

# PID in research metadata

SIST 2022, June 15



[jonathan.schaeffer@cnsr.fr](mailto:jonathan.schaeffer@cnsr.fr)



# Introduction

---

# The context of research data activities management

Research metadata need to reference external objects

- › scientific publications
- › datasets
- › data providers
- › people and contact
- › organizations
- › instruments

In a perfect world (for a computer engineer), every object must be identified reliably in the long term. And the identifier would come with some valuable attributes for this object.

Example: identifying a person by it's name is not reliable :

- › there are people with same name
- › people can change their name
- › email can change also, as phone numbers, etc.

# The F-1 of FAIR

From Go-Fair website<sup>1</sup>:

*Principle F1 is arguably the most important because it will be hard to achieve other aspects of FAIR without globally unique and persistent identifiers.*

---

<sup>1</sup><https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/>

# Solutions provided by Persistent Identifiers

- › identifies uniquely an object
- › object's attributes are embedded with the identifier

Each type of identifier comes with an attribute schema trying to describe the object and people involved in it.

For instance, DataCite metadata [[The Metadata Working Group, 2021](#)] emphasis attribution to persons and organizations involved in the object management (dataset, article, ...).

# Digging the PID system

---

# A registry for PID

A registry provides a service to :

- › register an object
- › describe the associated metadata
- › link the PID to the object
- › means to access the metadata associated with the PID

# Example of registries for research activities

- › DataCite to get a DOI
- › ePIC registers PIDs handles
- › ORCID to register ORCID ID
- › IDRef for people in french education/research
- › RE3DATA registering data repositories with a R3D id
- › ROR registering organizations with ROR ID
- › ScanR : french **search engine** for all research and innovation activities, putting a ScanRID on each object
- › HAL : french open archive repository mints HAL-ID
- › RRID : identifies key biological resources

And many more ...



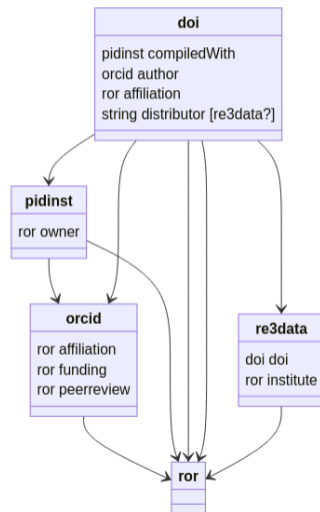
# Focus on some PID for research metadata

# Overview

- › DOI for publications, datasets, any academic material
- › ORCID ID for people
- › ROR ID for organizations
- › PIDINST for instruments
- › RE3DATA : for data repositories
- › Funder Registry (crossref) : for referencing project funders

# Links

PID attributes can be used to establish links between objects. This schema shows some of the possibilities.



# Focus on ROR

Identifying institutes. <https://ror.org/>  
<https://ror.readme.io/docs/ror-basics>

Initialized from GRID database (so your institute might be registered without you knowing it, take a look)

- GRID has been retired (Q4 2021) and passes the torch to ROR
- ROR metadata is very minimalist and easy to manage : 5 required metadata fields and a landing page.
  1. name
  2. organization types
  3. addresses
  4. status (active or not)
  5. country

So, go for it !

# Focus on PIDINST

Recent production from the RDA  
Persistent Identification of instruments  
WG<sup>a</sup>

Identifying a measuring instrument is  
useful for:

- › publishing contextual information
- › reproductibility and automatic processing
- › interpretation of the data produced

---

<sup>a</sup><https://www.rd-alliance.org/groups/persistent-identification-instruments-wg>

Metadata associated :

- › has been implemented in the ePIC infrastructure.
- › is possible to match with DataCite metadata.

# PIDINST mapping on DataCite metadata

- ▶ PIDINST schema has a few attributes described in [Krahl et al., 2022]
- ▶ There is a mapping with DataCite schema (to assign a DOI to your instrument)<sup>a</sup>  
This way, an instrument can be assigned with a DOI or a Handle.

<sup>a</sup><https://rda-pidinst.readthedocs.io/en/latest/datacite-cookbook/metadata.html>

Attribute	Obligation
Identifier	M
SchemaVersion	M
LandingPage	M
Name	M
Owner	M
Manufacturer	M
Model	R
Description	R
InstrumentType	R
MeasuredVariable	R
Date	R
RelatedIdentifier	R
AlternateIdentifier	R

# Personal humble general recommendations for PID service provider [1/2]

So you need to provide PIDs for your community's data ?

- ▶ Do not manage your PIDs manually  
when you have to make dozens of them, the problems arise :
  - ▶ maintaining the metadata
  - ▶ maintaining the landing pages

I do not know of any free and generic PID management software

# Personal humble general recommendations for PID service provider [1/2]

So you need to provide PIDs for your community's data ?

- ▶ Do not manage your PIDs manually  
when you have to make dozens of them, the problems arise :
  - ▶ maintaining the metadata
  - ▶ maintaining the landing pages

I do not know of any free and generic PID management software

- ▶ Delegate metadata content to the appropriate person.  
There can be a lot of political issues in the content. It is wise to identify the appropriate person to manage the metadata



# Personal humble general recommendations for PID service provider [2/2]

So you need to provide PIDs for your community's data ?

- ▶ Publish the DOI even if the metadata content is minimal.
  - ▶ use schema values for unknown information (:unav, :tba)<sup>2</sup>
  - ▶ if the PI has some interest in it, she will provide the information

---

<sup>2</sup><https://support.datacite.org/docs/schema-values-unknown-information-v43>

<sup>3</sup><https://jamstack.org/>

# Personal humble general recommendations for PID service provider [2/2]

So you need to provide PIDs for your community's data ?

- ▶ Publish the DOI even if the metadata content is minimal.
  - ▶ use schema values for unknown information (:unav, :tba)<sup>2</sup>
  - ▶ if the PI has some interest in it, she will provide the information
- ▶ As a DOI service provider, manage a landing pages system by yourself:
  - ▶ Do not point to landing pages you have no control of (out of your domain)
  - ▶ The best is to generate landing pages dynamically and host them in your domain
  - ▶ Wise use of the registries API and a jamstack<sup>3</sup> website generator can be easy and reliable
  - ▶ Link your thematic metadata to standard ID on standard registries (ORCID, ROR, ...)

---

<sup>2</sup><https://support.datacite.org/docs/schema-values-unknown-information-v43>

<sup>3</sup><https://jamstack.org/>

# Takeaway message

PID ecosystem has grown vast and complex, but is now mature and stable.  
To get the best out of the system, link PIDs between them.

# Bibliography

---

# References I

R. Krahl, L. Darroch, R. Huber, A. Devaraju, J. Klump, T. Habermann, M. Stocker, and R. P. W. Members. Metadata Schema for the Persistent Identification of Instruments, Mar. 2022. URL <https://doi.org/10.15497/RDA00070>.

The Metadata Working Group. Datacite metadata schema documentation for the publication and citation of research data v4.3. *DataCite*, 2021. doi: 10.14454/7XQ3-ZF69. URL <https://schema.datacite.org/meta/kernel-4.4/>.