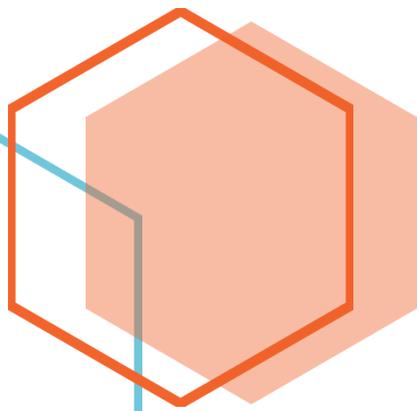




2019 Data ID Service Workshop

Summary Report

Findings and recommendations from the September 2019 Data ID Service Workshop
hosted by the Office of Scientific and Technical Information in Oak Ridge, TN.





2019 Data ID Service Workshop

Executive Summary

In September 2019, the U.S. Department of Energy (DOE) [Office of Scientific and Technical Information \(OSTI\)](#), within the [Office of Science \(SC\)](#), held a two-day workshop to gather detailed feedback on the data community's research and data management needs. The workshop hosted over 40 participants representing 16 DOE national laboratories and/or government agencies. Attendees included DOE-funded researchers, data managers, DOE Scientific and Technical Information Program (STIP) managers, DOE Data ID Service and Interagency Data ID Service clients, and other key data stakeholders. While diverse perspectives were revealed across various user communities and scientific disciplines, several consensus needs emerged to inform improvements to OSTI's data tools and services. This report includes information on the purpose, background and structure of the workshop, as well as the outcomes and actions defining the path for addressing these needs.

OSTI planned and convened its first DOE Data ID Service workshop at the SLAC National Accelerator Laboratory in April 2016. One significant outcome of this first workshop was an almost unanimous request to continue hosting similar workshops, providing an interactive environment for various communities to gather and discuss data-related topics. Subsequent workshops have been held biennially at OSTI in Oak Ridge to allow individuals creating and managing scientific data to convene and discuss issues and needs within the data community and strategies for future improvements to data management and preservation.

Purpose

The purpose of the 2019 Data ID Service Workshop hosted by OSTI was to share information about OSTI's data registration services and gain a deeper understanding of researchers' managers' data-related challenges, concerns, and needs and to consider how to help address them. This report summarizes the findings which will lead to targeted actions to better serve the needs of the various user communities.

Workshop Structure

The 2019 two-day workshop was structured to enable open discussion and practical learning opportunities through short presentations and breakout discussion sessions. Session topics included metadata submission and the Digital Object Identifier (DOI) assignment process, research workflows and data creation, data Collections, interlinking data and other research, DOIs for non-traditional research objects, [DOE Data Explorer](#) and data discovery, and the use of other persistent identifiers. Participants from DOE labs and other agencies were also given the opportunity to share their approach to data management and use of the Data ID Services that OSTI provides.



Each workshop session began with a short presentation to set the stage for breakout sessions. Breakout discussion groups were kept small to encourage conversation as groups talked through a series of predetermined questions. The workshop format allowed attendees to engage with each other and OSTI participants about pertinent issues that affect their work with data. OSTI collected feedback throughout the sessions, which is summarized in the Session Summaries section of this report. This feedback will inform improvements and enhancements to the tools and services offered to users.

Session Summaries

These summaries contain a short description of the introductory presentations, followed by bullet points that cover the key takeaways of the breakout discussions.

DOI Assignment Process

The session began with short presentations covering OSTI's two data DOI registration services, the [DOE Data ID Service](#) and the [Interagency Data ID Service \(IAD\)](#). The DOE Data ID Service presentation covered an overview of the service, metadata creation, the metadata submission process, and updates to the corporate ingest tool, E-Link, based on client and stakeholder feedback. The IAD presentation covered an overview of the service, current IAD clients, the differences between the two services, and enhancements to the submission process.

Key Takeaways:

- Most participants use the Application Programming Interface (API) to obtain DOIs. However, some API fields and submission requirements were unclear to researchers.
- Additional suggested data types include Instrument, Sample, Grant, Sensor, Beamline, and Proposal. Researchers would like the ability to suggest multiple data types and add different document formats.
- API documentation should be updated often and enhanced by providing more examples and improving overall readability.
- Additional guidance and education regarding data management plans (DMPs), data permanence, and the data submission process are needed.
- Some participants were not aware of the data status and password expiration email notifications OSTI provides.

Research Workflow

This session was composed entirely of breakout discussion.

Key Takeaways:

- Generally, DOIs are assigned at the end of the research workflow, either at the time of metadata submission or article publication.



- Researchers pointed out that some issues blocking data from becoming publicly available include an unwillingness to share data, and the length of the metadata submission process prohibits researcher motivation.
- Several labs use local data repositories to host data, while others use general repositories, institutional repositories, and cloud storage.
- There was a general lack of awareness and confusion about reserving a DOI.
- Best practices are needed for DOI granularity and versioning.

Data Collections

The session began with a brief overview of data Collections, including the history and expressed need for the ability to register several datasets with a single DOI. Benefits of data Collections such as ease of citation in a journal article, interlinking research, and improved reuse of data were discussed. In addition, enhancements to improve data Collection functionality based on client feedback were covered. A second presentation was given by a current data client, the [Environmental Molecular Sciences Lab \(EMSL\)](#), about their experience working with OSTI to improve data Collections to better associate data across collaborations.

Key Takeaways:

- Researchers were interested in using Collection metrics to determine how research is being used and accessed.
- Participants felt that Collections may be organized by research package or based on level of anticipated reuse. Participants expressed that providing some relational metadata, such as related DOIs or collaborations, would be beneficial to users.
- Collection sustainability is concerning for researchers. Educating users on the ability to edit Collections after creation may help with this issue.
- More education is needed about the role of Collections and Collection curation, ownership, and maintenance.

DOIs for “Non-Traditional” Objects

The session kicked off with a presentation covering the current objects that are commonly assigned DOIs as well as research objects that could be assigned DOIs that many do not consider, such as supplemental information, proposals and awards, or instruments. The ability to link these “non-traditional” objects to common STI through the assignment of a DOI by using the related identifier field in E-Link and then subsequent display in OSTI’s search tools was demonstrated.

Key Takeaways:

- Most attendees were unaware of the related identifier field, and those who were aware stated that researchers consider it a burden to submit and manage related identifier information.



- Many participants expressed a need for clearer documentation for how to include related identifier information when using the API to submit.
- DataCite relation types (controlled list of values to describe the relationship between two resources) are vague, and they need better definitions and examples, especially for the isCitedBy and isReferencedBy.
- There was interest around versioning – when to issue a new DOI and how to link the various versions together.
- Many participants expressed an interest in being able to see how relationships would look in DOE Data Explorer in a test environment prior to final submission.
- Attendees had confusion over the related identifier tab nomenclature in the search tools, specifically around difference between “Other Related Research” and “Related Works.”

Data Discovery

The session started with a brief presentation covering a variety of topics, beginning with an overview of OSTI’s data discovery tool, DOE Data Explorer (DDE). Search and navigation features, the hierarchy of data organization, and additional features such as the News and Data ID Services information pages were demonstrated. Google Dataset Search, the FAIR Guiding Principles, and data metrics were also discussed.

Key Takeaways:

- DOE Data Explorer (DDE) is not currently used as a primary data searching tool. Google and OSTI.GOV are more common resources.
- The usability of DDE could be improved by refining tabs and filters and updating the information organization. Showing more horizontal relationships along with organizational graphics would be beneficial to users.
- The consensus was that the Projects section of DDE was unnecessary. The Collections section may instead be used to fill this role.
- Participants expressed an interest in citation and download metrics to illustrate the impact of research data.

Using Other Identifiers

The final session of the workshop covered the major persistent identifier services and OSTI’s relationship with each service. The presentation explained how OSTI assigns DOIs to data through DataCite, DOIs to technical reports through Crossref, and the current ORCID iD integration available through OSTI.GOV. The upcoming DOE ORCID consortium and benefits to membership were discussed, as well as the new Award DOI Service that OSTI is developing to assign DOIs to grants/awards/awarded time through Crossref.

Key Takeaways:

- ORCIDs are commonly used and required by some organizations. The identifiers are useful for publication tracking and employee reviews.



- Many participants were unfamiliar with Research Organization Registry (ROR) identifiers, but there was interest in using persistent identifiers for organizations if concerns about organization specificity and granularity are addressed.
- DOIs were suggested for additional research outputs, including posters, collaborations, and instruments. Obtaining DOIs for software was also of interest, but guidance and best practices are needed to enhance this process.
- There was mixed interest in obtaining DOIs for awards. Some participants already have plans for assigning DOIs to awards, while others had questions about award and grant distinction and award ownership.

Conclusion

Throughout the workshop, consensus needs and views were expressed along with many needs unique to certain disciplines and/or labs. This report enumerated many of the significant outcomes surrounding the topics of DOI assignment and the submission process, data creation and research workflows, data Collections, DOIs for “non-traditional” objects, data discovery, and the use of other persistent identifiers. Informed by these outcomes, OSTI has identified strategies for improving its tools and services by enhancing documentation, organizing more communication venues for feedback and education, providing technical support in response to client feedback, and improving current tools and implementing new features.

For more information, please visit the following resources or contact DOEDataID@osti.gov.

- [2019 Data ID Service Workshop](#)
- [Data ID Services](#)
- [E-Link API Documentation](#)
- [E-Link 241.6 Web Submission Interface Instructions](#)
- [DOE Data Explorer \(DDE\)](#)
- [Definitions](#)