ISSN 2161-3974
Full-Length Paper

Journal of eScience Librarianship 13 (3): e964
DOI: https://doi.org/10.7191/jeslib.964

Journal of eScience Librarianship
putting the pieces together: theory and practice

Persistent Identifiers for Instruments and Facilities: Current State, Challenges, and Opportunities

Matthew Mayernik, National Center for Atmospheric Research/University Corporation for Atmospheric Research, Boulder, CO, USA, mayernik@ucar.edu
Andrew Johnson, University of Colorado Boulder, Boulder, CO, USA
Renaine Julian, Florida State University, Tallahassee, FL, USA
Matthew Murray, University of Colorado Boulder, Boulder, CO, USA
Claudius Mundoma, Stanford University, Stanford, CA, USA
Aditya Ranganath, University of Colorado Boulder, Boulder, CO, USA
Greg Stossmeister, National Center for Atmospheric Research, Boulder, CO, USA

Abstract

Objective: Persistent Identifiers (PIDs) are central to the vision of open science described in the FAIR Principles. However, the use of PIDs for scientific instruments and facilities is decentralized and fragmented. This project aims to develop community-based standards, guidelines, and best practices for how and why PIDs can be assigned to facilities and instruments.

Methods: We hosted several online and in-person focus groups and discussions, cumulating in a two-day in-person workshop featuring stakeholders from a variety of organizations and disciplines, such as instrument and facilities operators, PID infrastructure providers, researchers who use instruments and facilities, journal publishers, university administrators, federal funding agencies, and information and data professionals.

Received: July 14, 2024 **Accepted**: November 13, 2024 **Published**: December 3, 2024

Keywords: persistent identifiers, instruments, facilities, data, metadata

Citation: Mayernik, Matthew, Andrew Johnson, Renaine Julian, Matthew Murray, Claudius Mundoma, Aditya Ranganath, and Greg Stossmeister. 2024 "Persistent Identifiers for Instruments and Facilities: Current State, Challenges, and Opportunities." *Journal of eScience Librarianship* 13 (3): e964. https://doi.org/10.7191/jeslib.964.

Data Availability: Focus group questions are available under the article Supplementary Files:
Appendix 1: Focus Group Questions

Abstract Continued

Results: Our first-year efforts resulted in four main areas of interest: developing a better understanding of the current PID ecosystem; clarifying how and when PIDs could be assigned to scientific instruments and facilities; challenges and barriers involved with assigning PIDs; incentives for researchers, facility managers, and other stakeholders to encourage the use of PIDs.

Discussion: The potential for PIDs to facilitate the discovery, connection, and attribution of research instruments and facilities indicates an obvious value in their use. The lack of standards of how and when they are created, assigned, updated, and used is a major barrier to their widespread use. Data and information professionals can work to create relationships with stakeholders, provide relevant education and outreach activities, and integrate PIDs for instruments and facilities into their data curation and publication workflows.

Introduction

Persistent Identifiers (PIDs) are central to the vision of open science described in the FAIR Principles. For open science to work, data and associated scientific resources must be Findable, Accessible, Interoperable, and Reusable (Wilkinson et al. 2016). The role of PIDs is explicitly recognized as the first principle within the "Findable" category of FAIR ("F1. (meta)data are assigned a globally unique and persistent identifier"), and beyond that, PIDs are also mentioned in three of the eleven FAIR sub-principles.

The assignment of PIDs to scholarly resources began in the 1990s to address a number of challenges associated with finding resources cited in bibliographic references of scholarly texts, and the relative impermanence of URLs for resources on the web (Lynch 1998). In recent decades, as open science has taken on greater visibility within scholarly research institutions, the use of PIDs has expanded significantly to encompass many purposes and resource types, including data sets, software, laboratory materials, physical samples, people, and organizations (ORFG PID Strategy Working Group 2024).

While the assignment of PIDs to research data is now a mature and well-defined practice, the PID ecosystem does not yet systematically encompass research instruments and facilities that are used to generate that research data.[1] While there are ongoing efforts to assign PIDs to research instruments and facilities, these activities remain decentralized and fragmented, and are not yet guided by community-wide standards, norms, and best practices. The task of developing widely accepted community-based standards and principles for connecting PIDs to instruments and facilities is therefore important, since it has the potential to systematically enhance the discoverability and traceability of the research instruments and facilities that generate scientific research data, which in turn could facilitate data reuse and reproducible research.

1 We will explicitly define key terms such as "instrument" and "facility" in the next section.

In this paper, we describe initial findings from a community-based project that aims to understand the current PID ecosystem with respect to instruments and facilities, and to develop community-wide norms and guidelines that can bring instruments and facilities into the FAIR research ecosystem. The project aims to achieve these goals by bringing together various stakeholder groups whose work is relevant to the project of facilitating reproducible and open science. This stakeholder network (comprising instrument and facilities operators, PID infrastructure providers, researchers who use instruments and facilities, journal publishers, university administrators, federal funding agencies, and information and data professionals) will document use cases for how and why PIDs can be assigned to facilities and instruments and provide concrete recommendations to various open science communities on advancing the goal of systematically cataloging, documenting, and citing research instruments and facilities.

In the following, we first provide relevant background and context for the project and offer relevant definitions. We also present a concise survey of some real-world examples of PID assignment to instruments and facilities. After situating the project conceptually and empirically, we turn to a presentation of our key findings at this stage of the project. In particular, we provide lessons learned in relation to the following questions:

1. What are the motivations and benefits, or use cases, of assigning PIDs to instruments and facilities? In other words, given that systematically assigning PIDs to instruments and facilities will require a meaningful investment of scholarly and financial resources, what are the prospective returns on that investment?
2. How do instruments and facilities fit within the broader PID provider landscape, and what are the PID systems best suited to instruments and facilities use cases?
3. How do the intrinsic properties of instruments and facilities (such as their modularity and tendency to change over time) pose challenges for PID assignment?
4. How can we reduce the level of effort for relevant stakeholders to participate in the PID ecosystem for facilities and instruments?

After exploring these questions, we turn to a more explicit consideration of the implications of our findings for data and information professionals in particular, before concluding with a brief discussion of the project's future direction and goals.

Background

Project Activities

The work described in this paper has been conducted by the FAIR Instruments and Facilities Research Coordination Network, which was funded in 2022 by grant awards from the NSF's Fair and Open Science (FAIROS) program. This is a multi-institution collaboration between the NSF National Center for Atmospheric Research, the University of Colorado, Boulder, and Florida State University.

The findings described below were developed through activities and discussions from the first year of the project. During this time, we developed a project website (https://ncar.github.io/FAIR-Facilities-Instruments) and began identifying relevant stakeholders, projects, information, and resources. We hosted two online focus groups in early 2023 that solicited feedback on the project topics from Earth science data facility providers and users. The focus group questions are provided in Appendix I. We also held informal discussions with several other groups and presented this topic at a variety of conferences relevant to key stakeholder communities.[2]

The capstone of the first two years were in-person workshops hosted at the University of Colorado, Boulder, and Florida State University in September 2023 and August 2024 respectively. Each included 35 participants from a variety of organizations and disciplines, representing a cross-section of the relevant stakeholders. The workshops consisted of presentations and breakout focus groups that allowed us to map out the existing landscape and identify next steps for the network. Workshop session topics and breakout group questions are provided in the workshop reports (Johnson et al. 2024; Julian et al. 2024). The findings described in the next section of this paper are based on lessons learned and documented through these project activities.

Definitions

Before proceeding, it is important to clearly define our key terms, as different stakeholders hold varying understandings of key terms that are at the heart of the project, such as "instrument," "facility," and even "identifier." We define a persistent identifier as a digital reference to a resource that satisfies certain properties, such as being globally unique, stable over time, and machine resolvable and processable (McMurry et al. 2017; National Science and Technology Council 2022). Typically, there is also a metadata schema associated with the persistent identifier. These key properties differentiate persistent identifiers from other identifiers (such as URLs) that merely point to web addresses or may expire or be removed over time (Juty et al. 2020). The key benefit of PIDs is that they act as a type of primary key in locating and identifying research resources, and also enable differentiation between similar entities, since identifiers are unique and no one identifier points to multiple resources (Bandrowski et al. 2015). Digital Object Identifiers (DOIs) are the most well-known type of persistent identifier (Meadows and Haak 2018). DOIs are unique and open identifiers designed for various outputs including book chapters, journal articles, datasets, conference proceedings, and more (Meadows and Haak 2018). Other persistent identifier frameworks, such as Research Resource Identifiers (RRIDs) are also in use, and relevant to instruments and facilities; we discuss various PID systems in more detail below.

We define an instrument as a "device used for making measurements, alone or in conjunction with one or more supplementary devices" (Joint Committee for Guides in Metrology 2012, pg. 34). Instruments may

2 These conferences included meetings of the American Meteorological Society, Earth Science Information Partners, Association of Biomolecular Resource Facilities, and Research Data Access and Preservation Association.

have configurations, settings, or sub-components that change over time. They can also vary greatly in size and complexity. An example of an instrument might be a microscope or a Lidar sensor.

Facilities are a broader category than instruments, to the extent that they include personnel and are often tied to a specific location. For example, a biotechnology laboratory would be considered a facility while a microscope within that laboratory might be considered an instrument. As with instruments, facilities vary greatly in size and complexity, and facilities look quite different across various disciplines and contexts (e.g. national laboratories versus university-based facilities).

Examples of PID Assignment for Facilities and Instruments

Having presented some basic definitions, we now turn to briefly documenting existing efforts to assign PIDs to instruments and facilities. Below, Table 1 shows a number of examples of the use of DOIs and RRIDs for the purposes of identifying facilities and instruments. These examples are necessarily anecdotal, as there is no clear way to systematically survey PID practices in this space. The purpose of the table is to offer a glimpse into the complexity and diversity of this space (which is driven by the existence of multiple PID systems and use cases for assigning PIDs to research instruments and facilities, which we discuss below), and to invite readers to survey and become acquainted with potentially unfamiliar terrain.

Please note that Table 1 does not capture how assigning PIDs to data is sometimes used for the same purpose as assigning PIDs to instruments directly, as demonstrated by the U.S. Department of Energy Atmospheric Radiation Measurement (ARM): "Assigning DOIs at the data product level allows ARM to use the same DOIs for new deployments of the same instruments for future ARM sites. As an example: the SONDEWNPN data collected at Southern Great Plains (SGP) and the recently deployed ARM Mobile Facility at the McMurdo Station, Antarctica (AMF2), uses the same DOI of 10.5439/1021460" (Prakash et al. 2016, pg. 4). In this case, one DOI is used to provide access to data from a number of research sites and instruments. In addition, Table 1 also does not capture how RRIDs have been assigned to a wide range of instrument models. "Models" here means general types of instruments, not specific instances. For example, the LI-COR Odyssey Classic Imager (RRID:SCR_023765) is a purchasable instrument. Each individual instance of this instrument does not have its own RRID, but the instrument model does. The RRID Portal lists around 2,300 such instrument models (https://rrid.site/data/source/nlx_144509-1/search?q=%2A&l=&facet[]=Resource%20Type%3Ainstrument%20resource).

Table 1: Selection of examples of PID application to research instruments and facilities.

Organization & Citation	PID Type	Resource(s)	Comment
University of Colorado Boulder Research Computing (2021a,b; 2023), RRID:SCR_019299	DOI, RRID	Supercomputer, computing clusters, large-scale storage service, campus core facilities	Both DOIs and RRIDs have been assigned to identify and track citation of their facilities in the published literature.
NSF National Center for Atmospheric Research (Aquino et al. 2017; UCAR/NCAR-Earth Observing Laboratory, 1990; 1994; Computational and Information Systems Laboratory 2023)	DOIs	Supercomputers, airplanes, radar systems, surface flux observatories, etc.	DOIs have been assigned to 25 observational facilities, and to multiple generations of high-performance computing systems.
Animal Physiology Facility of the French National Research Institute for Agriculture, Food and Environment (PAO 2018)	DOI	Research facility	This DOI was assigned to a research facility as a whole, encompassing many constituent activities.
U.S. Geological Survey (Falgout, Gordon, Williams, and USGS Advanced Research Computing 2015)	DOI	Supercomputer	This DOI is for a single computing resource.
University of Cape Town (University of Cape Town, Carr, and Lewis 2023)	DOI	High performance computing facility	This DOI was assigned via Zenodo, a generalist repository. In this case, the University of Cape Town used Zenodo as a DOI registration service, as the DOI resolves only to a minimal Readme file.
Florida State University (Ruhs et al. 2018), RRID:SCR_011228	RRID	Research instruments, campus core facilities	More than 1700 pieces of large equipment (e.g. microscopes, MRI scanners) have been assigned RRIDs.
Stanford University, RRID:SCR_011538	RRID	Campus core facilities, software, other campus resources	More than 150 campus facilities and other resources have been assigned RRIDs.
Oregon Health and Science University, RRID:SCR_009665	RRID	Campus core facilities, other campus entities	More than 50 campus facilities and other resources have been assigned RRIDs.

Findings from Project Activities

Our first-year efforts resulted in the percolation of four main topics that require further exploration and coordination if PIDs are to be widely adopted for instruments and facilities. We now turn to a discussion of these topics.

Topic 1 - Clarifying the Use Cases for Assigning PIDs to instruments and facilities

It is essential to be explicit about the motivations and prospective benefits, or use cases, associated with assigning persistent identifiers to research instruments and facilities. One fundamental use case identified by participants in our project is that assigning PIDs to facilities and instruments would potentially allow these resources to be more easily tracked and cited. This was an initial goal of the University of Colorado, Boulder, and NSF National Center for Atmospheric Research work in this space as well (Mayernik and Maull 2017). Proponents of PIDs have touted their ability to enable research facilities to measure impact as well as to create interconnectivity that supports collaboration across institutions and facilities (Cousijn et al. 2021). Many participants from our project, including in the focus groups and workshop discussions, noted that PIDs assigned to instruments and facilities are often not cited or given attribution in scientific publications by researchers using these resources. This lack of citation and attribution has many debilitating effects. One effect is an inability for directors and operators to demonstrate a return on investment or track the impact of their respective facilities and instruments. In focus groups and the workshop, we heard that the ability to track use and impact is important for facility and scientific equipment administrators because these resources are costly and, in many cases, funded through grants where measuring a return on investment is important. Persistent identifiers have the potential to better enable research facilities and centers to measure and predict the impact of their facility, while simultaneously creating an environment of interconnectivity that supports collaboration across institutions and facilities (Cousijn et al. 2021). Lack of citation also works against progress in advancing open science and FAIR Data Principles (Juty et al. 2020; Wilkinson 2016).

The second main use case we heard from participants involves reproducibility. In this case, PIDs associated with the instruments would help scientists who are attempting to reproduce findings to assess whether or not different results in subsequent studies are due to differences involving instrumentation. PIDs could make it easier for these scientists to replicate the original methodology of published work. This was a particularly important use case for workshop attendees who come from the biomedical field, as precisely reproducing laboratory experiments via the same instruments, reagents, cell lines, etc., was seen as important to ensure confidence in laboratory results.

A third use case for instrument PIDs that we have encountered centers on provenance. This use case is related to the reproducibility use case noted in the previous paragraph but focuses on transparency and understandability of data and results that were produced via the use of specific facilities or instruments, rather than on actively reproducing those data or results. Workshop participants discussed how PIDs could provide a convenient mechanism to establish provenance traces between facilities and instruments and

subsequent research outcomes. This is a particularly central use case for data repositories who provide access to data that came from research facilities or instruments. "Instrument papers" that are published in journals like the Journal of large-scale Research Facilities (JLSFR) are another example of this provenance use case (Stocker et al. 2020). These papers describe how instruments were utilized in research and are similar to the concept of "data papers" published in journals such as Scientific Data, which describe original datasets and their potential uses. Instrument papers advance open science and FAIR principles by helping users and researchers better understand how instruments are used in experiments; in addition, these papers also provide researchers with a deeper knowledge of the instrument itself (Stockman et al. 2020).

The last use case for instrument PIDs noted by workshop attendees is to facilitate equity in the research ecosystem. For example, assigning PIDs to facilities and instruments might make it easier for researchers from under-resourced institutions to discover and access instruments they need for their own work. In other words, assigning PIDs to instruments and facilities could lead to a more equitable allocation and use of resources across the research landscape. This benefits all researchers, but it might be especially beneficial for researchers from institutions with fewer resources. This use case is particularly important for operators of core facilities on university campuses, who want to ensure that their facility is known by all possible users.

The use cases identified by contributors to our project all align with prior literature and a recent report funded by research funders in the UK (de Castro et al. 2023). Provenance and reliability are important factors for researchers and facility/equipment operators to consider when deciding to undergo the additional work of assigning and using PIDs. Dappert et al. (2017) claim that widespread creation of PID infrastructure can accelerate open science by creating a technical environment with seamless discovery of resources, clear attribution to contributors, traceable provenance, and citations in scholarly communication. They also demonstrate the impacts that persistent identifiers have had in enabling follow-on research to be conducted much sooner, leading to quicker solutions to scientific problems.

The literature around PIDs also discuss how the benefits of adopting PIDs can be amplified by leveraging their metadata to create a "PID graph" that represents how various research entities interconnect within the scholarly ecosystem. Connecting persistent identifiers and their associated metadata within the framework of a PID graph enables researchers, labs, and institutions to access new insights and information. Cousijn et al. (2021) claim that by identifying and resolving objects, persistent identifiers and their associated metadata can ensure that research entities and infrastructure are accessible and discoverable. The persistent identification and upkeep of research data and metadata about its associated instruments and facilities can also improve the level of trust in which other researchers can reuse such data (De Smedt et al. 2020). This interconnectivity also makes it more possible to automate the flow of information between systems to streamline reporting and sharing of research outputs. In sum, the interconnectivity of persistent identifiers and their associated metadata enhance reproducibility by making it easier to verify scientific claims (Cousijn et al. 2021).

Topic 2 - Navigating the PID Landscape

Participants across project activities, including focus groups and the workshop, expressed a need for better understanding of the PID systems used for instruments and facilities. Across the research ecosystem, there are many PID systems in use, and as discussed earlier, several are already being applied to instruments and facilities. These uses are disconnected and fragmented, a challenge that is compounded by the fact that most PID systems were not designed for instruments and facilities. Participants therefore expressed the need for a better understanding of the infrastructure behind these systems, as well as clear rationale as to which systems are best suited to instruments and facilities.

Figure 1 provides a visual representation of several services that exist in the broader PID landscape. As noted here, multiple services either already play a role or could play a role in offering services that are relevant to the goal of systematically assigning PIDs to research instruments and facilities. Because the resources depicted in the figure are connected to each other in complex ways, the PID systems being used across this ecosystem have interdependencies, conceptually if not technically.

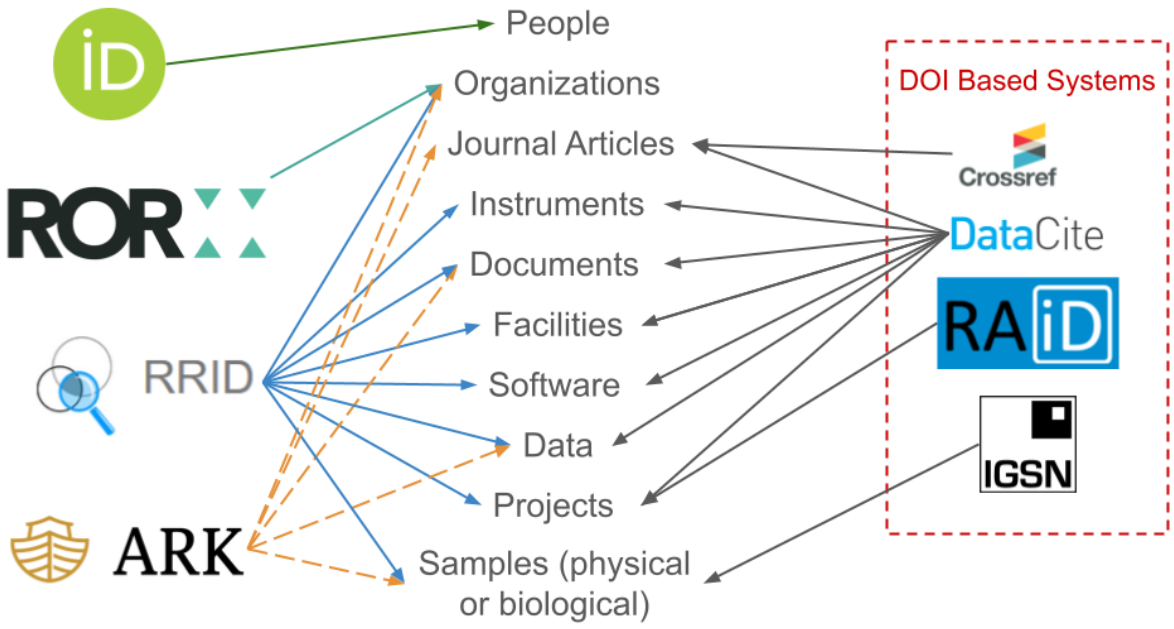

Figure 1: PID landscape - Three columns with arrows indicating association. Column 1 features four PID systems (ORCID, ROR, RRID, ARK), column 2 features generic entities (people, organizations, journal articles, instruments, documents, facilities, software, data, projects, samples (physical or biological)), column 3 features four "DOI Based Systems" (Crossref, DataCite, RAiD, and IGSN).

Each PID system has its own characteristics and functionality. As one example, some PIDs function mainly as locators for any set of content at a particular URL, while others are aptly considered to be identifiers for particular entities (Duerr et al. 2011). As noted in the examples we identified from our focus groups, workshop presentations, and the literature, the two persistent identifiers used most commonly for instruments and facilities are RRIDs and DOIs. RRIDs are assigned to research resources like antibodies, model organisms, software, and core facilities to facilitate methodological transparency in research practices and are most commonly used within the biomedical literature. RRIDs resolve to standardized web pages hosted by the SciCrunch organization (Bandrowski et al. 2015). RRIDs are created by a number of different services, depending on the type of resource being identified. For research facilities, the Core Marketplace service (https://coremarketplace.org) creates RRIDs. Using RRIDs for this purpose is a fairly recent development (Bandrowski 2022).

The DOI system functions differently than RRID. DOIs provide persistent links to research materials, wherever they may reside. Initially started as a system of persistent identifiers for scholarly literature, DOIs are now assigned to a large variety of resource types, including everything depicted in Figure 1 (with the exception of people). DOIs function primarily as locators and sources of metadata as they resolve to particular web pages designated by the DOI assigner. Since instruments and facilities are generally physical objects, not web-based digital resources like data sets or software, DOIs should resolve to web pages that describe the object.

Given that both DOIs and RRIDs are already being used for instruments and facilities, focus group and workshop participants indicated that more work is needed to articulate their relative merits and drawbacks. The two systems have no explicit interoperability at present; their metadata schemas are different, and their resolvers are distinct. As we found in the use cases discussed in our project activities as well as in the literature, the communities using them are also largely separate, with biomedical and other campus-based research facilities using RRIDs while DOIs are used for the purposes of identifying research facilities, computing systems, and other research instruments by a more heterogeneous set of organizations, including some government agencies.

Another source of variation among PID providers is in their associated metadata schema. Several project participants referred to efforts within the Research Data Alliance (RDA), where a working group called Persistent Identification of Instruments (PIDINST) created a community-driven metadata schema specifically focused on instrument PIDs (Stocker et al. 2020). The DataCite metadata schema version 4.5 made updates to include elements of the PIDINST schema. RRIDs, ARKs (Archival Resource Keys), and other PIDs represented in Figure 1 have different metadata approaches. As we saw in the use cases identified by our project activities, particularly at the workshop, this variation is a source of disconnect between the services, as well as a point of differentiation between them.

Trust in PID systems is a key factor and is essential to facilitate consistent use (Clark 2021). Organizations and individuals who seek to assign PIDs must trust in any individual PID system to survive, trust that the systems will be adopted and work as promised, trust that the various PID-related services will interconnect, and trust that the ecosystem of PIDs as a whole is the right approach. A recent inter-organizational working group within the US released a report (ORFG PID Strategy Working Group 2024) that presents the following list of "Desirable characteristics of PIDs":

- Open availability of core metadata
- Use of well-established resolver services
- Documentation of identifier policies
- Monitoring and reporting services
- Ease of assignment/metadata creation and curation
- Standardized structures, metadata, and services that allow for community input
- Extensibility
- Community governance

These characteristics are also aligned with what we have heard from facility and instrument providers who have participated in our project activities. In particular, the "ease of assignment" criterion has been consistently highlighted in our discussions; we will return to this topic at greater length in our exploration of how the level of effort for PID assignment can be lowered (Topic 4 below).

Topic 3 - Intrinsic Challenges: Variation, Modularity, Evolution and the Ambiguities of PID Assignment to Instruments and Facilities

The intrinsic properties of instruments and facilities pose important challenges for PID assignment and citation. One of the central features of instruments and facilities is that these entities have unclear boundaries and change over time. This topic came up in every one of our project activities regardless of whether we asked about it directly. To be sure, these features are not unique to instruments and facilities, and a key challenge when considering PIDs for any type of entity is to develop a systematic way to account for the evolution and change of the underlying resources to which PIDs are linked. This is a fundamental challenge for data citation (Mayernik 2013) as well as for software, organizations, and even people (e.g. name changes). Similarly, instruments and facilities are frequently reconfigured, recalibrated, and changed during the course of their regular use and maintenance, which raised important questions for our project participants across all of our focus groups and the workshop, such as the following:

- How can a persistent identifier account for changes in instrument configurations or sub-components over time?
- At what point do changes in an instrument accumulate to the point that it's a different instrument entirely? (The modern science version of the "Ship of Theseus" paradox, Furner 2009)
- What might the implications of these changes be for the persistent identifier that is attached to such an instrument?

A related class of challenges concerns the modularity of instruments. In particular, instruments are often comprised of multiple subcomponents and interchangeable parts and can be disaggregated in various ways. Several focus group and workshop participants provided use cases demonstrating this. What then is the relevant unit of analysis for purposes of instrument or facility citation? For example, remotely sensed data requires (among other things), a sensor that can detect electromagnetic radiation reflected from the Earth's surface, and a platform to carry the sensor (such as a satellite or balloon). Should each of these components receive its own PID, with the expectation that they will be cited individually when referencing remotely sensed data? Or should the sensor and platform be viewed as an integrated instrument that merits a single PID? If the apparatus is viewed as a unified instrument with a single PID, should this PID be changed when a sensor is swapped out? More generally, the relationship between facilities and instruments poses its own granularity problem. For example, if a microscope from a lab is used to produce a research publication, what should be cited? Should the microscope be cited using its persistent identifier? Or should the broader lab facility be cited? Both? There are no clear-cut answers to such questions in the context of facilities and instruments but focus group and workshop participants expressed a strong need for guidance in this area since attribution is a key concern for many stakeholders. Data citation recommendations typically discuss a distinction between "major" and "minor" version changes (ESIP 2019), but based on discussions within our project activities, it is unclear if these concepts apply to instrumentation. Indeed, some of these issues may not have elegant conceptual solutions and may need to be addressed through a "learning by doing" approach; participants indicated that a good rule of thumb would be to start with simple use cases and then move to more complicated use cases as needed.

Topic 4 - Systemic Challenges: Reducing Level of Effort and Creating Incentives for Participation in the PID ecosystem for instruments and facilities

While a pragmatic "learning by doing" approach may be the best way to navigate the challenges described in the previous subsection, an important barrier to such an approach is that building and engaging with the PID ecosystem for instruments and facilities is costly. Instrument and facility providers we spoke to in our project activities indicated that they face significant resource constraints and limitations that make assigning and updating PIDs challenging even if it is something they recognize as important and want to do. This is true regardless of the strengths and weaknesses of the different PID systems discussed above. Lowering these barriers to adopting and using PID systems for instrument providers is therefore an important challenge that must be addressed.

One way to lower the level of effort for instrument and facilities providers is to improve and streamline relevant infrastructure; doing so may involve up-front costs but will pay dividends over time by increasing the ease of use of PID systems. For example, in the realm of citation services, project participants who are instrument and facility providers repeatedly expressed a need to have easy ways to use PIDs to gather relevant citation metrics. The ecosystem for PID-based citation metrics is still relatively new and remains fragmented. As project participants indicated, the primary interest is in tracking citations to instruments and facilities within journal articles to demonstrate the impact of the facility or instrument in the published literature (e.g., to funders); however, facilities and instruments are cited in a variety of different places within articles (e.g., methods sections, acknowledgments, references) that can affect how citations to PIDs are tracked. Several services can provide citation metrics for DOIs, including CrossRef and DataCite, both of which provide citation counts for DOIs registered via their services, typically only when citations occur in reference lists. Other citation indexes are less consistent in providing citation counts, such as Elsevier Scopus and the Clarivate Web of Science. Outside of DOIs, there has been demonstrated success in citation services for RRIDs, which participants in the workshop shared are able to track RRIDs regardless of where those are used within an article. Workshop participants shared that these services have been particularly effective for tracking citations to antibodies in biological research. Over ten years of gradual adoption and use, approximately 300,000 RRIDs for antibodies have been used in 45,000 publications within 2,000 journals (Bandrowski et al. 2023). It is not yet clear whether RRIDs for research instruments have or will receive the same level of adoption or be indexed in the same way.

Focus group and workshop participants emphasized that it is also essential to encourage researchers and PIs to engage with the PID ecosystem for instruments and facilities by lowering barriers to participation (Dappert et al. 2017; Macgregor, Lancho-Barrantes, and Pennington 2023). One approach that we heard from project participants is to ensure that relevant processes and workflows for discovering instrument and facilities PIDs, and for citing them, are as clear and efficient as possible. Bandrowski et al. (2015) examined authors' behavior in order to learn more about their participation, performance, identifiability, and utility regarding the citation of RRIDs for research resources, such as software/tools, primary antibodies, and model organisms; among other findings, this RRID pilot project concluded that authors are usually willing to adopt new styles of citation for research resources if the process is clear (Bandrowski et al. 2015).

While it is important to lower the level of effort for researchers when engaging with the PID ecosystem, it is important to note that simply lowering barriers is not likely to be enough. This is particularly the case for researchers, who are less likely than instrument and facilities providers to be familiar with the benefits of assigning and citing instrument and facility PIDs, according to project participants. It is therefore important to communicate the benefits of adopting and using PIDs through systematic community engagement and outreach work. Moreover, focus group and workshop participants identified the need to build and maintain a robust incentive structure that encourages researchers to develop a vested interest in participating in the instruments and facilities PID ecosystem. For example, publishers could potentially require the submission

of PIDs for the equipment used in research projects, and facilities and funders can do the same, which could increase uptake in PID use. Plomp (2020) also suggests that incorporating FAIR Principles into the promotion and tenure process, as well as evaluating research proposals with respect to the FAIR Principles, would increase the adoption of PIDs.

Implications for Data and Information Professionals

While the four key topics described in the previous section include direct implications for some of the stakeholder groups in our community of practice (e.g. instrument and facility operators, PID infrastructure providers, researchers, publishers, etc.), we will focus, in this section, on lessons learned that could be of particular value to data and information professionals.

As is often the case in data librarianship and related areas, individuals in such roles could play a valuable role as connectors and translators across the other stakeholder groups needed to advance the goals identified by our project. For example, data professionals often already interact with both operators and users of research instrumentation at their own institutions as well as with external groups like journal publishers and PID infrastructure providers. As our project continues to develop value statements that explicitly clarify the interests and incentives of diverse stakeholders in this space, data and information professionals could leverage these existing interactions to communicate the specific value of PIDs for facilities and instruments within and across groups.

In addition, data and information professionals bring extensive experience with assigning and working with PIDs for many other parts of the research ecosystem (e.g. ORCIDs for researchers, CrossRef DOIs for articles, DataCite DOIs for datasets, etc.). This experience is directly applicable to challenges with implementing PIDs for facilities and instruments. For example, data and information professionals can lend expertise to facility and instrument operators at their institutions in developing workflows for assigning PIDs, understanding the differences between PID systems, and navigating decisions around versioning and granularity of PIDs, which all have direct parallels to existing efforts related to PIDs for datasets. These professionals also bring experience advocating for the use of PIDs in citations to outputs like datasets and software within research articles, so they are well-positioned to encourage the researchers they collaborate with to include PIDs for instruments and facilities in article manuscripts as well. Furthermore, data and information professionals regularly work with PID infrastructure providers to improve metadata and ensure interoperability across systems, so they can play a key role in promoting the adoption of metadata elements specific to instruments and facilities in existing PID systems.

Data and information professionals can also apply some of our project findings directly to their own work with institutional repositories. For example, data and information professionals can incorporate metadata elements and documentation guidance specific to PIDs for instruments and facilities within repository curation and publication workflows. Similarly, data and information professionals can ensure that facility

and instrument PIDs are included in the underlying metadata in PID systems for datasets and other outputs (e.g. in DataCite DOI metadata). Some of this work could be done retroactively in repositories, in a process Habermann calls "re-curation" (2023). Efforts like these will help promote PID adoption and use while strengthening the connections among stakeholders in the research ecosystem at the heart of open and networked science.

Conclusion

The expansion of the PID ecosystem to encompass research instruments and facilities would meaningfully augment the value of this already critical ecosystem, and facilitate the discovery, connection, and attribution of research entities for a variety of purposes central to open science. In this paper, we have introduced a project that aims to develop a network of stakeholders that can coordinate their activities and mobilize a concerted effort to develop systematic community standards and guidelines for the assignment and citation of PIDs to facilities and instruments. In the first year of our project, we identified four key topics that this community must address; in the remaining time of the project, we will take steps, including hosting additional workshops, to bring the community together to make progress on the topics identified in the first year of activities. This will culminate in clear recommendations and best practices for adopting and implementing PIDs for instruments and facilities that we hope can help various stakeholders and contribute to national level initiatives such as the recent effort to develop a US National PID Strategy (ORFG PID Strategy Working Group 2024). In the meantime, data and information professionals can develop relationships with instruments and facilities providers at their institutions, engage in education and outreach activities that communicate the value of citing instruments and facilities to the researchers they collaborate with, and integrate PIDs for instruments and facilities into their data curation and publication workflows.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgements

We thank our project advisory committee members, and the participants in our 2023 and 2024 project workshops. This project was funded by NSF Awards #2226396, 2226397, 2226398.

This project was reviewed and approved for human subjects research by the institutional review boards of the University Corporation for Atmospheric Research, University of Colorado Boulder, and Florida State University.

This material is based upon work supported by the NSF National Center for Atmospheric Research, which is sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

The content of this article is based on the presentation entitled "Community-based interoperability for research facility and instrument persistent identifiers: Enabling connection throughout the data ecosystem" originally presented at RDAP Summit 2024, available from Open Science Framework: https://osf.io/9zk2c.

Data Availability

Focus group questions are available under the article Supplementary Files:

Appendix 1: Focus Group Questions

References

Aquino, Janine, John Allison, Robert Rilling, Don Stott, Kathryn Young, and Michael Daniels. 2017. "Motivation and Strategies for Implementing Digital Object Identifiers (DOIs) at NCAR's Earth Observing Laboratory – Past Progress and Future Collaborations." *Data Science Journal* 16 (2017): 7. https://doi.org/10.5334/dsj-2017-007.

Bandrowski, Anita. 2022. "A Decade of GigaScience: What Can Be Learned from Half a Million RRIDs in the Scientific Literature?" *GigaScience* 11 (2022): giac058. https://doi.org/10.1093/gigascience/giac058.

Bandrowski, Anita, Matthew Brush, Jeffery S. Grethe, Melissa A. Haendel, David N. Kennedy, Sean Hill, Patrick R. Hof, et al. 2015. "The Resource Identification Initiative: A Cultural Shift in Publishing." [version 2; peer review: 2 approved]. *F1000Research* 4 (2015): 134. https://doi.org/10.12688/f1000research.6555.2.

Bandrowski, Anita, Mason Pairish, Peter Eckmann, Jeffrey Grethe, and Maryann E Martone. 2022. "The Antibody Registry: Ten Years of Registering Antibodies." *Nucleic Acids Research* 51 (D1): D358–D367. https://doi.org/10.1093/nar/gkac927.

Bunakov, Vasily. 2019. "Metadata for Large-Scale Research Instruments." In *Metadata and Semantic Research (MTSR 2018)*, edited by Garoufallou, E., Sartori, F., Siatri, R., Zervas, M. *Communications in Computer and Information Science* 846. https://doi.org/10.1007/978-3-030-14401-2_30.

Bunakov, Vasily, Simon Lambert, Brian Matthews. 2018. "Persistent Identifiers for Facilities Research: Current Practices and Opportunities." Paper presented at *International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2018), Moscow, Russia, October 9-12, 2018*. https://ceur-ws.org/Vol-2277/paper32.pdf.

CERN Scientific Information Service (SIS). 2020. *Why use persistent identifiers?* Retrieved September 22, 2023. https://sis.web.cern.ch/submit-and-publish/persistent-identifiers/why-pids.

Clark, Jonathan. 2021. "Open Science—A Question of Trust." *Data Intelligence* 3 (1): 64-70. https://doi.org/10.1162/dint_a_00078.

Computational and Information Systems Laboratory. 2023. *Derecho: HPE Cray EX Cluster*. UCAR/NCAR. https://doi.org/10.5065/QX9A-PG09.

Cousijn, Helena, Ricarda Braukmann, Martin Fenner, Christine Ferguson, René van Horik, Rachael Lammey, Alice Meadows, and Simon Lambert. 2021. "Connected Research: The Potential of the PID Graph." *Patterns* 2 (1): 100180. https://doi.org/10.1016/j.patter.2020.100180.

Dappert, Angela, Adam Farquhar, Rachael Kotarski, and Kirstie Hewlett. 2017. "Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research." *Data Science Journal* 16 (2017): 28. https://doi.org/10.5334/dsj-2017-028.

De Castro, Pablo, Ulrich Herb, Laura Rothfritz, and Joachim Schöpfel. 2023. *Persistent Identifiers for Research Instruments and Facilities: An Emerging PID Domain in Need of Coordination*. Bristol, UK: Knowledge Exchange. https://doi.org/10.5281/zenodo.7330372.

De Smedt, Koenraad, Dimitris Koureas, and Peter Wittenburg. 2020. "FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units." *Publications* 8 (2): 21. https://doi.org/10.3390/publications8020021.

Duerr, Ruth E., Robert R. Downs, Curt Tilmes, Bruce Barkstrom, W. Christopher Lenhardt, Joseph Glassy, Luis E. Bermudez, and Peter Slaughter. 2011. "On the Utility of Identification Schemes for Digital Earth Science Data: An Assessment and Recommendations." *Earth Science Informatics* 4 (2011): 139-160. https://doi.org/10.1007/s12145-011-0083-6.

ESIP Data Preservation and Stewardship Committee. 2019. "Data Citation Guidelines for Earth Science Data, Version 2." Earth Science Information Partners. https://doi.org/10.6084/m9.figshare.8441816.

Falgout, Jeff T., Janice Gordon, Brad Williams, and USGS Advanced Research Computing. 2015. *USGS Yeti Supercomputer*. U.S. Geological Survey. https://doi.org/10.5066/F7D798MJ.

Furner, Jonathan. 2009. "Interrogating 'Identity': A Philosophical Approach to an Enduring Issue in Knowledge Organization." *Knowledge Organization* 36 (1): 3–16. https://doi.org/10.5771/0943-7444-2009-1-3.

Habermann, Ted. 2023. "Connecting Repositories to the Global Research Community: A Re-Curation Process." *Journal of eScience Librarianship* 12 (3): e739. https://doi.org/10.7191/jeslib.739.

Johnson, Andrew, Renaine Julian, Matt Mayernik, Claudius Mundoma, Matthew Murray, Aditya Ranganath, and Greg Stossmeister. 2024. *FAIR Facilities and Instruments Workshop #1 Report: Exploring Persistent Identifier Needs, Barriers, and Incentives*. NCAR/TN-577+PROC. Boulder, CO: NSF National Center for Atmospheric Research. https://doi.org/10.5065/ZGSX-2D06.

Joint Committee for Guides in Metrology (JCGM). 2012. *International Vocabulary of Metrology - Basic and General Concepts and Associated Terms, 3rd Edition*. Sèvres Cedex, France: BIPM. https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf.

Julian, Renaine, Andrew Johnson, Matt Mayernik, Claudius Mundoma, Matthew Murray, and Aditya Ranganath. 2024. *FAIR Facilities and Instruments Workshop #2 Report: Recent Progress, Remaining Challenges, and Emerging PID Strategies*. NCAR/TN-586+PROC. Boulder, CO: NSF National Center for Atmospheric Research. https://doi.org/10.5065/jea7-yf24.

Juty, Nick, Sarala M. Wimalaratne, Stian Soiland-Reyes, John Kunze, Carole A. Goble, and Tim Clark. 2020. "Unique, Persistent, Resolvable: Identifiers as the Foundation of FAIR." *Data Intelligence* 2 (1-2): 30-39. https://doi.org/10.1162/dint_a_00025.

Lynch, Clifford. 1998. "Identifiers and Their Role in Networked Information Applications." *Bulletin of the American Society for Information Science and Technology* 24 (2): 17-20. https://doi.org/10.1002/bult.80.

Macgregor, George, Barbara S. Lancho-Barrantes, and Diane Rasmussen Pennington. 2023. "Measuring the Concept of PID Literacy: User Perceptions and Understanding of PIDs in Support of Open Scholarly Infrastructure." *Open Information Science* 7 (1): 20220142. https://doi.org/10.1515/opis-2022-0142.

Mayernik, Matthew. 2013. *Bridging Data Lifecycles: Tracking Data Use via Data Citations Workshop Report*. NCAR/TN-494+PROC. Boulder, CO: NSF National Center for Atmospheric Research. https://doi.org/10.5065/D6PZ56TX.

Mayernik, Matthew S., and Keith E. Maull. 2017. "Assessing the Uptake of Persistent Identifiers by Research Infrastructure Users." *PLOS ONE* 12 (4): e0175418. https://doi.org/10.1371/journal.pone.0175418.

McMurry, Julie A., Nick Juty, Niklas Blomberg, Tony Burdett, Tom Conlin, Nathalie Conte, Mélanie Courtot, et al. 2017. "Identifiers for the 21st Century: How to Design, Provision, and Reuse Persistent Identifiers to Maximize Utility and Impact of Life Science Data." *PLOS Biology* 15 (6): e2001414. https://doi.org/10.1371/journal.pbio.2001414.

Meadows, Alice, and Laure Haak. 2018. "How Persistent Identifiers Can Save Scientists Time." *FEMS Microbiology Letters* 365 (15): fny143. https://doi.org/10.1093/femsle/fny143.

National Science and Technology Council. 2022. *Guidance for Implementing National Security Presidential Memorandum 33 (NSPM-33) on National Security Strategy for United States Government-Supported Research and Development: A Report by the Subcommittee on Research Security*. Joint Committee on the Research Environment. https://www.whitehouse.gov/wp-content/uploads/2022/01/010422-NSPM-33-Implementation-Guidance.pdf.

ORFG PID Strategy Working Group. 2024. *Developing a US National PID Strategy*. Zenodo. https://doi.org/10.5281/zenodo.10811008.

PAO. 2018. *Animal Physiology Facility*. INRAE. https://doi.org/10.15454/1.5573896321728955E12.

Plomp, Esther. 2020. "Going Digital: Persistent Identifiers for Research Samples, Resources and Instruments." *Data Science Journal* 19 (2020): 46. https://doi.org/10.5334/dsj-2020-046.

Prakash, Giri, Biva Shrestha, Katarina Younkin, Rolanda Jundt, Mark Martin, and Jannean Elliott. 2016. "Data Always Getting Bigger—A Scalable DOI Architecture for Big and Expanding Scientific Data." *Data* 1 (2): 11. https://doi.org/10.3390/data1020011.

Ruhs, Nicholas, Claudius Mondoma, Renaine Julian, Annie Glerum, Mark Lopez, and Michael Meth. 2018. "Universal Scientific Equipment Discovery Tool (USEDiT): If You Used It...You Should Cite It." Poster presented at the *Association of Biomedical Research Facilities Conference 2018*. Florida State University Digital Repository. http://purl.flvc.org/fsu/fd/FSU_libsubv1_scholarship_submission_1540926633_9b1fba26.

Stocker, Markus, Louise Darroch, Rolf Krahl, Ted Habermann, Anusuriya Devaraju, Ulrich Schwardmann, Claudio D'Onofrio, and Ingemar Häggström. 2020. "Persistent Identification of Instruments." *Data Science Journal* 19 (2020): 18. https://doi.org/10.5334/dsj-2020-018.

UCAR/NCAR - Earth Observing Laboratory. 1990. *NCAR Integrated Surface Flux System (ISFS)*. UCAR/NCAR - Earth Observing Laboratory. https://doi.org/10.5065/D6ZC80XJ.

UCAR/NCAR - Earth Observing Laboratory. 1994. *NSF/NCAR Hercules C130 Aircraft*. UCAR/NCAR - Earth Observing Laboratory. https://doi.org/10.5065/D6WM1BG0.

University of Cape Town, Timothy Carr, and Andrew Lewis. 2023. *UCT HPC Facility*. Zenodo. https://doi.org/10.5281/zenodo.10021613.

University of Colorado Boulder Research Computing. 2021a. *Blanca Condo Cluster*. University of Colorado Boulder. https://doi.org/10.25811/V32C-GY42.

University of Colorado Boulder Research Computing. 2021b. *PetaLibrary*. University of Colorado Boulder. https://doi.org/10.25811/81NC-WV41.

University of Colorado Boulder Research Computing. 2023. *Alpine*. University of Colorado Boulder. https://doi.org/10.25811/K3W6-PK81.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (2016): 160018. https://doi.org/10.1038/sdata.2016.18.