Australian Research Data Commons

Best Practices: PIDs for Instruments

i4iOZ (Identifiers for Instruments in Australasia)

Australian Research Data Commons

Contents

Purpose

Instruments are essential to the activity of research, and there is a growing need to accurately describe and identify instruments digitally using a globally unique digital Persistent Identifier (PID). The purpose of this document is to explore use cases and outline best practices for assigning PIDs to research instruments. This document is an output of the Identifiers for Instruments in Australasia Community of Practice (i4iOZ).

Rationale

PIDs facilitate the linking of research components (instruments, data, people, organisations, funding) with outputs (metadata, publications, data sets, software, workflows, calibrations). Potential benefits that may arise from better linkage include:

- Metrics that quantify the use of instruments and the rationale for future funding
- Improved connection between data outputs and the instruments that generated them
- Facilitation of interoperability and open data sharing, especially in advancing technologies that foster sharing of instruments
- Improved discoverability and visibility of instruments and their data, published on the web
- Improved reproducibility of scientific results
- Support of appropriate authentication and access to sensitive data
- Improved tracking/locating of systematic errors generated by an individual instrument in a dataset/data collection, particularly where a survey has used several instruments/sensors.

These best practices recognise that there are a variety of definitions for an instrument. This document recognises that an instrument may be any and all of the following:

- A single instance of a tool, sensor or a device
  Eg: Rigaku SmartLab 9kW X-Ray Diffractometer
- A sensor or individual measuring component on a single, complex instrument
  Eg: The 9kW Cu rotating anode source (component of the Rigaku Smartlab above)
- A network, system or group of separate instruments that may be co-located or geographically distributed in space, but are connected through a single observation epoch
  Eg: Sensor array system

This document recognises that the definition of an instrument is likely to be domain-specific.

Scope

This document draws on the work of the Research Data Alliance PIDs for Instruments (PIDINST) Working Group and presents best practices for use of the schema and existing PID services. The PIDINST group has collaborated internationally (including with i4iOZ) to address the need for a globally standardised method of describing research instruments.

The i4iOZ best practices should be read in conjunction with the recommendations of the Research Data Alliance FAIR for Software Working Group and the Persistent Identifiers for Instruments Working Group Metadata Schema

The recommendations in this document cover best practices for:

- Base-level technologies
- Calibration data[1]

These best practices are drawn from use cases provided by members of thei4iOZ.
Specific use cases come from:

- Commonwealth Scientific and Industrial Research Organisation (CSIRO)
- The University of Auckland – Waipapa Taumata Rau
- Microscopy Australia
- The University of Queensland
- The University of New South Wales (UNSW Sydney).

Best Practice Recommendations

DOI

The recommended PID for instrument description is a Digital Object Identifier (DOI).
There are various DOI Registration Agencies available globally that research organisations can use to mint DOIs. We recommend using DataCite as the Registration Agency because they are making changes to their metadata schema to better support instrument description. The current DataCite Metadata Schema (version 4.4) has a general Resource Type that can be used now for this purpose. The next release of the DataCite Schema (version 4.5) will contain a specific Resource Type for Instruments.

The ability to mint a DOI for an instrument will depend on the relationship that each institution has to a DOI Registration Agency such as DataCite. For example, in Australia ARDC is a Consortium Lead for DataCite DOIs. Therefore Instrument DOIs may be registered using DataCite infrastructure (both a web interface and an API are available) through the ARDC's DataCite DOI service, which is provided free of charge to Australian research universities, institutes and organisations. Alternatively, organisations can join DataCite directly.

Where is it not possible to mint a DOI a Handle should be used

[1] Haller, A., Janowicz, K., Cox, S. J. D., Lefrançois, M., Phuoc, D. L., Lieberman, J., et al. (2019). The Modular SSN Ontology: A Joint W3C and OGC Standard Specifying the Semantics of Sensors, Observations, Sampling, and Actuation. Semantic Web, 10(1), 9–32. https://doi.org/10.3233/SW-180320

Landing Page/ Discovery

- An instrument PID must resolve to a landing page which contains a description of the instrument or a link to a description of the instrument. (Landing pages improve discovery. Please see FAIR guidelines for more information)
- The landing description of the instrument should be in a standardised format that takes into account the current DataCite Metadata Schema. Discovery of the instrument may be via an organisational repository or similar.
- Repository metadata may be harvested into Research Data Australia or similar for discovery purposes. A PID should return a machine-readable response for these purposes.

Metadata

- i4iOZ recommends describing instruments using the PIDINST Schema which aligns with the DataCite Metadata Schema.
    - Where a crosswalk or similar is needed please refer to the DataCite Mapping resources
- Calibration data for instruments should be stored appropriately and identified by a Handle due to its working data nature. (See REF[2])
    - For example a calibration report may be stored in a file that is given a Handle and noted in the instrument record
- Modifications and maintenance of an instrument should be recorded as appropriate to the discipline.
- For Australian organisations, Handles may be minted via the ARDC Handle service

Regarding versioning: Best practice is to use versioning with instruments, but the details of how this is to be managed is to be determined on a case-by-case basis in accordance with the host institution's standard practice.

For recommendations around FAIR and versioning please see the article Ten simple rules for making a vocabulary FAIR.

[2] Haller, A., Janowicz, K., Cox, S. J. D., Lefrançois, M., Phuoc, D. L., Lieberman, J., et al. (2019). The Modular SSN Ontology: A Joint W3C and OGC Standard Specifying the Semantics of Sensors, Observations, Sampling, and Actuation. Semantic Web, 10(1), 9–32. https://doi.org/10.3233/SW-180320

Appendix 1. Use cases

Precursors and contexts

Early work in Instrument Identifiers was done by The Centre for Advanced Imaging (CAI) and Qld NIF (National Imaging Facility) node based at The University of Queensland.

The CAI and UQ NIF Node had a clear need for a means to identify and reference instruments so worked in conjunction with the UQ library to create Research Data Australia records for a number of machines and these records were assigned Handles. Those published records can be found below and have formed the template for the new records being created for many of the use cases that have informed our best practices.

https://researchdata.ands.org.au/3t-magnetom-prisma/1305787

https://researchdata.ands.org.au/bruker-biospec-9430-usr-mri/938315

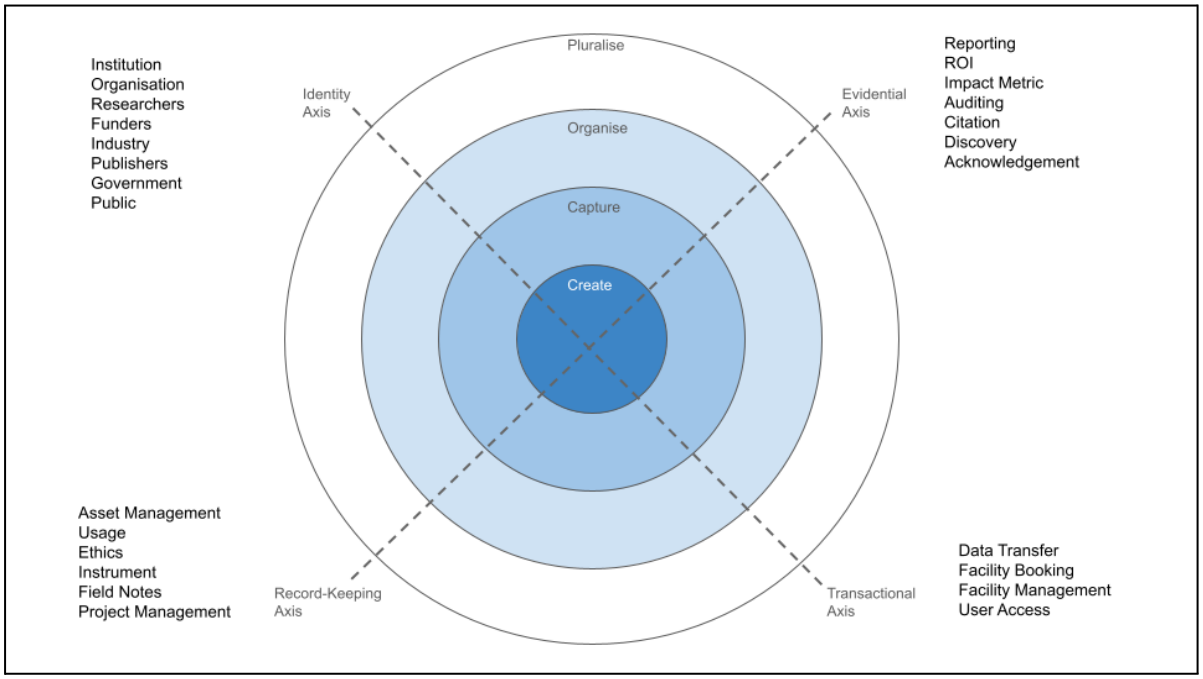
https://researchdata.ands.org.au/bruker-biospec-9430-usr-mri/938276

https://researchdata.ands.org.au/bruker-avance-164t-mri/1305748

https://researchdata.ands.org.au/x-ray-diffraction-oxidized-state/314514

https://researchdata.ands.org.au/7t-magnetom/1305790

This document also takes as context the idea of the data life cycle and records/archive continuum as effective plans for assessing and stewarding research components.
The terms found in Fig.1 are used as tags for the use cases in this document. Each case is tagged with one to many of these.



[3]

Figure 1. The Records continuum

These tags are:

- Field Notes
- Users' access and project management
- Booking and asset management
- Data transfer
- Auditing
- Discovery
- Citation
- Ethics
- Reporting (including return on Investment, impact metrics)

Audiences:

- Institutions
- Facilities/organisations
- Researchers
- Funders
- Industry
- Publishers
- Government

[3] Bates, M.J., & Maack, M.N. (Eds.). (2015). Encyclopaedia of Library and Information Sciences (3rd ed.). CRC Press. https://doi.org/10.1201/9780203757635

- Public

*More tags may be used/added as this document is updated

Use case 1:

Creation of a DOI for a service at The University of New South Wales

Tags: discovery, citation, Booking and asset management, Reporting
Audience: Institutions, Researchers

Context

UNSW's Pro Vice Chancellor-Research Infrastructure (PVC-RI) contacted the Library about the creation of a DOI for a service provided to students and academics called Katana and supported by a shared computational cluster at UNSW. Katana is designed to provide easy access to computational resources for groups working with non-sensitive data.

They requested a DOI for Katana to enable greater discovery, but also so the DOI and a citation could be added to publications when authors had used the service as a research instrument.

Approach

UNSW Library examined various tools such as Dryad which handles datasets, and Figshare that allows various output and resource types to be described and a DOI can be created. Internal tools were also looked at, including Research Outputs System (ROS) which is the University's research output management system for publications that is built on the Symplectic Elements platform. ROS was able to handle equipment records, and one type listed was service. However, ROS couldn't create a DOI. The final platform investigated was that of Research Data Australia that could be used as a good discovery platform and it allows for service records.

Through its membership of the Australian DataCite DOI Consortium led by ARDC, UNSW had access to DataCite Fabrica which can be used for the creation of DOIs. Certain metadata is required for the DOI creation and this needs to be investigated and related back to the metadata that could be created for the service record.

Outcome

UNSW Library created a record for the service using the metadata fields as presented in ROS and using these to create the record in Research Data Australia. Datacite Fabrica was used for minting the DOI.

A spreadsheet detailing the metadata requirements was sent to the team in PVC-Research Infrastructure to complete and return to the library. Fields included description, location, URLs, protocols and access rights.

For the creation of a record in Research Data Australia you require a 'key' for the record which is usually a handle. This handle (URL) is also needed by Datacite Fabrica for the creation of the DOI. UNSW have a handle creation tool, so this was used to create the handle for the service record that could be used as the key in Research Data Australia.

Upon receiving the completed spreadsheet of metadata for the service record, this could be added to Research Data Australia and components used for Datacite Fabrica. As the handle URL is used as the key in Research Data Australia

and the handle is used in Datacite Fabrica when creating the DOI, this creates a link between the created DOI and the record in Research Data Australia. Therefore, once the Library had inputted all the required metadata in the different platforms, the DOI was created and this linked back to the record in Research Data Australia.

The service record for Katana can be found at this DOI: https://doi.org/10.26190/669x-a286

This DOI is used by researchers to acknowledge the use of Katana.

The DOI has since been used by PVC-Research Infrastructure and the concept is being expanded to other instruments and services at UNSW.

Use case 2:

Creating Services records in Research Data Australia for instruments at The University of Queensland

Tags: discovery, asset management and booking, auditing
Audience: institutions, researchers, facilities/organisations

Context

The University of Queensland (UQ) is a large university that is home to a variety of research facilities and instruments. The UQ PVC-RI is investigating identifiers for the facilities and instruments housed at UQ and connecting them to other research systems such as the Research Data Management System (UQ RDM) using a standard identifier like DOIs or Handle. In doing so, the management of these facilities and instruments becomes part of the research project management ecosystem, in particular, the facilities and instruments can be cited in research outputs to enhance reproducibility and re-usability. The management team of these facilities and instruments also get acknowledgement and recognition.

Standard identifiers assigned to the UQ facilities and instruments improve the ability to track metrics on the usage of instruments, and promote UQ services and instruments available to researchers to generate data.

UQ is a member of the ARDC-led DataCite DOI consortium and as such is able to mint DOIs via its UQ eSpace system. UQ eSpace is an in-house-developed repository hosting all UQ research outputs including datasets. It mints DOIs for research outputs such as datasets. UQ researchers publish their datasets to UQ eSpace and DOI is automatically minted as part of the publishing process, which are automatically harvested by Research Data Australia on a daily basis

However, currently UQ eSpace is not able to publish Services records, because the proper Services record form has not been set up in UQ eSpace. This issue has been raised and discussed, and it's on the UQ eSpace development roadmap. It is planned that In future, there will be a metadata form set up in UQ eSpace to allow researchers to create and publish Services records for their facilities and instruments with DOIs minted. And these records will be automatically harvested by Research Data Australia, in the same way as the UQ eSpace data records.

Approach

To work around the issue for now until UQ eSpace has the Services record form available, UQ decided to create Services records within the RDA system and mint Handles to those records via ARDC Handles service. This is done by UQ data librarians on an ad hoc basis on request.

This approach was actually adopted many years ago (see Case 0) when a data librarian was contacted by Center for Advanced Imaging (CAI) to publish the instruments records in UQ eSpace. The data librarian in the end created six Services records in RDA manually and minted Handles for those records.

Outcomes

The Centre for Microscopy and Microanalysis (CMM) contacted the UQ data librarian with information about their instruments and facilities at the end of 2021. They supplied information about 77 instruments. The data librarian who has RDA admin rights created five Services records as the test.

https://researchdata.edu.au/rigaku-smartlab-9kw-ray-diffractometer/1796445

https://researchdata.edu.au/hitachi-tm4000plus-tabletop-microscope/1796322

https://researchdata.edu.au/hitachi-scanning-electron-su3500-b/1955369

https://researchdata.edu.au/jeol-jcm-5000-neoscope/1955372

https://researchdata.edu.au/renishaw-invia-raman-microscope/1955354

Similar records will be created for the rest of instruments from CMM if requested.

Use case 3:

CSIRO in-situ instrument records

Tags: field notes, asset management, discovery, citation, reporting
Audience: researchers, institutions, organisations

Context

Researchers utilise many different instruments to undertake observations and measurements in the field. In-situ observations need context to be interpreted and used appropriately. Frequently we find ourselves using many different systems and service providers for collecting, managing, and accessing data from sensors and instruments.

Metadata collected may include records relating to: field notes, maintenance records, deployment configurations, locations, instruments, sensors, quality control and calibrations.

This can present challenges where different services have different features and/or use different schemas. Different protocols and formats can also present an interoperability challenge for discovery, access and processing.

User stories arising from this context:

- **As a** field technician **I want to** know where instruments are, what state they are in and who the custodian is **so that** they can be accessed, retrieved, serviced. **For this I need** instruments to be uniquely identified so records can be related.
- **As a** researcher **I want to** be able to cite the use of a particular instrument **so that** I can identify the equipment that was used.
- **For this I need** instruments to be described and versioned so the measurement techniques are documented and it is possible to relate records such as location, calibration, software, servicing.
- **As a** researcher operating a platform **I want to** identify each instrument individually **so that** I can trace the configuration of my platform over time. **For this I need** to be able to identify the platform and each of its instruments and note points in time when the configuration was changed.
- **As a** researcher **I want to** be able to relate observations and measurements to instruments **so that** I can understand how and where the values were measured. **For this I need** instruments to have metadata data and a reference/identifier in order to unambiguously relate observations to instruments.
- **As an** institute **I want to** use research equipment effectively **so that** funds can be well spent. **For this I need** to know what instruments have been procured, their availability and who to contact so that researchers know what resources are available (and an interface to interact with this information).
- **As a** data manager **I want to** relate different instrument data streams with different tags/groups **so that** appropriate access to data is granted/denied. **For this I need** to know which instrument is being used by which project and when.
- **As a** data publisher / institute **I want to** provide citations to data and instruments **so that** credit can be attributed and recognised. **For this I need** a system of records relating projects, instruments, deployments and observations.
- **As an** institute and industry/community member **I want to** be able to use and contribute to solutions (standards, software, etc.) that are interoperable and reusable **so that** greater gains may be made and shared by the collective effort. **For this I need** a documented, extensible solution that facilitates uptake and contribution.

Approach

CSIRO has a pilot project experimenting with decoupling metadata records from the services hosting the observation data and using persistent identifiers and linked data to enable integration of the records with the data hosted by various applications.

Integration includes the capacity to construct knowledge graphs linking many records to create context rich observation data. This approach is complemented by an internal pilot PID service (based on the Permanent Identifiers for the Web service: https://w3id.org/).

Outcomes

The pilot has been able to experiment with metadata schemas for the various types of related records. This includes trialling existing schemas (such as PIDINST for instruments), extending them and developing new schemas. The decoupling of metadata from bespoke and proprietary services, along with the use of linked data, is providing flexibility to develop records (schemas) to support multiple requirements. We anticipate greater stability (through persistence) to metadata records as they exist independently of the services and platforms hosting the data as well as

providing greater agility in being able to migrate data from one hosting solution (storage, platform, service, etc.) to another if needs change. For example, support for different instruments, budget constraints, archiving observations to different systems, replicating records to another system to facilitate access or analysis.

Use case 4:

University of Auckland, Use of Instrument PIDs in an instrument data repository

Tags: discovery, citation, reporting

Audience: Institutions, Researchers, Facilities/organisation

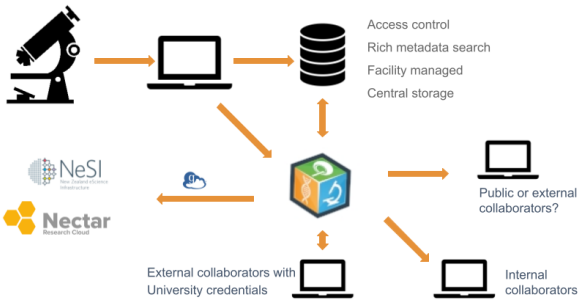
Context

Research instrumentation is generating data at an exponentially-increasing rate, with researchers and facility managers undertaking time-consuming administrative and experimental tasks, needing to find, reuse and share their data and to provide mechanisms for verification and validation of research findings (integrity and reproducibility).

Generally, the instrument data landscape is fragmented (data is spread across multiple devices, infrastructure and facilities), metadata and file formats are often not transferable, and there is minimal cohesion. This adversely affects sustainability institutionally, across people, research capacity, and storage infrastructure and makes finding data more difficult than it needs to be.

Approach

We are working towards a centralised instrument data service and repositories that align instrument data storage practices. This aims to reduce the impact of a burgeoning need for storage, meet ethical, legal, funder and third-party requirements, and allow data to be found and shared (potentially for reuse) which is increasingly required by publishers, and have use of services and facilities connected to outputs. This involves:



1. Aligning **transparent data management and retention policies**, and facilitating early planning of emerging data and infrastructure needs facilitating planning and best practice
2. Providing a service which at its foundation includes **persistent identifiers** particularly RAiD, PIDInst, ORCID, an extended ROR, and DOIs for the data outputs and promoting their use by researchers. This facilitates the development of reporting, and research outputs tracking for facilities and instruments through integration and automation.
3. Meeting FAIR and CARE principles through establishing minimum **metadata standards** on data ingested into the instrument data repository.

4. Providing the capability to connect the use of facilities and instruments to research data and publication outputs to **enabling data-driven reporting and investment**. This will provide a rich data-driven picture of usage and impact to assist in strategic decision making.
5. **Automating workflows and information capture** thereby enhancing the integrity and verifiability of instrument data and saving valuable researcher time.

Outcome

MyTardis Instrument Data Service

Given the University of Auckland's project centric approach to research, we are using a hierarchical instrument data repository which allows for a replication of the research project with many experiments, each of which generates sets of data. The workflow associated with ingestion into MyTardis, likewise is a hierarchical process, requiring that a project is defined prior to experiments being attached to this project.

The use of PIDs for a research activity/project (in this case RAiD) allows for an unambiguous link between experiment and project. When data is generated on an instrument, its parent experiment is identified (though a workflow co-designed with researchers and facility managers), which in turn is linked to a parent project. Minimum metadata requirements are placed on the data in order for it to be successfully ingested into the repository at this stage, with facility managers notified where data was not ingested due to incomplete metadata.

The instrument that the data was generated from is also captured by way of an identifier (soon to be a DOI with PIDInst schema) such that there is a chain of connections that stretches from the project to the instrument and the facility.

Example facilities and instruments include - gene sequencing instruments (e.g. nanopore), confocal and advanced light microscopes, and mass spectrometers.

As part of this process we have a series of user stories which can be used within an implementation:

- **As a** facility owner **I want to** be able to see where researchers have used the instruments and published outputs that connect with it (link to publications (eventually)) **so that** I can report on usage and advocate for researcher instrument requirements. **For this I need** instrument identifier citations using PIDInst.
- **As a** researcher **I want to** book an instrument easily **so that I** do not need to enter my project data multiple times. **For this I need** a connected ecosystem that can pull in my project data.
- **As a** researcher **I want to** be able to cite the use of a particular instrument (at a particular time) (FAIR) **so that** I can unambiguously identify the equipment that was used and other researchers can reproduce the conditions. **For this I need** instruments to be described and versioned with a record of important changes such as calibrations or software updates.
- **As a** repository owner **I want to** be able to augment the information in the repository and uniquely identify instruments to link to data **so that** I have an accurate record of the data in my repository and information/data can be easily found for a given instrument, project or technique. **For this I need** well described instruments and the use of PIDs such as ORCID, RAiD and PIDInst.
- **As an** institute **we want** to have findable instruments which live in a connected ecosystem **so that** data that flows seamlessly. **For this we need** to understand the information that we need to record and share and to link services and data, and create a catalogue of instruments
- **As an** institute **we want** to have findable instruments which live in a connected ecosystem **so that** we have a record of usage patterns and our capabilities. **For this we need** the ability to identify instruments, researchers output and impact and the ability to identify what instrument PIDs are associated with our institution.

- **As an** institute **we want** to have findable instruments which live in a connected ecosystem **so that** researchers can easily find and use the tools they need. **For this we need** to have a standardised methods of collecting and describing instruments, defined workflows for instrument usage and data publication, and a catalogue of instruments
- **As an** industry member **I want to** find specificalised research equipment **so that** I can undertake a specialised query to answer my question. **For this I need** a catalogue that is searchable, and a summary that I can understand with contact information for more details.
- **As a** facility owner/institute **I want to** make my specialised research equipment findable by potential external users/collaborators **so that** I can maximise usage of my equipment. **For this I need** a catalogue that is searchable, and a summary that external users can understand with contact

Use case 5:

Microscopy Australia

Tags: discovery, citation, acknowledgement, asset management, booking, reporting
Audience: Institutions, facilities/organisations, researchers, government

Context

Microscopy Australia is a consortium of nine university-based microscopy facilities; namely: the Advanced Imaging Precinct at The Australian National University; Flinders Microscopy and Microanalysis at Flinders University; the Monash Centre for Electron Microscopy and the Ramaciotti Centre for Cryo-Electron Microscopy at Monash University; Adelaide Microscopy at The University of Adelaide; the Electron Microscope Unit at The University of New South Wales; the Centre for Microscopy and Microanalysis at The University of Queensland; the Future Industries Institute at the University of South Australia; Sydney Microscopy and Microanalysis at The University of Sydney; and the Centre for Microscopy, Characterisation and Analysis at The University of Western Australia.

Until recently, there was no common or coordinated approach to the collection or creation of metadata and PIDs or to the management of research data in general across Microscopy Australia. Microscopy Australia is now looking actively into ways to promote the FAIR data principles across its network of facilities and users. In particular, Microscopy Australia has started a pilot program on PIDs for instruments with five volunteer facilities (Centre for Microscopy and Microanalysis, Centre for Microscopy, Characterisation and Analysis, Flinders Microscopy and Microanalysis, Adelaide Microscopy and Sydney Microscopy and Microanalysis).

Approach

The pilot will use the PIDINST schema and the PIDINST metadata properties will be mapped onto the DataCite metadata schema. The pilot will focus on two main aspects:

1. Determination of what an instrument is and cases when an instrument requires a new PID (*e.g.* multi-component instrument, change in functionalities, upgrade of a component and recalibration); and
2. Determination of the procedure or workflow at the facility and/or at university to register instruments and mint DOIs (or Handles), and whether current procedures in place are sustainable.

The pilot will also be an opportunity to give feedback on the PIDINST schema.

Outcomes

The definition of instruments and cases when instruments require new PIDs will be shared across Microscopy Australia as guidelines. The guidelines will also be shared with the National Imaging Facility (NIF) to see if Microscopy Australia and NIF can share common guidelines. This would be helpful to some facilities that are also NIF nodes.

Ultimately, Microscopy Australia will require all instruments (co)-funded by Microscopy Australia be registered and associated with Microscopy Australia's ROR ID and other PIDs (*e.g.* grant ID).

Instrument PIDs could play a role beyond enabling FAIR data for researchers. They could be used for the unambiguous identification of assets at facilities, for internal reporting within facilities and to Microscopy Australia (*e.g.* return on investment). Booking of instruments and the collection of various types of metrics could be associated with instrument PIDs.

Glossary:

FAIR
CARE
DOI
Handle
RAID
Research Data Australia
Research Data Alliance

References

Chue Hong, Neil P., Katz, Daniel S., Barker, Michelle, Lamprecht, Anna-Lena, Martinez, Carlos, Psomopoulos, Fotis E., Harrow, Jen, Castro, Leyla Jael, Gruenpeter, Morane, Martinez, Paula Andrea, Honeyman, Tom, Struck, Alexander, Lee, Allen, Loewe, Axel, van Werkhoven, Ben, Jones, Catherine, Garijo, Daniel, Plomp, Esther, Genova, Francoise, … RDA FAIR4RS WG. (2022). FAIR Principles for Research Software (FAIR4RS Principles) (1.0). https://doi.org/10.15497/RDA00068

Cox SJD, Gonzalez-Beltran AN, Magagna B, Marinescu M-C (2021) Ten simple rules for making a vocabulary FAIR. PLoS Comput Biol 17(6): e1009041. https://doi.org/10.1371/journal.pcbi.1009041

Botts, M, Robin, Alexandre, Hirschorn, Eric , OGC SensorML standard, 2019, http://docs.ogc.org/is/12-000r2/12-000r2.html

RDA Persistent Identification of Instruments (PIDINST) working group, PIDINST white paper , 2022, https://docs.pidinst.org/en/latest/white-paper/index.html#

Stocker, M, Darroch, L, Krahl, R, Habermann, T, Devaraju, A, Schwardmann, U, D'Onofrio, C and Häggström, I. 2020. Persistent Identification of Instruments. Data Science Journal, 19: 18, pp. 1–12. DOI: https://doi.org/10.5334/dsj-2020-018)