2008-08-13T14:34:33Z

The DSpace Course - Identifiers

Lewis, Stuart

http://hdl.handle.net/10673/62

Downloaded from DSpace Repository, DSpace Institution's institutional repository
Module: Identifiers

Module overview:
Persistent Identifiers are an integral feature of repositories to assist with resource identification and preservation. This module will look at what persistent identifiers are, how they work and the benefits to using them in a DSpace repository. The module will then explore handles, the identifiers used for persistence in DSpace. The module will conclude with a look at how to apply for and configure a handle.

Module objectives:
By the end of this module you will:

1. Understand what persistent identifiers are, how they work and the benefits to using them in a DSpace repository environment
2. Understand what a handle is – the persistent identifier currently used in DSpace
3. Have an overview of how to apply for a handle
4. Have seen a handle in use
5. Understand how to set up the handle server

Note
For the practical exercise, please refer to your sheet ‘Local instructions’ for details of the following:

- How to launch a web browser
Persistent Identifiers

- The use of location based identifiers such as the Uniform Resource Locator (URL) often leads to problems in accessibility to resources with time.
- Often when accessing a resource via a hyperlink users receive a “404 - page not found” error.
- Persistent identifiers are an attempt at solving the issues surrounding resource identification and long term preservation.
- A persistent identifier allows the resource to be uniquely identified in a way that will not change if the resource is renamed or relocated.

Persistent Identifiers

The use of location based identifiers such as the Uniform Resource Locator (URL) often leads to problems in accessibility to resources with time. Often when accessing a resource via a hyperlink users receive a “404 - page not found” error. This is caused by the resource being removed, relocated or renamed while external links to the resource remain unchanged.

Persistent identifiers are an attempt at solving the issues surrounding resource identification and long term preservation of access to online digital materials. A persistent identifier allows the resource to be uniquely identified in a way that will not change if the resource is renamed or relocated and will persist regardless of the protocol used to access it. This means that a resource can be reliably referenced for future access by humans and software. An important part of persistence is organisational policy, not just the adoption of technical solutions, in essence the persistence of an object is only effective if the organisation maintains and manages this persistence. In this manner even a URL could be considered persistent so long as it is maintained by some administrative service.
Persistent Identifiers

This means that a resource can be reliably referenced for future access by humans and software

Caveat: Persistence is heavily dependant on organisation policy i.e. persistence of an object is only effective if an organisation maintains and manages this persistence

Different systems in use for persistent identifiers
- Persistent Uniform Resource Locators (PURLs)
- Digital Object Identifiers (DOI)
- Handle – Used by DSpace

Persistent Identifiers

Different systems in use for persistent identifiers

- **PURL**
  The Persistent Uniform Resource Locator takes the concept of the URL and adds a resolution service context. Instead of the URL pointing directly to the resource in question, it references an intermediate PURL resolution service. This service is used to look up the actual address of the URL before redirecting to the appropriate resource.

- **Handle**
  In this system the resource address is identified by a unique handle assigned by a common registration service. When the browser receives a handle, it is sent to the global registration service for resolution and redirection to a local handle server which in turn can resolve the local part of the identifier to the resource in the repository.
• DOI

The Digital Object Identifier makes use of the Handle System as a resolution service but gives additional value added services such as descriptive metadata about the digital object that is being represented. DOIs are mainly used by publishers to identify journal articles, but are starting to be used by repositories to identify documents or data.

Further Resource

PILIN Project https://www.pilin.net.au/Welcome/Welcome.htm
The PILIN project is an initiative to research into and promote the use of global persistent identifiers within e-Research communities.

PADI, an initiative of the National Library of Australia provides information and resources on digital preservation.
The Handle

- In a handle system, resource address is identified by a unique handle assigned by a common registration service

<table>
<thead>
<tr>
<th>Registration Service</th>
<th>Handle Prefix</th>
<th>Local Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://hdl.handle.net">http://hdl.handle.net</a></td>
<td>2160</td>
<td>568</td>
</tr>
</tbody>
</table>

The Handle

1. In a handle system, resource address is identified by a unique handle assigned by a common registration service
2. When the browser receives a handle, it is sent to the global registration service for resolution
3. The global registration service performs a lookup using the handle prefix and redirects to the local handle server on the DSpace repository
4. The local handle server then resolves the local part of the identifier to the resource in the repository
- Handle resolution: http://hdl.handle.net/2160/568
Practical: Using a Handle

- Navigate to Aberystwyth's DSpace repository – Cadair
- Select an item from a collection and note the handle address
  
  Please use this identifier to cite or link to this item: http://hdl.handle.net/2160/568

- Open this address in a new browser window

- The handle will resolve an redirect back to your original item
Configuring the Handle Service

First a few facts to clear up some common misconceptions:

You don't have to use CNRI's Handle system. At the moment, you need to change the code a little to use something else (e.g. PURLs) but that should change soon.

You'll notice that while you've been playing around with a test server, DSpace has apparently been creating handles for you looking like hdl:123456789/24 and so forth. These aren't really Handles, since the global Handle system doesn't actually know about them, and lots of other DSpace test installs will have created the same IDs.

They're only really Handles once you've registered a prefix with CNRI (see below) and have correctly set up the Handle server included in the DSpace distribution. This Handle server communicates with the rest of the global Handle infrastructure so that anyone that understands Handles can find the Handles your DSpace has created.
Configuring the Handle Service

Once the configuration file has been generated, you will need to go to http://hdl.handle.net/4263537/5014 or http://www.handle.net/service_agreement.html to upload the generated sitebndl.zip file. The upload page will ask you for your contact information. An administrator will then create the naming authority/prefix on the root service (known as the Global Handle Registry), and notify you when this has been completed. You will not be able to continue the handle server installation until you receive further information concerning your naming authority.

Note that since the DSpace code manages individual Handles, administrative operations such as Handle creation and modification aren't supported by DSpace’s Handle server.
Generating the sitebndl.zip

The Site Bundle is an archive which contains information about your DSpace installation and is used to generate your handle.

To generate the sitebndl.zip run the command:

```
[dspace]/bin/dsrun net.handle.server.SimpleSetup
[dspace]/handle-server
```

- You will be required to complete a series of questions.
- Once completed the sitebndl.zip can be found:

  `[dspace]/handle-server/sitebndl.zip`

- Complete the registration and upload the sitebndl.zip.

Generating the sitebndl.zip

When running this script you will be asked a series of questions further information on generating the sitebndl.zip can be found at: [http://www.handle.net/hs_manual/index.html](http://www.handle.net/hs_manual/index.html)

When running:

```
[DSpace]/bin/dsrun net.handle.server.SimpleSetup [DSpace]/handle-server
```

Ensure that:

```
[DSpace]/handle-server matches whatever you have in DSpace.cfg
```
Configuring the handle server

Once registration is complete, a handle should be returned from CNRI

- Edit the [dspace]/handle-server/config.dct to include the lines in the "server_config" clause:
  "storage_type" = "CUSTOM"
  "storage_class" = "org.dspace.handle.HandlePlugin"

- Update all references to YOUR_NAMING_AUTHORITY to your assigned handle:
  300:0.NA/YOUR_NAMING_AUTHORITY -> 300:0.NA/2097

Configuring the handle server

config.dct tells the Handle server to get information about individual Handles from the DSpace code.
Updating the Handle Prefix

If you need to update the handle prefix on items created before the CNRI registration process, you can run the [DSpace]/bin/update-handle-prefix script. You may need to do this if you loaded items prior to CNRI registration (e.g. setting up a demonstration system prior to migrating it to production). The script takes the current and new prefix as parameters. For example:

```
[DSpace]/bin/update-handle-prefix 123456789 1303
```

will change any handles currently assigned prefix 123456789 to prefix 1303, so for example handle 123456789/23 will be updated to 1303/23 in the database.
Starting the Handle Server

- Finally start the handle server
  
  \[\text{[dspace]}/\text{bin}/\text{start-handle-server}\]

- A script will be required to automate the starting of the handle server upon a server boot

- Once configured the handles should resolve as the practical demonstrated earlier in this module

Starting the Handle Server

See the practical below on the result of implementing a handle servers
Practical: Using a Handle

For the practical exercise, please refer to your sheet ‘Local instructions’ for details of the following:

How to launch a web browser

- Launch a web browser window
- Navigate to Aberystwyth’s repository Cadair (http://cadair.aber.ac.uk/)
- Navigate to a community of your choice
- Navigate to a collection of your choice
- Select an item from the recent items menu
- Make a note of the handle address (http://hdl.handle.net/xxxx/xx)
- Open a new tab and enter the chosen handle address
- The handle should resolve and take you back to the original item
Credits

- These notes have been produced by:
  - Stuart Lewis & Chris Yates
  - Repository Support Project
    - http://www.rsp.ac.uk/
  - Part of the RepositoryNet
  - Funded by JISC
    - http://www.jisc.ac.uk/