Data, Software Citation and PIDs

Frances Madden, Research Identifiers Lead
British Library, https://orcid.org/0000-0002-5432-6116
Aims for this morning

By lunchtime you will be able to:

• Understand the importance of persistent identifiers within research
• Cite data and software in research publications
• Understand some of the issues you may encounter
• Develop strategies to manage these citations in your own work
Housekeeping

• Please ask questions throughout the morning and be prepared to participate and answer my questions.

• Tell me to slow down if I am speaking too quickly!
Introductory Exercise
What is a Persistent Identifier?

A persistent identifier is a long-lasting reference to a resource.

Persistent Identifier

An organisation has made a commitment to keep it alive
Globally unique string
How PIDs work?

The short description….

Generally PIDs are backed by a registry which indicates what is being identified. These can have varying levels of metadata describing the resource, both depending on the PID and the who was registering it.

Often PIDs are expressed as URLs today and the registry describes where the PID should resolve, that way even if the URL changes the PID will always point to the correct item.
What types of things have a PID?

- Journal articles
- People
- Datasets
- Software
- Organizations
- and many more....
Why use PIDs?

- Trustworthiness
- Disambiguation
- Linking
- Making outputs FAIR

Photo by Stephanie Ronquillo on Unsplash
Trustworthiness

- Threshold to become a registration agency – reputable
- Contractual requirements
- Longstanding trust
Disambiguation

Robin Dasler

ORCID iD
https://orcid.org/0000-0002-4695-7874

Also known as
RH Dasler, RL Dasler, RL Howard, Robin Howard

Other IDs
ResearcherID: N-9035-2013

<table>
<thead>
<tr>
<th>ORCID ID</th>
<th>First name</th>
<th>Family name</th>
<th>Other names</th>
<th>Affiliations</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="https://orcid.org/0000-0001-8410-6010">https://orcid.org/0000-0001-8410-6010</a></td>
<td>Christine</td>
<td>Ferguson</td>
<td>Crls Ferguson, C. L. Ferguson, Christine Ferguson</td>
<td>Murray State University, Furman University, Mississippi State University, University of Tennessee, University of Richmond</td>
</tr>
<tr>
<td><a href="https://orcid.org/0000-0002-2261-6290">https://orcid.org/0000-0002-2261-6290</a></td>
<td>Christine</td>
<td>Ferguson</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="https://orcid.org/0000-0002-9317-6819">https://orcid.org/0000-0002-9317-6819</a></td>
<td>Christine</td>
<td>Ferguson</td>
<td>Christine A Ferguson</td>
<td></td>
</tr>
</tbody>
</table>
References


Making Research FAIR

Findable Accessible Interoperable Reusable
How to get a persistent identifier?

Zenodo

Institutional Repository

Github

ORCID

Figshare
When to use a PID?

- When talking about your work, Twitter, blogs etc.
- Assign PIDs to all kinds of outputs e.g. posters, presentations etc.
- Citing others work
- Add ORCID to presentations, email signatures etc.
- Everywhere else you can think of!

Resources
- PIDs for Researchers [https://www.pidforum.org/c/knowledge-hub/pids-for-researchers/34](https://www.pidforum.org/c/knowledge-hub/pids-for-researchers/34)
Have a break!
Exercise

• Read the article – at least to page 6

• Note where you think there should be a citation

• How would you cite the dataset used in this article?
Why cite resources?

- Inform reader about the original source of information
- Gives authority to your work
- Gives credit to other researchers
- Reproducibility
Format of a dataset citation

• Author
• Publication Date
• Title
• Location
• Publisher
• Also
  • Version
  • Identifier
  • Resource Type

Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014
https://doi.org/10.25490/a97fegyk

Example Data Citations

APA

Chicago

Zenodo recommended format
https://doi.org/10.7717/peerj-cs.86
Checklist for an author

• Have I *identified the software*, including my own, which makes a significant and specialised contribution to my academic work, and therefore should be cited?

• Have I checked if the software has a *recommended citation* from the creators and used it if available? If this is to a paper, have I also cited the software directly?

• Have I *created as complete a citation as possible* if no recommended citation is given? Does it include who created the software, when it was created, the title of the software (and version if available) and where the software can be accessed (preferably via a persistent identifier to an archival repository)?

• Have I *referenced the software appropriately* in my academic work, in compliance with any citation formatting guidelines?

Example Software Citations


Potential Issues

- **Embargoed/Closed** Data
- Supplementary Materials
- Granularity
- Versioning
- Evolving datasets
- Bi-directional linking
- Finding a creator/copyright issues

Resources

What are you going to do now?
Evaluation

https://bit.ly/2qIfekl
Thank you!

Frances Madden, Research Identifiers Lead
British Library, https://orcid.org/0000-0002-5432-6116
frances.madden@bl.uk