Dataset Identification & Citation:

DataCite and EZID

Laine Farley

Executive Director

California Digital Library

November 30, 2011
Dataset Identification & Citation

Life before DataCite
The research data problem
Each repository is an island...

A life cycle approach

DataCite
“Helping you find, access and reuse data.”

EZID
Easy creation and management of DataCite DOIs and other identifiers.

Next steps
## The research data problem

<table>
<thead>
<tr>
<th>Journal Articles</th>
<th>Research Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniquely and persistently identified</td>
<td>Nope</td>
</tr>
<tr>
<td>Concept of “publish”</td>
<td>Not really</td>
</tr>
<tr>
<td>Multiple copies</td>
<td>Typically one</td>
</tr>
<tr>
<td>Easily findable</td>
<td>Difficult</td>
</tr>
<tr>
<td>Additional services: impact metrics, citation tracking, etc.</td>
<td>Nope</td>
</tr>
</tbody>
</table>

- **Second-class citizens in scholarly record**
Before DataCite...
With DataCite...

Publishers

Data centers and Repositories
DataCite
Creating a global citation framework for data
DataCite

German National Library of Economics (ZBW)
German National Library of Science and Technology (TIB)
German National Library of Medicine (ZB MED)
GESIS - Leibniz Institute for the Social Sciences, Germany
Australian National Data Service (ANDS)
ETH Zurich, Switzerland

Canada Institute for Scientific and Technical Information (CISTI)
Technical Information Center of Denmark
Institute for Scientific & Technical Information (INIST-CNRS), France
TU Delft Library, The Netherlands
The Swedish National Data Service (SNDS)
The British Library, UK
California Digital Library (CDL), USA
Office of Scientific & Technical Information (OSTI), USA
Purdue University Library, USA
DataCite example

International DOI Foundation

DataCite/TIB

CDL

Repository, Library or Data Publisher

Researcher

DOI resolver

registration service

URL plus id

registration request

data + metadata

(opt) CDL-hosted EZID id service
A life cycle approach

**CDL Curation and Publishing Services**

http://www.cdlib.org

Create, edit, share, and save data management plans

Open source add-in for Microsoft Excel as a data collection tool

Create and manage persistent identifiers

Curation repository: store, manage, and share research data

Open access scholarly publishing services: papers, journals, books, seminars & more

An infrastructure to publish and get credit for sharing research data
• Get identifiers
• Add location
• Add metadata
• Update location
• Update metadata

*take control of the management and distribution of your research, share and get credit for it, and build your reputation through its collection and documentation*
## DataCite Metadata V. 2.2

<table>
<thead>
<tr>
<th>Required properties</th>
<th>Optional properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>Identifier</strong> (with type attribute)</td>
<td>6. Subject (with schema attribute)</td>
</tr>
<tr>
<td>2. <strong>Creator</strong> (with name identifier attributes)</td>
<td>7. Contributor (with type &amp; name identifier attributes)</td>
</tr>
<tr>
<td>3. <strong>Title</strong> (with optional type attribute)</td>
<td>8. Date (with type attribute)</td>
</tr>
<tr>
<td>4. <strong>Publisher</strong></td>
<td>9. <strong>Language</strong></td>
</tr>
<tr>
<td>5. <strong>PublicationYear</strong></td>
<td>10. <strong>ResourceType</strong> (with description attribute)</td>
</tr>
<tr>
<td></td>
<td>11. <strong>AlternateIdentifier</strong> (with type attribute)</td>
</tr>
<tr>
<td></td>
<td>12. <strong>RelatedIdentifier</strong> (with type &amp; relation type attributes)</td>
</tr>
<tr>
<td></td>
<td>13. <strong>Size</strong></td>
</tr>
<tr>
<td></td>
<td>14. <strong>Format</strong></td>
</tr>
<tr>
<td></td>
<td>15. <strong>Version</strong></td>
</tr>
<tr>
<td></td>
<td>16. <strong>Rights</strong></td>
</tr>
<tr>
<td></td>
<td>17. <strong>Description</strong> (with type attribute)</td>
</tr>
</tbody>
</table>
EZID (easy-eye-dee) makes it easy to create and manage unique, persistent identifiers.

- You can create identifiers for any kind of entity — physical, digital, abstract, etc. For identifiers of objects on the web, you can use EZID to maintain their current locations so that people who click on the identifiers are correctly forwarded.

- You can also store citation metadata with identifiers to aid in interpreting and maintaining them. Several citation formats are possible.

- EZID supports a number of persistent identifier technologies, including ARKs (Archival Resource Keys) and DOIs (Digital Object Identifiers).

- Visit our website for more information about EZID identifiers, features, and benefits.

To operate EZID use the tabs above or the programming interface. You will need a login account to create identifiers, but anyone can view identifier metadata. Please contact us if you are interested in setting up an account.
http://n2t.net/ezid

EZID (easy-eye-dee) makes it easy to create and manage unique, persistent identifiers.

- You can create identifiers for any kind of entity — physical, digital, abstract, etc. For identifiers of objects on the web, you can use EZID to maintain their current locations so that people who click on the identifiers are correctly forwarded.

- You can also store citation metadata with identifiers to aid in interpreting and maintaining them. Several citation formats are possible.

- EZID supports a number of persistent identifier technologies, including ARKs (Archival Resource Keys) and DOIs (Digital Object Identifiers).

- Visit our website for more information about EZID identifiers, features, and benefits.

To operate EZID use the tabs above or the programming interface. You will need a login account to create identifiers, but anyone can view identifier metadata. Please contact us if you are interested in setting up an account.
We hope that most of this interface is easy to use. You can use the form below to create test identifiers exactly as you would under the Create tab. Test identifiers are the same as long-term identifiers except they are deleted after 72 hours.

Clicking the Create button below (leaving the Remainder blank) will generate a preservation-ready test identifier for you.

A namespace is the set of all possible identifiers that begin with a certain prefix, such as ark://99999/4k4. EZID always supplies the prefix, and usually the remainder as well. We strongly advise letting EZID generate the identifier for you. This way it will be short, opaque, globally unique, and transcription-safe (EZID adds a "check character").

On the Manage tab, you can add metadata ("data about data") for an identified object. Editable fields change color to yellow as you mouse over them. The "URL" field stores the current location of the identified object. When you first create your identifier, EZID supplies a default value for this field, namely the URL of the Manage tab for that identifier. We recommend that you replace this with the actual current location of the object.

On the right is a pull-down for selecting a "citation profile." The profile determines the metadata elements you can complete for the identifier. Depending on the type of identifier you create, EZID will select the appropriate profile as the default. At present there is no automatic mapping between profiles.
http://n2t.net/ezid
http://n2t.net/ezid
What this means...

When using this data, please cite the original article:


Additionally, please cite the Dryad data package:


Dryad File Identifier: doi:10.5061/dryad.1030/4

Files in this item

<table>
<thead>
<tr>
<th>Files</th>
<th>Size</th>
<th>Format</th>
<th>View</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gurgel LSmeans.xls</td>
<td>17.89Mb</td>
<td>Microsoft Excel</td>
<td>View/Open</td>
</tr>
</tbody>
</table>

To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.
What this means...
Next Steps

DataCite
• Metadata v. 2.3
• Dublin Core application profile
• Content Service

EZID
• UI redesign
• Automated link checking
• Exposure for citations

By Nicola Whitaker http://www.flickr.com/photos/nicolawhitaker/111009156/
For more information

EZID
EZID application: http://n2t.net/ezid/
EZID website: http://www.cdlib.org/services/uc3/ezid/

DataCite
DataCite Home: http://datacite.org/
DataCite Metadata Schema:
   http://schema.datacite.org/meta/kernel-2.2/index.html
DataCite Metadata Search: http://search.datacite.org

Contact EZID Service Mgr: uc3@ucop.edu
Questions?

©
by Horia Varlan
http://www.flickr.com/photos/horiavarlan/4273168957/in/photostream/