Characteristics of usable PID services
laura.rueda@datacite.org
The problem!

- We need reliable and unambiguous access to data:
  - collaboration
  - reuse
  - reproducibility
  - attribution
  - grant credit
  - faster progress
  - consistent funding
  - feed future researchers
Our mission

- DataCite develops and support methods to:
  - locate
  - identify
  - cite

data and other research objects to:
  - establish easier access
  - increase acceptance
  - foster reuse
A quick snapshot

- Not-for-profit global initiative since 2009
- > 35 members worldwide
- > 730 data centres
- > 8,200,000 DOIs created
- > 8,000,000 resolutions/month
All around the world
Principles

- Through collaboration, we:
  - support **researchers** by helping them to find, identify, and cite research data and other research objects with confidence
  - support **data centres** by providing persistent identifiers for datasets, workflows and standards for data publication
  - support **publishers** by enabling research articles to be linked to the underlying data/objects
Down in the weeds

1. Take a dataset

2. Describe it

3. Assign a DOI

4. Reuse and reference!

5. Enjoy the benefits!

ATLAS Collaboration, “Data from Figure 7 from: Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC: $H \rightarrow \gamma\gamma$,”
http://doi.org/10.7484/INSPIREHEP.DATA.A78C.HK44

- Unique
- Persistent

- Findability
- Reusability
- Track citations
- Measure impact

10.1234/exampledata
Metadata

• Mandatory, recommended, optional

• DataCite Metadata Schema 4.0 launching this month!

• Key improvements:
  • resource type becomes mandatory
  • new funding information

• [http://schema.datacite.org](http://schema.datacite.org)
*METADATA*
*METADATA*
Metadata

• Mandatory, recommended, optional

• DataCite Metadata Schema 4.0 launching this month!

• Key improvements:
  • resource type becomes mandatory
  • new funding information

• http://schema.datacite.org
What can you do with a DOI?
What can you do with a DOI?

- **R245fa Adsorption on MIL-101**

  Pete McGraw

  Numeric Data published 2013 via DOE Geothermal Data Repository; Pacific Northwest National Laboratory (PNNL), Richland, WA (United States)

  Adsorption isotherms for R245fa on metal organic heat carrier candidate MIL-101

  [http://doi.org/10.15121/1148803](http://doi.org/10.15121/1148803)

- **Data center**

  - WIP - Renewable Energies: 623
  - EPFL Infoscience: 40
  - University of British Columbia: 37
  - Technische Universität Berlin - Universitätsbibliothek: 32
  - ETA-Florence Renewable Energies: 17
What can you do with a DOI?

- **Resource type**
  - Text: 772
  - Dataset: 42
  - Software: 3
  - Collection: 2

- **Publication year**
  - 2010: 127
  - 2013: 111
  - 2009: 98
  - 2012: 96
  - 2011: 91
  - 2015: 90
  - 2008: 86
  - 2014: 76
  - 2007: 9
  - 2006: 8

- **Data center**
  - WIP - Renewable Energies: 623
  - EPFL Intoscience: 40
  - University of British Columbia: 37
  - Technische Universität Berlin: 32
  - Universitätsbibliothek: 17

**Energy carrier**

841 Results

**R245fa Adsorption on MIL-101**

Pete McGrail

Numeric Data published 2013 via DOE Geothermal Data Repository; Pacific Northwest National Laboratory (PNNL), Richland, WA (United States)

Adsorption isotherms for R245fa on metal organic heat carrier candidate MIL-101

http://doi.org/10.15121/1148803

Add to ORCID record

**ORCID**

Laura Rueda

ORCID ID: orcid.org/0000-0001-5952-7630

Also known as: Laura Rueda Garcia

**FOR RESEARCHERS**

- Education (3)
- Employment (4)
- Works (5)

**FOR ORGANIZATIONS**

**ABUO**

**FOR IN**

**Preferred source**

A comparative analysis of disciplinary data management workflows


DOI: 10.1109/call.2014.6970880

Source: CrossRef Metadata Search

D.2.1: Artefact, Contributor, and Organisation Relationship Data Scheme - Appendix B
What can you do with a DOI?

<table>
<thead>
<tr>
<th>Relation types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bookmarks</td>
</tr>
<tr>
<td>Cites</td>
</tr>
<tr>
<td>Compiles</td>
</tr>
<tr>
<td>Continues</td>
</tr>
<tr>
<td>Corrects</td>
</tr>
<tr>
<td>Discusses</td>
</tr>
<tr>
<td>Documents</td>
</tr>
<tr>
<td>Downloads</td>
</tr>
<tr>
<td>Has metadata</td>
</tr>
<tr>
<td>Has part</td>
</tr>
<tr>
<td>Is cited by</td>
</tr>
<tr>
<td>Is compiled by</td>
</tr>
<tr>
<td>Is continued by</td>
</tr>
<tr>
<td>Is corrected by</td>
</tr>
<tr>
<td>Is derived from</td>
</tr>
<tr>
<td>Is discussed by</td>
</tr>
<tr>
<td>Is documented by</td>
</tr>
<tr>
<td>Is downloaded by</td>
</tr>
<tr>
<td>Is identical to</td>
</tr>
</tbody>
</table>

Data from: Rise of the machines – recommendations for ecologists when using next generation sequencing for microsatellite development.

Michael G Gardner, Alison J Fitch, Terry Bertozi, Andrew J Lowe, Michael G Gardner, Alison J Fitch, Terry Bertozi, Andrew J Lowe
DataPackage published 2011 via Dryad Digital Repository

http://doi.org/10.5061/DROID.F3CB2

Examples Stats Search

Europe PMC http://doi.org/10.1073/PNAS.1205856110
Europe PMC http://doi.org/10.1371/JOURNAL.PONE.0084559
PLOS http://doi.org/10.1371/JOURNAL.PONE.0084559
Europe PMC http://doi.org/10.3732/APPS.1200295
Europe PMC http://doi.org/10.1371/JOURNAL.PONE.0040861
PLOS http://doi.org/10.1371/JOURNAL.PONE.0040861
Funder mandates

b. Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants.

c. Investigators and grantees are encouraged to share software and inventions...

...all scientific publications on the results of publicly funded research must be freely available. It also must be able to optimally reuse research data...

Scientific research data should be easily:
1. Discoverable
2. Accessible
3. Assessable and intelligible
4. Useable beyond the original purpose
5. Interoperable
How to tackle it

<table>
<thead>
<tr>
<th>Category</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical infrastructure</td>
<td>• building a reliable underlying layer</td>
</tr>
<tr>
<td>Support services</td>
<td>• applying technology to the real world</td>
</tr>
<tr>
<td>Integration</td>
<td>• making it useful &amp; usable for researchers</td>
</tr>
<tr>
<td>Metrics</td>
<td>• tracking impact and making it visible</td>
</tr>
<tr>
<td>Funding</td>
<td>• working with funding agencies</td>
</tr>
<tr>
<td>Training</td>
<td>• sharing best practices</td>
</tr>
</tbody>
</table>
Get involved

- Two-way community involvement is our core component
- We want to hear about requirements, use cases, and stories from researchers, funders, publishers, universities, data-centres and beyond
- Ambassadors will spread the word within their own communities, encourage best practices and drive adoption
  - http://project-thor.eu/become-an-ambassador/
Take part!

• PIDapalooza: Reykjavik, 9-10 November!

• If you’re doing something interesting with PIDs (or you want to!) come and share your ideas with a crowd of innovators

• Submit a short proposal: www.pidapalooza.org
Help us make research better, become member!

https://www.datacite.org/members

support@datacite.org