Dataset Citation and Identifiers

DOI, ARK, and EZID

Joan Starr
California Digital Library
April, 2012
@joan_starr
Dataset Citation & Identifiers

Data Citation
Identifiers 101
Dataset Identification with EZID
Choosing an Identifier
Life Cycle Data Management
Looking ahead

By Brain farts (Joschua) http://www.flickr.com/photos/brainfarts/97676505/
Partnership between CDL | 10 UC campuses | Peer institutions

Provide solutions, services, resources for digital assets

Pool & distribute diverse experience, expertise, & resources
Data Citation

By barryegan (Vitor Leite) http://www.flickr.com/photos/vixon/116447718/
What?

• Key identifying elements
• Emerging recommendations
• Variation among the domains
How?

• Key identifying elements
• Emerging recommendations
• Variation among the domains
• In common: Persistent identifier
What this means...
What this means...
Dataset Citation & Identifiers

Data Citation

Identifiers 101

Dataset Identification with EZID

Choosing an Identifier

Life Cycle Data Management

Looking ahead

By Brain farts (Joschua) http://www.flickr.com/photos/brainfarts/97676505/
Identifiers 101
What is an identifier?

**What you see:** alphanumeric string *(never changes)*

**Associated with:** location of object *(such as a URL)*

**Optional:** who, what, when, etc *(i.e. metadata)*

By Joelk75: [http://www.flickr.com/photos/75001512@N00/2728233597/](http://www.flickr.com/photos/75001512@N00/2728233597/)
Identifier example

string: doi:10.9999/FK40K2GTV
html version: http://dx.doi.org/10.9999/FK40K2GTV
location: http://www.bologna.edu/biology/xfg/123.xls
metadata
creator: Dr. Felix Kottor

title: Data for chromosomal study of catfish (Ictalurus punctatus)
publisher: University of Bologna
date: 8/31/2011
Identifier example

string: doi:10.9999/FK40K2GTV
html version: http://dx.doi.org/10.9999/FK40K2GTV
location: http://www.state.edu/ecology/783sdr/123.xls
metadata

creator: Dr. Felix Kottor

title: Data for chromosomal study of catfish (Ictalurus punctatus)
publisher: Dryad Data Repository
date: 10/01/2011
Identifiers 201

By Christi Nielsen http://www.flickr.com/photos/christinielsen/476326980/
Identifiers 201

- string: doi:10.9999/FK40K2GTV

  "prefix"  "suffix"
Dataset Citation & Identifiers

Data Citation
Identifiers 101

Dataset Identification with EZID
Choosing an Identifier
Life Cycle Data Management
Looking ahead

By Brain farts (joschua) http://www.flickr.com/photos/brainfarts/97676505/
EZID: long-term identifiers made easy

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

Primary Functions

1. Create persistent identifiers
2. Manage identifiers over time
3. Manage associated metadata over time
DataCite
Creating a global citation framework for data
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 DOIls (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

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 2 weeks.

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:/99999/0k4.... 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
Dataset Citation & Identifiers

Data Citation
Identifiers 101
Dataset Identification with EZID
Choosing an Identifier
Life Cycle Data Management
Looking ahead

By Brain farts (Joschua) http://www.flickr.com/photos/brainfarts/97676505/
DOIs and ARKs

• both can work like regular hyperlinks.
• both can refer to a subset or portion of a resource.
• both become persistent when the target URL is maintained.
DOIs vs ARKs

• Case sensitive
• Special feature supports granularity
• Informative
• Less costly
DOIs vs **ARKs**: suffix pass-through

- **string**: `ark:/99999/Big4/*`
- **location**: `http://x.y.z/foo/Big4/db/*`
DOIs vs **ARPs**: suffix pass-through

- string: `ark:/99999/Big4/table/cell/45-8.txt`
- location: `http://x.y.z/foo/Big4/db/table/cell/45-8.txt`
DOIs vs ARKs

- Established brand in publishing
- Indexed by major A&I citation databases
- Cannot be deleted
- More costly

[DataCite and Crossref logos]
Dataset Citation & Identifiers

Data Citation
Identifiers 101
Dataset Identification with EZID
Choosing an Identifier
Life Cycle Data Management
Looking ahead

By Brain farts (joschua) http://www.flickr.com/photos/brainfarts/97676505/
The Life of Data

By jfcherry http://www.flickr.com/photos/67272961@N03/6123892769/
A life cycle approach

**CDL Curation and Publishing Services**

[http://www.cdlib.org](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
Identifiers and the data life cycle

Track your results

Organize your data

Get more citations

Meet funder requirements
Dataset Citation & Identifiers

Data Citation
Identifiers 101
Dataset Identification with EZID
Choosing an Identifier
Life Cycle Data Management
Looking ahead
1. New User Interface.

By Leonard John Matthews http://www.flickr.com/photos/mythoto/3964995003/
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), DOIs (Digital Object Identifiers), and URNs (Uniform Resource Names).
- 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.
### Recent Activity

<table>
<thead>
<tr>
<th>ID</th>
<th>Date created</th>
<th>Title</th>
<th>Creator</th>
</tr>
</thead>
<tbody>
<tr>
<td>ark:/13030/m5j5061l</td>
<td>03/07/2012 12:36:42 AM</td>
<td>ark:/13030/m5j5061l created 03/07/2012 12:36:42 AM</td>
<td></td>
</tr>
<tr>
<td>ark:/13030/6t88x3d1lw</td>
<td>03/07/2012 12:27:17 PM</td>
<td>ark:/13030/6t88x3d1lw created 03/07/2012 12:27:17 PM</td>
<td></td>
</tr>
<tr>
<td>ark:/13030/qt9g722e1</td>
<td>03/07/2012 12:27:17 PM</td>
<td>ark:/13030/qt9g722e1 created 03/07/2012 12:27:17 PM</td>
<td></td>
</tr>
<tr>
<td>ark:/13030/m5s65k38</td>
<td>03/07/2012 12:21:31 PM</td>
<td>ark:/13030/m5s65k38 created 03/07/2012 12:21:31 PM</td>
<td></td>
</tr>
<tr>
<td>ark:/13030/m5769v1</td>
<td>03/07/2012 12:11:53 PM</td>
<td>ark:/13030/m5769v1 created 03/07/2012 12:11:53 PM</td>
<td></td>
</tr>
<tr>
<td>ark:/13030/m5k0722t</td>
<td>03/07/2012 11:50:48 AM</td>
<td>ark:/13030/m5k0722t created 03/07/2012 11:50:48 AM</td>
<td></td>
</tr>
<tr>
<td>ark:/13030/q6h6v4zxc</td>
<td>03/07/2012 11:30:27 AM</td>
<td>ark:/13030/q6h6v4zxc created 03/07/2012 11:30:27 AM</td>
<td></td>
</tr>
<tr>
<td>ark:/13030/q7660w0x</td>
<td>03/07/2012 11:30:25 AM</td>
<td>ark:/13030/q7660w0x created 03/07/2012 11:30:25 AM</td>
<td></td>
</tr>
<tr>
<td>ark:/13030/q76k08zh3</td>
<td>03/07/2012 11:27:17 AM</td>
<td>ark:/13030/q76k08zh3 created 03/07/2012 11:27:17 AM</td>
<td></td>
</tr>
<tr>
<td>ark:/13030/q7xw9s01g</td>
<td>03/07/2012 11:27:17 AM</td>
<td>ark:/13030/q7xw9s01g created 03/07/2012 11:27:17 AM</td>
<td></td>
</tr>
</tbody>
</table>

### Identifiers

<table>
<thead>
<tr>
<th>ID</th>
<th>Date created</th>
<th>Title</th>
<th>Creator</th>
</tr>
</thead>
<tbody>
<tr>
<td>ark:/13030/hb000000k</td>
<td>09/09/2011 04:14 PM</td>
<td>Jug, brown earthenware for brandy.</td>
<td>United States. District Court (California : Northern District). Land case. 128 (Related Name Indexing Term)</td>
</tr>
<tr>
<td>ark:/13030/hb000000mb</td>
<td>09/09/2011 04:29 PM</td>
<td>Bidwell's Bar, 1856</td>
<td></td>
</tr>
<tr>
<td>ark:/13030/hb0000002w</td>
<td>09/13/2011 12:38 AM</td>
<td>[Diseno del Rancho Cabeza de Santa Rosa : Calif.]</td>
<td>United States. District Court (California : Southern District). Land case. 370 (Related Name Indexing Term)</td>
</tr>
<tr>
<td>ark:/13030/hb000002gs</td>
<td>09/12/2011 11:41 PM</td>
<td>Bosquejo del sitio llamado Sta. Anna a cuatro leguas NE de S. Juan Bautista pa. acompañar la solicitud de el hace el Licdo. Luis del Castillo Negrete : [Rancho Santa Anna y Queen Sabe, Calif.]</td>
<td>United States. District Court (California : Southern District). Land case. 258 (Related Name Indexing Term)</td>
</tr>
<tr>
<td>ark:/13030/hb000002ks</td>
<td>09/14/2011 06:31 AM</td>
<td>[Diseno del Rancho Cañada de Jonwe : Sonoma Co., Calif. / Jaspar O'Farrell, maker?]</td>
<td>United States. District Court (California : Northern District). Land case. 46 (Related Name Indexing Term)</td>
</tr>
</tbody>
</table>
2. Growing user community

http://www.cdlib.org/services/uc3/ezid/clients.html

Thanks to Scott Edmunds, GigaScience Journal for input
2. Growing user community
3. A&I Indexing
For more information

EZID
EZID application: http://n2t.net/ezid/
EZID website: http://www.cdlib.org/services/uc3/ezid/
EZID on Twitter: @ezidCDL

Joan Starr: uc3@ucop.edu @joan_starr