PIDs 101

Helena Cousijn (DataCite) & Ivo Wijnbergen (ORCID)
3 March 2020 - Lyngby, Denmark

What is a persistent identifier?

persistent identifier

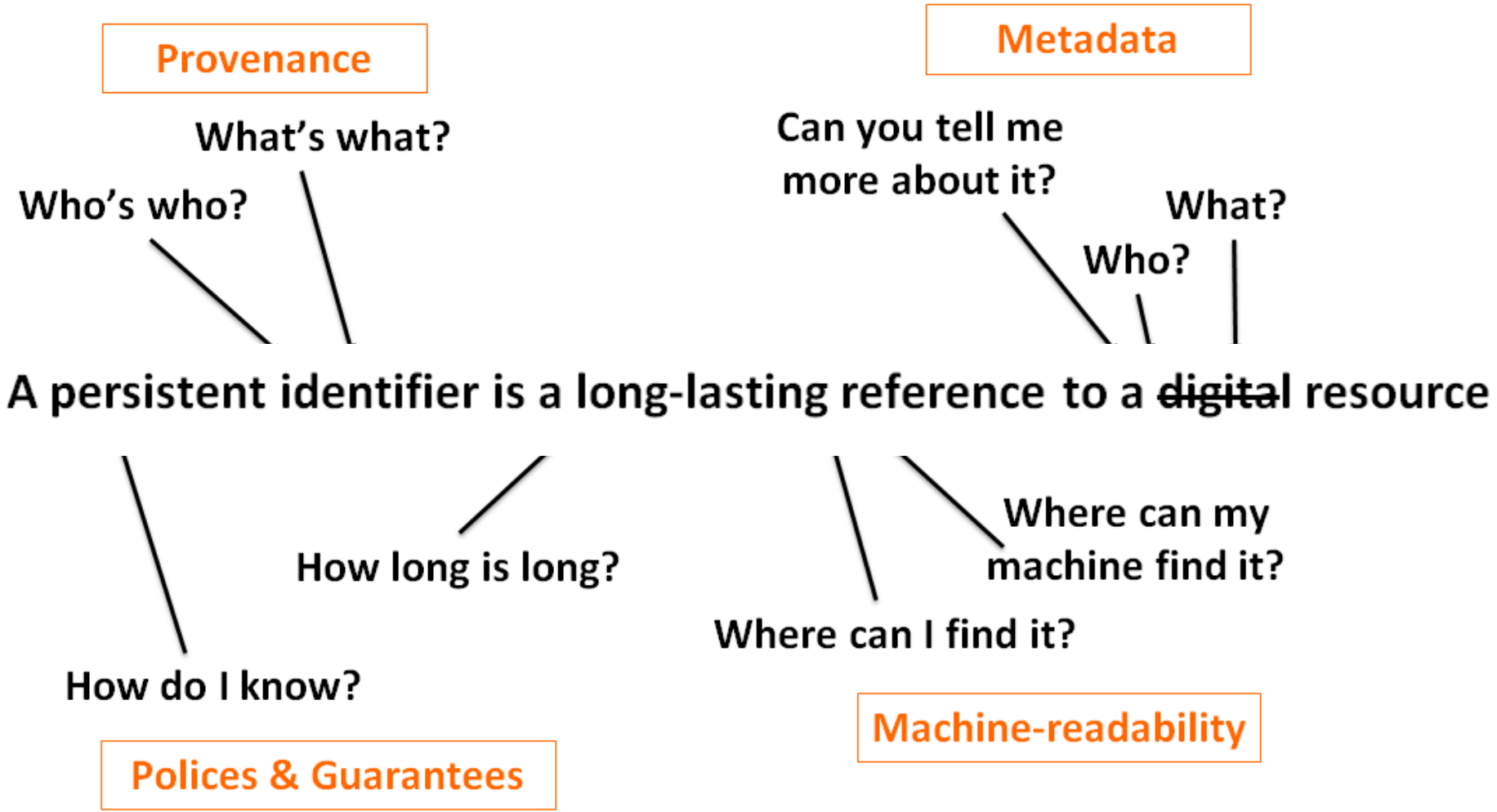
an organization
made a promise
to keep it alive

globally unique
string

(known as PIDs to their friends)

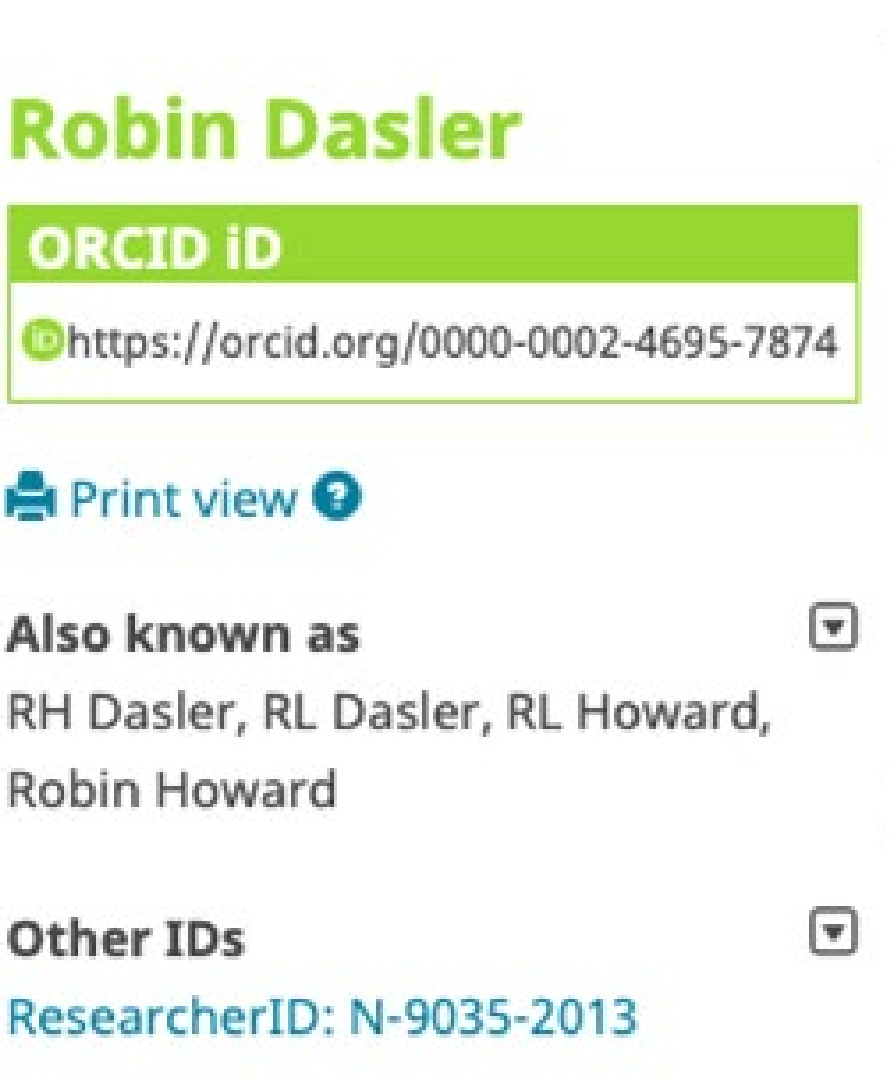
A trustworthy PID system must

- be maintained by a dedicated and reliable team,
- be based on a transparent sustainable business model,
- be provided by a non -profit organisation,
- be subject of regular quality assessments by external parties,
- be governed by international boards,
- be based on open standards,
- be based on a redundant and secure architecture,
- support a huge address space (comparable or even larger than IPv6) and
- support an openly documented API optimally supporting accepted data models.


Provenance
What's what?
Who's who?
Metadata
Can you tell me more about it?
What?
Who?
A persistent identifier is a long-lasting reference to a ~~digital~~ resource
How long is long?
How do I know?
Where can my machine find it?
Where can I find it?
Polices & Guarantees
Machine-readability

. . . but what can PIDs *do*?

PIDs Disambiguate



PIDs Link

https://doi.org/10.1016/j.arabjc.2017.05.011

References

Abd Ellah and Abouelmagd, 2016 N.H. Abd Ellah, S.A. Abouelmagd
Surface functionalization of polymeric nanoparticles for tumor drug delivery: approaches and challenges
Expert Opin. Drug Deliv., 1–14 (2016),
10.1080/17425247.2016.1213238
Google Scholar

Abouelmagd et al., 2016 S.A. Abouelmagd, F. Meng, B.-K. Kim, H. Hyun, Y. Yeo
Tannic acid-mediated surface functionalization of polymeric nanoparticles
ACS Biomater. Sci. Eng. (2016), p. 6b00497,
10.1021/acsbiomaterials.6b004
Google Scholar

Ahmed et al., 2016 S. Ahmed, S. Annu, S.S. Yudha
Biosynthesis of gold nanoparticles: a green approach
J. Photochem. Photobiol. B: Biol., 161 (2016), pp. 141-153,
10.1016/j.jphotobiol.2016.04.034
Article Download PDF View Record in Scopus
Google Scholar

Akhavan et al., 2011 O. Akhavan, R. Azimirad, S. Safa, E. Hasani

PIDs can make research FAIR

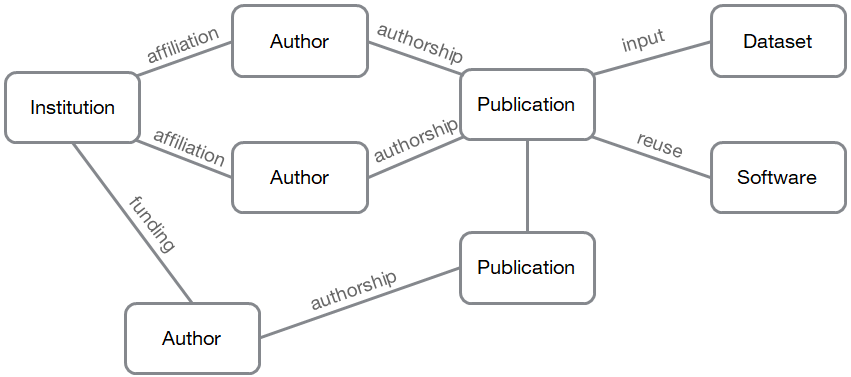
Data should be **Findable**	F1. (meta)data are assigned a globally unique and persistent identifier (DOI) F2. data are described with rich metadata F3. metadata clearly and explicitly include the identifier of the data it describes F4. (meta)data are registered or indexed in a searchable resource
Data should be **Accessible**	A1. (meta)data are retrievable by their identifier using a standardized communications protocol A1.1 the protocol is open, free, and universally implementable A1.2 the protocol allows for an authentication and authorization procedure, where necessary A2. metadata are accessible, even when the data are no longer available
Data should be **Interoperable**	I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (meta)data use vocabularies that follow FAIR principles I3. (meta)data include qualified references to other (meta)data
Data should be **Reusable**	R1. meta(data) are richly described with a plurality of accurate and relevant attributes R1.1. (meta)data are released with a clear and accessible data usage license R1.2. (meta)data are associated with detailed provenance R1.3. (meta)data meet domain-relevant community standards

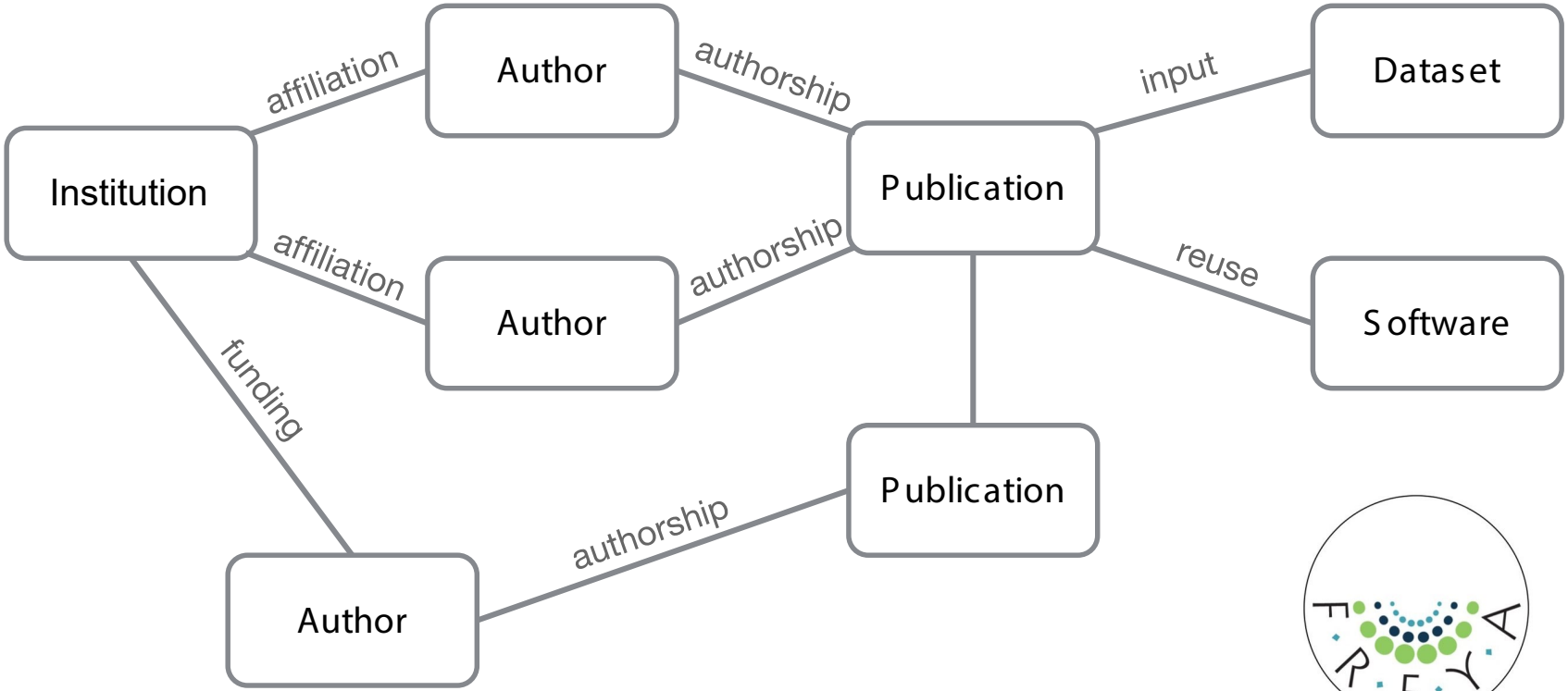
https://doi.org/10.1038/sdata.2016.18

Good start, but we want more

By connecting everything, you can see the true power of PIDs

Researchers, institutions, publications, datasets, and more are already interconnected in real life, and this can be reflected and tracked through PIDs

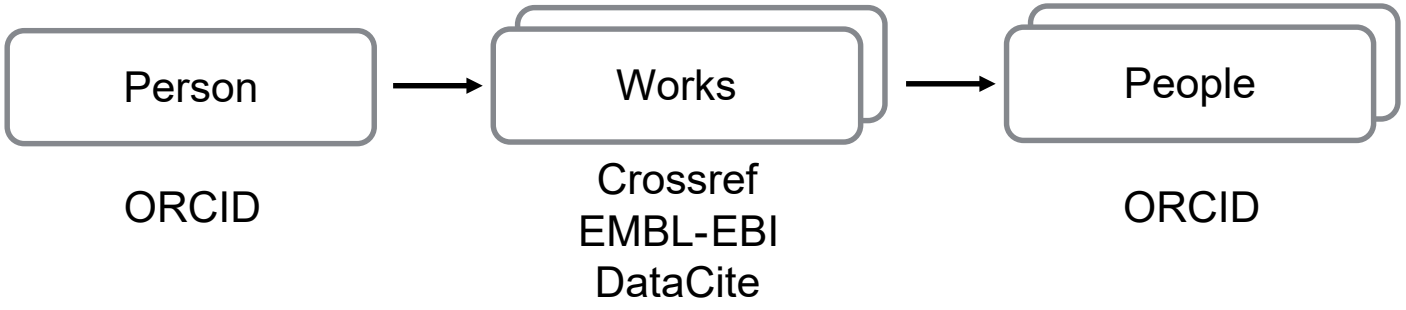




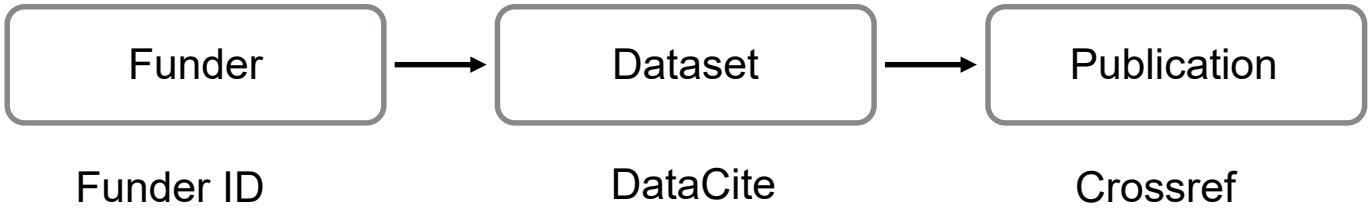
Connected PIDs form a graph

https://www.project-freya.eu/en

Who are all the co-authors of a given researcher?



Show all datasets funded by the European Commission that have been cited by a journal article



Which can be used to answer new questions

PIDs for (almost) everything

People, places, things

Persistent identifiers for people include ISNI and ORCID’s

Persistent identifiers for places include GRID and ROR

Persistent identifiers for things include Crossref and Datacite

ORCID

ORCID
Connecting Research and Researchers

- The Open Researcher and Contributor ID is used to uniquely identify scientific and other academic authors and contributors throughout their career, across disciplines, locations and time.
- ORCID iDs are in fact 16-digit alphanumeric ISNI codes.
- ORCID maintains an open and independent registry;
  - intended for contributor identification in research and academic publishing
  - accessible by systems through an API
- A not for profit, non-proprietary organization, sustained by the community

Our vision

A world where all who participate in research, scholarship, and innovation are uniquely identified and connected to their contributions across disciplines, borders, and time.

DataCite DOIs

DataCite
FIND, ACCESS, AND REUSE DATA

Non-profit membership organization working with 1000+ academic institutions in the world to provide DOIs for research data and other research outputs.

Connecting research, identifying knowledge.

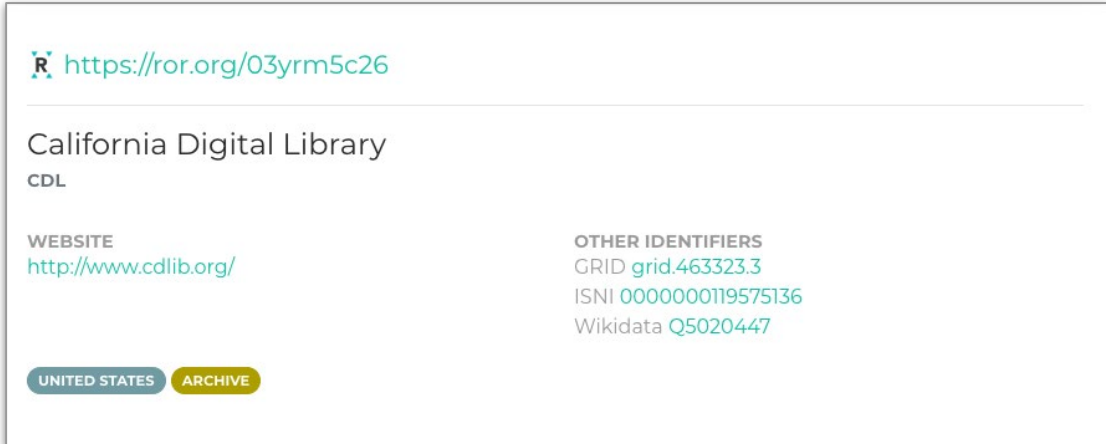
Mission:

DataCite's mission is to be the world's leading provider of persistent identifiers for research. Through our portfolio of services, we provide the means to create, find, cite, connect, and use research. We seek to create value and develop community - driven, innovative, open, integrated, useable and sustainable services for research.

https://datacite.org/

ROR

ROR is a **community-led project** to develop an **open**, sustainable, usable, and unique identifier for every **research organization** in the world.



https://ror.org/

What you should do after this workshop

Step 1: Give PIDs to your stuff

Get an ORCID iD for yourself→ https://orcid.org

Give DOIs to your data and software→ https://datacite.org , https://guides.github.com/activities/citable -code/

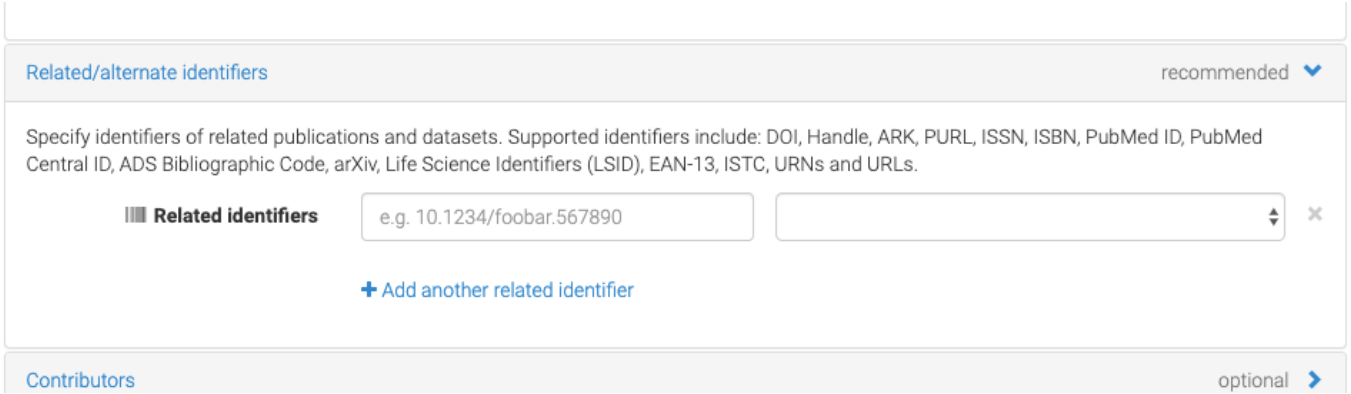
Put your reports and white papers into a repository that gives out PIDs→ https://repositoryfinder.datacite.org or your institutional repository

Make sure you include theRORof your institution in the metadata

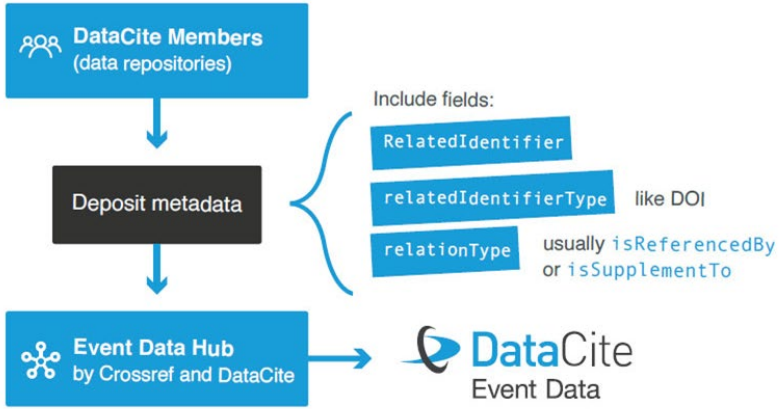
Step 2: Tell your PIDs about your other PIDs

Include relevant related PIDs in the metadata for your software, dataset, and paper PIDs, even if your repository says they're optional.

In Zenodo (for example), it looks like this:



Step 3: Share these connections with the community



Interested in using this information? Find out more at: https://support.datacite.org/docs/eventdata-guide

Thank you!

hcousijn@datacite.org
i.wijnbergen@orcid.org

Appendix

PID types overview and maturity

Research entity	PID types used[5]	Maturity of PID Infrastructure
Publication	DOI, Accession number, Handle, URN, Scopus EID, Web of Science UID, PMID, PMC, arXiv Identifier, BibCode, ISSN, ISBN, PURL	Mature
Citation	OCI (secondary aggregation of information)	Emerging
Conference	DOI, Accession number	Emerging
Researcher (or Scholar)	ORCID iDs, ISNI (also DAIs, VIAFs, arxivIDs, OpenIDs, ResearcherIDs, ScopusIDs)	Mature
Organization	DOI; ISNI, GRID, Ringgold IDs	Emerging

Source: FREYA deliverable D3.1 Survey of Current PID Services Landscape May 2018 https://www.project -freya.eu/en/deliverables/freya_d3-1.pdf

PID types overview and maturity

Research entity	PID types used[5]	Maturity of PID Infrastructure
Data	DOI, Accession number, Handle, PURL, URN, ARK	Mature
Data repository		Immature
Grants	DOI, PURL	Emerging
Project	local identifier, accession number, RAiD	Emerging
Experiment	none	immature
Investigation	DOI, Accession number	Emerging
Analysis	Git gist	Immature

Source: FREYA deliverable D3.1 Survey of Current PID Services Landscape May 2018 https://www.project -freya.eu/en/deliverables/freya_d3-1.pdf

PID types overview and maturity

Research entity	PID types used[5]	Maturity of PID Infrastructure
Software	DOI, SHA-1 hash	Emerging
Computer Simulation	UUID	Emerging
Software License	none	Immature
Equipment		
Instrument, Device, Sensor, Platform, Research Facility	DOI, RRID, UID	Emerging
Archival/Storage facility	URI, DOI, UUID	Emerging
Field Station	none	Immature

Source: FREYA deliverable D3.1 Survey of Current PID Services Landscape May 2018 https://www.project -freya.eu/en/deliverables/freya_d3-1.pdf

PID types overview and maturity

Research entity	PID types used[5]	Maturity of PID Infrastructure
Sample		
Geological or Biological Sample	Accession number, RRID, DOI, IGSN	Emerging
Cultural artefact	DOI, URN, Accession number	Emerging
Historical or mythical person	URI	Emerging
Temporal period & historical place	ARK, URI, accession number	Immature

Source: FREYA deliverable D3.1 Survey of Current PID Services Landscape May 2018 https://www.project -freya.eu/en/deliverables/freya_d3-1.pdf

PID types overview and maturity

Research entity	PID types used[5]	Maturity of PID Infrastructure
Study registration		
Clinical trial; non-clinical registration	accession number; DOI	Immature
Data Management Plan	DOI	Immature
Workflow	URI, DOI	Immature
Protocol	DOI	Immature

Source: FREYA deliverable D3.1 Survey of Current PID Services Landscape May 2018 https://www.project -freya.eu/en/deliverables/freya_d3-1.pdf