from two Hungarian organisations: Hungarian Academy of Sciences (MTA)³⁵ and National Széchényi Library³⁶. Andras Micsik from the MTA described efforts underway to collect user requirements and to select a PI system for the new National Digital Data Archive (NDDA)³⁷ in Hungary. The NDDA holds over 300,000 cultural and scientific assets and is currently looking into the application of PIs to enable access to services and various metadata records for the objects in the archive. Micsik also emphasised the need to assign persistent identifiers to geographic places, events, organisations and even individuals. For instance, one street in Budapest has been known by at least four different names over the past fifty years. By assigning a PI to place names, such changes would become less problematic and confusing. Their research has identified some basic key requirements for any PI system that the NDDA might implement. These include that the system be both OAI compliant and open-source. This work is ongoing and no system has yet been selected but the organisation has

³³ CARMEN-AP4 <http://www.mathematik.uni-osnabrueck.de/projects/carmen/index.en.shtml>.

³⁴ EPICUR <http://www.persistent-identifier.de>.

³⁵ Hungarian Academy of Sciences <http://www2.mta.hu/index.html>.

³⁶ National Széchényi Library http://www.oszk.hu/index_en.htm.

³⁷ NDDA <http://www.nda.hu/engine.aspx?page=englishversion>.

recognised the importance of carrying out a requirements analysis and their findings could be of benefit to similar organisations.

Adam Horvath then presented the use of the URN:NBN within the National Széchényi Library. Following the guidelines of the Technical Subgroup of CENL, the National Széchényi Library has created a plug-in resolver to enable access to Hungarian digital resources. The system is simple and user friendly and enables resolution from URL to URN or vice versa. The system only allows the owner of a document to request a URN and as such helps to establish a certain level of authenticity. The system uses a three-step process to assign URNs. The first involves the document owner sending a request to the central server. The server then checks to ensure that the document exists in an HTML format and that no previous URN has been assigned to it. The second step assigns the URN to the document. This is done by the requester and serves to validate the owner of the document. The third step finalises the request by verifying again that the document contains a URN, is in HTML format and does not duplicate any other URN entry. If the document meets these criteria, the URN is accepted into the central archive database. This process is fully automated relying on open source applications such as Java, PHP and PostgreSql. Currently, the production database only has a Hungarian interface. However, the development database also provides an English interface to help make the URNs in the archive more widely accessible.

Eva Müller from Uppsala University Library provided the last view of PIs from the library perspective. In 2000, the first version of the Digitala Vetenskapliga Arkivet (DiVA)³⁸ publishing system was implemented at the University of Uppsala and in ten other universities across Denmark, Norway and Sweden. The aim of DiVA was to provide technical solutions and workflows to facilitate the full-text publishing, storage and dissemination of university research outputs. One specific goal of the project was to ensure that the universities' research output is available for re-use far into the future.

The DiVA infrastructure consists of a central archive and several local archives. The local archives agree to adhere to recommendations regarding metadata, storage formats, and PIs set out by the central archive. The central archive is also responsible for ensuring the quality control of the materials within the infrastructure. Müller stressed throughout her presentation that PIs must form part of an overall preservation strategy to ensure permanent access to digital resources. The DiVA project identified some basic requirements for the application of PIs. These include that the PI system should be non-proprietary, should be easily administrated and maintained, should be associated with a preservation version of a given research output, should be low-cost and should integrate automation wherever possible. The URN:NBN was selected as the best choice to meet these requirements. In addition to persistent identification, the URN:NBN also fulfils several other roles within the DiVA publishing system. They supply the object with a unique ID, provide a naming convention and can be incorporated into disseminated metadata. As the URN supports legacy naming schemes – including DOIs and ARKS - to be incorporated into the naming structure, several are supported within the DiVA project. DiVA stores multiple copies of research output in several different formats (including XML and PDF) and locations to minimize the risks of data loss.

DiVA has created a temporary file format registry to assist with the long-term

³⁸ DiVA Portal <http://www.diva-portal.se/about.xsql>.

preservation of the resources in their archives. DiVA will eventually switch to one of the large-scale file format registry projects currently being developed in the US and the UK. File format information is also assigned a PI within the DiVA publishing system. The cooperation between several large, trusted libraries across Denmark, Norway and Sweden helps to guarantee the sustainability of DiVA should any one institution become no longer viable. A new version of the resolution service was launched in May 2004 and enables harvesting of URN:URL from across several distributed repositories. The harvesting model used by the DiVA project makes the updating of the mapping registry relatively straightforward and simple. Additionally, the resolution service offers resolution of URN to actionable links. Müller concluded by stating that although the URN:NBN is meeting the needs for the DiVA project, work must still be carried out on the development of a global resolution mechanism that works with and across all persistent identification systems.

The Publishing Perspective – John Wiley & Sons Ltd.

Cliff Morgan of John Wiley & Sons Ltd.³⁹ provided the publishing perspective for the application and use of PIs. The publishing world has used the application of unique and permanent identifiers such as ISBNs and ISSNs to facilitate rights management and commercial transactions for many years. As ISBNs and ISSNs started to run out and more and more information was being made accessible electronically, the publishing community recognised that they needed to look for other permanent naming options. This need has been the major driver for development of PIs to date. Research into the application of International Standard Text Code (ISTC) was undertaken to provide a long, intelligent string for electronic texts. However, progress has stalled over the past four years due to problems with developing policies and management structures for the system. DOI development has been more successful within the publishing community and to date over 300 publishers have been involved in collaborative efforts to push the DOI forward as the publishing PI standard. The DOI owes much of its success within the publishing world to CrossRef who adopted the system to link citations to source materials. The multiple resolution capabilities of DOIs allows the publishing world to identify versions, formats as well as individual chapters and pages of books. DOIs are easy to implement and allow the publishers to provide value-added services.

Morgan then went on to describe the importance of metadata within any PI system. The publishing world favours ONIX, but is aware of the need to enable interoperability with other standards such as Dublin Core⁴⁰, OAI-PMH⁴¹ and IEEE/LOM/SCORM⁴². Morgan – like Paskin - suggested that data dictionaries and metadata mapping tables would be essential to enable interoperability across domains. Morgan stressed that the publishing community has supported the application and use of PIs largely due to the fact that they have been proven to increase revenue generation. More work needs to be carried out in other domains to identify and illustrate potential benefits associated with the application of PIs. The

³⁹ John Wiley & Sons Ltd. <http://wiley.co.uk>.

⁴⁰ Dublin Core Metadata Initiative <http://dublincore.org/>.

⁴¹ The Open Archives Initiative Protocol for Metadata Harvesting
<http://www.openarchives.org/OAI/openarchivesprotocol.html>.

⁴² IEEE/LOM and SCORM <http://ltsc.ieee.org/wg12/>; <http://www.adlnet.org/>.

publishing community has done more than any other community to establish user requirements and standards for PIs. Other communities must follow the example of the publishing sector to work towards the creation of both user requirements and illustrated benefits before PIs will become widely adopted in other domains.

The Higher Education Perspective – JISC Report on Digital Object identifiers for the Publishing and e-Learning Community, The Stationary Office (TSO)

Robin Wilson of The Stationary Office (TSO)⁴³ presented the higher education (HE) and academic publishing perspective. Based on a pilot study undertaken between May and August 2003, this report aimed to provide support and recommendations for the implementation of identifiers for digital objects in the higher and further education and the publishing communities. The study was funded by the Joint Information Systems Committee (JISC) as part of their Publishers and Library/Learning Solutions (PALS)⁴⁴ Metadata and Interoperability Group. The report provides background information on the evolution of persistent identifiers and offers a useful overview of some of the major strategies including Handles, DOIs, PURLS and Extensible Resource Identifiers (XRI)⁴⁵.

Wilson stressed that metadata plays an integral role in any persistent identification strategy as it provides the context and sequence for the object identified. In addition, metadata is vital for allowing people to know what they are finding – images, text, or aggregate learning objects. INDECS⁴⁶, CORES⁴⁷ and ARK are potential metadata standards that could be used by the HE and academic publishing communities but it is not yet clear whether metadata should be built into the identifier hierarchically or maintained separately. Wilson then explained how persistent identifiers could be used within the publishing and e-learning communities and described some potential use cases. A set of minimum requirements for digital object identifiers are proposed in the report including that identifiers are actionable, extensible, allow for multiple resolutions and enable backward compatibility. Wilson urged organizations to carefully consider what they want identifiers to identify, what they need them to return and for how long they need to be persistent. Wilson concluded with a call for participants to consider and share their specific user requirements. This is the only way that PI system developers can produce strategies that meet the needs of as many user communities as possible. He also recommended further research into the long-term management of and metadata for persistent identifiers.

⁴³ The Stationary Office <http://www.tso.co.uk>.

⁴⁴ Publishers and Library/Learning Solutions (PALS)
http://www.jisc.ac.uk/index.cfm?name=programme_pals.

⁴⁵ Extensible Resource Identifiers (XRI)
<http://www.oasis-open.org/committees/download.php/3063/wd-xri-specification-07.pdf>.

⁴⁶ Interoperability of data in e-commerce systems (INDECS) <http://www.indecs.org/>.

⁴⁷ CORES Element Subgroup (part of PREMIS)
http://www.oclc.org/research/projects/pmwg/core_elements.htm.

The Scientific Perspective: World Data Centre for Climate (WDCC)

Michael Lautenschlager of the WDCC⁴⁸ provided an insight into the use of PIs from the scientific perspective. Funded by the German Research Foundation (DFG)⁴⁹ and in conjunction with CODATA⁵⁰, the WDCC began research into making climate data permanently accessible in 2001⁵¹. The major problem in enabling re-use of scientific data is the lack of contextual information associated with the dataset. To remedy this situation, the WDCC and CODATA investigated publishing primary research data with the application of PIs to datasets and contextual referencing. Then metadata at the scientific level was integrated into library catalogues to facilitate linking from the primary data to related published results. Essentially, the metadata acts as an interface between the data archives and the libraries.

The WDCC utilise both DOIs and URNs to name the scientific data. A DOI is assigned by the data archive and is then sent to Die Deutsche Bibliothek (DDB) as a URN to enable searching through the library catalogues. The URN is recorded using the same structure as the DOI. The German National Library of Science and Technology (TIB/UB)⁵² in Hanover is now registered as a DOI registration agency for scientific and technical data (STD-DOI). This is the first example of its kind in the world. A Handle server has been installed at TIB and is used to register and resolve DOIs while the URNs are resolved through the DDB. This hybrid approach is interesting and shows that the application of more than one PI system may be needed to meet an organisation's or discipline's individual needs.

The WDCC has assigned PIs and metadata descriptions for over 400 data sets so far. At the moment, the PIs point directly to the data itself, but ongoing research is being carried out to include pointers to related services and analytical tools as well. By the end of 2004, it is expected that over 10,000 data sets will be accessible through the central library database. The project aims to incorporate datasets from other scientific disciplines over the next year and plans to establish the TIB/UB as the central registration agency for primary scientific data in Germany by 2006.

The Government Perspective: Persistence and Access to Information in the Government of Canada

Cecil Somerton of the Treasury Board for the Government of Canada provided a government perspective for the use of PIs. The Government of Canada is responsible for both ensuring that the general public has access to government information (under the Freedom of Information legislation) and that the privacy rights of its citizens are protected. The newly-formed Library and Archives Canada⁵³ has joint responsibility for leading development in the area of information management. Information management is integral to enabling access to information and persistent

⁴⁸ WDCC <http://www.mad.zmaw.de/wdcc/>.

⁴⁹ German Research foundation (DFG) <http://www.dfg.de/en/index.html>.

⁵⁰ CODATA <http://www.codata.org>.

⁵¹ project web site <http://www.std-doi.de>.

⁵² German National Library of Science and Technology (TIB/UB) <http://www.tib.uni-hannover.de>.

⁵³ Libraries and Archives Canada <http://www.collectionscanada.ca/>.

access to information depends upon reliable identification of resources. Persistence in the Government of Canada means that information remains unaltered, accessible and useable over time regardless of technological changes. Stability and persistence are major factors in building trust and accountability within the Government of Canada. The management of Government Information Policy requires among other things, that departments must describe the institution's information, provide users with timely and convenient access to information, ensure its usability over time, ensure that information of enduring value to the Government of Canada or to Canadians is available for current and future use and optimize its sharing and re-use. To meet these policy requirements, the government will require the application of interoperable persistent identifiers.

Somerton stressed the need to consider what users want to remain persistent. Is it access to the information itself, the location of the information or simply the identity of the information that is required? Somerton pointed to the fact that PIs could potentially be used to identify tools, programmes and services, objects or even individuals. Accordingly, careful consideration regarding what kinds of information need to be persistent must be taken into account. Another factor to reflect upon is whether persistence of the resource format is necessary or just the intellectual content. All of these issues are subjective and meanings could vary greatly from one government department to another. As such, work must be done to establish a common understanding of terms.

While Somerton acknowledged that there are generally costs involved with the implementation and management of all PI systems, the costs of constantly updating broken URLs is equally time consuming and expensive. There is a certain amount of risk involved with the adoption of any PI system due to their lack of maturity. These must be weighed up against the risks the organisation faces if information becomes inaccessible. The Government of Canada has set up a PI working group under the Library and Archives Canada. While the group have been actively researching PIs for some time, no decision has yet been made regarding a single solution. Currently, investigation into XML Namespace Management within the Government of Canada is being carried out.

Somerton recommended that government managers and staff from a wide range of backgrounds, departments and agencies be brought together to raise awareness of these issues and to promote viable solutions, strengthen government standards and guidelines and secure funding for PI pilot projects. Extensibility and repository models will also be important for long-term persistent access to digital resources. Somerton emphasised that without a commitment to long-term persistence on the part of the resource creator, no PI strategy will ever work. All current PI strategies require human commitment to ensure long-term persistence and resolvability. As has been mentioned before, this will involve philosophical change at the organisation level. Somerton concluded by restating the need to engage a wide range of stakeholders to develop user requirements and a coordinated approach to persistent identifiers.

Conclusions

The presentations revealed that each of the PI systems has specific benefits and potential drawbacks. While the URN has been widely adopted and progressed by the European library community, it often requires a plug-in to resolve from the PI to the item itself. The plug-in requirement was considered by many participants to be a major drawback. However, as the DiVA project demonstrated, it is possible to offer directly actionable links through a URN:NBN resolution service. ARKs offer a unique organisational commitment statement that assists users in determining the level of guaranteed persistence associated with a given item. However, as this is a relatively new strategy, it may still need to be proven effective by a range of users before widespread uptake is observed. PURLS offer greater flexibility than many of the other systems and are freely available to download and customise. However, the one-to-one resolution facilitated by PURLS may lack the necessary granularity that some organisations require. Handles offer a distributed architecture and are less dependent on web-based environments. But, the costs associated with implementation could deter some potential users. The DOI extension of the Handle System offers access to value added services and its widespread adoption by the publishing world lends it credibility that some of the others systems lack. Again, the costs associated with the implementation and management of this system may limit its potential user-base. Ultimately, there is no one system that currently meets the needs of all users. This is largely due to a lack of clearly defined user requirements from a range of domains. As was mentioned throughout the course of the two days, this is something that requires further investigation.

During the breakout sessions it became clear that a reference model is needed to help potential users decide what system best meets their needs. The development of some kind of comparative table was suggested and is currently being investigated by ERPANET. Participants also felt that more information on specific roles and responsibilities would benefit organisations in making informed choices regarding PI systems. The majority of participants agreed that persistent identification and resolving is fundamentally a social problem and that without the necessary philosophical changes within an organisation, no PI system could ever be completely effective.

Without clearly-defined user requirements, PI systems cannot be developed to fulfil a variety of purposes across a wide range of disciplines. To date, the publishing industry is the only one to have made any real efforts in this area. Robin Wilson, Stuart Weibel and Larry Lannom urged the participants to consider and draft their requirements so that PI development would become user-driven in other domains as it has been in the publishing community. The concept of a universal resolution system was discussed but participants questioned whether such a system and common naming syntax could ever be achieved. Due to the level of European interest in the URN, many participants wondered if the development of a URN infrastructure to replace the required resolution plug-in could be beneficial. Further interaction between PI system developers, user communities and standards bodies will be necessary to push research in this area forward. Participants acknowledged here that greater motivation to collaborate would help to promote communication across various sectors.

Improved communication between R&D and other stakeholders is something that all of the participants felt would be beneficial for ensuring interoperability between naming schemes and across disciplines and nations. It was suggested that ERPANET and the new JISC-funded Digital Curation Centre might be best placed to

coordinate further research in this area. More illustrations of the potential uses for PIs must be provided for a range of disciplinary perspectives before uptake will increase dramatically. For instance, at the recent ERPANET workshop on the Preservation of Born Digital Art⁵⁴, it was revealed that accurate information regarding the original and subsequent locations of a piece of networked art can have a major impact on the object's provenance and authenticity. The application of persistent identifiers to born digital art could offer a viable solution to establishing and maintaining provenance.

While much more work needs to be done on developing persistent identification requirements and systems, this seminar has been a good starting point. It is now clear that improved communication between all stakeholders will be key in developing any viable solutions. Encouragingly, communications via the DC-Persistent Identifiers listserv have been very active since the workshop and many of the workshop participants have joined the debate. This debate can be accessed and contributed to via the DC-Persistent Identifiers⁵⁵ listserv. Most of the speakers at the event are regular contributors to this list and welcome input from as wide a range of potential users as possible. A working group listserv has also been set up for participants and speakers from the Cork event.⁵⁶

⁵⁴ <http://www.erpanet.org/events/2004/glasgowart/index.php>.

⁵⁵ to register for this list, go to JISC Mail at <http://www.jiscmail.ac.uk>.

⁵⁶ *Ibid.*

Appendix One: Seminar Programme

Programme – Day One

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 – 12:00 | Panel Discussions: Related Issues
-Overview of the URN, Kathrin Schroeder (DDB)
-Overview of OpenURL, Tony Hammond (Nature Publishing Group)
-Overview of InfoURI, Stuart Weibel (OCLC)
-Update from the DCMI working group on persistent identifiers, Robin Wilson (TSO) and Stuart Weibel (OCLC)
-Update from the CENDI working group on persistent identifiers, Larry Lannom (CNRI)
-Question period |
| 13:30 – 14:00 | Overview of ARK
Mary Heath (California Digital Library) |
| 14:00 – 14:30 | Overview of PURL
Stuart Weibel, (OCLC) |
| 15:00 – 15:30 | Overview of Handle System
Larry Lannom, (CNRI) |
| 15:30 – 16:00 | Overview of Digital Object Identifiers (DOI)
Norman Paskin, (IDF) |
| 16:00 – 17:00 | Breakout session |
| 17:00 – 17:30 | Report back |
| 17:30 - 19:00 | Reception |

Breakout Session – Suggested Questions:

What are the major features, benefits and drawbacks for each of the major strategies?

Which sectors would benefit most from using each of the major strategies?

Programme – Day Two

Friday, June 19

09:00 – 09:30	EPICUR Project, Kathrin Schroeder (Die Deutsche Bibliothek)
09:30 – 10:00	National Digital Data Archive of Hungary Andras Micsik (Hungarian Academy of Sciences) NBN:URN generator and resolver Adam Horvath(OSZK)
10:00 - 10:30	DiVA Project, Eva Muller (Uppsala University)
11:00 -11:30	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

Breakout Session - Suggested Questions:

How can the application of persistent identifiers fit into an overall digital preservation strategy?

What are the main roles and responsibilities involved in ensuring that the persistent identifiers are implemented and maintained?

Appendix Two: List of Participants

Ingela Algotsson	National Tax Board, Sweden
Reinhard Altenhoener	Die Deutsche Bibliothek, Germany
Mike Clarke	Higher Education Academy, UK
Joy Davidson	ERPANET, UK
David Dawson	MINERVA Europe; The Council for Museums Archives and Libraries, UK
Monica Duke	UKOLN, UK
Jim Downing	Cambridge University, UK
Maria Guarini	Ministero delle Comunicazioni, Italy
Yvette Hackett	Library and Archives Canada, Canada
Tony Hammond	Nature Publishing Group, UK
Mary Heath	California Digital Library, USA
Adam Horvath	OSZK, Hungary
Jon Juliusson	Swedish Tax Agency, Sweden
Max Kaiser	Austrian National Library, Austria
Madeleine Kiss	Office for Official Publications, Luxembourg
Larry Lannom	CNRI, USA
Michael Lautenschlager	Director World Data Center for Climate (WDCC), Germany
John McDonough	National Archives, Ireland
Peter McKinney	ERPANET, UK
Andras Micsik	Hungarian Academy of Sciences, Hungary
Cliff Morgan	John Wiley & Sons Ltd., UK
Eva Muller	Uppsala University, Sweden
Kevin O'Neill	CCLRC, UK
Mared Owen	National Library of Wales, UK
Norman Paskin	International DOI Foundation, UK
Adrian Price	Danish Electronic Research Library, Denmark
Carol Quinn	UCC, Ireland
Seamus Ross	ERPANET, UK
Kathrin Schroeder	Die Deutsche Bibliothek, Germany
Brack Serge	Office for Official Publications, Luxembourg
Fabio Simeoni	Centre for Digital Library Research (CDLR), UK
Cecil Somerton	Treasury Board of Canada Secretariat, Canada
Marco Streefkerk	Library of the Universiteit van Amsterdam, The Netherlands
Stuart Weibel	OCLC, USA
Robin Wilson	TSO, UK

Appendix Three – Additional Seminar Reports

At the Event: ERPANET Seminar on Persistent Identifiers
Monika Duke, UKOLN
<http://www.ariadne.ac.uk/issue40/erpanet-ids-rpt/>

ERPANET Persistent Identifier Symposium Stuart Weibel, OCLC
DC-PERSISTENT-IDENTIFIERS@JISCMail.AC.UK, June 30, 2004

A report on the ERPANET Seminar on Persistent Identifiers
Fabio Simeoni, CDLR
<http://hairst.cdlr.strath.ac.uk/documents/Erpanet%20Training%20Seminar%20on%20Persistent%20Identifiers.pdf>

Appendix Four – Sponsoring Organisations

University College Cork

<http://www.ucc.ie>

University College Cork (UCC) is one of four constituent universities of the federal National University of Ireland.

Founded in 1845, UCC is one of Ireland's oldest institutes of higher learning and was originally established as Queen's College. At the same time two other Queen's Colleges were established in southern and northern Ireland-one in Galway, the other in Belfast.

The original site chosen for the University is believed to have a connection with the patron saint of Cork, St Finbarr who set up a monastery and school of learning in close proximity to the University. Hence the University's motto "Where Finbarr Taught, Let Munster Learn."

University College Cork is situated in south-west Ireland and is a progressive, dynamic university of over 14,500 students. The University has a current enrolment of over 12,000 degree candidates. In addition, 2,500 students are enrolled in one or more of the University's postgraduate courses. UCC's widely diverse student-body includes over 1,000 international students representing 60-plus countries worldwide. 3,000 people work at UCC, including more than 800 faculty. UCC is Ireland's leading research institute. It is the highest research income earner in the Irish state. The University offers an innovative, research-led curriculum, taught by world-class academics who consistently attract a quality student intake. The University's internal research reputation spans all of its faculties where it offers over 120 degree and professional programmes through seven schools and 27 departments.

MINERVA

<http://www.mivervaeurope.org>

MINERVA is a network of Member States' Ministries to discuss, correlate and harmonise activities carried out in digitisation of cultural and scientific content for creating an agreed European common platform, recommendations and guidelines about digitisation, metadata, long-term accessibility and preservation.

Due to the high level of commitment assured by the involvement of EU governments, it aims to co-ordinate national programmes, and its approach is strongly based on the principle of embeddedness in national digitisation activities.

It also establishes contacts with other European countries, international organisations, associations, networks, international and national projects involved in this sector, with a special focus on actions carried out in the DigiCult action of IST. The project has organised an advisory Group, relying on existing actions to identify

and integrate best practices in a pan-European framework, to facilitate the adoption of the Lund action plan.

Digital Curation Centre

<http://www.dcc.ac.uk>

Scientists and researchers across the UK generate increasingly vast amounts of digital data, with further investment in digitisation and purchase of digital content and information. The scientific record and the documentary heritage created in digital form are at risk, by technology obsolescence and by the fragility of digital media. Working with other practitioners, the Digital Curation Centre will support UK institutions to store, manage and preserve these data to ensure their enhancement and their continuing long-term use. The aim of the Centre is to provide a national focus for research into curation issues and to promote expertise and good practice, both national and international, for the management of all research outputs in digital format.

The DCC partners bring world class expertise. All the partners are contributors to the National e Science Centre. Edinburgh also contributes service expertise via EDINA, five-star RAE-rated Informatics research, and several e Science partners with real life experience of curation issues. Glasgow also brings HATII, with strengths in training and advisory services, and international cutting edge links in digital preservation, together with relevant expertise from the archives and records management world. CCLRC brings major expertise in standards and agenda setting activities in e Science, including OAIS; it also hosts large-scale long term repositories and has links to thousands of scientists. UKOLN is also involved in standards and agenda setting for the JISC, including digital preservation, and has major strengths in communication and dissemination.

The Centre has two aims: to be an organisation that is research proficient and to be one that is service orientated; each internationally respected, and to a standard that would warrant leadership and advocacy across disciplines.

ERPANET

<http://www.erpanet.org>

Increasing amounts of Europe's cultural and scientific heritage is being created or represented in digital form. The preservation and reuse of these digital assets forms both the cornerstone of future economic growth and development, and the foundation for the future of memory. This material represents Europe's heritage and is its future intellectual capital. The Electronic Resource Preservation and Access Network (ERPANET) widely recognises the benefits of using digital information and a result of its prevalence is the emerging vision of Europe as an information rich society whose record is just waiting to be harvested and processed by the technology-enabled

researcher of the future or by emerging eContent industries. Ensuring this vision depends upon the survival of digital data in accessible and usable form.

The fast pace of change in the technological landscape makes ensuring technological advances preserving digital assets cannot happen as an after-thought, it needs to be planned. Policies, technical methods and strategies are required because media degrade (e.g. magnetic particles lose their properties and dye layers on optical media break down), technological developments make systems obsolete, or information is rendered inaccessible by changes in encoding formats.

ERPANET addresses the lack of awareness, fragmentation of knowledge and skills amongst the stakeholder communities about how to handle existing digital preservation problems, and how to plan effectively for the future. ERPANET tackles the lack of identification and focus on core research areas and brings coherence and consistency to activities in this area.