Persistent identifiers and the B2Handle service
Outline

PIDs in EUDAT

Presentation of the API

Demo (web interface)

Demo (terminal)

Questions
The EUDAT data domain handles registered data
  - Each digital object must have a persistent identifier

This persistent identifier is used for:
  - Replica identification
  - Identification of the repository of record (in the case of replication)
  - Querying of additional information
  - Checksum (time stamped)

Actionable URLs
Handle system used within EUDAT
  - EPIC service to be more precise
Why use a service such as EPIC?

- In principle possible for every repository to run its own PID service
  - but not every organisation is willing or able to do that
  - also there is an advantage of increased reliability by replicating services
  - etc..
- DataCite & EPIC offer services based on HS for data PID since 2010
  - Both offer APIs for creating and managing PID handles
  - Different business models
Why use a PID service?

- DataCite targeted towards complete data-sets and also includes a specific metadata scheme for data-set publication
- EPIC targeted more at data management of individual resources allowing association of extra data with the PID: checksum, link to flexible metadata, ...
- EPIC is only a steward for the PIDs, no lock-in
Services using B2Handle

- B2DROP
- B2SHARE
- B2SAFE
- B2STAGE
- B2FIND
EUDAT offers integration of the EPIC API into iRODS via a python script.

- This comes out of the box with the B2SAFE service
- Complexity of the structure hidden

The script takes your credentials as input

- Supplied on the command line
- Stored in a configuration file (iRODS or local fs)

The script supports the following actions

- Searching
- Resolving
- Creation
- Modification
B2Safe workflow
B2Safe workflow
B2Safe workflow
B2Safe workflow
Repository of Record

The Repository of Record (RoR) is the repository where the digital object originated from. This repository is considered the “owner” of the digital object.

▶ Joining EUDAT
  ▶ The community center is typically regarded as the RoR. This is also the place where permissions are managed.

▶ Using EUDAT
  ▶ The EUDAT data center where data is first ingested is considered the RoR. This implies relying on the EUDAT infrastructure in order to manage the digital object and its permissions.
EUDAT Handle Record administration

1839/00-0000-0000-0001-53C5-2

| URL  | http://www.mpi.nl/1 |

Map of Europe with a red dot labeled 'DO'.
EUDAT Handle Record administration

1839/00-0000-0000-0001-53C5-2
URL http://www.mpi.nl/1

11096/00-0000-0000-0001-53C5-3
URL http://www.rzg.mpg.de/1
EUDAT Handle Record administration

- **1839/00-0000-0000-0001-53C5-2**
  - URL: http://www.mpi.nl/1

- **11096/00-0000-0000-0001-53C5-3**
  - URL: http://www.rzg.mpg.de/1

- **11112/00-0000-0000-0001-53C5-4**
  - URL: http://www.sara.nl/1
EUDAT Handle Record administration
EUDAT Handle Record administration

<table>
<thead>
<tr>
<th>Handle</th>
<th>URL</th>
<th>ROR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1839/00-0000-0000-0001-53C5-2</td>
<td><a href="http://www.mpi.nl/1">http://www.mpi.nl/1</a></td>
<td></td>
</tr>
<tr>
<td>11096/00-0000-0000-0001-53C5-3</td>
<td><a href="http://www.rzg.mpg.de/1">http://www.rzg.mpg.de/1</a></td>
<td>1839/00-0000-0000-0001-53C5-2</td>
</tr>
<tr>
<td>11112/00-0000-0000-0001-53C5-4</td>
<td><a href="http://www.sara.nl/1">http://www.sara.nl/1</a></td>
<td></td>
</tr>
</tbody>
</table>
EUDAT Handle Record administration

<table>
<thead>
<tr>
<th>Handle</th>
<th>URL</th>
<th>ROR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1839/00-0000-0000-0001-53C5-2</td>
<td><a href="http://www.mpi.nl/1">http://www.mpi.nl/1</a></td>
<td>1839/00-0000-0000-0001-53C5-2</td>
</tr>
<tr>
<td>11096/00-0000-0000-0001-53C5-3</td>
<td><a href="http://www.rzg.mpg.de/1">http://www.rzg.mpg.de/1</a></td>
<td>11096/00-0000-0000-0001-53C5-3</td>
</tr>
<tr>
<td>11112/00-0000-0000-0001-53C5-4</td>
<td><a href="http://www.sara.nl/1">http://www.sara.nl/1</a></td>
<td>11112/00-0000-0000-0001-53C5-4</td>
</tr>
</tbody>
</table>
EUDAT Handle Records

RoR Record:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>123/abc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>URL</td>
<td>1</td>
<td>15-09-28 00:00:00Z</td>
<td><a href="http://www.example.com/1">http://www.example.com/1</a></td>
</tr>
<tr>
<td>10320/LOC</td>
<td>2</td>
<td>15-09-28 00:00:00Z</td>
<td>&lt;locations&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHECKSUM</td>
<td>3</td>
<td>15-09-28 00:00:00Z</td>
<td>3801e6832b73e0fc4d39b3630712cd</td>
</tr>
<tr>
<td>HS_ADMIN</td>
<td>100</td>
<td>15-09-28 00:00:00Z</td>
<td>...</td>
</tr>
</tbody>
</table>
**RoR Record:**

<table>
<thead>
<tr>
<th>123/abc</th>
<th></th>
<th></th>
<th><a href="http://www.example.com/1">http://www.example.com/1</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>URL</td>
<td>1</td>
<td>15-09-28 00:00:00Z</td>
<td><a href="http://www.example.com/1">http://www.example.com/1</a></td>
</tr>
<tr>
<td>10320/LOC</td>
<td>2</td>
<td>15-09-28 00:00:00Z</td>
<td>&lt;locations&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;location id=&quot;0&quot; href=&quot;http://www.example.com/1&quot; weight=&quot;0.5&quot; /&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;location id=&quot;1&quot; href=&quot;http://hdl.handle.net/456/abc&quot; weight=&quot;0.5&quot; /&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;/locations&gt;</td>
</tr>
<tr>
<td>CHECKSUM</td>
<td>3</td>
<td>15-09-28 00:00:00Z</td>
<td>3801e6832b73e0fc4d39b3630712cd</td>
</tr>
<tr>
<td>HS_ADMIN</td>
<td>100</td>
<td>15-09-28 00:00:00Z</td>
<td>...</td>
</tr>
</tbody>
</table>

**Replica Record:**

<table>
<thead>
<tr>
<th>456/abc</th>
<th></th>
<th></th>
<th><a href="http://www.example.com/1">http://www.example.com/1</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>URL</td>
<td>1</td>
<td>15-09-28 00:00:00Z</td>
<td><a href="http://www.example.com/1">http://www.example.com/1</a></td>
</tr>
<tr>
<td>10320/LOC</td>
<td>2</td>
<td>15-09-28 00:00:00Z</td>
<td>&lt;locations&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;location id=&quot;0&quot; href=&quot;http://www.replica.com/1&quot; weight=&quot;1&quot; /&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;/locations&gt;</td>
</tr>
<tr>
<td>CHECKSUM</td>
<td>3</td>
<td>15-09-28 00:00:00Z</td>
<td>3801e6832b73e0fc4d39b3630712cd</td>
</tr>
<tr>
<td>ROR</td>
<td>4</td>
<td>15-09-28 00:00:00Z</td>
<td>hdl:123/abc</td>
</tr>
<tr>
<td>HS_ADMIN</td>
<td>100</td>
<td>15-09-28 00:00:00Z</td>
<td>...</td>
</tr>
</tbody>
</table>
API URI Space

«root»/
  discovery/
  NAs
    «NAsegment»/
    handles/
      «LNsegment»/
        (...)
    profiles/
    «profile»
  status/
    «id»
  template
API (1)

- «root»/NAs/

  GET returns a collection of all NamingAuthorities hosted by this web service, indexed by «NAsegment»

- «root»/NAs/«NAsegment»/handles/

  GET Returns a collection of Handles indexed by «LNsegment»

  (Optional) Supports searches
API (2)

▶ 「root」/「NAsegment」/handles/「LNsegment」/
  GET  Returns a Handle
  PUT  Submits a Handle (create or modify).
  DELETE  Deletes the handle.
  POST  Accepts a Value Set which **MUST NOT** contain a handle member.

▶ In this case, the 「LNsegment」 is interpreted as a suffix template. A suffix template is a suffix containing exactly one unescaped ‘*’ character. This character will be replaced with a unique string by the server, resulting in a new and unique handle.

▶ Creates a new handle with the provided metadata. An **HTTP/1.1 201 Created** status is returned upon success, with the location of the new resource in the **Location:** response header. The new handle is returned in an **X-Handle:** response header, encoded as per RFC5987 if necessary.
API (3)

- HTTP content negotiation supported
  - All implementations must support JSON
  - Text and XHTML are encouraged
- Documentation
  - https://github.com/pidconsortium/EPIC-API-v2/blob/master/API.md
Web Interface Demo (1)

- `https://epic-test.pdc.kth.se/v2/handles/`
  - List all prefixes under this EPIC server

- `https://epic-test.pdc.kth.se/v2/handles/11432/`
  - List all handles under the 11432 prefix. The response might limit the number of handles returned

- `https://epic-test.pdc.kth.se/v2/handles/11432/D3BDBB10-6100-11E5-9653-0022190F5141`
  - Query the handle record for handle: `11432/D3BDBB10-6100-11E5-9653-0022190F5141`
Web Interface Demo (2)

- [Web Interface Demo](https://epic-test.pdc.kth.se/v2/handles/11432/?EMAIL=john.doe@example.com)
  - Search in the 11432 prefix for all handles with a type **EMAIL** with value = john.doe@example.com
  - Powerful feature providing added value on top of the HANDLE system
  - Allows reverse lookup a.k.a search by **URL** = ....
  - Allows for lookup by checksum value (if checksums are stored)
Example command

curl -s -D - --fail --user 11432:$PASSWORD --data @new_pid.json --header "Content-type: application/json" "https://epic-test.pdc.kth.se/v2/handles/11432/" -o /dev/null

Response

HTTP/1.1 201 Created
Date: Wed, 11 Nov 2015 14:26:30 GMT
Content-Length: 2188
Location: https://epic-test.pdc.kth.se/v2/handles/11432/3394ade0-8880-11e5-9653-0022190f5141
ETag: "1B2M2Y8AsgTpgAmY7PhCfg"
Last-Modified: Thu, 01 Jan 1970 00:00:00 GMT
Connection: close
Content-Type: text/plain; charset=UTF-8

For full scripts see https://github.com/WillemElbers/epic-api-v2-demo
Questions?