PIDs in the Context of Open Science

Helena Cousijn (DataCite), Paloma Marín-Arraiza (ORCID)

National EOSC Tripartite Event and Czech Open Science Day

June 1st, 2023









National Centre for Persistent Identifiers

A new team of the National Library of Technology formed in 2023

Mission: To support and coordinate the implementation of internationally recognised identifiers in the Czech R&D&I environment

National ORCID Centre



launching on June 1st, 2023

National DOI Centre



launching on June 1st, 2023

National ISSN Centre





Speakers

Helena Cousijn



Community Engagement Director, DataCite

Paloma Marín-Arraiza



Engagement Manager for Global Consortia, ORCID



Agenda

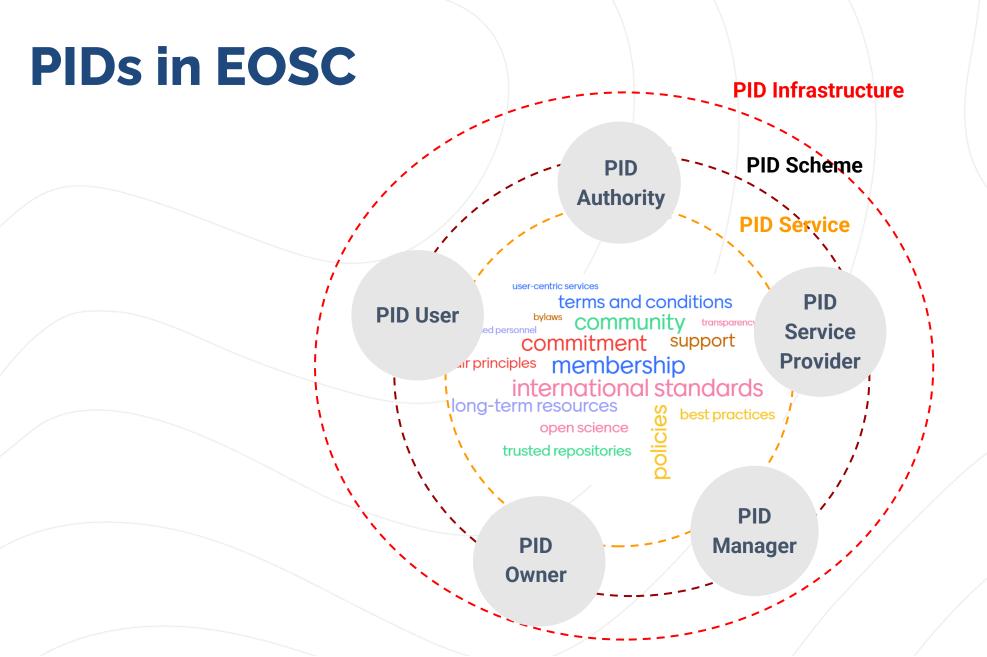
- Introduction to PIDs
- ORCID / DataCite Overview
- PIDs in the research cycle
- Examples from abroad

Q&A via Menti menti.com 8242 5205

Workshop is being streamed and recorded



An introduction to Persistent Identifiers



PIDs

What is a persistent identifier (PID)?

https://doi.org/10.5061/dryad.708g

Special URL that's registered in a known system, like DOI, ORCID or ROR



https://datadryad.org/stash/dataset/doi:10.5061/dryad.708gr

Always points to the same resource (or a metadata representation)

DOIs for scholarly outputs https://doi.org/10.5281/ zenodo.3630248



ORCID iDs for people https://orcid.org/ 0000-0001-6622-4910



ROR IDs for research organizations https://ror.org/01y2jtd41



PIDs for places, people, and things

PIDs for people (researchers) include ISNIs and ORCID iDs



PIDs for places (research organizations) including ROR



PIDs for things (research outputs) include DOIs, handles, IGSNs, ARKs, and more



Why are we talking about this?

DOIs (digital object identifiers) and other PIDs are an important part of the digital scholarly infrastructure that includes *all* types of resources/outputs and *all* disciplines.

Why are we talking about this?

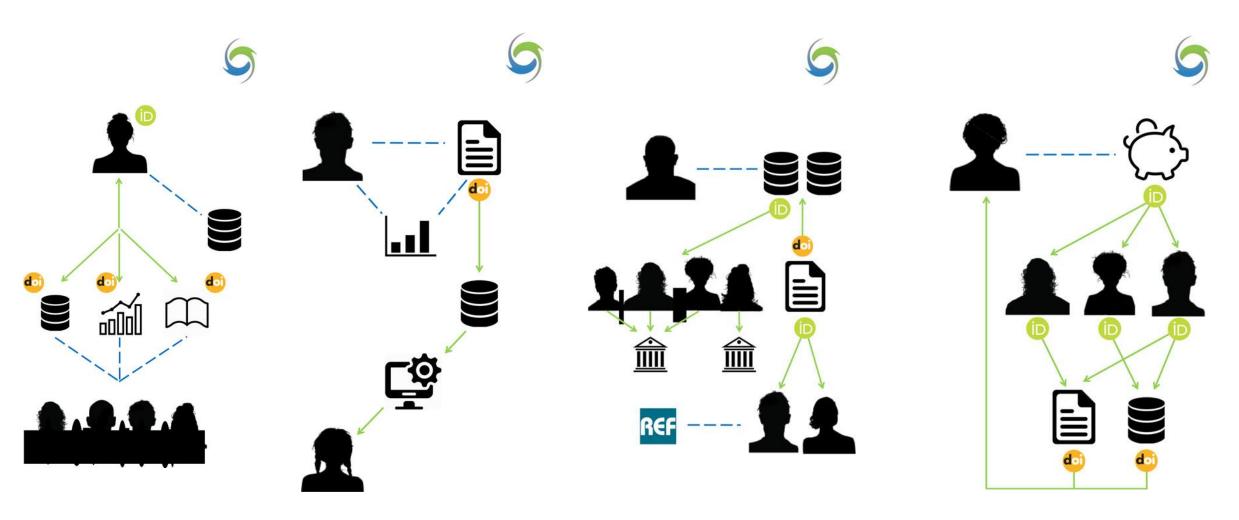
PIDs like DOIs, ORCID iDs and ROR IDs increase discovery, access, citation, reuse, and recognition of resources



PIDs are for everything!

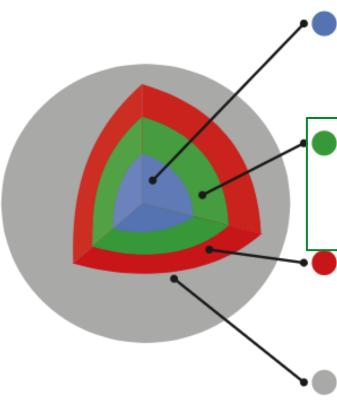
All disciplines, all resource types and all processes

PIDs for every thing and everything



Source: Dappert, A., Farquhar, A., Kotarski, R. and Hewlett, K., 2017. Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research. Data Science Journal, 16(0), p.28.DOI: https://doi.org/10.5334/dsj-2017-028

PIDs are a key part of FAIRification



DIGITAL OBJECT

Data, code and other research outputs

At its most basic level, data or code is a bitstream or binary sequence. For this to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and documentation. These layers of meaning enrich the object and enable reuse.

IDENTIFIERS

Persistent and unique (PIDs)

Digital Objects should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and supports citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders and associated research resources (RRIDs).

STANDARDS & CODE

Open, documented formats

Digital Objects should be represented in common and ideally open file formats. This enables others to reuse them as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code use to process and analyse the data.

METADATA

Contextual documentation

In order for Digital Objects to be assessable and reusable, they should be accompanied by sufficient metadata and documentation.

Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the objects were created. To enable the broadest reuse, they should be accompanied by a plurality of relevant attributes and a clear and accessible usage license.



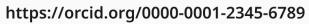
ORCID's Mission

ORCID's mission is to enable transparent and trustworthy connections between researchers, their contributions, and their affiliations by providing a unique, persistent identifier for individuals to use as they engage in research, scholarship, and innovation activities.



The ORCID iD







An ORCID record





Application Programming Interfaces (APIs)





Application Programming Interfaces (APIs)





Application Programming Interfaces (APIs)





The ability to disambiguate researchers is critical for a trustworthy scholarly record

What's in a name?

Most names are not unique



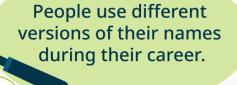
Many people have the same or similar name, and names may change through marriage or other circumstances.



Individuals and organizations can use diferent alphabets, abbreviations, or naming conventions.







Peta-electron volt gamma-ray emission from the Crab Nebula

The LHAASO Collaboration*,†, Zhen Cao, F. Aharonian, Q. An, Axikegu, L. X. Bai, Y. X. Bai, Y. W. Bao, D. Bastieri, X. J. Bi, Y. J. Bi, H. Cai, J. T. Cai, Zhe Cao, J. Chang, J. F. Chang, B. M. Chen, E. S. Chen, J. Chen, Liang Chen, @ Liang Chen, Long Chen, M. J. Chen, M. L. Chen, Q. H. Chen, S. H. Chen, S. Z. Chen, T. L. Chen, X. L. Chen, Y. Chen, N. Cheng, Y. D. Cheng, S. W. Cui, X. H. Cui, Y. D. Cui, B. D'Ettorre Piazzoli, B. Z. Dai, H. L. Dai, Z. G. Dai, Danzengluobu, D. della Volpe, X. J. Dong, K. K. Duan, J. H. Fan, Y. Z. Fan, Z. X. Fan, J. Fang, K. Fang, C. F. Feng, L. Feng, S. H. Feng, Y. L. Feng, B. Gao, C. D. Gao, L. Q. Gao, O. Gao, W. Gao, M. M. Ge, L. S. Geng, G. H. Gong, O. B. Gou, M. H. Gu, F. L. Guo, J. G. Guo, X. L. Guo, Y. O. Guo, Y. Y. Guo, Y. A. Han, H. H., He, H. N. He, J. C. He, S. L. He, X. B. He, Y. He, M. Heller, Y. K. Hor, C. Hou, X. Hou, H. B. Hu, S. Hu, S. C. Hu, X. J. Hu, D. H. Huang, Q. L. Huang, W. H. Huang, X. T. Huang, X. Y. Huang, Z. C. Huang, F. Ji, X. L. Ji, H. Y. Jia, K. Jiang, Z. J. Jiang, C. Jin, T. Ke, D. Kuleshov, K. Levochkin, B. B. Li, Cheng Li, Cong Li, F. Li, H. B. Li, H. C. Li, H. Y. Li, Jian Li, Jie Li, K. Li, W. L. Li, X. R. Li, Xin Li, Xin Li, Y. Li, Y. Z. Li, Zhe Li, Zhuo Li, E. W. Liang, Y. F. Liang, S. J. Lin, B. Liu, C. Liu, D. Liu, H. Liu, H. D. Liu, J. Liu, J. L. Liu, J. S. Liu, J. Y. Liu, M. Y. Liu, R. Y. Liu, S. M. Liu, W. Liu, Y. Liu, Y. N. Liu, Z. X. Liu, W. J. Long, R. Lu, H. K. Lv, B. Q. Ma, L. L. Ma, X. H. Ma, J. R. Mao, A. Masood, Z. Min, W. Mitthumsiri, T. Montaruli, Y. C. Nan, B. Y. Pang, P. Pattarakijwanich, Z. Y. Pei, M. Y. Qi, Y. Q. Qi, B. Q. Qiao, J. J. Qin, D. Ruffolo, V. Ruley, A. Saiz, L. Shao, O. Shchegoley, X. D. Sheng, J. Y. Shi, H. C. Song, Yu. V. Stenkin, V. Stepanov, Y. Su, Q. N. Sun, X. N. Sun, Z. B. Sun, P. H. T. Tam, Z. B. Tang, W. W. Tian, B. D. Wang, C. Wang, H. Wang, H. G. Wang, J. C. Wang, J. S. Wang, L. P. Wang, L. Y. Wang, R. N. Wang, Wei Wang, 💿 Wei Wang, X. G. Wang, X. J. Wang, X. Y. Wang, Y. D. Wang, Y. D. Wang, Y. P. Wang, Z. H. Wang, Z. X. Wang, Zhen Wang, Zheng Wang, D. M. Wei, J. J. Wei, Y. J. Wei, T. Wen, C. Y. Wu, H. R. Wu, S. Wu, W. X. Wu, X. F. Wu, S. Q. Xi, J. Xia, J. J. Xia, G. M. Xiang, D. X. Xiao, G. Xiao, H. B. Xiao, G. G. Xin, Y. L. Xin, Y. Xing, D. L. Xu, R. X. Xu, L. Xue, D. H. Yan, J. Z. Yan, C. W. Yang, F. F. Yang, J. Y. Yang, L. L. Yang, M. J. Yang, R. Z. Yang, S. B. Yang, Y. H. Yao, Z. G. Yao, Y. M. Ye, L. Q. Yin, N. Yin, X. H. You, Z. Y. You, Y. H. Yu, Q. Yuan, H. D. Zeng, T. X. Zeng, W. Zeng, Z. K. Zeng, M. Zha, X. X. Zhai, B. B. Zhang, H. M. Zhang, H. Y. Zhang, J. L. Zhang, J. W. Zhang, L. X. Zhang, Li Zhang, P. F. Zhang, P. P. Zhang, R. Zhang, S. R. Zhang, S. S. Zhang, X. Zhang, X. P. Zhang, Y. F. Zhang, Y. L. Zhang, Yi Zhang, Yong Zhang, B. Zhao, J. Zhao, L. Zhao, L. Z. Zhao, S. P. Zhao, F. Zheng, Y. Zheng, B. Zhou, H. Zhou, J. N. Zhou, P. Zhou, R. Zhou, X. X. Zhou, C. G. Zhu, F. R. Zhu, H. Zhu, K. J. Zhu, X.

Trivalent NDV-HXP-S Vaccine Protects against Phylogenetically

Distant SARS-CoV-2 Variants of Concern in Mice

Authors: Irene González-Domínguez . Jose Luis Martínez, Stefan Slamanig . Nicholas Lemus, Yonghong Liu, Tsoi Ying Lai, Juan Manuel Carreño, Gagandeep Singh 💿, Gagandeep Singh 💿, Michael Schotsaert, Ignacio Mena, Stephen McCroskery, Lynda Coughlan 💿, Floria mmer 💿, Adolf cía-Sastre 💿, Peter Palese 🖼, Weina Sun

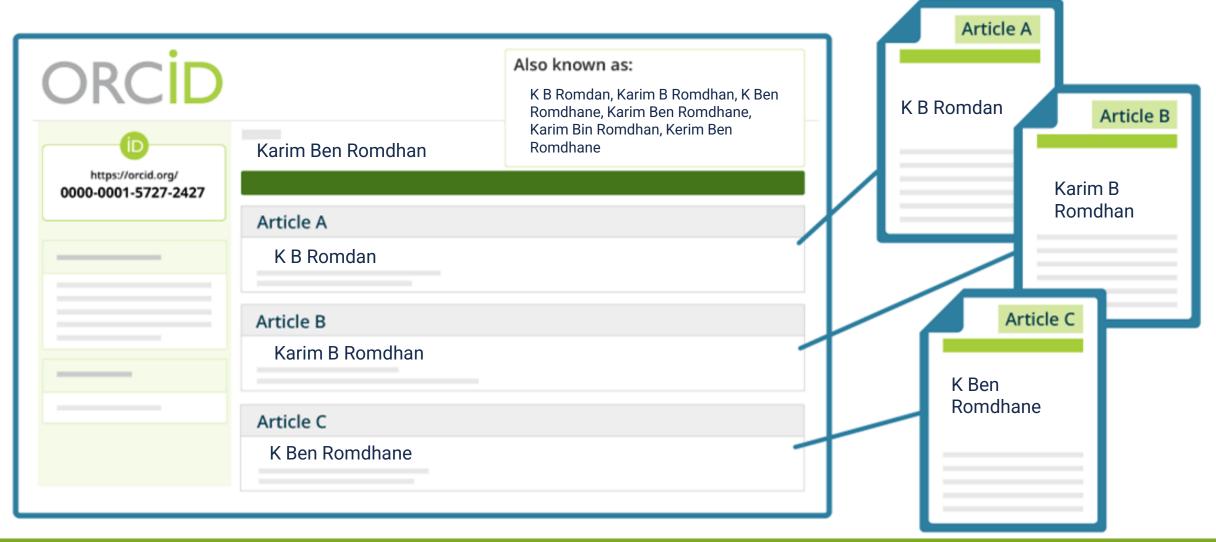


8 | Virology | Research Article | 06 June 2022



f y in M

ORCID also solves name variation, the iD never changes

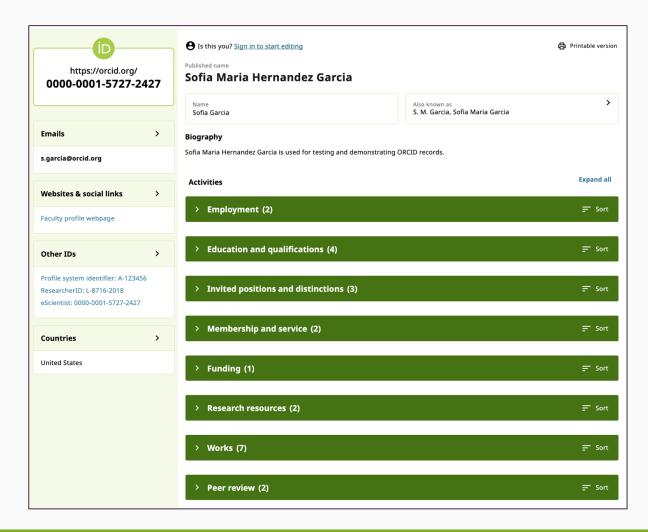




Information on ORCID records can help

Entries that can be added:

- Affiliations
- Professional activities
- Funding information
- Outputs
- Peer reviews
- Research Resources
- Website URLs
- Other non-sensitive personal data



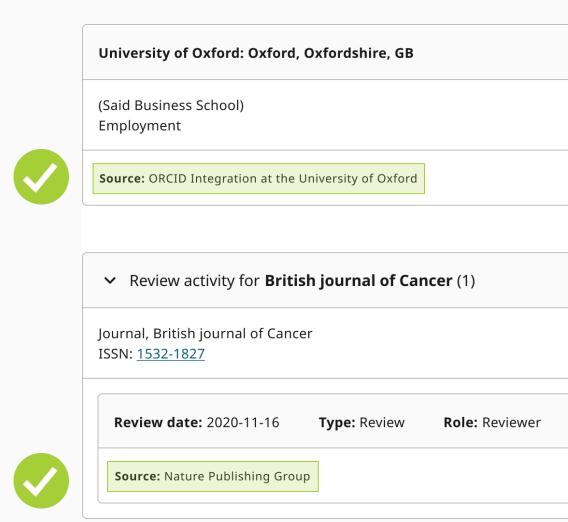


ORCID member organizations add validated information

When an ORCID member updates an ORCID record, the source (provenance) of that update is captured for re-use:

- Research organizations add affiliations
- Publishers add outputs and reviews
- Funders add funding items

These provide 'trust markers' that can be used to help in decision making.





We aim to balance researcher control and data quality

Researchers:

- Own their own records
- Control who accesses their information
- May change access preferences at any time





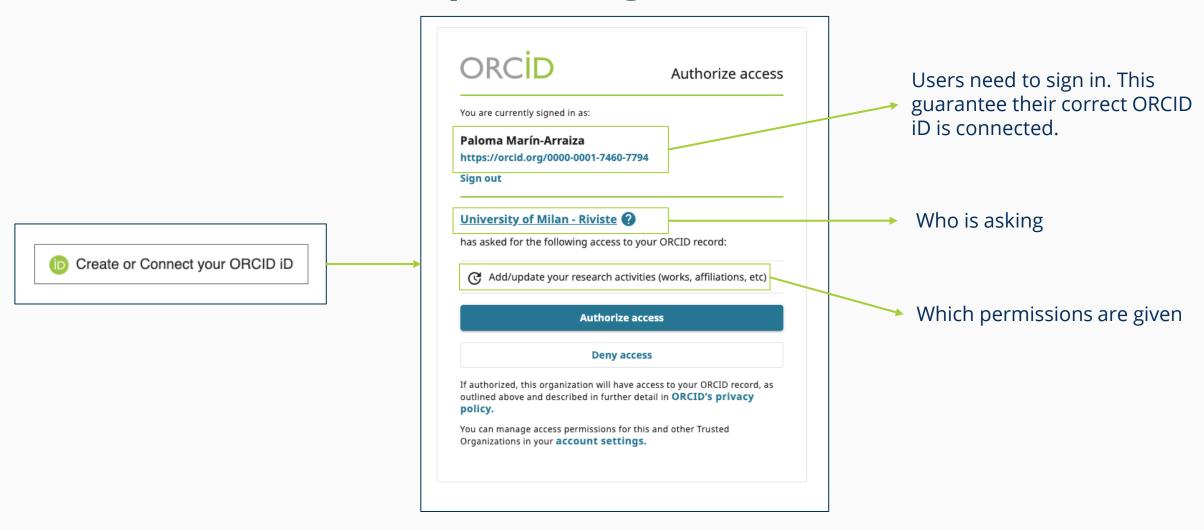
- Terms of use prevent misrepresentation
- False data in records may be disputed by anyone in the community
- Machine-learning algorithm detects obvious spam
- Disputed and spam records are removed from use, pending correction or withdrawal

Organizations may only add information to ORCID records with the researcher's permission, and may only update or delete information that was added by them



Trust in an ORCID record accumulates over time as reliable and trustworthy data sources add information (with the record holder's permission)

The authentication process guarantees data control

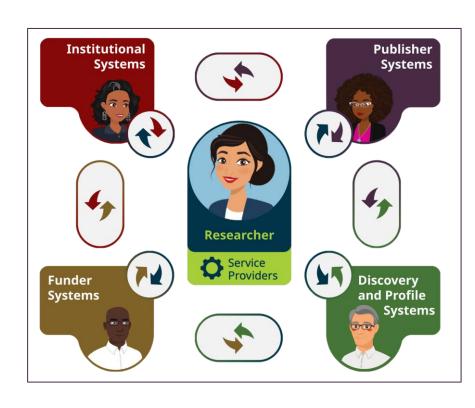




ORCID is increasingly a hub for an immense amount of profile activity with over 4700 interconnected systems

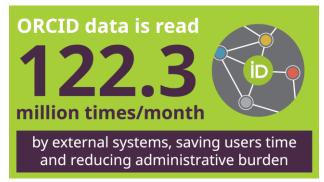
Software connected with ORCID:

- ePrints
- Open Journal Systems
- DSpace 7.x
- DSpace CRIS
- SmartSimple
- Proposal Central
- ..









Constant exchange with the global community



Members Connect (Monthly Newsletter)



Friends of ORCID channel



Member Mixers (coming soon)





About DataCite



Global non-profit membership organization working with 2800+ repositories in the world to provide DOIs for research outputs and resources.

We are a global community that share a common interest: to ensure that research outputs and resources are openly available and connected so that their reuse can advance knowledge across and between disciplines, now and in the future.



Connecting research, identifying knowledge

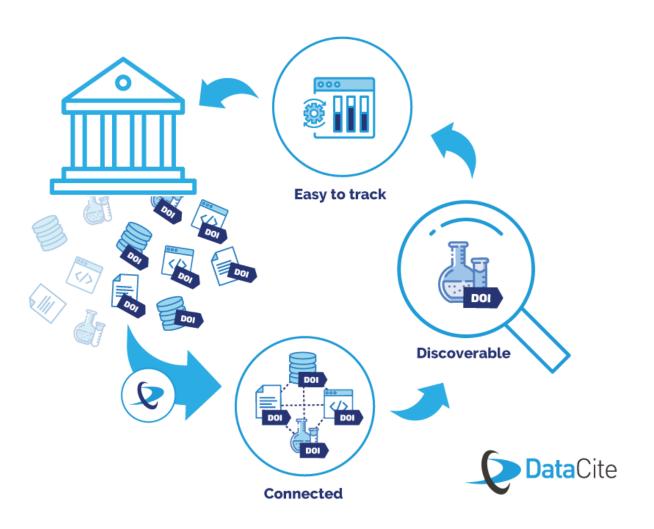
How do we connect research?



DOIs & metadata for a wide range of research outputs:

- 1. Research datasets and collections, associated workflows, software, images, models, samples, DMPs
- 2. Grey literature such as theses, dissertations, reports, unpublished conference papers, newsletters, preprint journal articles, technical standards, and specifications for which the institutional repository is the primary publication point.





Our value





Registering DataCite DOIs makes your research outputs discoverable.

- · A DOI makes your research outputs uniquely identifiable.
- · Metadata that you register with DataCite is in a central location, harvestable by anyone.
- · Metadata for our Members' research outputs appear in other search engines.



DataCite services make it easy to follow best practices.

- · We make research data management easy: you register your first DOI in less than 1 minute.
- · DataCite DOIs and metadata help you make your research FAIR.
- · We connect you to the DataCite Member community, which is full of passionate people who share experience and continue to support best practice.
- Our metadata schema is extensive and has been adopted by other PID service providers globally.



DataCite services help you track and report on your research.

- · A DOI enables easy tracking of your research outputs through simple user interfaces.
- · DataCite services make institutional reporting simple.
- DataCite services support data citation and usage analytics



Our community











2800+

280+

50

47m+

1300+

Repositories

Members

Countries

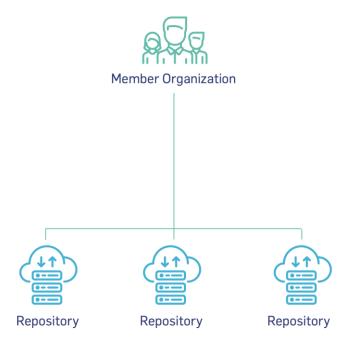
DOIs

Organizations

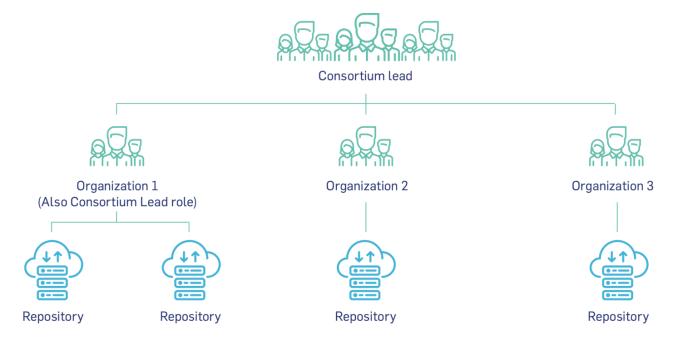
DataCite Membership



Direct Membership



Consortium Membership



Createand Manage DOIs

DataCite membership allows you to create and manage DOIs for all of your repositories. You can do this through:

- Our primary REST API that supports JSON and enables automated DOI registration
- Our manual interface that enables you to register DOIs in less than a minute.
- Registered DataCite Service providers that provide a platform where you can register DOIs.



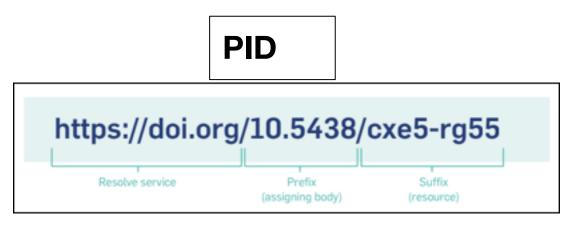
Services



We sustain open infrastructure services to ensure that research outputs and resources are easily findable, accessible, interoperable, and reusable (FAIR).

	Create DOIs	Integrate Workflows	Enable Discovery	Promote Reuse
Why?	Make research outputs and resources discoverable and citable for the long term	Enhance research workflows and services through integration	Enable discovery of research outputs and resources and reporting with advanced tools and analytics	Promote reuse with flexible, state-of-the art tools and technology
What?	 DOI & metadata management services Link checker 	APIs to automate system integrationUsage tracker	 Metrics & relational metadata files Search & analytics dashboards Data Citation Corpus 	 Content negotiation & citation formatter ORCID auto updates & claiming
How?	Register DOIs and provide rich metadata in standardized formats	Automate DOI registration and/or retrieve relevant metadata through system integration	Make outputs resolvable and discoverable through metadata including connections between outputs/resources and other entities	Enable reuse through good citation practices and embedding DOIs in scholarly infrastructure workflows

PIDs are an integral piece in solving the FAIR puzzle



INDENTIFIER

URL/LANDING PAGE

TITLE

CREATOR

PUBLISHER

RESOURCE TYPE

ENRICH WITH OPTIONAL METADATA



• (meta)data are assigned a globally unique and persistent identifier



- (meta)data are retrievable via an identifier using a standardized protocol
- metadata are accessible, even when the data are no longer available



- (meta)data use a formal, accessible, shared, and broadly applicable language knowledge representation.
- (meta)data include qualified references to other (meta)data

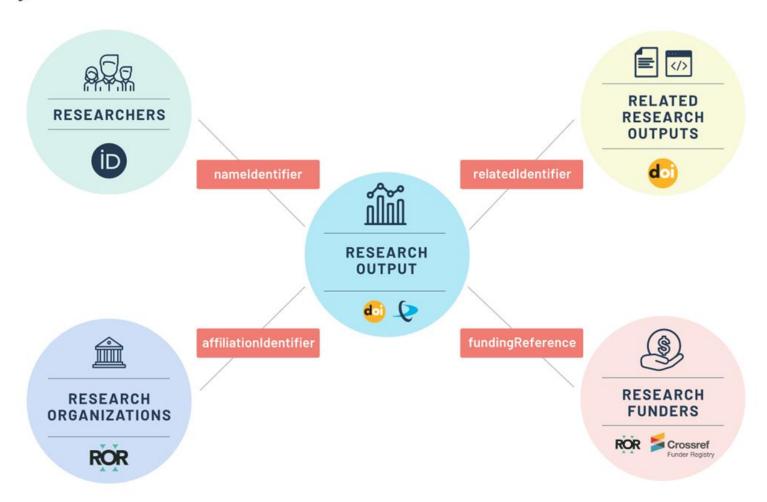


- meta(data) are described with a plurality of accurate & relevant attributes
- (meta)data are associated with detailed provenance

METADATA

DataCite Connection Metadata

Connect DataCite DOIs to every part of research ecosystem

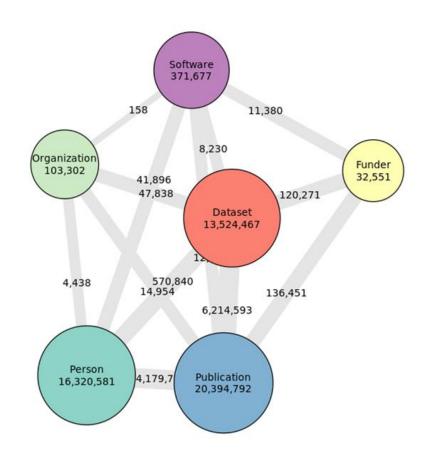


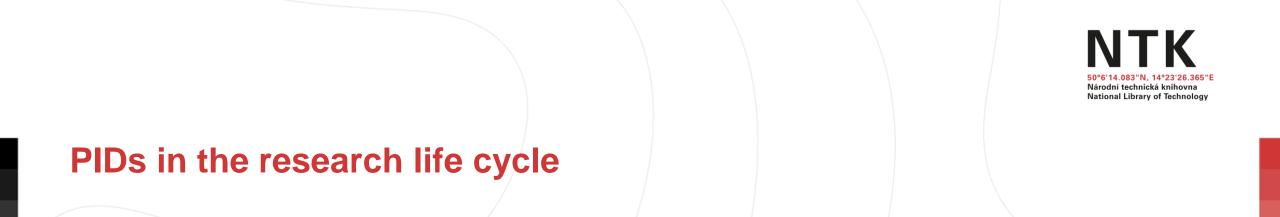
Connected PIDs



PID Graph

Number of nodes and connections (7 March 2023)

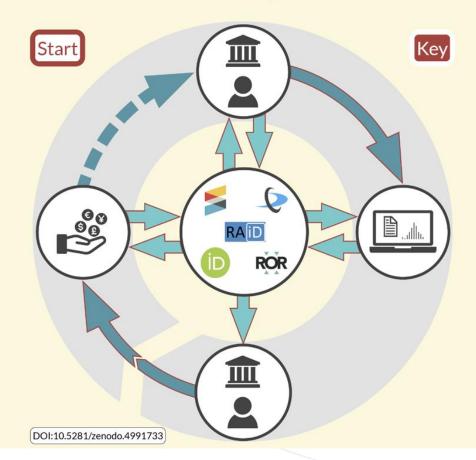




The PID optimised research cycle

MOREBRAINS

The PID-optimised research cycle





Actions to take in the research process

4. Research design

Researcher to develop research design, metadata model and experimental protocol.

3. Submit Data Management Plan

Researcher to develop and submit DMP, use DMP-ID in subsequent reporting.

2. Register Grant

Funder to register grant and assign Grant ID, provide instruction for usage.

1. Grant Application

Researcher to use relevant PIDs (team member, affiliation, funder) in the grant application process.



5. Experiment conducted

Researcher to conduct experiments proposed in pre-registration according to protocol.



6. Research code, data, other outputs deposited

Researcher to deposit protocols, data, conference posters, etc.

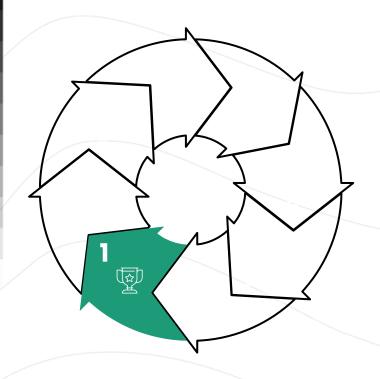
7. Preprint submitted

Researcher to share manuscript on preprint server.

8. Research article published

Researcher to select appropriate publishing channel and submit manuscript to journal, journal to register articles and submit metadata to Crossref

Grant Application

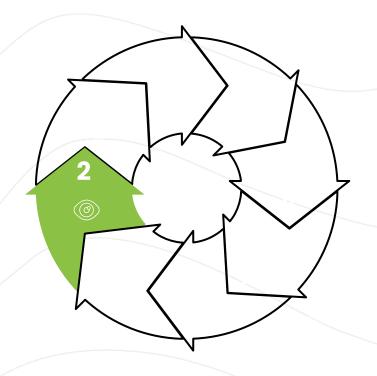


Researchers to obtain an ORCID and use their ORCID iDs during the application process (and subsequent research process).

Researchers to include the ROR ID of their affiliated institute/organization during the application process.

Key activities

Registering Grant

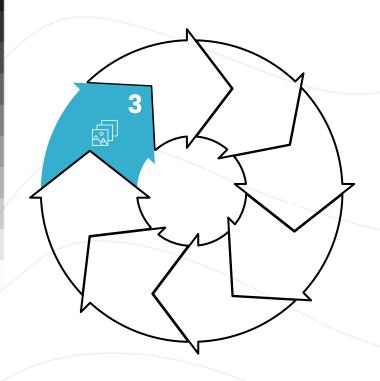


Funder to use an organization PID to self-identify.

Funder to assign Grant ID once the grant is awarded, and to provide the Grant ID to the awardees and instruction for the application of Grant ID in reportable outputs.

Key activities

Data Management Planning



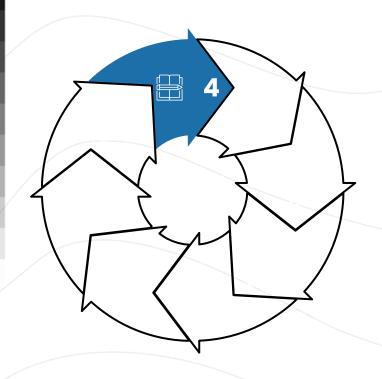
Researcher to develop and share DMP and register a DMP-ID.

Researcher to associate outputs with the DMP.

Researcher or research administrator to consider registering project identifier (e.g. RAiD).

Key activities

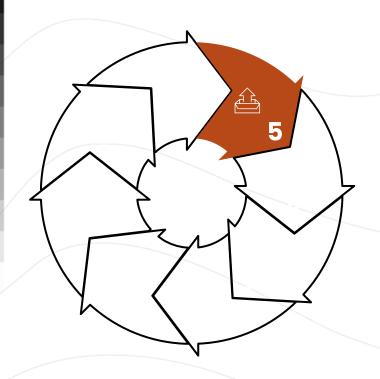
Research design



Researcher to consider submitting a registered report or preregistration to a platform that registers DOIs.

Researcher to create rich metadata for research activities by keeping metadata records and using connection metadata.

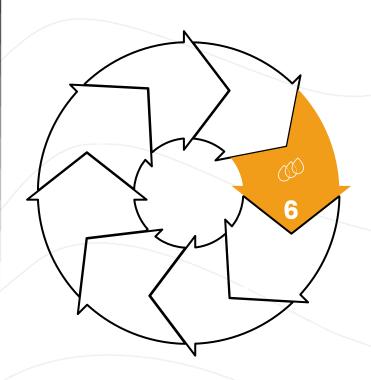
Conducting experiments



Researcher to build experimental protocols, making new protocols publicly available or reusing existing protocols connected through metadata.

Researcher to publicly communicate the progress and interim outputs on platforms registering PIDs.

Deposition of code, data, and other outputs

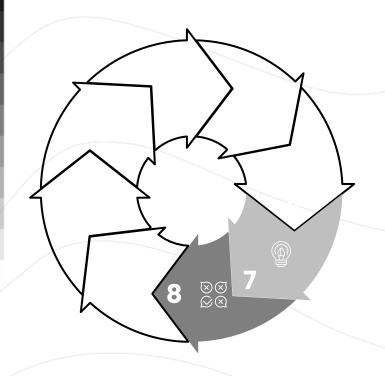


Researcher to prepare and share dataset with PIDs and metadata.

Researcher to prepare and share data analysis scripts/software with PIDs and metadata.

Researcher to connect these to the grant/project/DMP

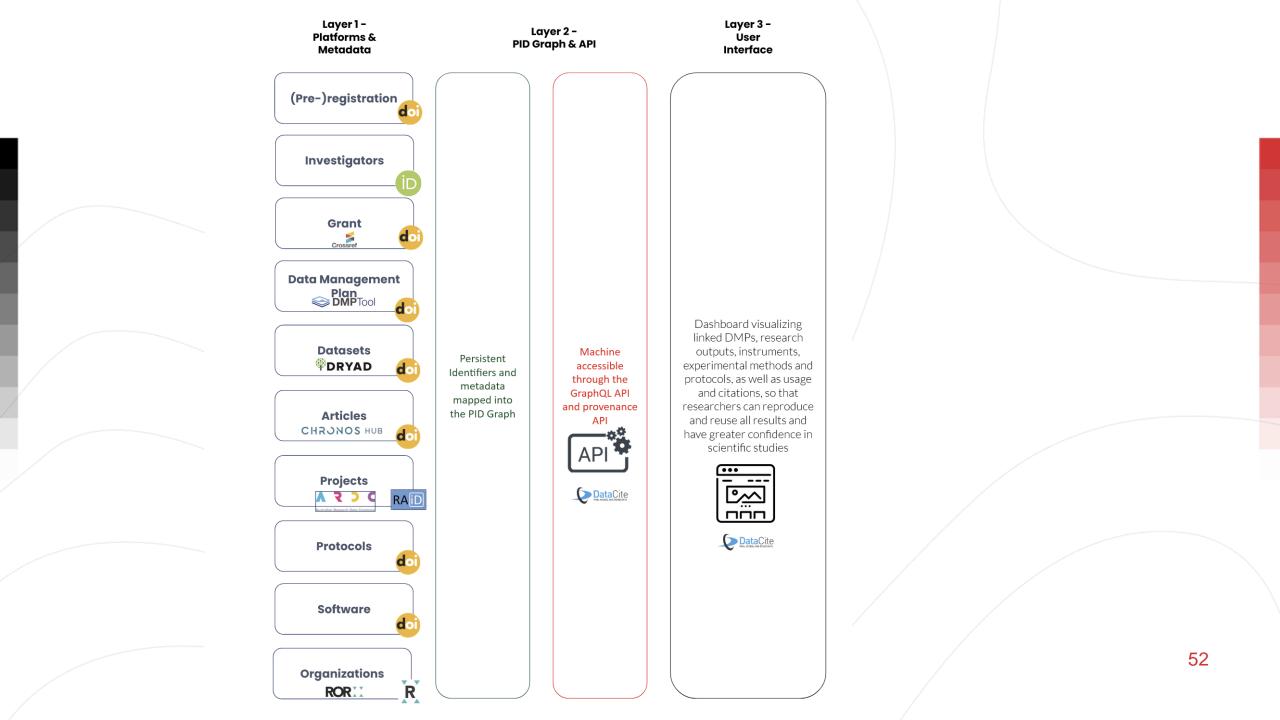
Preprint submission and article publication



Researcher to submit the manuscript to a preprint server.

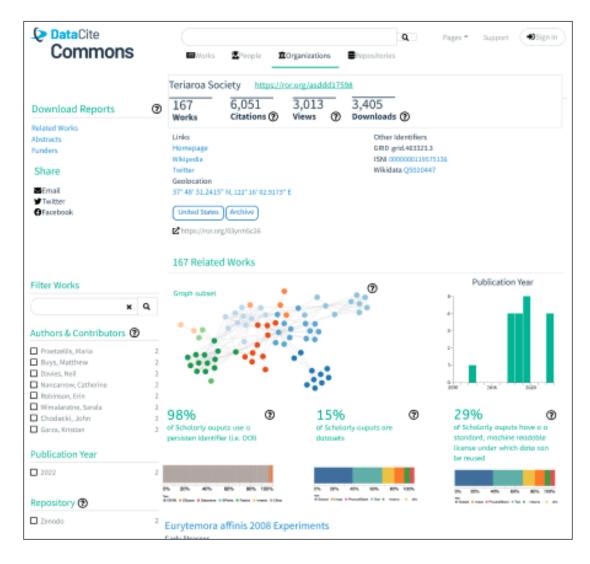
Researcher to submit the manuscript to a journal.

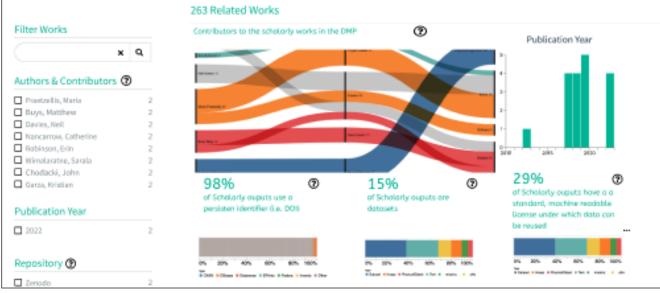
Journal/preprint server to register DOIs and make connection metadata available.



Connected research









International approaches to PIDs Implementation

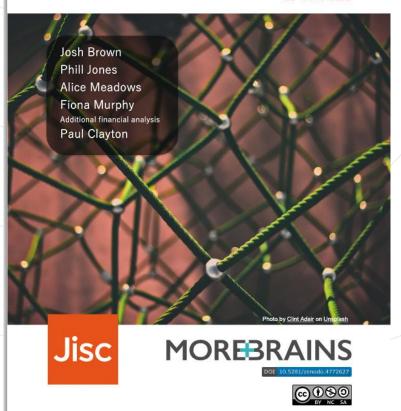
PID cost-benefits analysis focus on metadata reuse, automation, aggregation and analysis



UK PID Consortium

Cost-Benefit Analysis

21st June 2021



MORESRAINS

Incentives to Invest in Identifiers

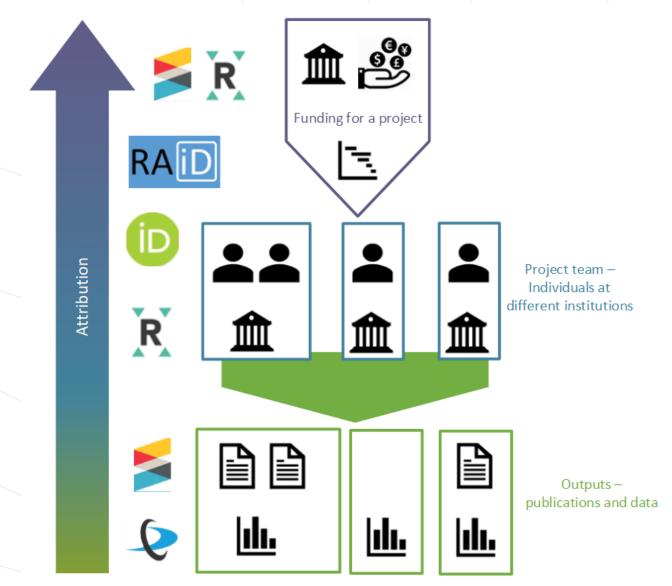
A cost-benefit analysis of persistent identifiers in Australian research systems



Sources:

PID National strategy in the UK





PID National Strategy in Germany









PIDs für Forschungsdaten







PIDs für Instrume nte



PIDs für kulturelle Objekte und deren Kontexte



PIDs für Organisat ionen und Projekte



PIDs für Personen



PIDs für physische Objekte



PIDs für PI
Repositorien put
&
Publikationsdien

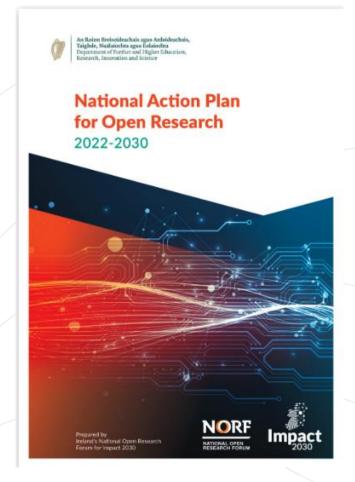
ste



PIDs für Textpublikatio nen PIDs für Software

Alignment with the National Action Plan for Open Research: the Irish case





Action	Description	Key stakeholders	Timeline
A4.4 Invest in Persistent Identifier infrastructure to enable consistent monitoring and improve interoperability.	A4.4.1 Support the Irish ORCID Consortium and encourage further development and adoption of ORCID according to international best practice by researchers and within the systems and processes of publishers, research performing organisations, research funding organisations, and infrastructures. ³⁴	NORF, RPOs, RFOs, research infrastructures (e.g. IReL)	Ongoing
	A4.4.2 Develop a national roadmap for the adoption of a range of Persistent Identifiers according to international best practice, such as ORCID, DOIs, RAiDs and ROR identifiers. Implement this roadmap to consolidate national coordination and accelerate the uptake and integration of priority identifiers.	NORF, RPOs, RFOs, research infrastructures (e.g. IReL)	2023-27

Alignment with the National Action Plan for Open Science: the Italian case





PIANO NAZIONALE PER LA SCIENZA APERTA

Green,

f. favorire il pieno utilizzo di ORCID-ID per ogni ricercatore, nelle sue funzionalità di abilitare e connettere servizi, anche legati alle citazioni;

ENG: encourage the full use of ORCID-ID for each researcher, in its functionality of enabling and connecting services, including citation-related services;

PIDs in policies



NATIONAL SCIENCE AND TECHNOLOGY COUNCIL



GUIDANCE FOR IMPLEMENTING NATIONAL SECURITY PRESIDENTIAL MEMORANDUM 33 (NSPM-33) ON NATIONAL SECURITY STRATEGY FOR UNITED STATES GOVERNMENT-SUPPORTED RESEARCH AND DEVELOPMENT

A Report by the

Subcommittee on Research Security

Joint Committee on the Research Environment

January 2022

GUIDANCE FOR IMPLEMENTING NATIONAL SECURITY PRESIDENTIAL MEMORANDUM 33 (NSPM-33) ON NATIONAL SECURITY STRATEGY FOR UNITED STATES GOVERNMENT-SUPPORTED RESEARCH AND DEVELOPMENT

Table of Contents

Foreword	v
Vision for Our Shared Research Environment	vii
Introduction to Implementation Guidance	ix
General Implementation Guidance	1
Disclosure Requirements and Standardization	2
Digital Persistent Identifiers	8
Consequences for Violation of Disclosure Requirements	11
Information Sharing	15
Research Security Programs	18
Appendix: Definitions	22

PIDs in policies





UKRI Open Access Policy

Contents:

Policy scope

Requirements for research articles

Requirements for long-form publications

Monitoring and compliance

Funding and policy implementation

Further information

Annex 1: Additional information on policy definitions and scope

Annex 2: Technical requirements for research articles

Glossary

Version Control

Technical requirements for journals and publishing platforms

- 4. To be considered compliant with UKRI's open access requirements, research articles and conference proceedings with an ISSN made available via journals and publishing platforms are required to meet the following technical requirements:
 - Persistent Identifiers (PIDs) for articles must be implemented according to international recognised standards, examples of international standards include Digital Object Identifiers (DOI), Uniform Resource Name (URN) or Handle
 - article-level metadata must be used according to a defined application profile that supports the UKRI Open Access Policy and is available, if possible, via a Creative Commons public domain dedication (CC0); the metadata standard must adhere to international best practice such as the Crossref schema and OpenAIRE guidelines
 - machine-readable information on the open access status and the licence must be embedded in the article metadata in a standard non-proprietary format
 - d. long-term preservation must be supported via a robust preservation programme such as CLOCKSS, Portico or an equivalent
 - e. openly accessible data on citations must be made available according to the standards set out by the Initiative for Open Citations (I4OC)
 - f. self-archiving policies must be registered in the SHERPA RoMEO database
 - g. common unique PIDs for research management information (for example identifiers for funders and/or organisations) are strongly encouraged; ORCID, the researcher identifier must be supported.

Policies for PIDs





AP 6 Persistent Identifiers

Institutional Model Policy for the Registration of Digital Object Identifiers (DOIs)

Andreas Ferus¹, Silvia Gstrein², Anna-Laetitia Hikl³, Christian Kaier⁴, Michael Kranewitter⁵, Paloma Marín Arraiza⁴, Adelheid Mayer³

- 1 Akademie der bildenden Künste Wien
- ² Universität Innsbruck
- 3 Universität für Bodenkultur Wien
- ⁴ Universität Graz
- ⁵ Johannes Kepler Universität Linz
- 6 Technische Universität Wien
- 7 Universität Wien

March 2019



The content of this publication is licensed under a Creative Commons Attribution
4.0 License. (http://creativecommons.org/licenses/by/4.0/)

DOI 10.25651/1.2019.0002

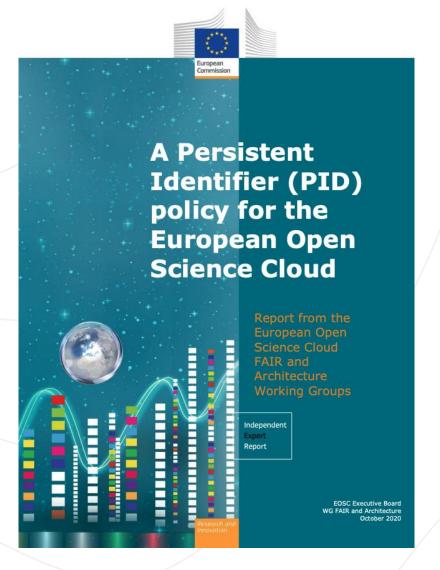
About E-Infrastructures Austria Plus

The project "e-Infrastructures Austria Plus" (2017-2019) is a project of nine Austrian universities funded by the Austrian Federal Ministry of Education, Science and Research (https://bildung.bmbwf.gv.at/). The aim of the project is the coordinated development of an Austrian network for the establishment and further development of common e-Infrastructures by bundling resources and existing knowledge.

62

PID policies





63

Thank you very much Děkujeme mnohokrát

Helena Cousijn, helena.cousijn@datacite.org, @HelenaCousijn/@DataCite

Paloma Marín-Arraiza, <u>p.arraiza@orcid.org</u>, @pmarrai/@ORCID_Org





