HELMHOLTZ
Metadata Collaboration

Towards a Helmholtz Data Space – adjusting responsibilities for metadata data by the use of PIDS

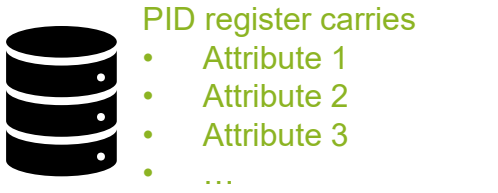
Emanuel Söding (GEOMAR), Andrea Pörsch (GFZ), Yousef Razeghi (UFZ)
Dorothee Kottmeier (AWI), Sören Lorenz (GEOMAR)

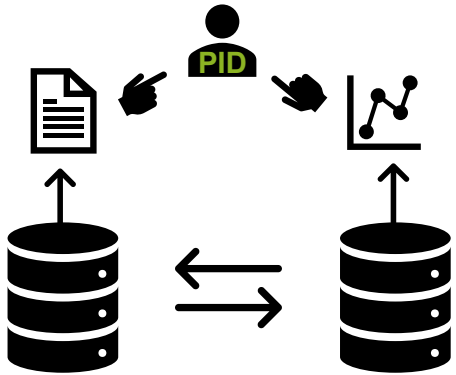
HMC Conference 2023

www.helmholtz-metadaten.de

PIDs Status quo … What is the definition and use of PIDs?

- A **persistent identifier** (**PI** or **PID**) is a long-lasting reference to a document, file, web page, or other object (from Wikipedia).
- PID registers typically store core metadata with the PID
- PIDs are used in data infrastructures to link to common information
- Harmonizing the use of PIDs may improve interoperability of datasets between data repositories



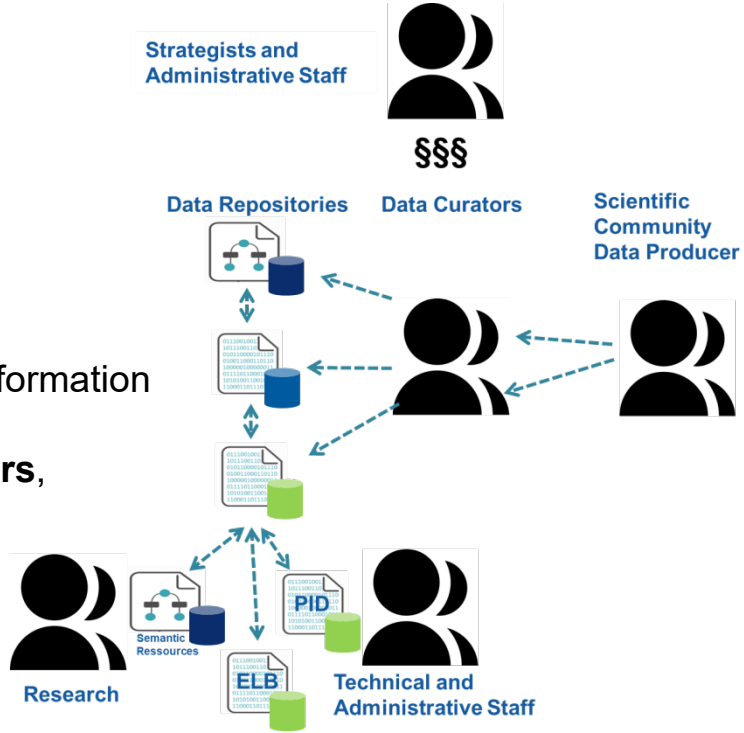


Responsible parties … What are the different roles in the use of PIDs?

Stakeholders to be involved:

- **Scientific community** producing data
- **Data stewards** & curators
- **Data infrastructures**
- **Technicians** and **administration** responsible for relevant information
- **Expert panels**, **strategists** and **administrative stakeholders**,

  who adopt recommendations and issue policies



PID Systems … What is the best way to evaluate them?

How can we evaluate, if PID systems are stable and ready for use?
We plan to use these criteria:

- ✓ PID is **globally unambiguously** registerable
- ✓ PIDs are **widely used**, e.g. recommended by national and international organizations
- ✓ PID registrar has **transparent governance**
- ✓ Handle system has **stable funding**
- ✓ PID system is **reliably available** and maintained
- ✓ PIDs have an **adequate metadata set**
- ✓ PID registration is **open** to all (vs. paying users only)
- ✓ PID is **machine readable** via API
- ✓ Further developments are **downward compatible** (with previous updates, content and technical)

Also: What **other PID systems** capture the same types of data?

Example ORCID … What could a recommendation look like in detail?

Recommendation M1.0: It is recommended that all **data infrastructures** and repositories **use ORCID** to identify people and contributors to resources in data infrastructures wherever and whenever possible.

Recommendation M1.1: It is recommended, that **center management** should

- Employ measures and incentives to **actively encourage all employees to register with ORCID** and to keep their metadata current and share the relevant parts of this data with the center.
- **Keep record of their staffs ORCIDs** and share them with the RDM teams, where appropriate, in order to record ORCIDs with their output.

Recommendation M1.2: It is recommended, that **employees** should

- **register with ORCID** if they haven't done so already and
- **keep their ORCID related metadata current** and
- **share the relevant parts** of this data with the center for use in the centers data systems in order to keep the information in the systems current.

Recommendation M1.3: It is recommended that **data infrastructures (data repositories, data bases)** should:

- **record an ORCID with any person** registered in conjunction with the metadata of datasets, publications, instruments and alike where possible.
- **treat ORCID metadata as the primary source of truth** and update the associated metadata accordingly.
- Optionally **inform persons** if they think the metadata registered with the ORCID is not accurate, or request permission to update the ORCID metadata (see also M1-1).

Example ORCID … What are the Challenges in using ORCID?

The use of ORCID would be slightly changed.

The use of ORCID in Helmholtz is originally **authors** of publications in publication repositories.
Here we intend to broaden the use to **anyone contributing to a dataset**, rather than just the role of author for ORCID.

ORICD is not present or inaccurate.

Institutions or repositories cannot rely on ORCIDs to be present or accurate for every person referenced with published datasets.
They also cannot create or maintain ORCIDs for people on their behalf, even if they know the data is outdated.

In cases where an ORCID is not available for a person, what can we do?

The use of ORCID may lead to data privacy issues.

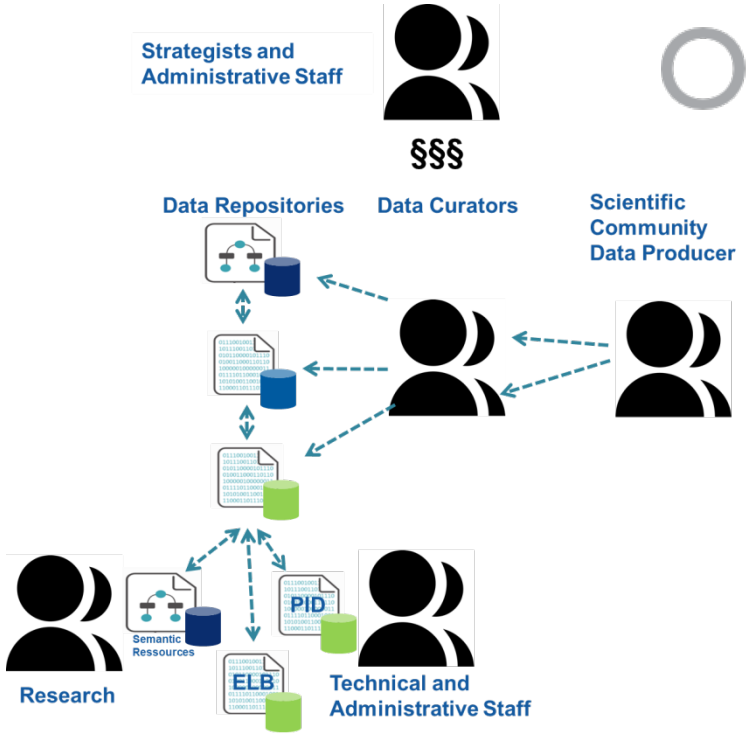
People's role and impact within the research community can be tracked by combining personal and bibliographical data.
Data privacy concerns should be discussed with the data privacy officers at the respective centers.

Example ORCID … What are the different roles in the use of ORCID?

Roles in maintaining ORCID

- **Centers administration** encourages and records
- **Employees** (could be anyone) register and maintain ORCID
- **Data stewards and curators** use orcid to describe data

➡ **Together we can make it work!**



Conclusions … What do we currently see as our roadmap?

Using common PID metadata in our data infrastructures is an important prerequisite to establish a FAIR data space.

Implementing PID metainformation systems requires to

- **Assess** each systems usability.
- **Define** the relevant stakeholders.
- **Recommend** roles and activities for different stakeholders.
- **Monitor** how the PID recommendations are implemented and **revise them if necessary.**


we are <HMC> HELMHOLTZ
Metadata Collaboration
Better metadata for better science
Together towards a FAIR Helmholtz data space
Contact us at hmc-hub-ee@geomar.de