

Persistent identifiers for the NRGL – decisive criteria

The document was produced in order to aid in the selection of the persistent identifiers for the National Repository of Grey Literature during solving of the project „Digital Library for Grey Literature – functional model and pilot implementation, which is supported by the Ministry of Culture. The project is being solved by the State Technical Library in cooperation with the University of Economics in Prague. See <http://nysl.stk.cz>

Introduction

Web technologies have created ample opportunities in the area of availability of electronic information and thus induced dramatic changes in the procedures of making various electronic documents accessible by libraries, archives and other providers of information to their users.

In this way the digital libraries can offer access to information in distributed systems in open network where the location of the users or information is unimportant, and thus secured by means of quite standard software on the part of the provider as well as of the user, who can only have a web browser.

Distributed digital libraries have two goals in this environment:

- to store and archive information;
- to keep and maintain access to the information, so that users could find them, the identifiers must lead to the documents even after a longer period of time.

A success of the digital library, and in particular a distributed one, is based on a long-term consistent links to the resources and persistence of the identifiers and links, which make these resources accessible.

The hyperlink URL (Uniform Resource Locator) is a common indicator, identifier and a link to the information in the web environment. However, the longer period passes, the more links become broken or irrelevant and do not access the original source.

Cause of these broken links can be:

- the information which the link made accessible has been removed for any reasons (the user does not assume it to be important, it does not reflect the current situation of the matter, it is not update and so on)
- the file was reasons removed for any reasons (reorganization of the web or file structure, change of the web domain, change of the owner of the organization)

A solution to these problems lies in organisational measures, minimising the risk of corruption of the links, in particular by using the system of persistent identifiers. There are various kinds of such systems (Handle, URI, URN, DOI, OAI, NBN, etc. – see hereinafter).

It must be taken into account that a system of the persistent identifiers does not solve the problem on its own, but it can be effective only in case it is maintained. Should the source of information be moved, the new location must be linked to the persistent identifier in the selected system. This function is usually a feature of a the so-called resolver database (database of persistent identifiers, mapping them to the current database of information resources).

Organisation that decides to use the system of persistent identifiers must internally support such a system, to create organisational rules regulations, and to answer many

questions and solve many problems on the practical level. The main objective is to design the identifier itself, and one of the main issues to consider is whether the organisation decides to connect to the existing implementation of the system of persistent identifiers, or implements its own system including the resolver. To define the identifier and to deal with this issue, many questions that this document brings up must be answered.

There are many requirements that a system for persistent identifiers must meet, regardless of the type of material.

Uniqueness

The need of uniqueness will be met by naming the central authority of the system adopted. The identifiers must be unique within the organisation's digital resources, but they can also be unique globally. Global uniqueness is met in case they use a system that is used worldwide and it is being guaranteed either 1) by a national authority in each country, or 2) by a global central authority.

Commitment to Persistence / Securing the persistence

The organisation must maintain the association of the current location of the resource with the persistent identifier. It is important that the resource identified by the persistent identifier should not be moved or removed without updating the location information in the register.

Extensibility

This system must be extensible and capable of accommodating all resources which require an identifier.

Flexibility

The identifier system will be more effective if it is able to accommodate the special requirements of different types of materials or collections. The access "one size for all" is not always the most sensible. The appropriate level of intelligence to support the processes and system for various uses may be incorporated where it is useful and omitted in cases where it serves no purpose.

Ease of Use

Although not absolutely critical, and not essential for machine-generated persistent identifiers, a system will generally be more successful if it is easy to understand and apply, and if it lends itself to short and easy to use citations. It is thus advisable to avoid long and unintelligible identifiers where possible unless they are intended purely for internal use or machine processing.

Decisive criteria for the choice of the persistent identifier (PID) for the NRGL

Type of identifier

A persistent identifier may be a random collection of characters with no associated semantics that contains no information about the object it identifies – a dumb number. On the other hand, it can be a system that is programmed for the purpose of identification. Such a system of intelligent identifiers may have various degrees of complexity and specificity. As a general rule, it is easier for humans to remember and use identifiers with built in mnemonics rather than a sequence of meaningless characters, although for machine processing this is irrelevant. Intelligent identifiers are also easier to make unique, especially amongst organisational units, but in particular also globally. Generating of string of random characters need to be carefully controlled and monitored across all of the organisation with the aim to ensure uniqueness, which is not secured outside the organisation. An intelligent identifier may include relation information whereby smaller components are identified by a reference to larger entities, or collections of which they form a part.

Options:

- a string of random characters (dumb number)
- Intelligent identifier

Hierarchy

One of the most common uses of intelligent identifiers in the library world is to incorporate the relational information that reflects the organisation and the hierarchies of the digital collections or aggregated resources. Smaller components, such as digital page images or lower-level aggregated resources, are identified by a reference to larger collections or entities, to which they belong. This form of identification is especially useful for the digital surrogates of physical collections, although it can be used to reflect the born hierarchy of the digital collection, such as journal issues and articles. This system of hierarchy is not necessary in case each level of a magazine / a journal, for example a title, volume, issue, page has its own identifier, so there is no need to have the information of the hierarchy provided by the syntax of the identifier at first sight. Most users do not think this way and the problem can be caused by the fact that the hierarchy was changed (not for the journals, but for example for artificially created collections of particular documents).

Options:

- to support
- not to support

Granularity (the level of a link)

The decisions concerning the level of detail of the link, at which the persistent identifiers will be assigned, depend on the perceived needs of the stored material. The granularity will be different for various applications and materials. For many purposes, needs can be met by citing a top level web page which serves as an entry point to a self referencing collection of web files, or by citing at the journal article, item or chapter level. However some applications may require a finer level of detail. The level of granularity must be defined by each institution for each type of document.

Example of monograph:

- As a whole

- Chapters
- Pages
- Images

Versions

Versions of the same document at first sight may differ in a variety of ways, can have different content, different format or different resolution in the same format. Each version of an object, when we monitor them, will require a separate persistent identifier. The relation between the versions can be reflected in the identifier by means of a version code, a date, or a version type, or in metadata. Therefore it is important to consider whether and how to reflect the relationships between the individual versions.

Options:

- Not to monitor the versions
- Content
- Format
- A different resolution in the same format

Securing authority / Naming authority

The securing authority is the most important for the implementation of a successful PID system that is to secure the organising. The securing authority is to ensure the persistency of the identifier whereby it bears the responsibility for its long-term persistence. The securing authority provides for administration of the resolver for the persistent identifiers, i.e. allocating namespace to various institutions and in particular the maintenance of the register of relations between PID and the digital objects and updating of this register. Allocating of the PID themselves can be secured by a central institution, or an institution having its own name space (see hereinafter). The securing authority does not have to specialize in grey literature, but it has to be able to secure the selected scheme of PID and its persistence.

Options:

- State Technical Library
- National Library of the Czech Republic
- Other

Recommendations of identifiers for a local repository

Owing to fact that the aim of the project is to propose recommendations for a local repository of grey literature, we have included it herein as well. The recommendations will take into account the need for the simplest implementation for the local repository of grey literature and the need to secure the compatibility with the NRGL.

Options:

- To share the identifiers with the NRGL
- To have one's own identifiers of any kind
- To have one's own identifiers according to the concept of the NRGL

Generating of identifiers for a local repository

The decision to secure the generating of identifiers is followed by recommendations for the identifiers for local repositories. depending on the selected scheme and the central authority.

Options:

- They will be given a list / A list will be given / allocated
- They will have a name space for generating of the identifiers allocated
- They will have to install SW for generating

Availability of a solution

This aspect is important in relation to the timetable of the project. A selected scheme of the persistent identifiers including the resolver must be available by the end of 2009 at the latest, depending on the plan of the implementation of SW.

Options:

- Immediately
- In the course of 2009
- A solution is being developed

Services of the resolver

The main function of the resolver is the maintenance of the relational register between PID and the digital objects and allocating of new unique PIDs. Further the resolver provides for other services such as search and delivery of metadata or a individual documents and so on.

Options:

- Search of a valid url address
- Delivery of metadata
- Delivery of digital objects
- Cooperation with local resolvers

Financial criteria

In respect of the long-term prospective, financial criteria are important and cannot be omitted, which we will consider within the budget limits.

Options:

- No purchase cost – free of charge from now on as it is at present
- Purchase and maintenance cost
- Regular yearly fees
- Free of charge

Financial criteria of local repositories

As for the local repositories, the financial criteria are also very important. Therefore we have included them into the advisory document.

Options:

- No purchase cost – free of charge from now on as it is at present
- Purchase and maintenance cost
- Regular yearly fees
- Free of charge

Research

BELINI, Emanuele; CIRINNA, Chiara; LUNGHI, Maurizio. Persistent Identifiers for Cultural Heritage [online]. Fondazione Rinascimento Digitale. [cit. 2008-11-09]. Retrieved from [www:](http://www.digitalpreservationeurope.eu/publications/briefs/persistent_identifiers.pdf)

http://www.digitalpreservationeurope.eu/publications/briefs/persistent_identifiers.pdf.

CUBR, Ladislav.; MELICHAR, Marek; HUTAŘ, Jan. State of implementation of persistent identifiers in the National Library of the Czech Republic and outlook to the future.

In Seminar to accessing of grey literature 2008 : 1st year of the seminar focused on the problem of storage and accessing of grey literature, 8. 10. 2008 [online]. Praha : State Technical Library, 2008. Retrieved from WWW:

http://nusl.stk.cz/images/PID_text.pdf. ISSN 1803-6015.

HILSE, Hans-Werner; KOTHE, Jochen. Implementing Persistent Identifiers : Overview of concepts, guidelines and recommendations [online]. London : Consortium of European Research Libraries; Amsterdam : European Commission on Preservation and Access, C2006. [cit. 2008-11-09]. Retrieved from [www:](http://webdoc.sub.gwdg.de/edoc/ah/2006/hilse_kothe/urn%3Anbn%3Ade%3Agbv%3A7-isbn-90-6984-508-3-8.pdf)

http://webdoc.sub.gwdg.de/edoc/ah/2006/hilse_kothe/urn%3Anbn%3Ade%3Agbv%3A7-isbn-90-6984-508-3-8.pdf.

NLA Guidelines for the Development and Application of a Persistent Identifier Scheme for Digital Resources : Appendix 1 [online]. Canberra : National Library of Australia, c2008. [cit. 2008-11-09]. Retrieved from [www:](http://www.nla.gov.au/initiatives/persistence/PIappendix1.html)

<http://www.nla.gov.au/initiatives/persistence/PIappendix1.html>.

NLA Guidelines for the Development and Application of a Persistent Identifier Scheme for Digital Resources : Appendix 2 [online]. Canberra : National Library of Australia, c2008. [cit. 2008-11-09]. Retrieved from [www:](http://www.nla.gov.au/initiatives/persistence/PIappendix2.html)

<http://www.nla.gov.au/initiatives/persistence/PIappendix2.html>.

Persistent identifiers [online]. Canberra : National Library of Australia, c2008. [cit. 2008-11-09]. Retrieved from [www:](http://www.nla.gov.au/initiatives/persistence.html) <http://www.nla.gov.au/initiatives/persistence.html>.

Persistent Identification Systeme : Report on a consultancy [online]. Conducted by Diana Dack. Canberra : National Library of Australia, 2001. [cit. 2008-11-09]. Retrieved from [www:](http://www.nla.gov.au/initiatives/persistence/PIcontents.html) <http://www.nla.gov.au/initiatives/persistence/PIcontents.html>.