

Persistent Unique Identifiers for Instruments and Facilities

Adi Ranganath, Matthew Murray, Andrew Johnson, Renaine Julian,
Matt Mayernik, and Claudius Mundoma



Enabling Open Science Through Persistent Identifiers for Research Instruments and Facilities

- FAIR Facilities and Instruments
- NSF FAIR Open Science (FAIROS)
- Research Coordination Network (RCN)

- NSF Awards #2226396, 2226397, 2226398

- <https://ncar.github.io/FAIR-Facilities-Instruments/>



What is an instrument?

- “A device used for making measurements, alone or in conjunction with one or more supplementary devices.” (*International Vocabulary of Metrology*)
- Instruments may have configurations or settings that change over time. They can range in size from small temperature sensors to more complex microscopes or RADAR and LIDAR devices.
- A specific supercomputing cluster might be considered an instrument.

What is a facility?

- Broader than “instrument”, in that facilities also refer to people and may be tied to a specific physical location. For example, a biotechnology laboratory.
- A facility may contain multiple instruments
 - Satellites often contain multiple instruments and generate multiple discrete data streams
- A research computing center might be considered a facility.

What are PIDs?

- (Digital) Persistent Identifiers
 - A string of letters and numbers
 - Globally unique
 - Persistent
 - Machine resolvable and processable
 - Have an associated metadata schema

Examples of PIDs

- RRID
- URN
- PMID
- DOI
- DMPID
- Handles
- ARK
- PURL
- Research Organization Registry IDs

Networked Science



DataCite



RRID

What is FAIR?

- **Findable** - How do we enable people to find relevant facilities or instruments?
- **Accessible**- How do we enable facilities and instruments to be accessible by wider audiences?
- **Interoperable** - How do we consistently capture relationships between persistent identifiers?
- **Reusable** - How can we incorporate information about facilities and instruments into data set provenance metadata more consistently?

Examples

Stanford Nano Shared Facilities

Stanford University

SNSF

- electron & ion microscopy **snsf-eim** (rrid:scr_023230)
- soft & hybrid materials facility **snsf-smf** (rrid:scr_023230)
- stanford nano shared facilities **snsf**** (rrid:scr_023230)
- x-ray & surface analysis facilities **snsf-xsa** (rrid:scr_023230)

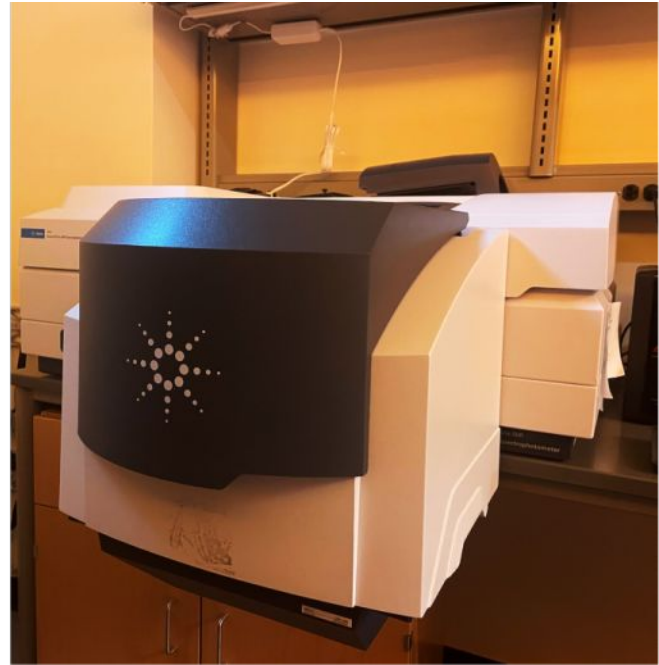
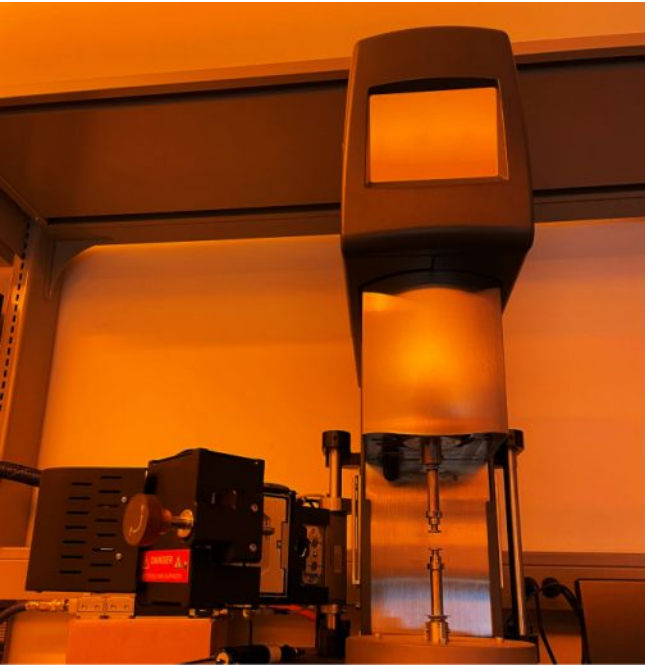
Electron & Ion Microscopy SNSF-EIM (RRID:SCR_023230)

Stanford Nano Shared Facilities SNSF (RRID:SCR_023230)

The image shows a map of Stanford University with several buildings labeled, including Poly Hall, Spruce Hall, Cedar Hall, Forsythe Hall, Cypress Hall, Via Ortega Garage, Shriram Center for Bioengineering & Chemical Engineering, SEQ Courtyard, Spilker Center for Engineering & Applied Sciences, and the Huang Engineering Center. A search bar at the top left contains the text 'SNSF' and a dropdown menu lists four facility options. A callout box on the map points to the location of the Electron & Ion Microscopy SNSF-EIM and the Stanford Nano Shared Facilities SNSF. The map also shows streets like Panama St, Via Ortega, and North-South Axis.

Polymeric and Optical Materials Characterization Shared Facility

RRID:SCR_022288



Alpine Supercomputing
Cluster, University of
Colorado Boulder Research
Computing
[https://doi.org/10.25811/k3w
6-pk81](https://doi.org/10.25811/k3w6-pk81)



NSF/NCAR's Lockheed Hercules C130 Aircraft

<https://doi.org/10.5065/D6WM1BG0>



The Project

What are our goals?

- Create and facilitate Research Coordination Network focused on the assignment of Persistent Identifiers (PIDs) to research facilities and instrumentation
- Gather use cases for why/how PIDs can be assigned to facilities and instruments
- Facilitate generation of expertise and guidance on the key topics of interest
- Provide recommendations and lessons learned targeted to the multiple communities involved

Why are we doing this

- There is, as of yet, no standard for using identifiers for scientific instruments and facilities
- Encouraging the use of PIDs for facilities and instruments advances open science allows for:
 - Increasing transparency
 - Promoting research reproducibility
 - Fostering equity
 - Supporting discoverability
 - Ensuring provenance of data
 - Providing credit to developers and providers

What we've done

- Organized focus groups & presentations for relevant groups
- Held an in-person workshop in Boulder, September 2023
 - Attendees from academic institutions, national labs, nonprofit orgs, publishers, and industry
 - Biomedical science, geological science, environmental science, space science, materials science, and more
 - Slide decks from presentations are available [on our website](#)
- Published: [Fair Facilities And Instruments Workshop #1 Report: Exploring Persistent Identifier Needs, Barriers And Incentives](#)
- Presented at events like this

Observations

Observation #1 - Need

- There is a need for PIDs for research instrumentation to encourage scientific reproducibility, ensure provenance of data, and provide credit for instrument developers and providers.
- *Community Responsibilities* - All stakeholders should encourage the assignment of PIDs to research instrumentation, as well as the citation of those instruments in scholarly work that results from their use.

Observation #2 - PID Systems

- The current use of PIDs for instruments is scattered and inconsistent in how and which PIDs are used. Multiple PID systems are already being used for the purposes of identifying research instrumentation.
- *Next Step For Our Project* - A thorough comparison of the existing PID systems should be conducted, to evaluate their applicability for identification of facilities, platforms, and instruments for the most common use cases.

Observation #3 - Adoption

- In general, the question of which PID system to use is less important than finding ways to lower the barrier for adoption of instrument PIDs and better communicate the value of using them.
- *Next Step For Our Project* - Develop guidance on how to navigate adoption considerations for the different PID options.
- *Community Responsibilities* - PID systems should continue to work to make connections, to enable interoperability in services, such as PID resolving and metadata, and to clarify the value of each respective system.

Observation #4 - Metadata

- Sometimes we need to think about metadata, in addition to PIDs. PIDs may not be appropriate to solve all challenges related to research traceability, transparency, and reproducibility.
- *Next Step For Our Project* - Working from the [PIDINST](#) metadata recommendations, develop a set of implementation guidelines for the scientific community that includes how to keep metadata relevant and up to date and connect instrument PIDs to resulting dataset metadata.

Observation #5 - Granularity and Evolution

- Granularity and evolution considerations can be very complicated for research instrumentation. A good rule of thumb is to start simple, and then move to more complicated approaches if needed.
- *Next Step For Our Project* - Granularity and instrument evolution challenges need to be better understood across the various disciplines involved in order to develop guidelines that will support the highest community priorities whether those are impact tracking, credit to developers/operators and/or scientific reproducibility.

Observation #6 - Resource Limitations

- Instrument and facility providers often face significant resource limitations that make assigning, managing, and promoting PIDs challenging.
- *Next Step For Our Project* - Continue to gather data from research facilities and assess challenges across disciplines to develop guidance for how instrument and facility providers can most easily and efficiently begin the process of assigning PIDs.

Observation #7 - Value

- Convincing users of the value of citing PIDs for instruments will be important to advance adoption.
- *Next Step For Our Project* - Develop value statements about how PIDs help instrumentation and facilities providers to contribute to research being “born FAIR.”

Observation #8 - Incentives

- While incentives can generally increase PID use and adoption, a lack of understanding of certain stakeholders' incentives can be a challenge. Different incentives are needed on both the researcher and administrative sides to ensure adoption of PIDs.
- *Next Step For Our Project* - Learn more about how PID use and citation can be made easier in hopes of encouraging use. Building better understanding of researcher and PI benefits and incentives for PID use and citation may reduce challenges.

What's next?

Workshop #2: Tallahassee/FSU

- August 20 - 22 at High Field Magnetic Laboratory
- Goals: Expand on workshop #1 by exploring pathways for PID implementation
- [Let us know if you want to attend!](#)
- We can support travel for some folks not at FSU.

What's next?

- PID comparison
 - Syntax, governance, cost, process, metadata, persistence, etc.
- Focus groups
 - Facility managers, instrument operators, campus administrators, grant organization staff
- Eating our own dogfood
 - Ensuring that the facilities and instruments at our institutions have PIDs and encouraging their use
 - Looking into adding PIDs to institutional repository metadata and README templates

Contact us / More Info

Project website:

<https://ncar.github.io/FAIR-Facilities-Instruments/>

Slack:

<https://fairosrcn.slack.com/>