

WORKING WITH PIDS IN TOOLS

Name of Proposed Interest Group:

WORKING WITH PIDS IN TOOLS

Introduction (A brief articulation of what issues the IG will address, how this IG is aligned with the RDA mission, and how this IG would be a value-added contribution to the RDA community):

The lack of interoperability between tools presents a significant barrier to streamlining research workflows throughout the research lifecycle. These gaps prevent the comprehensive collection and incorporation of research data and metadata into the research record captured during the active research phase. This is a particular issue in relation to persistent identifiers (PIDs). Furthermore, it limits the scope for passing this data and metadata, including PIDs, on to data repositories, thus undermining FAIR data principles and reproducibility. This IG is proposed as:

- A place to address interoperability between tools, leveraging PID and metadata infrastructure.
- A platform for research-supporting service providers and PID/ open infrastructure organizations to outline use cases, explore challenges, and poll resources to provide reference and guidance to chart ways forward.
- A forum for dissecting the interoperability challenges through discussion of concrete case studies involving application of PIDs to tools in multiple communities of practice.

An initial discussion about these issues was held in a FORCE 11 Community call in February, 2022. Since then conversations around these issues between those involved in the Community Call and others have continued, and several of the participants are engaged in projects involving implementation of these ideas. The group thinks there is a need to create an ongoing forum for discussion, and an RDA IG seems like the ideal vehicle for that.

This group is directly and closely aligned with the RDA mission of building the social and technical bridges that enable open sharing and re-use of data. It will do that by galvanizing community interest in an emerging area of interest and concern that has arisen with the increased adoption of PIDs, namely how they can be

incorporated into research tools. Addressing this issue is critical to ensuring that the data sharing and reuse potential of PIDs is realized in practice.

User scenario(s) or use case(s) the IG wishes to address (what triggered the desire for this IG in the first place):

As one of the core components of the open scholarly infrastructure, PIDs are widely adopted among research organizations, publishers, and the researcher community. Increasing numbers of research supporting tools and platforms are starting to design and integrate interoperability features that contribute to and benefit from the rich, open, and standardized metadata underlying the PIDs, to consolidate the research workflows.

In the Implementing FAIR Workflows project, integration of PID registration and metadata sharing workflows on systems that support the open sharing of research outputs throughout the research lifecycle is emphasized as a core component for reinforcing FAIRness, simplifying tracking and reporting of resources and outputs, as well as boosting impact.

We plan to start the IG to focus on 3 specific use cases where multiple research tools are brought together through the integration of PIDs.

- User scenario/use case 1: Geophysical sample data

| Tools | PIDs |
|------------------------------|---------------|
| FAIMS RSpace Dataverse | IGSN ORCID |

Samples are collected in the field where there is no connection to the internet. Sample data is collected in the FAIMS field collection notebook, and associated with DataCite IGSNs.

The data and the DataCite IGSNs are subsequently passed to the RSpace electronic lab notebook/inventory management system, where they are incorporated into the record of the experiments in which they have been used.

The sample data and the experimental data are then exported to the Dataverse repository, along with the DataCite IGSNs and the ORCIDs of involved researchers

- User scenario/use case 2: Biological sample data

| Tools | PIDs |
|--------|------|
| RSpace | RRID |

| | |
|---|--|
| Scicrunch - Research Resource identification portal | |
|---|--|

Biological samples are processed in a series of experiments, and data about the samples is entered into and structured in the Inventory management module of the RSpace ELN/Inventory management system.

The researcher(s) running the experiments can from within RSpace use Scicrunch to search for metadata including RRIDs to associate with the samples.

The RRIDs are associated with the sample data and can be exported to relevant domain registries and repositories.

- User scenario/use case 3: Data management plans

| Tools | PIDs |
|--------------------------------|---------------------|
| DMPTool RSpace Dataverse | DOI ORCID ROR |

A researcher creates a data management plan, to which is associated their ORCID and the appropriate ROR. The plan and the PIDs are subsequently ingested into the RSpace electronic lab notebook.

The DMP is associated with the datasets created during the project, and the datasets and PIDs are subsequently exported to Dataverse, which assigns a DataCite DOI to the DMP, with a resolvable URI in the DMP DataHub.

Objectives (A specific set of focus areas for discussion, including use cases that pointed to the need for the IG in the first place. Articulate how this group is different from other current activities inside or outside of RDA.):

The overall aim of this group is to build a community around the topic of fostering interoperability between research tools by leveraging the open PID and metadata infrastructure. The IG will bring together service providers, research software developers, researchers, and funders to work on the following objectives:

- Consolidate the language used for communicating interoperability between tools.
- Discuss interoperability goals based on user needs.
- Clarify specifications for incorporation of PIDs and share best practices.
- Gain different perspectives by engaging with various communities of practice, which are using different kinds of PIDs.

Other RDA groups, and external projects, e.g. EOSC projects, focussing on PIDs cover issues relating to PID policy, infrastructure and governance, delivery and development. With the increasing pervasiveness of PIDs and the accelerating production and use of PIDs a new 'problem' has emerged, namely that PIDs are often deployed and used in the context of research tools, yet there is very little experience with how to do this. This group aims to be a forum for exploration of practical issues that arise when PIDs are incorporated in individual tools, as well as issues relating to transfer of PIDs between tools used in different parts of the research lifecycle.

Participation (Address which communities will be involved, what skills or knowledge should they have, and how will you engage these communities. Also address how this group proposes to coordinate its activity with relevant related groups.):

The IG will be open and welcome all RDA members to participate, to contribute or adopt the outputs / recommendations developed through the IG.

Communities of interest

- PID and Open Infrastructure organizations
- Research facilitation platforms
- Repositories
- Reporting systems
- Publication systems
- Communities of practices

Ways of participation

- RDA IG page
- Google group

Outcomes (Discuss what the IG intends to accomplish. Include examples of WG topics or supporting IG-level outputs that might lead to WGs later on.):

Besides the implementation of the technical integrations involved in the proposed use cases, we plan to work on the following document as outcomes of the group to pave way for a future WG to develop actionable guidelines:

- Report of group activities
- Case studies of interoperability between tools build on the PID infrastructure
- Guiding principles for PID workflow integration

Mechanism (Describe how often your group will meet and how will you maintain momentum between Plenaries.):

The group operates following the RDA guiding principles of openness, consensus, inclusiveness, harmonization, community-driven, and non-profit & technology-neutral.

All discussions and outputs will be open to all RDA members and beyond. Outputs and results will be derived by consensus within the IG.

The IG will organize Quarterly meetings and plenary sessions.

Timeline (Describe draft milestones and goals for the first 12 months):

| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| IG call | | | | | | | | | | | | |
| RDA | | | | | | | | | | | | |
| Output | | | | | | | | | | | | |

We plan to organize quarterly meetings and 2 plenary sessions in the first 12 months of the group’s establishment (the months in the gantt chart are for demonstration purposes. The starting month will depend on the completion of the RDA review process.) We also plan to review the progress and interim output at the end of the first year to keep the group on track.

Our objective is to develop sufficient interest and information about the issues in the first 12 months to consider setting up a Working Group, which would implement specified functional infrastructure to facilitate streamlined incorporation of PIDs into research tools.

Potential Group Members (Include proposed chairs/initial leadership and all members who have expressed interest):

| FIRST NAME | LAST NAME | EMAIL | TITLE |
|------------|-----------|----------------------------|---|
| Rory | Macneil | rmacneil@researchspace.com | CEO, Research Space. Chair |
| Xiaoli | Chen | xiaoli.chen@datacite.org | FAIR Workflows project lead at Datacite. Chair |

| | | | |
|-------------|-------------|--------------------------------|--|
| Christopher | Erdmann | cerdmann@michaeljfox.org | Associate Director for Open Science. Chair |
| Sonia | Barbosa | sbarbosa@g.harvard.edu | Manager for Data Curation, Harvard Dataverse |
| Maria | Praetzellis | maria.praetzellis@ucop.edu | Product Manager for research initiatives, California Digital Library |
| Rowan | Cockett | rowan@curvenote.com | CEO, Curvenote |
| Anita | Bandrowski | abandrowski@health.ucsd.edu | Specialist, Neuroscience, UCSD |
| Connie | Clare | connie.clare@rdafoundation.org | Community Development Manager, RDA |
| Jens | Klump | Jens.Klump@csiro.au | Geosciences Analytics Team Leader, Mineral Resources Group, CSIRO. Chair |