Enhancing Scientific Research Through Instrument Metadata and PIDs – Tackling the "How Might We" Challenges

Cotober 31st, 2024 Event Metadata 7 min Read

https://doi.org/10.5438/34wz-er65

The <u>Persistent Identifiers (PIDs) for Instruments Community Dialogue</u>, hosted by DataCite and Metadata Game Changers, brought together a diverse group of experts to explore how PIDs can transform the identification of scientific instruments and the data they collect. With over 45 attendees from life sciences, astronomy, and other domains, the discussion highlighted the shared need for improving reproducibility, traceability, and metadata practices in scientific research and for clarity on roles and responsibilities for building and maintaining capacity.

This blog post summarizes the discussions and focuses on the key "**How Might We**" themes from the presentations and breakout groups to <u>help focus future efforts</u>. These themes point to practical steps the research community can take to address the challenges of creating and managing connected instrument metadata.

Key Presentations: Insights from the Experts

The event opened with presentations from experts across several fields, setting the stage for the deep-dive discussions that followed:

- Matt Mayernik (UCAR) emphasized the importance of making scientific instruments FAIR (Findable, Accessible, Interoperable, and Reusable). He stressed the need for traceability and provenance in instrument metadata to capture an instrument's history and ensure that experiments are reproducible. Mayernik also identified several use cases that rely on the unambiguous identification of instruments.
- Anita Bandrowski (SciCrunch) discussed how Research Resource Identifiers (RRIDs) have been used since 2014 to track biological instruments across the scientific literature. She introduced SciBot, a tool that automatically identifies RRIDs in PubMed papers, and noted the importance of creating standardized identifiers for "off-the-shelf" instruments to improve discoverability and tracking.
- 3. David Elbert (Johns Hopkins) shared his experiences in using PIDs to manage instruments in automated labs, where robots can perform tens of thousands of experiments daily. Elbert focused on the need to link datasets with instruments and emphasized how creating landing pages for instruments can motivate the adoption of PIDs.
- 4. <u>Rolf Krahl</u> (Helmholtz Zentrum Berlin) highlighted the challenges of managing metadata for custom-built and portable instruments in beamlines and experimental stations. He discussed the importance of using PIDs to track changes in instrument configurations and their role in establishing data provenance.
- 5. <u>Ted Habermann</u> (Metadata Game Changers) provided an overview of how the DataCite Metadata Schema has been extended to include instruments. Habermann demonstrated the ideas of instrument descriptions at multiple levels of detail and using related identifiers, such as "IsCollectedBy," to link instruments with datasets, ensuring proper citation and discoverability.

Recordings of these presentations are available on the DataCite YouTube channel. Slides and notes are also available.

Cross-Cutting Themes: What Can We Learn from the Community?

Following the presentations, breakout groups discussed several themes and formulated "**How Might We**" questions to guide future efforts in improving instrument metadata. These questions reflect common concerns across domains and provide a framework for moving forward.

1. Instrument Discoverability and Usage Tracking

How might we find references to instruments in papers?

One of the main challenges highlighted was the difficulty in locating instruments mentioned in scientific literature. Participants agreed on the need for a centralized system that tracks where instruments are used, who has used them, and what research outputs they have contributed to. **PIDs** can play a vital role in identifying instruments, making them more discoverable, and linking instruments to datasets and publications.

How might we track the impact (ROI) of instruments and facilities?

For large-scale research facilities, tracking the return on investment (ROI) of expensive instruments is crucial. PIDs can help measure an instrument's impact by linking it to datasets, citations, and publications, enabling institutions to demonstrate the value of their infrastructure.

2. Instrument Metadata and Data Linkages

How might we link datasets to instruments?

Ensuring that datasets, both observations and calibration data, are accurately linked to the instruments that generated them was a major focus of the discussions.

How might we ensure reproducibility through instrument metadata?

Reproducibility depends on capturing detailed metadata about an instrument's configuration, operational modes, and calibration history. Participants emphasized the need for better systems that track this information and allow researchers to replicate experiments under similar conditions.

3. Classification and Standardization of Instruments

How might we develop a common classification system for instruments across research domains?

The wide range of instruments across different scientific fields presents a challenge when creating standardized metadata. Participants explored the possibility of developing a unified instrument classification system that allows researchers to search for and compare instruments more easily. This would also simplify the process of documenting instruments in metadata records.

How might we document the use of off-the-shelf instruments?

Creating standardized metadata for commercially available instruments could streamline their use in research. Participants suggested that such metadata could track instrument usage across various experiments and publications, improving traceability.

4. Collaboration and Community Involvement

How might we involve the broader research community in creating effective metadata systems?

Participants recognized that the success of PIDs and instrument metadata systems depends on involving core facility managers, researchers, and even manufacturers. These stakeholders have the expertise to define how instruments are used and provide input into how metadata should be structured.

How might we encourage journal adoption of PIDs for instruments?

Journal mandates requiring PIDs for instruments in published research could be a game-changer in promoting adoption. Incentives for researchers and institutions to use PIDs could help drive widespread uptake and ensure that metadata systems are consistently implemented.

5. Calibration and Data Reuse

How might we track calibration histories and link them to datasets?

Calibration data is essential for understanding the quality of the data produced by an instrument. Without proper metadata linking calibration records, it becomes difficult for others to validate or reproduce the data and to use it effectively. Participants discussed the need for systems that record calibration histories and link them to datasets, allowing future researchers to assess the validity of the data.

How might we reuse instrument data across different studies?

Instruments often generate data that can be repurposed for new studies, particularly in fields like environmental monitoring. Participants explored how PIDs and metadata systems can facilitate the reuse of instrument data by providing the necessary context about how the data was collected.

6. Linking Instruments to Larger Systems

How might we link instruments to broader platforms and research facilities?

Many instruments are part of larger research platforms, such as telescopes, automated labs, or oceanographic vessels. There is a need to create connected PID systems that link individual instruments to the platforms they operate on. This would allow researchers to understand the role of each instrument within larger systems and its contribution to scientific outcomes.

Moving Forward: Addressing the "How Might We" Questions

Several key actions were suggested during the discussions to ensure that PIDs for instruments become a core part of research workflows:

- 1. Workshop Concrete Examples: Developing practical examples of how PIDs can be used for different instruments will help clarify the most pressing challenges and create pathways to implementation. These examples will serve as a foundation for future discussions and solutions.
- Create Shared Resources: Developing shared resources, such as public repositories and notebooks, can help the community "level up" together. These resources will showcase real-world applications of PIDs and serve as templates for others to adopt.
- 3. Align Responsibility and Capacity: Those engaged in pursuing next steps must remain mindful of who is responsible for what solution components and what their capacity for long-term engagement, management and upkeep might be.
- 4. **Support for Small Labs**: Ensuring that smaller labs and institutions can use PIDs effectively is critical for encouraging broad adoption of this technology. Opportunities for facilitating access should be explored by the community.
- 5. **Engage Publishers**: Encouraging publishers to require PIDs for instruments in research papers will drive adoption and make the system more standardized across disciplines.

Conclusion: Tackling the Challenges of Instrument Metadata

The **Persistent Identifiers for Instruments Community Dialogue** underscored the potential of PIDs in improving instrument metadata by addressing the "**How Might We**" questions raised during the discussions. Collaboration across disciplines and communities, combined with practical examples, meaningful incentives and clearer roles and responsibilities, will help ensure that PIDs become an integral part of the future of scientific research. Stay involved in these ongoing conversations by contributing to the <u>discussions on the DataCite Suggestions GitHub forum</u>, participating in <u>the PID Forum</u>, or joining the RDA Persistent Identification of Instruments Working Group. For more details about the dialogue and the recording, visit the <u>Persistent Identifiers for Instruments Community Dialogue page</u>.