

PERSISTENT

IDENTIFIERS (PIDs):

RECOMMENDATIONS

FOR INSTITUTIONS

**PERSISTENT
IDENTIFIERS (PIDs):
RECOMMENDATIONS
FOR INSTITUTIONS**

edited by

ATHENA WP3
Working Group
“Identifying standards
and developing
recommendations”

texts by

Gordon McKenna,
Collections Trust (UK)
Roxanne Wymys,
Royal Museums of Art
and History (Belgium)

The text of this booklet
is included
in deliverables 3.4 and 3.5,
of the ATHENA project
and is based on a survey
of the content that
ATHENA partners
contracted to provide
to Europeana through
the ATHENA project.

ATHENA

general co-ordinator
Rossella Caffo

design

mt milani,
geo graphic sdf

ATHENA logo

Susan Hazan

web version

<http://www.athenaeurope.org/index.php?en/110/promotional-material>

This work by ATHENA
project is licensed under
a Creative Commons
Attribution
Non-Commercial
Share Alike Licence
(CC-BY-NC-SA)
<http://creativecommons.org/licenses/by-nc-sa/3.0/>



table of contents

Foreword

Introduction

1. Persistent identifiers
A briefing note
2. Persistent identifier policy
in context
3. Standards landscape
 - 3.1 Physical objects in museums
 - 3.2 Digital objects
 - 3.3 Services
 - 3.4 Collections in museums
 - 3.5 Institutions
4. Managing organisations
5. Persistent identifier systems

Foreword

It is the aim of the ATHENA project to support especially museums in providing object data for publication in Europeana. Thus ATHENA is about access to digitised or digital cultural heritage held in museums and other institutions.

In order to ensure that information about an object and the object itself, its digital copies can be related to each other and can be retrieved easily at different points in time and from different places it is necessary to use “persistent identifiers” (PIDs).

This booklet is intended as a short introduction why “persistent identifiers” are needed and what systems are currently available.

We hope it is helpful to the community!

Monika Hagedorn-Saupe

WP2 - Awareness and dissemination:
Enlarging the network and promoting the service

1. Persistent identifiers

A briefing note

The basics

Although the subject of persistent identification (PIDs) can seem like a technical area of an organisation's work, it is actually straightforward. It is about:

- **Identification** – Using agreed strings of alphanumeric text (identifiers) to provide access, like a key, to information in paper-based, in-house computer, and online systems. They also provide access to physical objects using attached marks or labels.
- **Persistence** – Managing the identifiers in order to maintain the access.

Cultural heritage organisations use persistent identifiers for:

Cultural entity identification

This concerns the persistent identification of physical objects, the information describing those objects (metadata), and their associated cultural entities (e.g. people, places and events).

A physical object and an organisation's own metadata about the object usually have the

same identifier. Surrogates for an object (e.g. photographs, digital images, and 3-D models) should have different, but perhaps related identifiers.

Organisations may wish to use the in-house identifiers for access to in-house created information about associated cultural entities. However some of these entities already may have published recognised identifiers (e.g. ISBN for books) which can be used. Organisations may also use the published information too.

Collections management identification

This usually covers the identification of three things:

- **Collection management events** usually associated with physical objects (e.g. acquisition, conservation, movement, IPR licensing, disposal, and exhibition). Surrogates may also have events associated with them.
- **Display and storage locations** within an organisation. This is used especially for the movement of physical objects, but can also be applied to locations in a computer file system for digital surrogates (e.g. the licensing of a photograph of an object).
- **The organisation itself**. Used to externally identify the organisation.

The benefits

Implementing and maintaining the use of persistent identifiers in an organisation is

a classic change-management process. It requires both investment, in time and other resources, and commitment from staff. The return on this investment, however, can be significant.

When constructing a business case consider the:

Direct benefits

- Ability to retrieve information, and physical objects, quickly and simply.
- Cost-savings in staff time spent handling objects or re-identifying information.
- Greater confidence in managing your information and objects.
- Improved access to information for all areas of curatorial expertise and other departments.
- Using a standards-based approach will support applications for funded projects (e.g. for the EC).

Indirect benefits

- Greater clarity to funders about the extent and content of your collections.
- Better-managed intellectual property leading to greater opportunities for use and commercial activity.
- Enhanced ability to publish information and to make your collections visible online.
- Ability to share your information through portals: local, regional, national, thematic and international (e.g. Europeana).
- Ensuring that information and knowledge is used effectively in the future even if local staff changes.

2. Persistent identifier policy in context

Policy for the management of persistent identifiers is part of the overall policy needs of a cultural heritage organisation. It operates within the wider environment of the management of the collection. Here we explore this context. Collections management is a major activity of any cultural heritage organisation. For it to be successful a cultural heritage organisation must seek to balance between:

- Giving access to collections and ensuring the preservation of collections;
- The needs of the collections and the needs of the people who want to use them;
- Organisational priorities: ranging from short term, to medium term, to long term.

In order to meet the challenge of balanced collections management, the British Standards Institute developed, with the help and sponsorship of cross-domain set of cultural heritage organisations, a Publicly Available Specification (PAS 197) on a *Code of practice for cultural collections management*¹. It was published in early 2009, and will be reviewed in a few years with the aim of it forming the basis for national, and possibly international, standard in this area.

¹ British Standards Institute. 2009. *PAS 197, Code of practice for cultural collections management*.

The *Code* aims to:

- Enable an organisation's top management to take a strategic and integrated approach to collections management.
- Provide a blueprint for creating strategies that are sustainable.
- Take into account the legal environment within which an organisation operates.

In order to understand the collections management framework² in the *Code of practise* it is perhaps best to begin by giving definitions some of terms:

2 The *Code of practise* uses the term 'framework' instead of 'system' in order to avoid confusion with "collections management system" – a computer database used to store information.

- ***Collections management***: The strategies, policies, processes and procedures of an organisation connected with: the development of; information held about; access to; and care of its collections.
- ***Top management***: The person or group of persons at the highest level of an organisation who direct and control its activities.
- ***Policy***: The overall intentions and direction of an organisation. These are formally expressed, e.g. in a written statement, by its management. It forms the starting point for the setting of objectives and taking actions.
- ***Process***: A set of activities which are interrelated or interact with each other, and that have inputs and outputs.
- ***Procedure***: The way, documented or not, an activity or process is carried out.

The framework is structured in a hierarchy:



Figure 1. *Collections management framework (in a museum)*

At the top level of the hierarchy is the:

- **Mission statement** – A strategic statement giving a cultural heritage organisation's fundamental purpose, especially with regard to its collection.

The mission statement informs the different areas of collections management policy which are based on four different strands of activity:

- **Collections information**
- **Collections development**
- **Collections access**
- **Collections care and conservation.**

These policies are met by (implemented) by processes and procedures which the organisation uses. These may be based on a standard, like *SPECTRUM* for museums, but must be documented in the form of a written manual adapted for the organisation.

It should be noted that the creation of the mission statement, policies, processes and procedures is not a one-time process. There must be a commitment to continual review and change of the framework.

3. Standards and services landscape

Each persistent identifier (PID) standard or service is described 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 included 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 deliverables edited within the project. 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.

3.1 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: <ul style="list-style-type: none">• 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.

3.2 Digital objects

Standards

There are three, interrelated, standards:

URI

Title	URI • Uniform Resource Identifier
Creator	T. Berners-Lee (W3C/MIT); R. Fielding (Day Software); L. Masinter (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)

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	<p>A URI (i.e. a string) that specifies:</p> <ul style="list-style-type: none"> • Where a resource is available; • The mechanism for retrieving it. <p>Form: <code>scheme://domain:port/path?query_string#fragment_id</code></p> <ul style="list-style-type: none"> • <code>scheme</code> – defines the namespace, purpose, and the syntax of the remaining part, examples: <code>http</code>, <code>https</code>, <code>gopher</code>, <code>wais</code>, <code>ftp</code>. • <code>domain:port</code> – gives the destination location for the resource (domain name or IP address). Port is optional, if absent the default is used (for <code>http</code> default port = 80). • <code>path</code> – used to specify and find the resource • <code>?query_string</code> – used to pass data to a piece of software to enable retrieval • <code>fragment_id</code> – used to specify a part or a position within the overall resource <p>E.g. http://www.athenaeurope.org/index.php?en/91/information-on-the-project (the 'About us' page on ATHENA project website)</p>
Subject	persistent identifier (Internet); persistent identifier (book); persistent identifier (periodical); persistent identifier (audiovisual);
Relation	URI (Uniform Resource Identifier) 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; NBN ³
Subject	persistent identifier (Internet); persistent identifier (book); persistent identifier (periodical); persistent identifier (audiovisual);
Relation	URI (Uniform Resource Identifier) URL (Uniform Resource Locator)

- 3** 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.

3.3 Services

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

PURL & 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	<p>A URL pointing to a resolver (e.g. Handle) which redirects to current URL; Resolver software (OCLC free). Form: Has 3 parts –</p> <ol style="list-style-type: none">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 <p>E.g. http://purl.oclc.org/oclc/oluc/32127398/1</p> <p>-----</p> <table><tr><td> </td><td> </td><td> </td></tr><tr><td>Protocol</td><td>resolver address</td><td>name</td></tr></table>				Protocol	resolver address	name
Protocol	resolver address	name					
Subject	persistent identifier (digital object)						
Relation	<p>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]</p>						

PURL & HANDLE SYSTEM

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]

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	<p>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.</p> <p>Form: Divided into two parts:</p> <ol style="list-style-type: none"> 1. Prefix – identifies the registrant of name; 2. Suffix – chosen by the registrant to identify the document associated with the DOI. <p>E.g. doi:10.345/document.identifier12345</p> <p>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.</p> <p>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:'</p>
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]

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)

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	<p>A URL scheme which can identify both physical and digital objects.</p> <p>Form: [http://NMAH/]ark:/NAAN/Name[Qualifier]</p> <p>NAAN = Name Assigning Authority Number - mandatory unique identifier of the organization that originally named the object</p> <p>NMAH = Name Mapping Authority Host - optional and replaceable hostname of an organization that currently provides service for the object</p> <p>Qualifier = optional string that extends the base ARK to support access to subcomponents of an object or its variants (e.g. version, language).</p>
Subject	persistent identifier (digital objects); persistent identifier (physical objects)
Relation	https://confluence.ucop.edu/display/Curation/ARK (webpage)

3.4 Collections in museums

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

MDA Codes (see below, p. 28), 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.

3.5 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 institutions.

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 ('-'): <ul style="list-style-type: none">• 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)

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	<p>An alphabetic string of usually five letters (some national museum have shorter codes). Form: The code is made up of two concatenated parts:</p> <ul style="list-style-type: none"> • [Part one – usually three letters] = representation of location of the institution • [Part two – usually two letters] = representation of institution's name <p>(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).</p>
Subject	persistent identifier (organisation); persistent identifier (collection)
Relation	http://www.collectionstrust.org.uk/spectrum (SPECTRUM download page)

4. 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 it 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, then an external user can be confident of its persistency. 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.*

5. 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 a system to be effectively supported the 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.**

Printed in the month
of April 2011
by Repro Stampa Ind. Grafica,
Villa Adriana - Tivoli (Roma)