User Facilities and Publications WG: Findings and Opportunities

Authors: Laurel L Haak, Erin Arndt, Benjamin Brown, Mark D. Doyle, Mariam Elsayed, Patricia Garvey, Terry Law, Kathleen Nasta, Robert Peters, Howard Ratner, Crystal Schrof, Steven Watson, Susan White-DePace
Date of Publication: 21 November 2017
Google Doc Link: https://docs.google.com/document/d/1tHloldKv3E1NsMXSoS7tDLv9-jBP6Y0aGP6ZnXkoAw8/edit#

Abstract: User facilities are specialized government-sponsored research infrastructure available for external use to advance scientific or technical knowledge. Researchers compete for access to these facilities and specialized equipment, but do not consistently acknowledge the contributions made by facilities when they publish the results of their work. Collection of these research outputs is necessary to enable an accurate assessment of the scientific impact of these public investments. Recently several user facilities have begun to explore innovative approaches to these challenges, including adoption of ORCID.

This report summarizes the discussions of the **User Facilities and Publications Working Group**, which focused on ascertaining what data would help agencies and facilities to map impact, and to determine whether and how ORCID could enable its collection in a manner that increases data capture and reduces reporting burden for researchers. The group had three objectives: to bring together publishers and facilities to better understand research, publication, and reporting workflows; to define terms to enable conversation; and to identify opportunities for working together to streamline and, where possible, automate impact reporting.

Table of Contents

Executive Summary

The User Facilities and Publications Working Group was convened by ORCID in 2017 to investigate how user facilities, publishers and scholarly infrastructure providers can collaborate to streamline the sharing of information between researchers, scientific user facilities, and publishers to improve the capture of outputs from these facilities and lessen reporting burdens on researchers. The *ad hoc* group grew out of a number of community conversations with the US Department of Energy (DOE), Society for Science at User Research Facilities (SSURF; formerly, the National User Facility Organization), CHORUS, and user facilities at Oak Ridge National Laboratory (ORNL) and Argonne National Laboratory (ANL).

The group first met in May 2017, with representatives from four DOE National Laboratories, publishers, US Federal agencies and infrastructure providers (see Appendix 1). There are many scientific user facilities around the world; the findings from this working group will be shared with other stakeholders to solicit input, ensure global applicability of findings, and encourage participation in pilot projects. Together, the group analyzed workflows, technologies, and standards in use by user facilities and publishers, and identified key points at which information could be shared or made available to downstream systems. Taken as a whole, the work of the group suggests a route to improved attribution and acknowledgement of research support, and better tools for the capture and analysis of that information.

The core findings of the group were that:

- Researchers do not often report on their use of facilities or specialized equipment[1] when they are publishing papers, datasets, or other research results. When they do, it is usually in the acknowledgements section of a paper and not in a standardized form.
- Publishers do not often collect information on user facilities, instead focusing on funding sources for the published work that are related to journal article access policies.
- User facilities require researchers to acknowledge use, but there is both a lag between use and publication and no general standard for citing usage.
- Differing understanding of terms such as "author", "user", and "funding" among stakeholders present barriers to effective reporting.
- All of the stakeholders see value in using APIs and persistent identifiers including ORCID iDs and DOIs to enable reporting processes.

The group identified several opportunities for user facilities and their sponsors, publishers, and researchers to improve the current capture and assessment of the impact of these facilities on the research enterprise. The following sections articulate actions that each stakeholder group could take that would move the research community toward this goal. The working group recognizes that the greatest impact will come from stakeholders working collectively and collaboratively.

---

[1] See, for example, Persistent Identification of Instruments. Research Data Alliance, 2017. Accessed 26 Oct 2017.

User facilities

User facilities - from research facilities housing specialized equipment, to repositories and field stations that house physical collections - and their government sponsors stand to benefit from improved acknowledgement of use and a better understanding of the impact generated by work undertaken with facility support. They can make the reporting process more straightforward and reliable by embedding identifiers and creating and sharing connections between users and their facility. There would be benefits to facilities and their sponsors and users from the following actions:

- collect ORCID iDs when researchers submit proposals, sample requests, and/or register for a user account and request permission to update their ORCID record if the proposal is accepted
- allow researchers to auto-populate form fields using data from their ORCID record
- obtain and use persistent organization identifiers for the lab and the facility
- issue a persistent and resolvable identifier (e.g., a DOI) for each awarded proposal
- post award information - including the researcher's ORCID iD and the award ID - in a publicly accessible database
- leverage ORCID API to enable automated post-award publication reporting for researchers

Publishers[2]

Publishers play a vital role in facilitating the attribution of funding and other kinds of research support. Publishers often use or provide specific standard fields (known as DTDs[3] such as the JATS standard - see Appendix 2) to store article information, such as funding and affiliation. Each article is tagged with a suite of fields that indicates how it is processed, indexed, licensed, and how Crossref - a non-profit organization dedicated to enabling citability and access of scholarly content - issues and manages a persistent digital object identifier for the article. Publishers and authors stand to benefit from improved discoverability and reporting, easier content licensing, and reduced data entry effort through the following actions:

- enable ongoing operational collection of validated ORCID iDs from all authors
- implement ongoing operational collection of information about relevant user facilities leveraging information from author ORCID records and other sources
- work to amend the JATS standard and other DTDs to standardize the capture and tagging of user facility use information (see Appendix 2)
- implement operational processes that enable inclusion of ORCID iD and user facilities metadata in the article and Crossref submission
- enable citation of relevant datasets during the manuscript publication process

---

[2] Publishers may be able to apply lessons learned from work on user facilities to the capture of funding information.

[3] DTD stands for Document Type Definition and describes an XML tagging scheme for the metadata and full content of a journal article. JATS is a specialized DTD used by several publishers.

Identifier and Information Infrastructure providers

Identifier and other infrastructure providers can help to capture, share and simplify connections between outputs and facilities. In addition to ensuring that their tools are openly available to facilities and supporting integration with facility and publisher systems, identifier providers would benefit from the following actions:

- ensure data models, application programming interfaces (APIs), and user interfaces can store, exchange, and display information about user facilities needed for the workflows described in this document
- work with user facilities on establishing protocols for assigning award identifiers
- ensure that user facilities can easily obtain and manage organization identifiers
- encourage publishers and repositories to include new identifiers in metadata registered with DOIs

Users

Researchers interact with many information systems during the course of their professional activities. Examples include applying for a grant, submitting a facilities use proposal, and reporting on research activities. Researchers are regularly asked to re-enter the same data into forms in different systems, or are cut out of the process altogether. They would benefit from the following actions that enable them to both reduce manual entry and better control whether and how their information is shared between systems:

- obtain and use an ORCID iD when publishing and submitting grant and facility use proposals
- enable automated record updates from publishers, funders, and facilities
- declare facility use and funding when submitting manuscripts describing their research

In taking these actions in a wider partnership with the research community, these groups have a chance to improve our understanding of the impact of the research work performed at user facilities, to establish new possibilities for the capture and description of relationships between organizations and the work they sponsor or support, and to support the research funding process.

Summary

The User Facilities Working Group undertook work to ascertain what data would help agencies and facilities to map impact, and to determine whether and how ORCID could enable its collection in a manner that increased data capture and reduces reporting burden for stakeholders. Our overarching goal in this project has been to streamline the sharing of resource allocation information across organizations, enabling each stakeholder to contribute to the process of connecting people, organizations, and publications while reducing the reporting burdens on researchers and facilities.

The Working Group initially focused on Department of Energy National Laboratories in an effort to start from shared experience and thereby facilitate conversation and consensus. Our intention from the start has been to share a summary of the group findings with the broader community to gather feedback, ascertain relevance of findings to other facilities, and to translate the discussions into practical solutions.

To ensure a common understanding of the information gaps and administrative challenges to efficient tracking of user facilities research outputs, the Working Group developed a detailed workflow model. The following section describes research workflows from time of application for facility use through research publication to reporting, identifies touch points with researchers, defines terms, and proposes opportunities for proof-of-concept pilot projects.

Workflows

The Working Group described an end-to-end workflow to enable citation of user facilities in research works. The flow starts at the user facility, at the time the researcher submits a proposal outlining their research goals and requesting access to the facility, proceeds through proposal acceptance, publication of a manuscript describing research findings, and reporting of the public research products.

Terms in **blue boldface** are defined within the context of the workflow model (see *Definitions* section, below). Terms in **black boldface** are suggested additions to the defined terms.

Step 1. Proposal Submission

Researcher has an idea for a project, and determines that specialized resources available at a **facility** is required to carry out the project. Researcher submits a **proposal** to the facility. The **facility manager** collects the proposal and assigns it to reviewers.

  A. The facility should collect **ORCID iDs** and leverage the ORCID record to obtain **researcher host institution** and **organization ID** from the **proposal submitter** [and all **co-proposers**].

B. The facility should consider collecting ORCID iD and leverage the ORCID record to obtain home organization name and organization ID from the **reviewer(s)**, with the objective of adding these activities to the reviewer's ORCID record.

C. The facility should request **permission** from the proposal submitter and co-proposers to update their ORCID record should the proposal be awarded.

D. The facility should assign a unique identifier to the proposal. The facility can decide whether to use an internal identifier for the proposal and/or a public **Proposal/Project Identifier** that may be shared externally.

Step 2. Proposal Acceptance

Based on the review of the scientific merit of the proposal, the facility decides to approve the project. The researchers are granted time at the facility and use of specialized equipment as detailed in their proposal. While there is no funding per se associated with the award, some facilities calculate a cash value equivalent. Researchers are supported by their home organization and/or a separate **research grant** from a **funding organization**.

A. The facility should obtain an external **Proposal/Project Identifier** (preferably a **persistent identifier**) for the **facility access award**.

B. The facility should post public information about the award in a public database [that they manage], such as title, abstract, **Proposal/Project Identifier**, date of award, **principal investigato**r, ORCID iD, and facility name.

C. The facility should use the ORCID permission token to post award metadata (**Proposal/Project Identifier**, **facility name** and **facility organization ID**, and if relevant **facility host institution name** and **host organization ID**) in the principal investigator's [and associates] ORCID record.

D. The facility should consider collecting ORCID iDs, researcher host organization IDs, and IDs for data management plans from researchers in the facility registration/access process, and store these IDs in the researcher's **facility user account**, to enable identification of the **facility user** (who may be different from the principal investigator), enable connections to experiments, equipment, and collections, and assist in managing user training and other user services. This could serve as an alternate step (in place of Step 1 and Step 2A-C) for those facilities that do not have a proposal process.

Step 3. Dataset curation and publication

The **user** performs research at a facility and generates a **dataset**. This dataset may or may not be hosted by the facility and or **facility host**. Regardless, the researcher will typically have described a **data management plan** detailing whether the dataset will be curated and how/when it will be made openly available. In cases where the researcher intends to curate a dataset,

A. The researcher should store the dataset in a permanent **repository** that issues persistent identifiers, such as **DOI**s.

B. The researcher and co-contributors should include their ORCID iD when depositing their dataset.
C. The repository should collect authenticated ORCID iDs and permission to update the depositor's ORCID record, and post information about the dataset (title, persistent identifier, deposit date) when the dataset is made public.
D. The repository should include information (via ORCID if not already included in the curation process) about the related proposal and facility where the data were generated.
E. The publisher should make it possible for the author to cite relevant datasets during the manuscript publication process, for example by allowing selection of the dataset information from the author's ORCID record and including identifiers as part of the registered metadata.

Step 4. Findings publication

The **author** prepares a **manuscript** summarizing research objectives and findings and submits it to a **publisher** for review and dissemination. The author declares the source(s) of funding for the work described so that the publisher may apply the appropriate open/public access policy.

A. The researcher should include their ORCID iD when submitting a manuscript to a publisher.
B. The publisher should collect authenticated ORCID iDs from all authors at time of submission and/or acceptance.
C. The publisher should collect host organization, funding, dataset, and facility information and associated IDs from authors via ORCID. When not available from the author's ORCID record, the publisher should provide a pick list of organizations from which to choose.
D. The publisher should include these metadata in their Crossref submission and consider minting DOIs at acceptance to facilitate linking, earlier updating of ORCID records, and early proof of research product benefits.
E. Indexers should consider including these metadata in their abstract and indexing services to enable search, discovery, and analysis.

Step 5. Reporting

Funding agencies and facilities need to be able to report on use of facilities and researchers that they support. These organizations may leverage persistent identifiers to streamline and automate much of the reporting process.

A. Facilities may use the ORCID API to receive notifications when a user publishes a dataset or paper, either generally or only those papers that include facility access award metadata.
B. Aggregators, such as CHORUS, can prepare reports for facilities and funders using ORCID iDs, DOIs, and organization IDs.

Definitions

These Definitions are provided to support shared understanding of concepts. and also action on implementing the workflows. We include both general terms (marked with *) and terms that may be relevant only in specific contexts. Our goal is to provide enough detail to allow for mapping across facilities and enable action on implementing the workflows.

Persistent identifiers

A persistent identifier is a digital code that is uniquely associated with a resource. It is long-lasting, hyperlinks to the source, and is usually associated with some additional data about the thing it refers to. Persistent identifiers that are relevant to this project include:

1. ***ORCID Identifier (iD)** – A unique identifier for a member of the research community, such as a Researcher, expressed as a URI with the format https://orcid.org/####-####-####-####
2. ***Proposal/Project Identifier** – An external persistent identifier assigned by the Scientific User Facility to a proposal/project for a Facility Access Award. Should resolve to a public view (landing page) of the project.
3. ***Digital Object Identifier (DOI)** – A unique identifier assigned by Crossref or DataCite used to identify a Publication or Data Set. Resolves to a public view (landing page) of the object being identified.
4. ***Organization identifier –** A unique identifier for an Organization. Should resolve to a to a public description (landing page) of the organization being identified.

Organizations

Organizations are key players in these workflows. Individuals may conduct research, but they are employed, supported, funded and their work published by corporate entities. The group has identified the following types of organizations as bringing significant value to the processes discussed:

1. ***Facility** - a research facility available for public use to advance scientific or technical knowledge. Includes Scientific User Facility, Field Station, and Scientific Collection as defined below.
   a. **Scientific User Facility (SUF)** – A government-sponsored research infrastructure available for external use to advance scientific or technical knowledge.
   b. **Field Station[4] -** living libraries and outdoor laboratories for students, researchers, and the general public interested in the environment. They vary greatly in form and purpose, and include both marine laboratories whose focus is offshore and terrestrial reserves dedicated to protecting key ecosystems.

---

[4] What are field stations? OBFS Web page http://www.obfs.org/what-s-a-field-station-. Accessed 26 October 2017.

    c. **Scientific Collection**[5] - collections of physical objects that have been gathered for research and other purposes, and are set aside for long-term preservation because of their potential use and impact in the future. Documents and data that are connected directly to these objects are considered part of the collections (e.g., field notes, maps, photographs and digital images).

2. ***Facility Host Institution** – An organization that administers or operates the Scientific User Facility, Field Station, or Scientific Collection; typically a national laboratory, government agency, or research university.
3. **Facility Steward Agency** – An organization that stewards the Facility; may or may not provide full or partial funding for the operations of the Facility.
4. **Facility Asset Funder(s)**- An organization that provides funding to run equipment or other facility asset (e.g., Eli Lilly or University of Chicago funding to run a particular beam line)
5. ***Funding Organization(s)** – An organization that provides intramural support or Research Grants enabling the user to conduct the user facility project/experiment.
6. ***Publisher** – An organization that reviews Manuscripts and processes and disseminates accepted Manuscripts as Publications.
7. ***Researcher Host Institution** – The home organization(s) of the Researcher(s) performing work at the Facility. For intramural researchers, this will be the same as the Facility Host institution, and in some cases may also be the Funding Organization. extramural researchers, the Researcher Host Institution will not equal the Facility Host Institution. The Researcher Host Institution will be specified on the Research Grant and/or Facility Access Award and should be available in the Researcher's ORCID Record.
8. ***Repository** - An organization and associated database that enables deposit, storage, and access to curated data sets.

Research Activities

A researcher might choose to put any of these research activities on their resumé. Receiving a funding award after a competitive process, crafting an output that describes a method or shares results, are building blocks of a research career. Activities can also provide a means for organizations to understand how they have helped to support research careers and the enterprise as a whole.

1. ***Facility Access Award** – Awarded proposal made to a Facility Proposal Lead or a research team for access to the Facility by a Facility Host Institution; typically awarded through merit or peer review.
2. ***Facility User Account** - Data record storing information about researchers who have requested access to a Facility. May contain public information about the researchers, but not usually a public document.
3. ***Manuscript** – Written record summarizing research findings. Not usually a public document.

---

[5] What are scientific collections. USFRC webpage. http://usfsc.grscicoll.org. Accessed 26 October 2017.

4. ***Proposal** – Request for research funding or facility access, part of all of which were derived from a research project performed at a scientific user facility. Not usually a public document, but associated with a unique identifier.
5. ***Research Grant** – Awarded proposal made to one or more Principal Investigator(s) of funding by a Grant Funding Agency to conduct research; typically involves peer review.
6. ***Publication** – A version of a manuscript that is made public, usually after peer review, either as a pre-print or final version.
7. ***Data management plan** - A description of how data will be collected, curated, stored, and how it may be accessed by the community. May be included in proposal.
8. ***Curated Data Set** – A (public) data record created by Researchers as they perform research at a Facility.

Individual roles

Individuals alone or, more commonly, in teams perform myriad activities that foster new knowledge and understanding. Too often, these contributions are lumped together as 'author'. This obscures real contributions, and limits our understanding of how a discovery was made. By capturing more granular roles in the research process, we can better describe the processes we seek to understand, support, and improve.

1. ***Researcher** – The individual performing work at the Facility. Researchers may have specific roles of Principal Investigator and/or Author and/or Data Curator. Researchers also have specific affiliations, for example as an employee of the Researcher Host Institution.
2. **Data Curator** – The individual(s) managing the data created during research performed at a Facility. This role may involve establishing data management plans, metadata annotations, and/or preservation plans.
3. **Facility Manager –** The position/individual responsible for managing access to a Facility.
4. ***Principal Investigator** – The named Researcher(s) on the Proposal or Facility Access Award from the Facility supporting the user project/experiment; may not be a User, but is almost always an Author.
5. **Facility Proposal Submitter** – The Researcher who submits a Proposal for a Facility Access Award; may not be the same individual as the Principal Investigator.
6. ***Co-Proposers** – Researchers listed on a proposal/project that may or may not have been users at a Facility.
7. ***User**[6] – A Researcher (individually or member of a research team) who is **granted access** to resources at a Facility.
8. ***Author** – Originator of any written work—anyone who appears in the author list of a Publication of a project/experiment that used a Facility as a resource; the author list may contain Users.
   a. **Corresponding or Primary Author** – The Researcher who submits a manuscript to a Publisher.

---

[6] In the DOE Office of Science, User is a strictly defined term.

        b. **Co-Author** – A Researcher who appears in the author list of the Publication.
9. **Publisher Editor** – An individual who manages the manuscript review process.
10. **Publisher Production Manager** – An individual who manages the manuscript production process.
11. **Facility (Host) Librarian** – An individual who post-publication archiving and may also manage pre-publication manuscript review and reporting processes.

Opportunities

Enabling interoperability - sharing of information about research - across individual, facility, and publisher systems not only streamlines reporting, it also underpins open research. For researchers, the ability to easily share research is intertwined with the necessity of getting credit for one's contributions. Facilities and publishers care about openness because it facilitates rigor and reproducibility, both of which are important for ensuring public trust in the research enterprise. The User Facilities and Publications Working Group has identified the importance of persistent identifiers in enabling the infrastructure for information sharing. The next step is to show how this can work in practice. The Working Group proposes that stakeholders collaborate on pilot projects to demonstrate proof-of-concept of the opportunities outlined in this paper.

Pilot Project 1: Integrate identifiers into user facility proposal submission and award processes.

Collect iDs. User facilities already have a standard process for collecting, reviewing, and making awards. Some are already exploring how to integrate ORCID iDs into this process. The first part of achieving this goal involves adding a step for the researcher to authenticate their ORCID iD, and storing this iD and a permissions token in the proposal system database. **Effort and Dependencies:** This integration is well-described and tested in a number of other systems. It should not provide a technical challenge but will require technical resources. Some policy work and outreach with users will be required.

Make the award citable. To be able to cite an activity, there needs to be some public record of it. Facilities can enable citation by creating a public awards database and using persistent identifiers for their awarded proposals. At a minimum, there should be a public persistent identifier[7] for the award that can be used by the facility to navigate to an internal record about the award; ideal would be a public database listing the award number, persistent identifier, title and abstract, with a list of the project personnel and associated ORCID iDs, and a means to attach links to related project products such as papers and datasets. **Effort and Dependencies:** Facilities will need to determine how to make proposal IDs public, what

---

[7] A number of funders and facilities are exploring working with Crossref to provision grant identifiers upon award.

identifiers to use, and whether and how to set up a public database.   This will involve policy decisions and some outreach and technical work.

Facilitate information sharing. Many facilities already request or require researchers to acknowledge facility use in a standard format, however, there is a long lag between the award, use, and publication and many researchers do not have award information or requirements in mind when submitting a manuscript. Facilities can assist the researcher by sharing award data with the researcher in a standard and easily shareable format. Facilities can post information into the awardee's ORCID record at the time the award is made - award title and number, facility name and organization ID, with the Laboratory name and organization ID as the information source.   This would enable the researcher/user/author to share the award information with a publisher during the manuscript submission and review process, and the embedding of the information in the paper itself. In turn, this facilitates the tracing of award to the related research products. **Effort and Dependencies:** Facilities have to first be collecting ORCID iDs at proposal submission, and will need to have access to the ORCID member API. From there it is not much more work to collect and store permissions to update the researcher's ORCID record. There will need to be put in place a process to push information when the award is made. This integration is well-described and tested in a number of other systems, and should not provide a technical challenge. ORCID will also need to examine their data model and make adjustments to accommodate distinctions between funding and facilities. This will in turn enable Pilot Project 2.

Reduce reporting burden. Researchers interact with proposal submission and reporting systems, filling in forms and keying data by hand. Much of these data are already stored in other systems. When facilities collect a researcher's ORCID iD, they can also pull information from that person's ORCID record - such as researcher employment, publications, and other funding awards.   Facilities can also use ORCID iDs and APIs as a means to track researcher publications, thereby reducing the burden and pain of reporting and instead allowing the researcher to focus on narrative. **Effort and Dependencies:** Facilities have to first be collecting ORCID iDs at proposal submission, and will need to have access to the ORCID member API. From there it is not much more work to collect and store permissions to pull information from the researcher's ORCID record. There will need to be put in place a process to map data fields and create an interface to allow researchers to select which items to share.

Pilot Project 2: Integrate award and facility IDs into the manuscript publication process.

Collect ORCID iDs at time of manuscript submission or acceptance.   There are published best practices and many examples (over 7000 journals) of how to collect iDs for the corresponding author. However, facilities users may or may not be the corresponding author, so to capture their iDs (and funding/facility use information) during the manuscript publication process, publishers will need to implement a process for collecting iDs from *all* authors. Some publishers are collecting co-author iDs at time of acceptance. Some are pulling iDs in through

submissions from collaborative writing platforms. ORCID is in the initial stages of testing functionality to collect iDs for groups.   **Effort and Dependencies:** While it is possible to collect iDs for all authors, it will take time and researcher demand for practice to expand from the current focus on corresponding author.

Couple collection of ORCID iDs with other data from the author's ORCID record. For those journals that already collect iDs, it is possible to extend the interaction with the ORCID record to collect other pieces of information. The idea is to collect facility use (and grant) information that has been posted into the record by facilities (and funders), and create an interface that allows the author to choose which facilities (and funding) is related to the manuscript.   **Effort and Dependencies:** The challenge is to implement functionality that adds value to the researcher and does not complicate the manuscript publication system. It is possible to use the ORCID API to pull in data from the author's record, present a list of facilities and/or grants, and request that the author check which ones pertain to the article being submitted. In the initial rollout it will be necessary to also couple this presentation with an "other" field that allows the author to select a facility or funder from a list. There are a handful of manuscript tracking and review systems that support thousands of journals. A pilot project could focus on implementing this workflow in one or a few of these systems.

Amend JATS standard to enable collection of facility use information. The working group considered the user facility objectives and current JATS specification, and determined that introducing a new JATS container element and tags specific to research facilities and non-monetary support would avoid the semantic confusion over the term "funding". For example, new tags could be <research-support-group> and <research-support support-type="user-facility">. This has the advantage of strongly identifying the semantics of the information and perhaps could be made general enough to support other kinds of non-monetary support.   **Effort and Dependencies:** Publishers have presented this finding to the JATS working group, and it is possible that the new tags could be included in a JATS version to be released early in 2018. This would enable a number of the other actions for publishers suggested by the Working Group, including the funding and facilities metadata in the article metadata and Crossref submission.

These pilot projects have the potential to leverage proven integration techniques in order to bring together our various communities of practice in a shared endeavour. Each group has a role to play in sharing data, capturing connections and recognising individual and organizational contributions to research. Each group can bring their position in the workflow to bear on facilitating and streamlining the flow of data. Finally, every participant in these projects stands to gain from the improvements they will generate.

Appendices

Appendix 1: User Facilities Working Group membership

The following individuals contributed time, effort and insight to the work of the group, culminating in this report:

Erin Arndt, Wiley, http://orcid.org/0000-0002-9012-1425
Benjamin Brown, Office of Science, US Department of Energy, http://orcid.org/0000-0002-4873-6362
Mark Doyle, American Physical Society, http://orcid.org/0000-0001-5919-8670
Mariam Elsayed, Office of Science, US Department of Energy, http://orcid.org/0000-0002-0483-3021
Patricia Garvey, Brookhaven National Laboratory, https://orcid.org/0000-0001-6056-2038
Laurel Haak, ORCID (Chair), http://orcid.org/0000-0001-5109-3700
Ted Hatcher, American Chemical Society
Matthew Hawkin, National Science Foundation
Terry Law, Pacific Northwest National Laboratory, http://orcid.org/0000-0001-8278-6729
Mark Martin, Office of Scientific and Technical Information, US Department of Energy
Kathleen Nasta, Brookhaven National Laboratory, http://orcid.org/0000-0001-6560-038X
Dan O'Brien, American Chemical Society
Robert Peters, ORCID, http://orcid.org/0000-0002-0036-9460
Howard Ratner, CHORUS, http://orcid.org/0000-0002-2123-6317
Crystal Schrof, Oak Ridge National Laboratory, http://orcid.org/0000-0002-8462-4579
Steven Watson, Elsevier, http://orcid.org/0000-0002-9990-7299
Susan White-DePace, Argonne National Laboratory, http://orcid.org/0000-0002-5031-4935

Appendix 2: User Facilities in Publications: A Metadata Perspective

How might we best accomplish the tracking of publications that result from research conducted at scientific user facilities? This research is conducted primarily by people extramural to the facility; there are also a significant number of intramural users. Access to the facility is granted through a formal proposal process. While no funding *per se* is awarded, there is a funding equivalent in terms of provision of time and facility resources, such as beam time, computational time, etc. Each proposal has an ID that could be used as a unique identifier.

How should user facility support be flagged in publications? Publishers use specific standard fields (DTDs such as JATS or other proprietary tagging schemes) to store article information, such as funding and affiliation. Each article is tagged with a suite of fields that indicate how it is processed, indexed, licensed, and how Crossref assigns funder IDs. After an initial set of conversations, the User Facilities and Publications Working Group decided:

- The affiliation tag was not appropriate
- Should facility usage should be considered semantically the same as "funding."
- Mark Doyle of the WG should consult with the JATS listserv for guidance.

What is JATS? JATS – Journal Article Tag Suite - is an application of NISO which defines a set of XML elements and attributes for tagging full-text journal articles. JATS is a standard used by a large number of publishers across disciplines to process metadata elements, such as authors, affiliation, and funding. JATS is managed by the US National Library of Medicine. This standard evolves based on user comment. User-suggested changes are collected and discussed by the Working Group who decide on action. Mark posed two options based on the WG discussions:

OPTION 1. Use existing <funding-group> and include an agreed-upon new value for the award-type attribute on <award-group>. This solution requires no changes to JATS, but may require additional facilities added to the Crossref Open Funder Registry. For a grant from the NSF and a project performed at the SNS, the JATS tags would look something like:

```
<funding-group>
      <award-group award-type="grant">
            <funding-source country="US">National Science Foundation</funding-source>
            <award-id>NSF DBI-0317510</award-id>
      </award-group>
      <award-group award-type="facility-support">
            <funding-source country="US">
            <funding source>Spallation Neutron Source</funding-source>
            <award-id>SPS 12345</award-id>
      </award-group>
</funding-group>
```

OPTION 2. Introduce a new JATS container element and tags that are more specific to facilities and non-monetary support to avoid the semantic confusion over the term "funding". For example, new tags could be <research-support-group> and <research-support support-type="user-facility">. This has the advantage of strongly identifying the semantics of the information and has the flexibility to allow for capture of both proposal and non-proposal facility activities, and a range of identifiers for the facility, from organization IDs to Research Resource IDs.

```
<research-support-group>
     <award-group award-type="approved-proposal">
          <research-support-source support-type="user-facility">
           <institution-wrap>
             <institution>Spallation Neutron Source</institution>
           </institution-wrap>
           <institution-wrap>
             <institution>Oak Ridge National Laboratory</institution>
             <institution-id institution-id-type="doi" vocab="open-funder-registry" vocab-identifier="10.13039/open_funder_registry">10.13039/100006225</institution-id>
           </institution-wrap>
             <research-support-description>Beam time and computing resources</research-support-description>
     </research-support-source>
       <principal-award-recipient>
         <name>Dr. Albert Einstein</name>
         <contrib-id contrib-id-type="orcid" authenticated="true">http://orcid.org/0000-0000-0000-0000</contrib-id>
       </principal-award-recipient>
     <award-id>SPS 12345</award-id>
   </award-group>
   <research-support-source support-type="research-materials">
      <research-resource-wrap>
        <research-resource>Slc9a4 (C05) tm1b Mus musculus</research-resource>
        <research-resource-id resource-id-type="rrid" vocab="Research Resource Identifier">RRID:IMSR_HAR:5669</resource-resource-id>
      </research-resource-wrap>
   </research-support-source>
 </research-support-group>
```

Discussion

Some JATS list members agreed with option 1, to keep user facility information within the JATS <funding-group>, but with specific tagging to better distinguish traditional funding from these other kinds of support. These could be divided into separate <funding-group> sections with an attribute distinguishing the types of support. Others supported Option 2, to go further to better define the container element, and add <research-support-group> to hold the activity - so long as this expansion was semantically separate from the existing funder tags; this would mean

publishers could continue to use the existing tags for monetary funding awards and would have flexibility in timeline of adoption of the new tags.

We need to think carefully about how to incorporate new research activities in a scalable way that allays concerns about an infinitely expanding list of tags. In support of Option 2, the new tags could facilitate tracking of many currently unexposed and highly important research activities – such as scientific collections – that need to be acknowledged. For example, USGS provides maps and USDA provides physical samples. The entities that fund/manage these collections want to be able to easily track publications arising from their use. List members suggested engaging funders in the discussion of what constitutes "funding", "support", "awards" and learning more about their existing metadata tracking.

Based on the WG discussion, Mark Doyle submitted option 2 as stated to the JATS working group for consideration. The comment can be found here.