

LINKING RESEARCH INFORMATION ACROSS DATA SPACES

15th International Conference on Current Research Information
Systems (CRIS2022)
Dubrovnik, Croatia
May 12-14, 2022

12 May 2022

Some reflections on the current PID landscape – with an emphasis on risks and trust issues

Pablo de Castro
euroCRIS Secretary

University of Strathclyde

pablo.de-castro@strath.ac.uk

<http://orcid.org/0000-0001-6300-1033>

Joachim Schöpfel
euroCRIS Board Member

University of Lille

joachim.schopfel@univ-lille.fr

Laura Rothfritz
HU Berlin

Ulrich Herb
Saarland Uni

The PID Study



June 23, 2021

Proposal Open Access

Risks and Trust in Pursuit of a Well-functioning Persistent Identifier Infrastructure for Research

Belsø, Rene; Matthiesen, Martin; Parland-von Essen, Jessica; Béquet, Gaëlle; KE Task & Finish Group for PID Risk & Trust

Persistent Identifiers (PIDs) and their infrastructures are argued to be of significant strategic importance, to the increasingly digital reality of modern-day research.

With this in mind, Knowledge Exchange (KE) intends to investigate PIDs in depth, aiming to better understand what is needed to build and exploit a well-functioning PID infrastructure for research. Our ambition is to identify, through investigation, analysis and recommendations, what could be the best possible strategic and operational paths to achieve a well-functioning PID infrastructure for Knowledge Exchange (KE) member states and beyond.

This scoping document serves as a starting point, by providing an overview of KE current PID ecosystems, focused on pinpointing issues that need to be addressed through the investigation. We propose to map these issues with KE's *Open Scholarship Framework*, aiding identification and structuring of potential solutions.

<https://zenodo.org/record/5018216>

The PID Study

Focus

<https://www.knowledge-exchange.info/news/articles/24-06-2021>

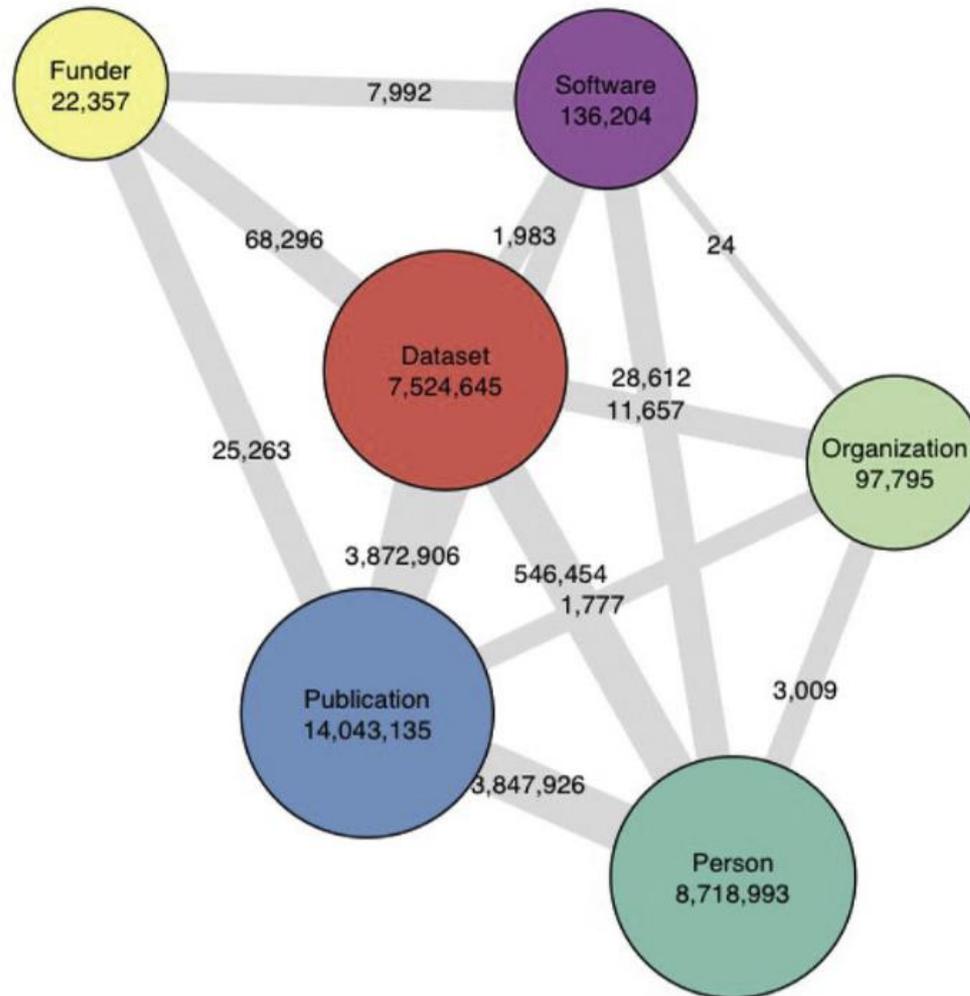
- An investigation on how to better understand what is needed to build and exploit a well-functioning PID infrastructure for research.
- To identify what could be the best possible strategic and operational paths to achieve a well-functioning PID infrastructure by ...
- ... considering well-known and consolidated sorts of PIDS (for publications, data, software, persons, organisations, archived objects) but also gradually emerging e-infrastructure (*eg* research equipment, facilities, conferences, medical or environmental science samples).

The PID Study

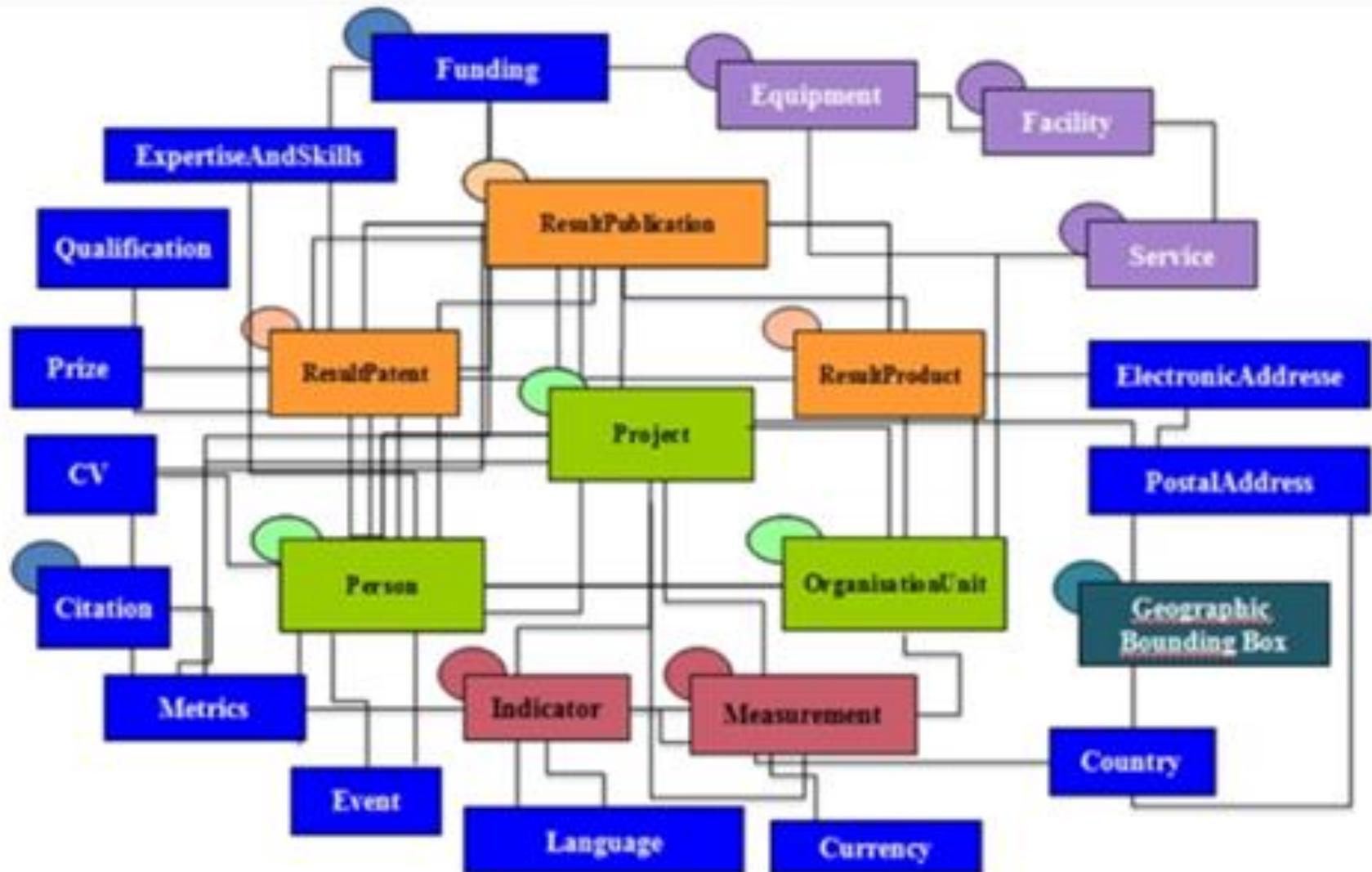
The consultants

- **Pablo de Castro:** Open Access Advocacy Librarian at the University of Strathclyde in Glasgow since Jan 2017. Technical Secretary of euroCRIS since Jan 2018. Member of the EOSC Association Task Force for PID Policy and Implementation
- **Dr. Ulrich Herb** (project lead): Open Access advocate and head of the Publication and Research Support Department at Saarland University, independent consultant
- **Laura Rothfritz:** Research assistant and PhD candidate at the Berlin School of Library and Information Science at Humboldt University Berlin
- **Dr. Joachim Schöpfel:** Professor for Information Science at the University of Lille and independent consultant.

The PID Graph



The CERIF data model



The (possible) role of CRIS systems

FRIS: Metadata model for research infrastructure

► Characteristics

→ 25 metadata fields

[Links to other
research objects](#)

- Identifier
- Federated identifier
- Name
- Acronym
- Description
- Keywords
- Type
- Location type
- Accessibility
- User modalities
- Starting date
- End date
- Location(s)
- Contact
- Website
- Technology classification (Fraunhofer-35)
- Research disciplines (FRDS)
- Data provider is consortiumcoordinator?
- Consortiumcoordinator
- Organisation(s) of consortiumpartners of infrastructure project
- Affiliations of consortiumpartners of the infrastructure project that provide data to FRIS
- **Link to funding project(s)**
- **Link to projects utilizing infrastructure**
- **Link to publications utilizing infrastructure**
- **Link to other infrastructure**

PID Study: interviews w/ experts & content analysis

Rich and partly controversial material

16 interviews with 18 experts from seven European countries, conducted between December 1, 2021 and February 3, 2022

- 15 general topics (typology, functionality, services, curation, community...)
- 4 risk levels (social, political, economic, technological)
- 7 trust dimensions (situation, structure, technology, organization, integrity...)

Expert experience and opinion: convergence and discrepancy

- Convergence
 - a. PIDs are here to stay
 - b. Means not goals (importance of purpose and services)
 - c. Competition, coexistence and (above all) coordination (interconnection)
- Discrepancy
 - a. Top-down and/or bottom-up?
 - b. Community (“people infrastructure”) is crucial, but which community?
 - c. Funding is required, but who should pay?

PID Study: interviews w/ experts & content analysis

Some general conclusions

PID landscape is not a marketplace

- The idea of PID assessment and selection with similar criteria is misleading
- No good or bad but functional or dysfunctional PIDs in a given system
- Towards interconnection (PID graphs)?

No one-size-fits-all solution (governance, business model...)

- Different communities, stakeholders, objects, purposes, interests, resources, practices...
- Strategic analysis of a given situation and exchange/coordination (forum) more important than general recommendations

No PIDified vision

- PIDs are not a solution for every problem
- Focus on research process and supporting services
- Efficient monitoring vs better science

PID Study: case studies

Digital Author Identifier (DAI – NL)

- The point is not whether it makes sense for a PID to be created at a national level but that at a time when neither ORCID nor ISNI were there yet, **this was the only way forward**. This is a lesson that could apply to other PIDs too – esp OrgIDs
- **77K DAIs** had been issued by 2014, including for deceased authors, which are not covered by ORCID. There was no time to build any services on top of the DAI layer
- The implementation process was very much **driven by universities and by SURF**, which little direct involvement from researchers themselves
- The replacement/superseding of DAIs by a mix of ISNIs and ORCIDs **did hence not have much impact on researchers** and was relatively smooth. The impact on trust would actually have been much greater should researchers have been directly involved (this risk is still there for other author IDs like the RAS).
- The main risk associated to this case study is the **loss of control** over the PID initiative by national-level stakeholders, who aim to address specific needs of theirs.

PID Study: case studies

Org IDs

- Significantly **more complex PID to implement** than author IDs. It's unclear who should own an OrgID for maintenance purposes.
- **ROR IDs** (initially based on GRIDs) are emerging as the default mechanism to address *top-level* OrgIDs, but **Ringgold** (a for-profit initiative) is there too. ROR is following well-established path followed by ORCID, but current implementation is still patchy.
- OrgIDs use cases well defined – but **ROR IDs only address the publisher use case.**
- "The Path to Department Level OrgIDs" project has defined the technical workflows to generate **multiple-level ROR IDs** – but has seen little uptake thus far.
- Several groups of national-level funders and institutions are already creating internal OrgID databases in their countries. **Best practice examples urgently needed.**
- The main perceived risk is the **lack of community involvement** – this is where top-down initiatives by national stakeholders could truly make a difference.

PID Study: case studies

PIDINSTs

- **Extremely complex area** (facilities and instruments are very different things) that requires researcher involvement both for PID creation and for their use
- Lack of clear definition for **use cases for having PIDINSTs**
- Multiple parallel initiatives going on in this domain with **little coordination** or contact w/ each other. DataCite supporting early efforts
- Role of funders will be critical. Same as for OrgIDs, there are early national-level attempts to put together databases. The **EU could play a key role via the EOSC**, since data management is among the clearest use cases
- Main perceived risk is **fragmentation** and lack of opportunities for coordination.

PID Study: case studies

Key role of funders

- Funders and publishers provide the **main use cases for PIDs**. Lack of involvement from the former means their case studies are neglected (see OrgIDs)
- **Not too many forums available** for possible coordination – Science Europe, cOAlition S, Global Research Council, CrossRef/FundRef not completely fit for purpose. Best hopes are best practice case studies like the NWO roadmap in NL
- Funders – like countries – **come in many different sizes and levels of relevance**. Same as publishers this is far from being a homogeneous group of stakeholders
- Funders have **traditionally relied on institutions** even for reporting purposes, many simply lack the knowledge and the staff
- Some (hegemonic) funders **are already leading**, either on their own or by teaming up with institutions in their countries. **FCT** and **FWF** (both outside the KE) are good examples, but NWO, the DFG and potentially UKRI could also play this role

PID Study: case studies

International Generic Sample Number IGSN

- **community need** to build a dedicated PID for samples/ physical objects/ collections
- uptake of IGSN made it a default solution for a globally unique identifier for physical samples & collections (also fostered by the **flexibility** of its metadata schemes)
- sought to gain trust through **technical-organizational alignment** with DataCite, as well as to create its own **brand value** through cooperation with the National Geographic Surveys (and later on DataCite)
- Success Factors: Meeting specific needs and knowing these needs of the community, flexible use cases and social, technical-conceptual openness, sustainability (e.g. through the partnership with DataCite, earlier through initial funding by NSF, Sloan foundation), technical reliability and up-to-dateness

PID Study: case studies

RePEc Author Identifier (RAS)

- RAS was launched 1999 (RePEc 1997), initially funded by JISC, of **significant relevance in economics** (e.g. as a data source for rankings, publication lists, CVs)
- RAS serves the **needs of a community**, whereas ORCID perhaps is perceived to serve the needs of libraries, universities, funders, publishers
- Risks: Unclear **community engagement**, highly **independent-minded** (no synchronization with external services, seemingly no interest in SCOSS funding or cooperation with ROR/ORCID)
- Trust: Sustainability given through **volunteer work** and **funding** by institutions, highly depend on the **RePEc creators & managers** (which also constitutes a risk), the **capability to meet its community's demands, openly available data**

PID Study: case studies

Failed PIDs and non-reliable PID implementations

- Different reasons why PIDs and PID implementations may fail or be unreliable: **organisational, financial, technical** are described in this case study
- Organisational: **Commitment** of the service provider is crucial and needs to be **proven to the user community** repeatedly, **contingency plans and escrow setups** need to be in place, there should be **no reliance on a single organisation** without Open Data
- Financial: Lack of financial resources leads to **reliance on third-party hosting organisations**, which can lead to errors in PID management
- Technical: **Publishers** may be 'the weakest link', there are many examples of **faulty CMS configurations, malfunctioning PIDs** or PIDs that do **not get registered** at all

A few recommendations

- More coordination needed across initiatives, stakeholders and countries
- It would be useful to have a PID observatory where an insight on the status and the developments in the field could be obtained
- Use cases for the various PIDs need to be clearly and comprehensively defined where not already available
- Public-sector entities need to step up their efforts to define the PID landscape they would like to achieve in a coordinated manner

the PID Forum

PIDapalooza is taking a break

■ PIDapalooza



alicemeadows

Oct '21

Hi all, in case you missed it, the PIDapalooza organizers shared the following message earlier this week, to let everyone know that there won't be a PIDapalooza in 2022. We'd love to hear your thoughts about the future of the event in this brief [survey](#) ² - or feel free to comment here. Thanks!

LINKING RESEARCH INFORMATION ACROSS DATA SPACES

15th International Conference on Current Research Information
Systems (CRIS2022)
Dubrovnik, Croatia
May 12-14, 2022

Thanks!
Questions?

Pablo de Castro

euroCRIS Secretary

Open Access Advocacy Librarian
University of Strathclyde

pablo.de-castro@strath.ac.uk

<http://orcid.org/0000-0001-6300-1033>

@pcastromartin