ECP-2007-DILI-517005

ATHENA

Assessment of requirements for persistent identification of objects, collections and institutions

Deliverable number	*D3.4*
Dissemination level	*Public*
Delivery date	*31 July 2010*
Status	*Final*
Author(s)	*Gordon McKenna, Collections Trust (UK), Roxanne Wyns, Royal Museums of Art and History (BE)*

e*Content*plus

This project is funded under the eContentplus programme1,
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

[1] OJ L 79, 24.3.2005, p. 1.

Table of Contents

1. Introduction

1.1 The purpose of work package 3

Work package 3 of the ATHENA project (WP3) is tasked with:

1. Reviewing the different standards in use by museums;
2. Facilitating the mapping of those standards to a common metadata standard;
3. Assessing the requirements for the persistent identification of digital objects and collections;
4. Producing tools to support the conversion of museums' data into the common harvesting format for ingestion into the main Europeana service.

WP3 also works together with other work packages in the project. In particular WP3 works closely with WP4 and WP7: feeding information about standards for their work. Also the survey which is the basis of this deliverable was extended to include collecting information on IPR issues for use within WP6.

1.2 Background to the deliverable

The deployment of system which delivers the persistent identification of its resources and entities is one of the keys to success for any information service. Europeana and its group of projects are creating such a service. Therefore it is natural that this area should be explored within the ATHENA project with the purpose of helping partners, and others, to participate effectively within the overall Europeana service. In particular there is a need to solve the issue of 'broken links' to resources being pointed to by Europeana on provider's website.

Also there is a broader aspiration for the Europeana information service – taking part in the Semantic Web and Linked Open Data worlds. Here persistent identification of resources and other entities, such as people, places, and events, is the fuel that powers the engine.

Even without the existence of Europeana the issue would need to be addressed by any organisation wanting to provide a reliable information service to its users.

1.3 Overview of the deliverable

Within the museum domain, this deliverable covers persistent identification for: physical objects; digital objects; collections; and institutions. Its aim is to provide:

- Information about the standards and services available for persistent identifiers.
- A review of persistent identifier use in the cultural heritage sector using a survey of the National Representatives associated with the ATHENA project.
- A set of requirements for the successful implementation of persistent identifiers in the CHS.
- Advice and best practice.
- Conclusions.

2. Persistent identifier standards and service landscape

2.1 Describing standards and services

Using a metadata scheme similar to that used in deliverable 3.1, we describe each persistent identifier (PID) standard or service in a Dublin Core (DC) derived format. 9 out of the 15 DC elements are used in the descriptions.

These elements are:

Title	The name (or names) under which the service or standard is known. Where there is an abbreviated and full name both are given.
Creator	The name of the organisation which originally created the service or standard.
Publisher	The name of the organisation that makes the service or standard publicly available.
Date	The date on which the service or standard was originally published.
Identifier	A number or other identifier under which a standard is published or a URL which points to the definition of the standard. Also include is a URL to a service's website.
Rights	Whether rights restrictions apply.
Description	A textual description explaining the service or standard and its usage.
Subject	Keywords that identify the nature of the service or standard.
Relation	Other services or standards that this one relates to, and associated websites.

The descriptions are aimed at a general reader. More technical details for the services and standards given can be found in the various references and links given in the records. The purpose of this section is to allow the reader to have an easy reference to the range of relevant persistent identifier services and standards in one place.

2.2 Physical objects in museums

There are no formal standards for persistent identifiers for physical objects in museums. Many organisations have their own internal systems which may or may not follow suggestions given by advisory bodies. See the section below on persistent identifiers for digital objects for the online identification of physical objects.

The importance of persistent numbers for physical objects is emphasised in the *SPECTRUM* standard where an *Object number* is described as:

Object number

Definition	A unique number identifying an object or specimens, including any separated parts.
How to record	The following points should be considered when assigning an *Object number*: Only use a single number to describe a group of objects if they are too numerous to number individually and either contained in a single container or separately accounted for, e.g. a box of sherds or an archaeological archive Do not include in the number any classificatory components as these may change Avoid alphabetical components
Examples	1992.1234; 1992.12.1
Use	Assign a unique *Object number* to each separated or separable part of an object.

Note in some systems this might be known as the inventory number, accession number, identity number, or just number.

2.3 Digital objects

Standards

There are three, interrelated, standards:

URI (Uniform Resource Identifier)

Title	URI (Uniform Resource Identifier)
Creator	Berners-Lee, T (W3C/MIT); Fielding, R (Day Software); Masinter, L (Adobe Systems)
Publisher	The Internet Society
Date	2005 (current standard) [original concepts in 1990]
Identifier	http://www.rfc-editor.org/rfc/rfc3986.txt (generic syntax)
Rights	[Open Standard]
Description	String of characters used to identify a name or a resource on the Internet. Form: The syntax of a URI is:

	[scheme name]:[scheme-specific part] • scheme name – includes examples as "http", "ftp", "mailto", file, or "urn" followed by a colon character, and then by a scheme-specific part • scheme-specific part – these are specified in the rules of the scheme. However they must conform to the general requirements for URIs. These include the rules on the use of particular characters. URLs and URNs are URIs.
Subject	persistent identifier (Internet)
Relation	URL (Uniform Resource Location); URN (Uniform Resource Name)

URL (Uniform Resource Locator)

Title	URL (Uniform Resource Locator)
Creator	T Berners-Lee (CERN), L Masinter (Xerox Corporation) & M McCahill (University of Minnesota) (Editors)
Publisher	Internet Engineering Task Force (IETF)
Date	1994 [original]
Identifier	http://tools.ietf.org/html/rfc1738
Rights	[Open Standard]
Description	A URI (i.e. a string) that specifies: • Where a resource is available; • The mechanism for retrieving it. Form: scheme://domain:port/path?query_string#fragment_id • scheme – defines the namespace, purpose, and the syntax of the remaining part, examples: http, https, gopher, wais, ftp. • domain:port – gives the destination location for the resource (domain name or IP address). Port is optional, if absent the default is used (for http default port = 80). • path – used to specify and find the resource • ?query_string – used to pass data to a piece of software to enable retrieval • fragment_id – used to specify a part or a position within the overall resource E.g. http://www.athenaeurope.org/index.php?en/91/information-on-the-project (the 'About us' page on ATHENA project website
Subject	persistent identifier (Internet); persistent identifier (book); persistent identifier (periodical); persistent identifier (audiovisual);
Relation	URI (Uniform Resource Identifier); URN (Uniform Resource Name)

URN (Uniform Resource Name)

Title	URN (Uniform Resource Name)
Creator	Network Working Group (ed. R Moats, AT&T)
Publisher	Internet Engineering Task Force (IETF) (syntax); IANA, the Internet Assigned Numbers Authority (namespace assignment).
Date	1997
Identifier	http://tools.ietf.org/html/rfc2141 (syntax)
Rights	[Open Standard]
Description	String acting as persistent, location-independent, resource identifiers, designed to make it easy to map other namespaces. Note that they do not point to a location and therefore might not be resolvable. Form: urn:<NID>:<NSS> <NID> is the Namespace Identifier, and <NSS> is the Namespace Specific String. The Namespace ID determines the syntactic interpretation of the Namespace Specific String. E.g. urn:isbn:0451450523 is URN for The Last Unicorn, identified by its book number. Example namespaces: ISBN; ISSN; ISAN; NBN2
Subject	persistent identifier (Internet); persistent identifier (book); persistent identifier (periodical); persistent identifier (audiovisual);
Relation	URI (Uniform Resource Identifier); URL (Uniform Resource Locator)

2 National Bibliography Number. These are identifiers used by national libraries for those documents (e.g. web pages) where there is no identifier given by the publisher (e.g. an ISBN). The URN namespace for NBNs is described in RFC 3188 (**http://tools.ietf.org/html/rfc3188**). Some national libraries have resolution services for these URNs.

Services

There a number of services which support the persistent identification of digital objects:

PURL (Persistent URL) & Handle System

Title	PURL (Persistent URL) & Handle System
Creator	OCLC (Online Computer Library Center)
Publisher	OCLC (Online Computer Library Center)
Date	1995
Identifier	http://purl.oclc.org/docs/help.html#overview
Rights	OCLC (Online Computer Library Center) (?)
Description	A URL pointing to a resolver (e.g. Handle) which redirects to current URL; Resolver software (OCLC free). Form: Has 3 parts – 1. Protocol - used to access the PURL resolver (Handle System). 2. Resolver's address – an IP address or domain name. (Resolved by the Domain Name Server (DNS)). 3. Name – assigned by the user E.g. http://purl.oclc.org/oclc/oluc/32127398/1 ---- ---------- ---------- | | | Protocol resolver address name
Subject	persistent identifier (digital object)
Relation	http://purl.oclc.org (PURL website); http://www.ietf.org/rfc/rfc3986.txt (Uniform Resource Identifier (URI): Generic Syntax); http://www.handle.net (Handle System website) [implementation];

Title	Handle System
Creator	Network Working Group
Publisher	Internet Engineering Task Force (IETF) [specifications]
Date	1994-2003
Identifier	http://www.ietf.org/rfc/rfc3650.txt (Handle System Overview)
	http://www.ietf.org/rfc/rfc3651.txt (Handle System Namespace and Service Definition)
	http://www.ietf.org/rfc/rfc3652.txt (Handle System Protocol (ver 2.1) Specification)
Rights	Internet Engineering Task Force (IETF) [specifications]
Description	Specification for a distributed computer system which assigns, manages, and resolves URLs. 'Handles' are the identifiers for digital objects. They are resolved into the information needed to locate and access the objects. Users are redirected to the current location. The information stored in the system has to be maintained with up-to-date information for the service to continue to work.
Subject	persistent identifier resolution
Relation	http://purl.oclc.org/docs/help.html#overview (PURL); http://www.handle.net (Handle System website) [resolution service]

DOI (Digital Object Identifier)

Title	DOI (Digital Object Identifier)
Creator	International DOI Foundation
Publisher	International DOI Foundation
Date	1998 (creation of International DOI Foundation)
Identifier	ANSI/NISO Z39.84 (Syntax for the Digital Object Identifier) [NB. DOI is about to become an ISO standard]
Rights	[Open standard] (definition); International DOI Foundation (implementation)
Description	A stored and maintained character string used to uniquely identify an electronic document (or other type of digital object). Associated with the DOI is metadata. This can include a location (e.g. a URL) where the referenced document can be found. The metadata is maintained to reflect changes in physical changes in the documents location. Form: Divided into two parts: 1. Prefix – identifies the registrant of name; 2. Suffix – chosen by the registrant to identify the document associated with the DOI. E.g. doi:10.345/document.identifier12345 The system is implemented by a federation of registration agencies, co-ordinated and controlled by International DOI Foundation. These pay to be a member of the federation and must agree to meet the contractual obligations associated with the system. A DOI 'name' may be resolved by inputting it to a DOI resolver (e.g. at the International DOI Foundation) or may be represented as an HTTP string by preceding the DOI name by the string 'http://dx.doi.org/' and omitting 'doi:'
Subject	persistent identifier (digital documents)
Relation	http://www.doi.org (DOI website); http://www.doi.org/about_the_doi.html (overview); http://www.handle.net (Handle System) [resolution service]

OpenURL

Title	OpenURL
Creator	Herbert Van de Sompel [original]
Publisher	OCLC (Online Computer Library Center) [standard maintainer]
Date	2000 (original); 2010 (standard)
Identifier	http://alcme.oclc.org/openurl/docs/pdf/openurl-01.pdf [original]; ANSI/NISO Z39.88 (The OpenURL Framework for Context-Sensitive Services)

Rights	[Open standard]
Description	A URL, with embedded metadata, which enables users to more easily find a copy of a resource. The metadata is used by the resolver service. It is often bibliographic in nature, and OpenURLs are commonly used by libraries. Form: In two parts: 1. Base URL for the resolver service; 2. Query string. E.g. [original version] http://www.springerlink.com/openurl.asp?genre=journal&issn=0942-4962 The new standard version is slightly more complicated in form.
Subject	persistent identifier (digital objects)
Relation	http://www.oclc.org/research/activities/openurl/default.htm (webpage)

ARK (Archival Resource Key)

Title	ARK (Archival Resource Key)
Creator	US National Library of Medicine (developer)
Publisher	California Digital Library (maintainer)
Date	2001
Identifier	**https://confluence.ucop.edu/download/attachments/16744455/arkspec.pdf?version=1**
Rights	[Open standard?]
Description	A URL scheme which can identify both physical and digital objects. Form: **[http://NMAH/]ark:/NAAN/Name[Qualifier]** **NAAN** = Name Assigning Authority Number - mandatory unique identifier of the organization that originally named the object **NMAH** = Name Mapping Authority Host - optional and replaceable hostname of an organization that currently provides service for the object **Qualifier** = optional string that extends the base ARK to support access to subcomponents of an object or its variants (e.g. version, language).
Subject	persistent identifier (digital objects); persistent identifier (physical objects);
Relation	**https://confluence.ucop.edu/display/Curation/ARK** (webpage)

2.4 Collections in museums

There are no formal international standards for persistent identifiers for collections in museums3.

MDA Codes (see below) can be used for part of an institution's collection, but this practice is rare.

See the section above on persistent identifiers for digital objects for the online identification of collections.

Collections can be thought of a 'super objects' and indeed some objects are 'naturally' collections (e.g. a ceramic dinner services made up of plates, serving dishes, cups and saucers). Such objects are often managed as the collection not as the individual parts. Therefore it is possible to use the same standard (URIs) and the same resolution services to define and manage PIDs for collections.

2.5 Institutions

There are no formal international standards for persistent identifiers specifically for museum institutions.

See the section above on persistent identifiers for digital objects for the online identification of institutions4.

Originally developed for libraries (but can be used for other types of organisation) is:

ISIL

Title	ISIL (International Standard Identifier for Libraries and Related Organizations)
Creator	International Organization for Standardization (ISO)
Publisher	International Organization for Standardization (ISO); ISIL Registration Authority (maintainer)
Date	2009
Identifier	ISO 2709:1996
Rights	[Open Standard]
Description	An alphanumeric string of up to 16 characters. Form: In two parts separated by a dash ('-'): • Prefix identifying the issuing authority. These can be country codes (two capital-letters, e.g. BE), or non-national codes for authorities that are international, e.g. OCLC • Identifier agreed with the institution.
Subject	persistent identifier (organisation)
Relation	http://biblstandard.dk/isil/ (webpages)

3 The DC-DI proposal (under evaluation) has a task is devoted to explore and possibly define persistent identification of digital collections.

4 DC-DI will also look at identifiers for institutions.

In addition some countries have systems for organisation identification. For example the UK has:

MDA Code

Title	MDA Code
Creator	Collections Trust [formerly Museum Documentation Association]
Publisher	Collections Trust
Date	1977-
Identifier	http://www.collectionstrust.org.uk/mdacodes (MDA Codes database)
Rights	Collections Trust
Description	An alphabetic string of usually five letters (some national museum have shorter codes). Form: The code is made up of two concatenated parts: • [Part one – usually three letters] = representation of location of the institution • [Part two – usually two letters] = representation of institution's name (For institutions in London Part one is 'LD' and Part two is three letters long. This is to allow for more codes) E.g, WINGM (Gurkha Museum in Winchester); TWCMS (Tyne and Wear County Museum Service); IWM (Imperial War Museum). MDA Codes pre-date the common use of computers and are used in the marking or labelling of physical objects (as a prefix to an internally unique object number).
Subject	persistent identifier (organisation); persistent identifier (collection)
Relation	http://www.collectionstrust.org.uk/spectrum (SPECTRUM download page)

3. Survey of persistent identifier use

3.1 How the survey was carried out

As part of the WP3's work on persistent identifiers (PIDs) a limited, paper-based, survey of their use in European cultural organisations was addressed to the national contact points within the ATHENA project. The questions asked in the survey can be found in the *Appendix* at the end of the deliverable. In overview the survey asked:

- On PIDs for digital objects:
  - If PIDs are used;
  - What the benefits of them are seen to be;
  - Which types of PID services and standards are being used;
  - Are other types of PID in use and any details.

- On identifiers for organisations and collections:
  - If these identifiers are used;
  - Who is assigning them?

- On future needs:
  - What information is required in this area;
  - Any other questions.
  - Which types of PID systems and standards are being used;

20 national representatives[5] responded to the survey. The rest of this section details the results of the survey.

3.2 Are PIDs used in a country?

15 out of the 20 (75%) national representatives answered 'Yes' to this question.

On first impression this is a very encouraging percentage. However from the answers to later questions on the use of PIDs systems need to be born in mind.

3.3 Why are PIDs used?

National representatives said:

- ***Czech Republic*** – "*Persistent identification of digital documents (files or intellectual entities) which could be used for linking to different services, referencing, digital preservation etc.*"

- ***Estonia*** – "*Persistent Identifier for discovering and locating resources on the World Wide Web relies on allocating an identifier to resources; specify the location of a resource by including a protocol, domain name and the actual name of the file within which the resource resides. Memory institution/owners and files are identified.*"

5 Replies to the survey were received from: Belgium; Bulgaria; Cyprus; Czech Republic; Estonia; Finland; France; Germany; Greece; Hungary; Italy; Latvia; Malta; The Netherlands; Poland; Romania; Russian Federation; Slovak Republic; Sweden; and the United Kingdom. Other countries were unable to reply to the survey in the short period allocated for the survey.

- ***Finland*** – "*Institutional involvement in the GBIF project*[6]."

- ***France*** – "*Facilitate the referencing and 'quotability' of digitised/digital documents; rely on a sustainable system of identification of objects even if collections management system and/or technologies evolve or change.*"

- ***Germany*** – "*PIDs are essential for archiving of and long-term-access to online resources*".

- ***Greece*** – "*Unique identification of digital objects (and their copies) which can be used for easy transaction control and management*".

- ***Hungary*** – "*Using persistent identifiers will ease the work of users: they do not have to maintain the URLs*".

- ***Italy*** – "*Long term unambiguous identification of resources*".

- ***Malta*** – "*To uniquely identify the item (or group depending on the type of description). Also, the identifier is used as a classification system so as to show object class in Heritage Malta's museums and archival funds at the National Archives of Malta.*"

- ***Netherlands*** – "*http://www.catchplus.nl/2010/presentatie-persistent-identifiers-online/ See this link: we need unique and persistent identifiers in our digital environment See also: http://www.den.nl/docs/20100122103440#Conclusies*"

- ***Romania*** – "*Avoids ambiguity.*"

- ***Sweden*** – "*Nordic Museum had since the 1970s a database where all museums could log in with their PID. They closed the database around 2000, but the PIDs have become actualized again in connection with K-samsök, a national aggregator.*"

- ***United Kingdom*** – "*Use is to preserve access of digital resources on the website of the organisation. However use of PIDs is limited in terms of: types of organisation (more in libraries, but not in museums) and in numbers of objects 'protected' (often limited to specific collections and sometimes related to projects). An unpublished study found that a substantial percentage of PIDs were not in fact persistent.*"

These responses show that:
- Organisations are aware of the basic reasons for using PIDs;
- PIDs may have a restricted use in an organisation.

6 Global Biodiversity Information Facility. See **http://www.gbif.org**

3.4 Types of PIDs in use

The table below gives the number of countries in the survey, and the percentage for the survey as a whole, indicating the use of a PID type. Note that it is possible for a country to use more than one type of PID and therefore the percentages do not add up to 100%.

PID type	Number of Countries in survey: %
[No use]	5 : 25%
URI – no resolution service	11 : 55%
NBN [URN]	9 : 45%
PURL	6 : 30%
ARK	2 : 10%
DOI	2 : 10%
OpenURL	2 : 10%
OCLC	2 : 10%

The **use of URIs in 75% of the countries surveyed**, with or without the use of a resolution service is one of the great success stories. However the low take up of the resolution services available is a concern. One wonders how the issue of persistence is being addressed. However anecdotal evidence and the results of the survey suggest that there is a commitment to guaranteeing persistence.

3.5 Other types of digital identifiers in use

Some countries indicated that they used some other types of digital identifiers. Below is what they said:

- ***Germany*** – *"Own systems:*
  - *For archives: on basis of EAG (Encoded Archival Guide) in the Bundesarchiv;*
  - *For museums: on basis of ISIL-number of museum enriched by ID of the real world object in the museum."*
- ***Hungary*** – *"Cool URIs."*
- ***Italy*** – *"ICCD codes for cultural heritage objects: The Central Institute for Catalogue and Documentation (ICCD) assigns unique identifiers, called NCT code, to each physical object described in the National CH Catalogue. The code NCT is composed of three parts, the Region code (NCTR), the Catalogue Number code (NCTN), and the Suffix, the latter is only used in specific cases."*
- ***Poland*** – *"OAI IP (only for libraries)."*
- ***Romania*** – *"GUID (UUID), i.e. Global/Universal Unique Identifiers."*
- ***Russia*** – *"ISBN for books: managed by the Russian Nations Agency of ISBN."*

- ***United Kingdom*** – "*Internal to the system or service, e.g. SCRAN. Special internal 'more persistent' URLs for object landing pages."*

The answers given refer to physical objects, or to digital objects, or sometimes to both.

3.6 Are PIDs used in for institutions or collections?

This question was aimed at the general use of identifiers for institutions and collections, whether physical or digital. 12 out of the 20 (**60%**) national representatives answered 'Yes'.

Again this is an encouraging result. In terms of who is managing the assigning PIDs:

Managing Organisation	**Number: Names of countries**
Ministry of Culture	5: Czech Republic; Estonia; Greece; Netherlands; Russian Federation
State Library	4: Germany; Hungary; Sweden; United Kingdom
State agency	3: Italy; Latvia; Romania
Museum or museum organisation	3: Germany; Sweden; United Kingdom
[no PID management]	8: Belgium; Bulgaria; Cyprus; Finland; France; Poland; Malta; Slovak Republic

In some countries state libraries issue PIDs for libraries whilst museums have theirs' issued by museum-related organisations.

This mixed picture is probably the result of the nature of the organisation of cultural heritage in different countries.

3.7 Future use of PIDs

9 out of the 20 (**45%**) national representatives indicated that their countries were looking at expanding the use of PIDs for physical and/or digital objects.

Here are the national responses:

- ***Belgium*** – "*Federal Belgian Institutes: URIs with inventory number of object number reference*"
- ***Czech Republic*** – "*There is a big demand on using some system of PIDs on the national level for libraries (URN:NBN) and other ALM institutions.*"
- ***France*** – "*There is a national project which aims at harmonizing the production of cultural data at national level: several major issues (common data model for metadata, vocabulary and persistent identifiers) are under discussion.*"
- ***Germany*** – "*Organisations are in the process of evaluating the different systems.*"
- ***Greece*** – "*The Hellenic Ministry of Culture is considering using PIDs in the implementation of a National Register of Monuments that is being planned. Other Greek cultural organisations could possibly consider using PIDs in planned new projects.*"

- ***Italy*** – "*A comprehensive survey would be needed in order to check if initiatives are being studied outside the ministerial and related domain.*"
- ***Latvia*** – "*We are interested by the aspect of the practical applicability, using PIDs.*"
- ***Malta*** – "*Country code (MT) followed by a 3 letter code for the institution. We are not sure whether we should include a code showing the collection type.*"
- ***Slovak Republic*** – "*URI is used by several cultural organisations for presentation of their collections.*"

3.8 Need for information on PIDs

To the question on what kind of information or documentation organisations wish to receive on the subject of PIDs, following answers were given:

- "*Any kind of information / documentation on the issue*".
- "*How to implement PIDs, what are the benefits for the local organisation, what are the economic implications for their implementation*".
- "*International best practices, technologies and description of adaptability to different scenarios*".
- "*Guidelines and recommendations taking into account specificities of each cultural field (archives, libraries, museums) and each type of document/object*".
- "*New developments. A survey on the experiences of European museums that use PIDs*".
- "*Advice of what to use as a PID and how to use – embed in systems in a simple guideline format so as to be used by inexperienced cultural organizations*".
- "*Description of the main PID standards and their use environment*".
- "*It would be good to have a guideline on recommended methodologies to assign organisation codes. It would also be very practical to have recommended guidelines on how to construct the subdivisions of the PID based on experience especially regarding the collection subdivision. After the assignment structure of the PID, it would be also helpful if the manual would contain a list of advantages and disadvantages on the different technological solutions available on the storage of PIDs, namely URIs etc*".

The authors of this deliverable hope that it meets some of these needs.

4. Requirements for persistent identification

Looking at the literature for PIDs many authors give a set of requirements for their successful implementation. Here are some examples:

Reference	Requirements
Bellini, Emanuele; Cirinnà, Chiara; and Lunghi, Maurizio. [IT & EUR] *Briefing Paper: Persistent Identifiers for Cultural Heritage.* Digital Preservation Europe. (2009). **http://www.digitalpreservationeurope.eu/publications/briefs/persistent_identifiers.pdf**	*"A CH institution should choose a PI infrastructure using the following system requirements as a guideline:* • *Global uniqueness* • *Persistence* • *Resolvability* • *Reliability* • *Authority* • *Flexibility* • *Interoperability* • *Costs"*
Hilse, Hans-Werner and Kothe, Jochen. [NL] *Implementing Persistent Identifiers.* Consortium of European Research Libraries (CERL). (2006). **http://www.knaw.nl/ecpa/publ/pdf/2732.pdf**	"... that documents can be identified unambiguously and located by those who need them."

Nicholas, Nick; Ward, Nigel; and Blinco, Kerry. [US] 'A Policy Checklist for Enabling Persistence of Identifiers' in *D-Lib Magazine*. January/February 2009. Volume 15 Number 1/2. Corporation for National Research Initiatives. **http://www.dlib.org/dlib/january09/nicholas/01nicholas.html**	*"... that well-managed resources remain available and accessible over the long term."* *"Persistence involves a guarantee to the user that the identifiers will be kept up to date, and this requires an ongoing commitment of resources. For that guarantee to be meaningful, identifier managers cannot undertake to identify everything in their domain: they need to decide on the resources for which they will provide persistent identifiers."*
Preserving Access to Digital Information (PADI). [AU] *Persistent identifiers*. National Library of Australia. (2002). **http://www.nla.gov.au/padi/topics/36.html**	[on ARKs]: "The scheme is underpinned by three requirements: " • A link from the object to a promise for stewardship; • A link from the object to metadata which describes it; • A link to the object itself (or appropriate substitute).
Tonkin, Emma. [UK] 'Persistent Identifiers: Considering the Options' in *Ariadne*, Issue 56. UKOLN. (July 2008). **http://www.ariadne.ac.uk/issue56/tonkin/**	Looks at: • Opacity • Authority and Centrality • Semantics, Flexibility and Complexity • Present-day Availability and Viability • Technical Solution versus Social Commitment

Perhaps most useful set here is that created by Digital Preservation Europe which we adapt and add to here to a set of 10 requirements. Some of these are addressed to PID system itself and others to the cultural heritage organisation.

4.1 Managing organisations

Some requirements are regarding the operations of the organisation which is considering using PIDs:

Uniqueness environment

A PID is label that is associated with something in a particular environment. On the Internet is should be globally unique, but may only be unique in combination with a limited name space. In the 'worse' case it may only be unique within an organisation's own systems.

- **Organisations should be clear, and make public, in which environment its PIDs are unique.**

Persistent

Persistence refers to lifetime of an identifier. During this lifetime it should not possible to reassign it another resource or to delete it. If an organisation can guarantee that a PID will be managed so that it will survive changes to ownership, and PID system them an external user can be confident of its persistent

Therefore:

- **Organisations should commit themselves to the persistence of their PIDs and make clear to others what they mean by 'persistent' and how this will be implemented.**

Resolvable

Choice to use PIDs does not imply that an external human user will be able to access anything that they can use effectively. Therefore:

- **Organisations should be clear, and make public, information about which, if any, their PIDs resolve to an available resource.**

Cost effective

Resources, particularly financial resources, are scarce in the cultural heritage sector. In addition organisations have a general mission to provide access to their items free of charge for non-commercial use. Therefore:

- **Cultural organisations should use PID systems that are free of charge, or very low cost in relationship to their available resources.**

Supported by policy

Collections management, which includes access to collections and collections access, is a balance between the competing needs of the organisation and its users. Also for anything to be successful it must be supported by the senior management who decide policy. Therefore:

- **The use of PIDs should be part of the written policy of the organisation.**

Managed by embedded processes and procedures

Having policies on PIDs is only the start in the implementation of a PID system (though an important part). The policy mandate must be made real by how an organisation operates. Therefore:

- **The management of an organisation's PID system should be part of the written processes and procedures of the organisation.**

These last two will be explored further in this work package's next deliverable: D3.5 – *Technical and policy infrastructure to support persistent identifiers.*

4.2 Persistent identifier systems

Other requirements are regarding the operations of the PID system being considered:

Reliable

For a PIDs system to function reliably these issues have to be assessed:

1. It should always be active (e.g. backed up, with redundant technology).
2. The register of PIDs should be updated (preferably automatically).

Therefore:

- **Organisations should evaluate and be assured of the technical reliability of a PID system (including their own) before adopting it.**

Authoritative

Some PID systems are dependent on responsible organisations who: manage the system, assign identifier; and resolve the identifiers to resources. Some services are provided by public institutions like national libraries and archives. For system to be effectively supported a system a responsible organisation must be able to demonstrate its commitment. Therefore:

- **Organisations should evaluate and be assured of the authority and credibility of a PIDs system's provider before adopting that system.**

Flexible

A PID system will work more effectively if it can handle the requirements of different types of collections. Parts of collections may be curated at different levels of 'granularity', from parts of

objects, to individual objects, to collections objects. The latter has an unbounded number of individual elements. Therefore:

- **Organisations should use PIDs systems that are flexible enough to represent the granularity their collections.**

Interoperable

This is vital to ensuring that cultural content can be shared and used by as a large a set of users as possible. Many PID solutions were designed for specific domains. Therefore:

- **Organisations should use intellectually open standards for the implementation of PIDs**

5. Advice and best practise

5.1 Physical objects

Advisory bodies, such as Collections Trust in the UK, give advice about numbering objects. What follows is based on their advice. It is not intended to be definitive but only to illustrate the range of options and issues.

The part of the *SPECTRUM Minimum Standard* for the *Acquisition* procedure which deals with the numbering objects accessioned into a collection requires that an organisation:

1. *Ensure that a unique number is assigned to, and physically associated with, all objects;*
2. *Ensure that accession registers are maintained, describing all acquisitions and listing them by number.*

The accession number is the number allocated in the accession register. This only applied to items formally acquired by the organisation. A unique number should be assigned to each object or group of objects. The organisation should have a policy for deciding the format of its numbers. This should be recorded in a *procedural manual*.

There are two common approaches for numbering physical objects:

1. A simple running number system, e.g. **14603**; **14604**; **14605**; **14606**; **14607**;
2. More common system is to use the year of accession followed by a running number, e.g. **1991.3**; **1991.4**; **1991.5**; **1991.6**; **1991.7**.

   Do not abbreviate the year to just two digits, as hopefully most museums will survive for more than a century. It can also cause confusion. For example, does '**64.68**' mean 1964 or 1968? Do not place the year element last as this will cause a problem for computerised sorting of records. So **1991.5** is correct rather than **91.5** or **5.1991** or **5.91**.?

Numbering individual items

Items that are given to the organisation individually should be given different accession numbers, then the accession number serves as the identity number. Four objects that were acquired separately from different sources would be numbered, e.g. **1999.1**; **1999.2**; **1999.3**; **1999.4**.

Numbering groups of items

Individual numbers for each item

Some organisations give a different number to each individual item in a group. For example a collection of four glass negatives could be allocated a separate number

e.g. **1999.21**; **1999.22**; **1999.23**; **1999.24**

This approach can break down if an organisation receives a large collection from a single source, for example a collection of several thousand glass plate negatives or the complete contents of a shoemaker's workshop. The organisation will be unable to allocate numbers to the next entry group until the large collection has been fully numbered. This can create a recording backlog.

Part numbers

A group of items could be allocated the same accession number and a suffix added to create a unique identity number. For example a group of four objects brought into the museum together might be accessioned as **1991.24**. Each individual item would then be numbered:

e.g. 1991.24.1; 1991.24.2; 1991.24.3; 1991.24.4 ….

One number for a group of items

Where large numbers of similar objects are physically grouped together they can be numbered as one 'object'. For example, a group of pottery sherds in a secure container the container may be numbered, the contents counted and the total recorded. A card of 20 buttons can be allocated one accession number, not 20.

Objects collected during fieldwork

Many objects entering organisations are the result of fieldwork events, such as an archaeological excavation, natural science expedition, or similar which were not carried out by that organisation.

There are two options for assigning identity numbers:

- *Using a museum accession number*
  This is frequently the best option for long-term curation, as it will fit in with existing museum systems and it means you will not have to re-mark objects. This may be the preferred method for museums which regularly receive archaeological archives from a number of different sources.

- *Using the collector's site code*
  The site code can be incorporated into the accession number. However, as site codes might be alphanumeric so organisations need to ensure that that their information systems can logically handle these codes. A site code usually includes the location and date of the excavation, which can be an advantage, using it also avoids further proliferation of numbers on an object.

5.2 Digital objects

As with the advice given for technical standards in the deliverable D3.1 appropriate advice and best advice for the standards can be found in the Minerva Project's:

Technical Guidelines for Digital Cultural Content Creation Programmes:
http://www.minervaeurope.org/interoperability/technicalguidelines.htm
[with links to various versions]

It is worth quoting the advice in full[7].

> *"Digitised resources **should** be unambiguously identified and uniquely addressable directly from a user's Web browser. It is important, for example, that the end user has the capability to directly and reliably cite an individual resource, rather than having to link to the Web site of a whole project. **Projects should make use of the Uniform Resource Identifier (URI) for this purpose, and should ensure that the URI is reasonably persistent.** Such URIs **should not** embed information about file format, server technology, organisation structure of the provider service or any other information that is likely to change within the lifetime of the resource.*

[7] p73 of the current (2008) English Language version.

*Where appropriate, projects **should** consider the use of OpenURLs, Digital Object Identifiers or of persistent identifiers based on another identifier scheme."*

The consideration of which external service (if any) an organisation will be using should be structured using the requirements given in *Section 4* above. Consider:

- Uniqueness environment of the PIDs;
- Persistent of the identifiers;
- Resolvability to give information;
- Cost effectiveness of the service;
- How supported by organisational policy;
- How managed by embedded in processes and procedures;
- Reliability of the service;
- Authoritative nature of the service supplier;
- Flexibility of solution with regards to granularity;
- Interoperability with other systems and services.

ICOM best practise for museums

At the time of writing CIDOC (International Committee of Documentation of the International Council of Museums) is in the process of proposing a recommendation on 'linked data' to ICOM which includes useful advice and best practice guidance on the creation and management of PIDs in museums[8].

Overview:

- Museums should take control of how the physical objects in their collections are identified online;
- Online, museum objects should be uniquely identified by suitable URIs;
- An object (or set of objects) should have one authority that assigns the URI for the object. This authority must be known to, or easily discovered by interested parties;
- The URI authority for an object should be the museum that curates the object;
- The URI should be derived in a simple way from the published object identifier (e.g. inventory number) used to manage the object;
- If the URI is part of an information service (e.g. linked open data service) it may be resolvable to a description of the object.
- If the ownership or curator of an object changes the published URI should not change. However it is possible create a new URI, that reflects that change, and to redirect the old URI to the new one

8 **CIDOC**. *ICOM recommendation on Linked Open Data for museums.* [draft] (2010).

Underlying principles:

- For those parties who are interested, the association between the object and the managing organisation should be: known; or easy to guess; or possible to find out.
- The managing organisation should be able (and willing) to resolve disputes about the relationship between the identifier and the objects.
- The managing organisation should not be in competition with another doing the same job for the same set of objects.

***How to create a URI for an object*:**

1. The museum should decide on a 'base URL' for the URI for objects. This may be:
    - Within the domain name of the museum's main website **or**
    - Separate distinct domain name.

    The second option has the advantage: of continuity if the main website domain name changes; and the ability of balance of server load for frequent requests;
2. Extend the base URL with the inventory number of the object.

A suggestion for the British Museum (http://www.britishmuseum.org) is that all objects of the Museum should be identified as the following:

http://collection.britishmuseum.org/object/[PRN number]

E.g. The Rosetta Stone has the PRN number: YCA62958. Therefore its URI would be:

http://collection.britishmuseum.org/object/YCA62958

Other organisations should make their own decisions on how to structure their URIs.

What if a museum is unable (or unwilling) to create and manage its URIs?

The *Recommendation* also foresees situations:

- Where a museum wishes to delegates the creation of URIs to another organisation (i.e. a PID service).

  In this situation parties should manage their relationship in such a way to ensure that the PIDs created follow agreed rules and are recorded by both.

- Where a museum does not create or manage PIDs.

  Here, for example, an aggregator may come forward to take over the role. To do this the aggregator should: publish its intentions; and be prepared adjust its management the affected objects and PIDs if the museum decides to takes over the creation or URIs at a later date.

5.3 Institutions

The basic advice is:

An institution should have a recognised, publically available, and unique identifier.

Therefore organisations, if they have not already done so, seek to comply with this guideline. A country may already have services[9] in place, and an organisation should join that service.

The Minerva *Guidelines* do not cover the persistent identification of institutions. However advice and best practice similar to that given for objects should be followed (see above) for institutions in the online environment.

5.4 Collections

Advice on PIDs for collections can be viewed as similar to objects and therefore similar advice can be given (see above). In the digital environment the Minerva *Guidelines* say[10]:

> ***"Projects may also wish to ensure that logical sets within the resources they are providing are uniquely and persistently addressable."***

There are few large scale aggregations of collections descriptions online. The only major example the authors are aware of is the multilingual service created by the MICHAEL project[11]. This had technical results of:

- *MICHAEL Data Model* – for multilingual digital cultural heritage inventories;
- Open source technical platform for national instances.
- Interoperability protocols for contributing data to the European service
- International portal
- Methodology and model for simple deployment of the system.

Organisations may consider using the experience of MICHAEL when considering setting up an information service for collection descriptions. However MICHAEL has not addressed the issue of the unique and persistent identification of the digital collections.

9 E.g. The ISIL website (**http://biblstandard.dk/isil/**) lists

10 Op cit. p74.

11 See: **http://www.michael-culture.org**

6. Conclusions

It is clear that cultural heritage organisations in Europe generally know what PIDs are and why they are important, especially in the digital environment. Where PIDs are employed most use the URI (URL/URN) standard, though its use is sometimes 'hidden' to the cultural organisation by a technology or service layer. This situation can be a good thing where there is only a limited technical knowledge within an organisation, but having basic information (which this deliverable presents) will improve the knowledge of the sector as a whole.

The situation for the resolution of PIDs to deliver a resource (e.g. a record on a portal) of interest to the user is less clear. As service providers (e.g. Europeana) have found the persistence of URIs is not always something that is maintained. The PID systems on offer have had a limited implementation. Most were developed within a library context and therefore offer services aimed at this domain. Museums have special needs to support a wide range of identifiers for physical objects, metadata descriptions, digital surrogates, collections and institutions. All these needs have also to be met within a limited resource budget. This often restricts their implementation by museums (and probably archives).

However there are reasons to be hopeful. Institutions are open to the use of PIDs and willing to implement them on their terms. There is recognition that where resources are being aggregated, e.g. in Europeana, national and regional portals, that the argument for PIDs becomes compelling and is essential if these aggregations are to work efficiently. Some national systems are in place or are planned, but it is also an issue that must be met by individual organisations.

The next deliverable of the work package will deal with the policy and technical infrastructure that needs to be in place.

Appendix 1: Use of persistent identifiers survey questions

A short survey was carried out by WP3. The questions were given to the National Contact Points for the countries in the ATHENA project. The questions asked were:

Question 1a Do you know if Persistent Identifiers are used by any cultural heritage organisations in your county? Please answer **YES** or **NO**.
If you answered 'NO' go Question 3a
Question 1b Please tell us what benefits these organisations think they get from using PIDs (if you know).
Question 2a Which of the following types of Persistent Identifiers are being used in your country? Please answer **YES** or **NO** to each: • Uniform Resource Identifiers (URI) • Persistent URL (PURL); the handle resolver system • Archival Resource Key (ARK) • Open URL • OCLC • National Bibliography Numbers (NBNs)
Question 2b Are there any other types of Persistent Identifiers are being used in your country? Please answer **YES** or **NO**.
Question 2c If you answered **YES** please give details.
Question 3a Does your country use codes to uniquely identify cultural heritage organisations and collections? Please answer **YES** or **NO**.
If you answered 'NO' go Question 4a
Question 3b Does your country use codes to uniquely identify cultural heritage organisations and collections? Please answer **YES** or **NO**.
Question 3c By whom are the assignments of these codes managed in your country (e.g. Ministry of Culture)

Question 4a If PIDs are not in use in cultural heritage organisations in your country, are organisations considering using them in the near future? Please answer **YES** or **NO**.
If you answered 'NO' go Question 5a
Question 4b Please provide some details about what is being considered.
Question 5a What kind of information/documentation do you wish Athena and/or Europeana to provide you on the subject of PIDs?
Question 5b If you have questions for us please ask them here.