Data, Software Citation and PIDs

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

BRITISH LIBRARY

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!

CITE THE DATA

University of Essex

UK Data Service

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

🖨 Print view ❓

Also known as ▾

RH Dasler, RL Dasler, RL Howard, Robin Howard

Other IDs ▾

ResearcherID: N-9035-2013

ORCID iD	First name	Family name	Other names	Affiliations
https://orcid.org/0000-0001-8410-6010	Christine	Ferguson	Cris Ferguson, C. L. Ferguson, Christine Ferguson	Murray State University, Furman University, Mississippi State University, University of Tennessee, University of Richmond
https://orcid.org/0000-0002-2261-6290	Christine	Ferguson		
https://orcid.org/0000-0002-9317-6819	Christine	Ferguson	Christine A Ferguson	

Linking

References

Alderslade R (1995) Human rights and medical practice, including reference to the joint Oslo statements of September 1993 and March 1994. J Public Health Med 17:335–342

Google Scholar

Al-Moujahed A, Alahdab F, Abolaban H, Beletsky L (2017) Polio in Syria: problem still not solved. Avicenna J Med 7:64–66. https://doi.org/10.4103/ajm.AJM_173_16

Article   Google Scholar

Asgary R (2015) Direct killing of patients in humanitarian situations and armed conflicts: the profession of medicine is losing its meaning. Am J Trop Med Hyg 92:678–680. https://doi.org/10.4269/ajtmh.14-0364

Article   Google Scholar

Banatvala N, Zwi AB (2000) Conflict and health. Public health and humanitarian interventions: developing the evidence base. BMJ 321:101–105. https://doi.org/10.1136/bmj.321.7253.101

Article   Google Scholar

Baskett PJ (1994) Ethics in disaster medicine. Prehosp Disaster Med 9:4–5

Article   Google Scholar

Ben Shahar I (1993) Disaster preparation and the functioning of a hospital social work department during the Gulf War. Soc Work Health Care 18:147–159. https://doi.org/10.1300/J010v18n03_14
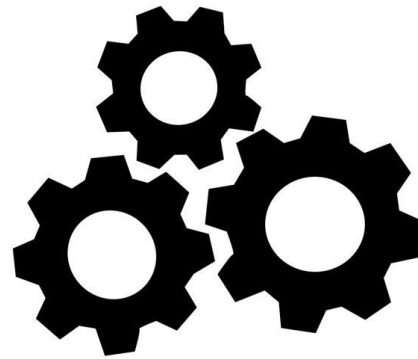
Article   Google Scholar

Making Research FAIR

F indable   A ccessible   I nteroperable   R eusable

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
- Why use PIDs? https://www.pidforum.org/t/why-use-persistent-identifiers/714
- PIDs for Researchers 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

Ball, A. & Duke, M. (2015). 'How to Cite Datasets and Link to Publications'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/how-guides

Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 https://doi.org/10.25490/a97f-egyk

Example Data Citations

APA

- Cool, H. E. M., & Bell, M. (2011). Excavations at St Peter's Church, Barton-upon-Humber [Data set]. doi: 10.5284/100038989

Chicago

- Haverals, Wouter. "Middle dutch syllabified words". Zenodo, July 31, 2018. doi:10.5281/zenodo.2402048.

Zenodo recommended format

- Eetu Mäkelä, Simon Hengchen, & Mikko Tolonen. (2019). Tweet IDs for the #brexit tweet dataset collected for the Helsinki Digital Humanities Hackathon 2019 (Version v1.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3380040

SOFTWARE CITATION PRINCIPLES

IMPORTANCE
Software should be considered a legitimate and citable product of research. Software citations should be accorded the same importance in the scholarly record as citations of other research products; they should be included in the metadata of the citing work, such as a reference list. Software should be cited on the same basis as any other research product such as a paper or a book.

UNIQUE IDENTIFICATION
A software citation should include a method for identification that is machine actionable, globally unique, interoperable, and recognized by at least a community of the corresponding domain experts, and preferably by general public researchers.

ACCESSIBILITY
Software citations should facilitate access to the software itself and to its associated metadata, documentation, data, and other materials necessary for both humans and machines to make informed use of the referenced software.

PERSISTENCE
Unique identifiers and metadata describing the software and its disposition should persist —even beyond the lifespan of the software they describe.

CREDIT AND ATTRIBUTION
Software citations should facilitate giving scholarly credit and normative, legal attribution to all contributors to the software, recognizing that a single style or mechanism of attribution may not be applicable to all software.

SPECIFICITY
Software citations should facilitate identification of, and access to, the specific version of software that was used. Software identification should be as specific as necessary, such as using version numbers, revision numbers, or variants such as platforms.

Smith AM, Katz DS, Niemeyer KE, FORCE11 Software Citation Working Group. 2016. Software citation principles. PeerJ Computer Science 2:e86
https://doi.org/10.7717/peerj-cs.86

Image courtesy of Datacite

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?

Chue Hong, Neil P., Allen, Alice, Gonzalez-Beltran, Alejandra, de Waard, Anita, Smith, Arfon M., Robinson, Carly, … Pollard, Tom. (2019, October 15). Software Citation Checklist for Authors (Version 0.9.0). Zenodo. http://doi.org/10.5281/zenodo.3479199

Example Software Citations

- Karsdorp, Folgert, Kestemont, Mike, & Riddell, Allen. (2019, December 4). Python Code in "Humanities Data Analysis" (Version 1.0). Zenodo. http://doi.org/10.5281/zenodo.3563075

- Giuseppe G. A. Celano. (2017, March 27). Tokenized and sentence-splitted CTSized Ancient Greek texts (v1.1.0) (Version v1.1.0). Zenodo. http://doi.org/10.5281/zenodo.438311

- Giuseppe G. A. Celano. (2017, March 20). Tokenized CTSized Ancient Greek texts. Zenodo. https://doi.org/10.5281/zenodo.598071

Potential Issues

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

Resources

- Ball, A. & Duke, M. (2015). 'How to Cite Datasets and Link to Publications'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/how-guides
- ANDS (n.d.). 'Data Citation for researchers. Melbourne. ANDS. Available online: https://www.ands.org.au/working-with-data/citation-and-identifiers/data-citation/data-citation-for-researchers

What are you going to do now?

Evaluation

https://bit.ly/2qIfekI

Thank you!

Frances Madden, Research Identifiers Lead

British Library, https://orcid.org/0000-0002-5432-6116

frances.madden@bl.uk