

Workshop on Persistent Identifiers, London, November 2018

Background

Persistent Identifiers (PIDs) are essential to support global open research and scholarly communications. They play a vital role in achieving the ideal of FAIR data (Findable, Accessible, Interoperable, Reusable)—indeed the FAIR principles in their expanded form explicitly require identifiers that are globally unique, eternally persistent and allow retrieval of metadata. Some PIDs are already well established parts of the research ecosystem and embedded into workflows and processes, but there is still much to be done. Development of new PID types (that is, standing for new types of entities, whether digital or real-world) and sustainability of the PID infrastructure are both pressing needs, as is enhanced cross-linking between PIDs to enable a richer variety of services.

An international approach is needed, as PID systems are global, and research is increasingly borderless in nature. A number of international initiatives and discussions are currently taking place. The Research Data Alliance has a number of Working Groups/Interest Groups focussing on PIDs, and PIDs have their own conference at the annual PIDapalooza event. The FREYA project (Horizon 2020 programme) aims to build PIDs into the European and global e-infrastructure, setting up a PID Forum as a stakeholder community, whose members collectively oversee the development of the infrastructure, as well as the PID Commons, defining the roles, responsibilities and structures for sustainability.

A previous workshop took place in Singapore in August 2018, coordinated by the Australian Research Data Commons in collaboration with ORCID. An important input to the workshop was a co-authored article “Mapping the PID Landscape”¹, which included the call “Let’s use the open identifier systems we already have effectively, consistently, and to mutual benefit.” The Singapore workshop participants envisaged it to be the beginnings of an internationally coordinated approach to PID adoption and use in research. That workshop was the first in a series of three, geographically diverse and with different emphases, of which this report summarises the second; the third is to be held in the US in 2019.

Workshop objectives and methods

The second workshop took place at the offices of Jisc on Fetter Lane in London on 12–13 November 2018. There were two full days of discussions, both plenary and working in groups. A small number of presentations were made by invited speakers to set the scene and stimulate discussion.

An objective of the workshop was to review the characterisation of “desirable features of a persistent identifier system” that emerged from the Singapore workshop. In advance, participants were asked to add ideas and thoughts to a collaboratively editable “brainstorming” document. At the previous workshop, the desirable features were classified into three groups: Governance, Operation of Service and Technology.

The brainstorming document raised questions and proposed statements under these headings, such as:

- (Governance) “Does it matter to the user community if there are multiple IDs for the same things?”
- (Operations) “Are able to provide technical resources to maintain and evolve the PID system and to ensure the system is robust”
- (Technology) “Avoid the use of semantic content in persistent identifier strings, especially temporary or transient meanings (an opaque identifier ...)”

¹ <https://orcid.org/blog/2018/06/21/mapping-pid-landscape>

A further objective was to select and analyse real-world workflows across the research lifecycle in which PIDs play an important role. The participants divided into groups, each analysing one of the workflows, with a view to identifying opportunities and gaps in the current PID landscape. Finally impacts and messages for different stakeholder groups were drawn out.

Outcomes

The selected workflows were labelled as:

- Data repository – Journal publisher
- Funder, grant award compliance monitoring
- Grant application including drafting initial Data Management Plan
- Instruments and equipment
- Research data

Drawing on the collective expertise of the participants, the workflows were expanded in detail, the roles of PIDs identified, and analyses presented under these headings:

Benefits – Obstacles – Achievability – Scale of community – Scale of impact – Candidates to make it happen

Standing back from the individual workflows, the impacts and messages were classified under four areas: Data – Equipment – Grant applications – OA compliance. Each was related to the following headings:

Stakeholders – Pain points – Benefits – Evidence of benefits – Key messages

As well as these detailed analyses, a number of general points were made in discussion. The FAIR Data Principles are an excellent example of what the PID world could aspire to: easily understood at the highest level but capable of being expanded in detailed implications. PIDs may arise either top-down or bottom-up (example of the latter being RRDs). When speaking of “PID systems”, distinguish between individual PIDs (such as ORCID), which have a “system” behind them, and the whole interoperating PID infrastructure.

Next steps

The next steps recommended by the Singapore workshop were considered to be sound, with the comments that alignment between initiatives is important, for example the PID Forum (FREYA) with Metadata 2020.

Potentially valuable next steps include:

- Identify and focus on the PIDs that can bring quick, widespread benefits across the landscape.
- Map the landscape - we need to understand where good work is being done toward this.
- Produce a coherent, clear value proposition for funders and policy makers.
- Leverage existing communities (such as RDA) and make coordinated use of our common communications channels.

The appropriate actions and forums depend on the maturity of the PID type and the current gaps.

The US workshop could help to prioritise goals and identify levers to create action, and consider how to fund the research that may be needed (e.g. economic analysis of benefits of PIDs).

Participants agreed on concrete contributions to maintain the momentum, listed in the table below.

Who?	What do you want?	How can you help?
Digital Science	To persuade the DS family of products to make the best possible use of PIDs	Deliver the argument to the product owners
Jisc	To be easy and seamless to use PIDs and help to safeguard their future and governance, help to improve the technical infrastructure	Produce evidence reports, reach out to policy influencers in the UK and beyond, advocate for PIDs in conversations with publishers, make the case to embed PIDs in Jisc services and systems, engage with RDA (in part via STFC as the UK node), reach researchers through meetings and events, work with UKRI/CASRAI etc to make the interoperability argument, help to articulate requirements and help to improve the way ORCID (and other PIDs) are used in systems, improving the technical infrastructure. Connect to standards and protocols.
STFC	To build PID graphs around PhD theses, explore workflows around facility awards, introduce instrument PIDs (either minting them or adopting)	Lead by example and communicating the value across many influential organisations
Fiona Murphy	Research ecosystem to work better, richer outputs, more effective researchers. Better communication and guidance to help researchers and eliminate friction	Work with funders, societies, RDA, open science projects (e.g. Scholarly Commons) to amplify and to take messages to those groups. Can bring research and consultancy. Consultant/advocate moving across many communities and spheres of influence.
EBI	To be able to connect datasets and data repositories. See fine tuning of the metadata around PIDs and improve integration at the data level, exposing links to articles, authors etc.	Act as community advocate for emerging best practices in linking data using PIDs, and encourage good practice in implementing PID linking.
ELIXIR	Interesting in PIDs (minting, resolution etc.) and identifying best practice for researchers and services	Share information about good practice around PIDs and metadata in EOSC, RDA and the ELIXIR community
ORCID	Researchers to be connected to all their activities and contributions across their career	Be responsive to the community to make sure it meets their needs and is easy to use
ARDC		Can advocate across the national level in Australia to make sure their systems and infrastructure are using PIDs well, and deliver materials that help folks to choose and use PIDs

Hindawi	Discovery and reuse of research materials to be as seamless as possible, breaking down the current culture of perverse incentives. Workflows cheaper to run, easier to integrate, provide customers with data about who is publishing what where, help researchers understand their impact. To make sure that PIDs, metadata and data are open, community governed and sustainable across the ecosystem.	Advocate for the PIDs they need, and help to demonstrate their value
BL	To understand what is really meant by “persistence”. As a research infrastructure provider, to integrate PIDs to better support their customers.	Bring the long term view, and help the community to understand persistence, through FREYA, exploring materials and training for the use of persistence, can bring an understanding of the needs of the arts and humanities community. Also work with DataCite UK serving data centres and repositories. Brings advocacy, service provision, and influence of the BL
DANS	Data to be sustainably accessible, provide a national portal for research information, PIDs enrich both missions.	Many connections in the social sciences and humanities, involved in EOSC and CESSDA and FAIR initiatives, so can connect to many other projects. Scope for advocacy within several communities
DARIAH	To support scholars to do their research more easily	Be an ambassador to reach out to the research communities, digital humanities scholars, service providers and policy makers.
ePIC	To support eResearch identification services, robust, sustainable and as standardised as possible.	Help to reach out to data centres, PID providers, help with standardisation, outreach to research communities, training and education activities, as a technology provider can adapt services to facilitate advances in PID practice. Help to find the most effective way of implementing PID infrastructures, and as a part of the PID ecosystem to help to advocate and shape negotiation and interoperation.
Elsevier (Pure)	To encourage PIDs and use them to support their customers	Advocate and communicate to their customer base, and respond to customer base requests for PID integrations and feedback to PID providers.
FREYA project	To see that PIDs are built by and for a broad community.	Bring the community together and build consensus. Strengthen the PID community (Forum, Commons) and ensure that developments are built on a community basis.

Further information

If you would like further information, please contact:

- Natasha Simons (Australian Research Data Commons) – natasha.simons@ardc.edu.au
- Josh Brown (ORCID) - j.brown@orcid.org
- Suenje Dallmeier-Tiessen (FREYA project) - sunje.dallmeier-tiessen@cern.ch
- Neil Jacobs (Jisc) - Neil.Jacobs@jisc.ac.uk