HELIOS + PID Community Partnership

Version: 1.1
Updated 5/7/2023
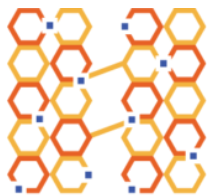
Prepared by:
- Zach Chandler (Stanford University) https://orcid.org/0000-0003-2402-9839

Contributors:
- Eric Olson (COS) https://orcid.org/0000-0002-5989-8244 ,
- Matthew Buys (DataCite) https://orcid.org/0000-0001-7234-3684 ,
- Anita Bandrowski (RRID) https://orcid.org/0000-0002-5497-0243 ,
- Yvonne Campfens (OA Switchboard) https://orcid.org/0000-0002-5406-5809 ,
- Shawna Sadler (ORCID) https://orcid.org/0000-0002-6103-5034 ,
- Chris Shillum (ORCID) https://orcid.org/0000-0002-1108-3660 ,
- Jennifer Kemp (Crossref) https://orcid.org/0000-0003-4086-3196 ,
- Amanda French (ROR) https://orcid.org/0000-0002-4325-1809 ,
- Maria Praetzellis (DMPTool) https://orcid.org/0000-0001-5047-3090 ,
- John Chodacki (California Digital Library) https://orcid.org/0000-0002-7378-2408

Contents

Overview

The purpose of this document is to highlight partnership opportunities in open scholarly communications infrastructure for the HELIOS community. The Higher Education Leadership Initiative for Open Scholarship (HELIOS) is a cohort of colleges and universities committed to collective action to advance open scholarship within and across their campuses. Leaders from US colleges and universities have joined this community of practice, working together to promote a more transparent, inclusive, and trustworthy research ecosystem.

HELIOS and its member constituents should support open infrastructure by partnering with persistent identifier (PID) service providers, most of whom are non-profit enterprises, with established community governance models.

Many HELIOS members already make use of these services; however, we can realize a higher order of benefit to open scholarship if we engage in collective action, and explore partnerships and integrations beyond using the basic services described below. As we consider various buy/build/partner possibilities, it is my hope that partnership becomes our default position, as no one wants to reinvent the wheel, and there is a lot of great work already happening in open scholarly infrastructure.

The Invest in Open (IOI) statement of support (https://investinopen.org/about/supporters/) provides a good foundation for understanding the economics at play in this space, and a jumping off point for examining how HELIOS can have an impact via joint investments (in funds, or in effort). For many non-profits, establishing a stable business model to sustain the organization past the initial years of grant funding is a challenge. For those that have distinguished themselves as reliable and trustworthy infrastructure providers, this could be an opportunity for HELIOS members to have an impact by shortening the time required to reach a sustainable model.

Below, we propose various collaboration opportunities with some of the key players in open scholarly communications infrastructure and persistent identifiers, as adoption is only the first step.

Potential Partners

(alphabetical order)

Crossref

https://www.crossref.org/

What they do:
Crossref is the largest DOI (Digital Object Identifier) registration agency for scholarly publishing, with over 17,000 global members (funders, publishers and publishing organizations) and 140+ million records. Journals, books and chapters, preprints, peer reviews, dissertations and theses, reports, standards, and grants can be registered and linked together as appropriate, for example for translations, publications linked to datasets. All records are openly available and widely used in a variety of systems, services and research efforts. Crossref works closely with other open research infrastructure organizations like DataCite and ORCID, including options for linking among records.

This Research Nexus vision of a network of relationships connecting research organizations, people, things, and actions is more than just PIDs and requires participation across the research community to properly reflect our shared work.

How universities can participate:
- University Presses, library publishers and OA publication outlets should consider improving discovery of their work and enhancing the scholarly record by registering and connecting outputs.
- Research administrators, libraries and others interested in tracking and describing outputs should consider using Crossref metadata to save time and improve accuracy and discoverability of publications and related outputs.

Business model: Crossref's fee principles are designed for sustainable participation. Membership fees are nominal for small outfits. Registration fees scale with the amount of records. Non-profit.
Contact: Jennifer Kemp, Head of Partnerships

DataCite

https://datacite.org/

What they do:
DataCite makes research more effective by connecting research outputs and resources–from data and preprint to images and samples. They support the creation and management of

metadata records including persistent identifiers, enhance research workflows with service integration, and enable the discovery and reuse of research outputs and resources.

They are a global community that share a common interest: To ensure that research outputs and resources are openly available and connected so that their reuse can advance knowledge across and between disciplines, now and in the future.

In particular, DataCite is well known as the primary provider of DOIs for research data published in any number of repositories (you may not be a member, but your repository of choice probably is). In addition to their role as a DOI registration agency, DataCite works on the underlying challenges of data citation best practices, from which HELIOS members can learn, as this is one of the biggest challenges in incentivizing data sharing. They maintain several web-based projects for making the current state and impact of data sharing more visible, notably DataCite Commons and the PID Graph.
https://commons.datacite.org/
https://blog.datacite.org/introducing-the-pid-graph/

Initiatives:
In line with their mission and vision, DataCite actively participates and leads various initiatives through collaboration with stakeholders in the community.
- Data metrics – they help further the adoption and implementation of responsible data metrics through leading the Make Data Count initiative.
- Identifier registries — they support a global, community-led registry of open persistent identifiers for research organizations, as one of the governing organizations for the Research Organization Registry (ROR).
- Repository discovery – they contribute to the re3data initiative through collaboration and financial support.

How universities can participate:
- Make research outputs and resources discoverable and citable by registering DOIs and metadata
- Enhance research workflows with service integration (most commonly with repositories or research information management systems), and consider using more of the possible attributes in DataCite metadata, e.g. explicit relationships between datasets
- Enable discovery with advanced tools and analytics (e.g. the PID Graph and DataCite Commons)
- Promote reuse with flexible tools and technology (e.g. Citation Formatter)

Business model: Sustainable through a cost-recovery membership model. Institutions can join the growing community of members and consortia across over 50 countries globally. Non-profit and signatory of POSI.
Contact: matt.buys@datacite.org

DMPTool

https://DMPTool.org/

What they do:
Managed and supported by the California Digital Library (https://cdlib.org), DMPTool provides templates for Data Management and Sharing Plans that meet the specifications of NIH and other federal agencies. Significantly, DOIs are minted for the plans themselves, and thus can be linked to the resultant datasets, and the ORCIDs of all team members. By minting DOIs, and persisting in an updateable online format, DMPTool has become the long-awaited "machine actionable DMP" described in best practices workshops such as this one:
https://www.arl.org/implementing-effective-data-practices/

How universities can participate:
- Start using DMPTool, adopt it as your preferred method of creating data management plans, and conduct outreach to your faculty and research teams. Encourage them to update their plans if circumstances change in their research, and edit the plan to include the DOI of the resultant dataset at award closeout.
- Appoint a local DMPTool admin at your university (you may already have one), and tune the forms and inline guidance to match your university policies, as well as federal requirements
- Connect with CDL to learn more about the future of the platform and how to get involved
- Encourage your researchers to attach their ORCID iD to their DMPs and to get DOIs for their DMPs

Business model: Non-profit, free to institutions and individual users.
Contact: Maria Praetzellis, maria.praetzellis@ucop.edu

OA Switchboard

https://www.oaswitchboard.org

What they do:
The OA Switchboard is a mission-driven, community-led initiative designed to simplify the sharing of information (metadata) between stakeholders about open access publications throughout the whole publication journey. It provides a standardized messaging protocol and shared infrastructure. It is built by and for the people who use it, and is leveraged with existing PIDs.

The three key stakeholder groups are **research funders, institutions and publishers** and the participants are part of a community, who believe the central information exchange hub grows stronger with every participant, becoming a better tool for all. The OA Switchboard simplifies many-to-many relationships through the power of an intermediary: efficiency and cost savings. It is complementary and supportive to the vendor systems offered to institutions and publishers. It is designed to operate and integrate with all stakeholder systems.

So far, two use cases are supported:

1. Open Access Reporting Made Easy
2. Matching Publication Costs with Publication Funds

How universities can participate:
- https://www.oaswitchboard.org/institutions (free trials possible)
- For research institutions, libraries and consortia who want to connect with their research and simplify their workflows: The OA Switchboard can be used by institutions to enhance their workflows, enable fulfillment of OA publication-level arrangements, make OA publishing highly visible, and simplify Open Access mandate compliance. Participation in the OA Switchboard will lead to enhancements in OA publication metadata and integration with the increasingly complex open access research and publishing ecosystem. It provides a safe space for publication metadata.
- Connect with your university librarian, or scholarly communications office, and their systems & data librarians to figure out how to take advantage of the OA Switchboard APi.
    - https://www.oaswitchboard.org/blog-post-12-september-2022
    - https://oaspa.org/wp-content/uploads/2022/10/OASWB_19_4_PPT_Oct_20_2022_YC_extended-poster-1.pdf
    - https://www.oaswitchboard.org/institutions

Business Model: The OA Switchboard operates on a self-sustaining business model, whereby the operational and development costs are supported by service fees (with transparent pricing) for participating funders, institutions and publishers. The fees are for development and maintenance of the infrastructure, and for participants to exchange information and communicate. Pricing is transparent and fees are kept as low as possible to offer an affordable solution. Fee structure is a sliding scale such that larger institutions pay more, smaller ones pay less (as little as $500/year)
Contact:

Open Science Framework (OSF)

https://osf.io/

What they do:
OSF, an open-source product of the Center for Open Science (COS - https://www.cos.io/), is a comprehensive management platform for researchers, including data and code sharing, integrations with 3rd party services, and specialized open science study design formats, such as pre-registration and registered reports.

OSF supports researchers throughout their entire project lifecycle. As a collaboration tool, OSF helps research teams work on projects privately or make the whole project publicly accessible for broad dissemination with DOIs for data, registrations, and preprints. As a workflow system, OSF enables connections to the many scientific tools researchers already use, streamlining their process and increasing efficiency.

How universities can participate:
- OSF is free for individual researchers, an institutional level view can help administrators see and share the data sharing activity of their communities across the platform
- A subscription fee enables single sign-on using verified institutional credentials, institutional affiliations on OSF content, aggregation of affiliated content on branded public interfaces, ROR IDs included in DOI metadata for affiliated content, insights into data sharing activity within the institution, and prioritized technical support

Business model: Non-profit, free for individuals. Moving to a subscription based model with modest annual fee (varies, but baseline is $2500/year), benefits include an institution-wide dashboard of activity on the platform.

Contact: Eric Olson, eric@cos.io

ORCID

https://orcid.org/

What they do:
ORCID is the canonical PID of record for individual researchers. Founded in 2012 to solve the problem of author disambiguation, ORCID helped define the persistent identifier space. Fully adopted by the publishing industry, ORCID is now included in two significant memos from the White House Office of Science and Technology Policy; NSPM-33 on Research Security and the Nelson Memo on Public Access to research. ORCID is the only PID that satisfies the requirements listed in OSTP's NSPM-33 implementation guidance document, and the recent Nelson memo includes ORCID in section 4.

Federal funding agencies are now citing ORCID in their policies, the NSF has officially listed ORCID as their persistent identifier for researchers (now published in the PAPPG 2023), and NASA requires its funded researchers to have an ORCID iD (recently published in the Scientific Information Policy for the Science Mission Directorate). Similar policies for ORCID adoption from the NIH and other funders are expected soon.

ORCID actively synchronizes data within the research community, specifically with universities, funders and publishers. There is a growing list of 3rd party integrations with ORCID often leveraging ORCID as a central record for scholarly activity. Paired with DOI, ORCID is the basis for establishing credit for data sharing.

How universities can participate:
- Join the ORCID-US Community - https://orcidus.lyrasis.org/
- Develop ORCID policies to normalize ORCID iDs with your researchers and staff
- Integrate ORCID into your local systems of record via the Member API. Systems include but not limited to,
  - Research Information Management systems
  - Electronic Research Administration systems
  - Human Resource systems
  - Repositories
  - Lab Scheduling and Management Systems
- Synchronize your administrative systems so university administrators receive timely data from the source - instead of asking researchers for this information
- Use ORCID data to track and measure the impact your researchers are making
- The more data in an ORCID record, the less time researchers have to spend re-entering the same data into forms like the Biosketch

There are greater expectations for researchers to disclose their professional research affiliations when applying for federal awards. Ideally, universities will write the following data to their researcher's ORCID records - this data in the ORCID record will then populate the relevant funding forms, including: Employment, Education (your graduating students), Visiting Scholars, Research Instruments, Funding, Awards, Service, Publications, etc.

Business model: Free to individuals, institutions can become members, which grants access to the Member API (and read/write access, 5 keys). ORCID recently achieved self-sufficiency via this model and no longer relies on grant funding. Membership ($5000) may already be covered by an existing consortium membership you have. If you exceed the initial 5 API keys bestowed with membership, additional keys maybe purchased (~$1k/ea)
Contact: Shawna Sadler **s.sadler@orcid.org**

ROR

https://ror.org/
https://api.ror.org/organizations

What they do:
ROR, the Research Organization Registry, is a global, community-led registry of open persistent identifiers for research organizations that launched in 2019 whose purpose is to make basic information about research organizations cleaner and easier to exchange among scholarly research systems – funder systems, university systems, publisher systems, etcetera.

As a community led and governed project whose records are in the public domain, ROR is preferable to competing systems that have restrictive licenses or are provided by for-profit

vendors. ROR's potential value scales with its adoption rate – it will likely soon be the preferred query parameter across a number of different systems to filter results on a per-institution basis. When a piece of research is granted a DOI by Crossref or DataCite, a ROR ID in the metadata helps link together all the research produced by researchers affiliated with that organization, as for instance in the ROR-powered research explorer for Virginia Tech at DataCite Commons.

How universities can participate:
- Look up your ROR record at https://ror.org/search
- If the information in your ROR record needs to be changed, submit a request
- Sign up for the ROR newsletter
- Email info@ror.org and ask to be added to the invitation list for bi-monthly ROR Community calls
- Check with university units / programs that self-publish research to see if their systems use ROR:
  - Research institutes
  - Institutional repositories and thesis repositories
  - Data repositories
  - Large laboratories
  - University presses

Business model: Non-profit, free service. ROR is not itself an organization and has no member fees, paid services, or premium technology tiers: it is supported by the operating budgets of Crossref, DataCite, and the California Digital Library, who consider ROR to be an essential part of scholarly communication infrastructure.

Contact: Amanda French - amanda@ror.org

RRIDs.org

Project pages: https://rrids.org
Get RRIDs: https://scicrunch.org/resources

What they do:
SciCrunch is the PID registrar for Research Resource IDs (RRIDs) an NIH funded project born in the UCSD FAIR Data Informatics Lab, now a stand-alone non-profit enterprise. RRIDs are used to specify which specific resources were used in a given study, such as antibodies, cell lines, organisms and other tools. which has a positive effect on rigor and reproducibility (improved identifiability: Bandrowski et al 2016; reduction in contaminated cell line usage: Babic et al, 2019). RRIDs are also used to differentiate research cores (aka "shared facilities") and improve citation of resources in journal articles. SciCrunch has a sophisticated data model, API, and both a human workflow as well as an AI crawler for extracting citations of RRIDs.

How universities can participate:
- Mint RRIDs for all research cores and shared facilities of note

- Encourage the use of RRIDs in labs for all cores, reagents, biological samples, etc.. including and especially in the Materials & Methods section of publications
- Use RRIDs to facilitate better impact reports and citation counts for resources
- Subscribe to keep tabs on what SciCrunch doing https://mailchi.mp/b1c222e2562a/rrid-update-15060661

Business model: free service, non-profit. There is a separate, for-profit entity, SciScore, which can generate university-wide rigor & reproducibility scores, leveraging data from RRID. This is a separate service, and one need not subscribe to use RRID.
Contact: Anita Bandrowski, Abandrowski@ucsd.edu (rii-help@scicrunch.org)

APIs and Integrations

Each of the above systems has its own Application Programming Interface (API), which means that integration possibilities abound. Part of what makes this ecosystem valuable is that they use each others' PIDs -- ORCID is used everywhere, DOIs are likewise ubiquitous, and ROR ids are increasingly being adopted in several of these projects. It's a community, and these orgs collaborate, to everyone's benefit.

Each API is an opportunity for HELIOS institutions to connect these systems with our own, either as standalone apps, or in complex integrations. We should explore the possibilities of each API independently, then consider them together. Some basic questions can help us ground this work:
- What data does this API provide?
- What are the allowed query parameters? (i.e. what fields can we filter on? Can we filter on institution(s) using ROR? Filter by individual using ORCID as a parameter? etc.)
- What systems that we control might benefit from integrations with these systems? What are the use cases that we are trying to solve?
- Are we interested in reading data out from these APIs? Or writing data into these systems? Or both?
- What is needed for API access? Is there a public API? A member API?

API Documentation

- Crossref : https://www.crossref.org/documentation/retrieve-metadata/rest-api/
- DataCite : https://support.datacite.org/docs/api
- DMPTool : https://github.com/CDLUC3/DMPTool/wiki/API-Overview
- OA Switchboard : https://bitbucket.org/oaswitchboard/api/src/master/
- OSF : https://developer.osf.io/
- ORCID : https://info.orcid.org/documentation/api-tutorials/
- ROR : https://ror.readme.io/docs/rest-api
- RRID : https://scicrunch.org/browse/api-docs/index.html?url=/swagger-docs/

Conclusion

All of the PID and infrastructure providers in this brief are community governed non-profits committed to open science, which is where HELIOS members should be investing our time and funds. Part of what makes these orgs trustworthy and successful is that they have community governance models, and avoid lock-in through open data and/or non-proprietary formats. Being non-profits also means that financial viability is an issue, especially once initial grant funding ends --they all need a path to becoming self-sustaining, which typically means fees for institutions, or for publishers. Many of them keep fees for individual researchers at or near zero.

HELIOS should consider how we might leverage collective action to support open infrastructure by leveraging persistent identifiers, and partnering with the organizations that support them. Some of the value proposition of using or integrating with persistent identifiers is non-obvious, and we should also partner with these service providers by exploring possible collaborations with them in an open-ended way.

Regarding the **Build vs Buy vs Partner** calculus, we should not think of these as mutually exclusive options. Our primary mode in open science should be to **Partner** -- and with the above service providers in particular. However, this belies the fact that some funding is required to partner, as is effort. All HELIOS members should do the work internally to marshal resources to cover subscription fees that will allow non-profit PID providers to exist, develop, and grow with us (so there is still a **Buy** component). And APIs are only one half of a handshake. To realize the full value of these services, there may be local app development and IT resource planning that we need to do to build meaningful integrations (so there is still a **Build** component).

There will be plenty of challenges ahead in building out the open science future, but we have lots of like minded people around us if we look, and we are better together.

//ZC

Active development of this work continues online in a separate V2 document, with additional co-authors and perspectives. Those interested in engaging with that effort can reach out to HELIOS (https://www.heliosopen.org) or the author directly (zchandler@stanford.edu).