



Persistent Identification of Instruments

Markus Stocker, Louise Darroch, Rolf Krah, Ted Habermann, Anusuriya Devaraju, Ulrich Schwardmann, Claudio D'Onofrio, Ingemar Häggström and the RDA WG PIDINST

markus.stocker@tib.eu | @markusstocker
TIB Leibniz Information Centre for Science and Technology

VIVO 2020 | June 6

A bit of history ...

- PIDapalooza, Reykjavik, November 2016
 - First presentation of Persistent Identification of Instruments
 - <https://doi.org/10.6084/m9.figshare.4246100.v1>
 - Yay! ... yapid.org
- THOR-ENVRIplus Bootcamp, Helsinki, March 2017
 - ORCID Integrations in Environmental Research Infrastructures
 - Met Louise Darroch and Alessandro Oggioni, mulling over the idea of RDA WG PIDINST

A bit of history ...

- September 2017: BoF at P10, Montreal
- December 2017: Case statement submission
- March 2018: RDA endorsement and kick-off at P11, Berlin
- October 2019: Wrap-up at P14, Helsinki
- December 2019: Submitted manuscript
- May 2020: Published paper

Lesson learned ...

What begins in Finland ends in Finland

Why it matters

- Instruments have an essential role in creating research data
- Instrument metadata needed to assess data quality and reuse
 - Borgman in *Big Data, Little Data, No Data*: “To interpret a digital dataset, much must be known about the hardware used to generate the data, whether sensor networks or laboratory machines.”
- Persistent linking of research data and instruments is important
- Instrument models typically mentioned in literature: could be cited
- Inventory, funding, etc.

What we did

- Collect use cases
- Identify common metadata about instrument *instances*
- Develop and publish the schema
- Implement community feedback to schema versions
- Catalyse schema implementation by *existing* PID infrastructure
- Prototype adoption by existing institutional instrument providers

Use cases, metadata analysis and schema

- 15 use cases collected between November 2017 and February 2019
- 60% in Earth Sciences
- 14 were complete and 10 timely for October 2018 metadata analysis
- Metadata analysis resulted in first version of the schema
- Schema was revised to account for community feedback
- <https://github.com/rdawg-pidinst/schema/blob/master/schema.rst>

Schema

Identifier	Unique string that identifies the instrument instance
LandingPage	A landing page that the identifier resolves to
Name	Name by which the instrument instance is known
Owner	Institution(s) responsible for the management of the instrument
Manufacturer	The instrument's manufacturer(s) or developer
Model	Name of the model or type of device as attributed by the manufacturer
Description	Technical description of the device and its capabilities
InstrumentType	Classification of the type of the instrument
MeasuredVariable	The variable(s) that this instrument measures or observes
Date	Dates relevant to the instrument
RelatedIdentifier	Identifiers of related resources
AlternateIdentifier	Other identifiers pertaining to the same instrument instance

Schema implementations

- DataCite
 - Based on a PIDINST - DataCite Schema mapping
 - Only partial, e.g. no measured variable, model name not included
 - Bending of terminology needed, e.g. creator for manufacturer, publisher, publication year, ...
 - Globally known PID infrastructure
- ePIC (pidconsortium.net)
 - Persistent Identifiers for eResearch
 - Full PIDINST schema implementation
 - Less well-known, more European-centric PID provider

Schema adoptions

- Helmholtz-Zentrum Berlin (HZB)
 - Tested the DataCite implementation
 - <https://doi.org/10.5442/NI000001>
 - <https://search.datacite.org/works?query=doi%3A10.5442%2FNI000001>
 - <https://api.datacite.org/doi/10.5442/NI000001>
- British Oceanographic Data Centre (BODC)
 - Tested the ePIC implementation
 - <https://doi.org/21.T11998/0000-001A-3905-F>
 - <http://hdl.handle.net/21.T11998/0000-001A-3905-F?noredirect>

Deliverables

- Journal article
 - Published by Data Science Journal Special Collection: Research Data Alliance Results
 - <https://doi.org/10.5334/dsj-2020-018>
- White paper
 - Includes more technical details, best practices, how-tos, etc.
 - Work in progress
 - <https://github.com/rdawg-pidinst/White-paper>
 - Living document, will be maintained and updated
 - Published with ReadTheDocs, Gitbook or similar

Next steps

- Complete the first version of the white paper by Summer
- Develop some adoption stories in the second half of 2020
- Present the adoption stories at P17
- Maintain the schema and implement required changes
- For instance, including identified instrument models
- Engage with the broader community, e.g. i4iOz

References

- Borgman, CL. 2015. Big Data, Little Data, No Data. MIT Press. DOI: <https://doi.org/10.7551/mitpress/9963.001.0001>