Maria Gould

People, places, and things

Persistent identifiers in the scholarly communication landscape

Persistent identifiers are a key element in the scholarly communication landscape. However, many scholarly communication librarians may not be familiar with persistent identifiers and what to do with them. I should know—I used to be one of these librarians.

When I worked as a scholarly communication librarian, I knew a few basic things about persistent identifiers. I knew what an ORCID iD was and I could explain to researchers why they should have one. I knew what a DOI was, but I did not know that there are different types of DOIs for different types of things, or that there's more to a DOI than the identifier itself. But I didn't know there were other persistent identifiers relevant to scholarly communication, like identifiers for affiliations or funders or grants. And I certainly didn't know that persistent identifiers can be commercialized and paywalled, just like research outputs.

Then I switched jobs and took on a portfolio focused exclusively on persistent identifier services for the University of California library system and beyond. This opened my perspective on scholarly communication and the infrastructure that underpins it.

As I became situated in my new role, I kept thinking about how I wish I had known more about persistent identifiers when I was a scholarly communication librarian. This realization motivated me to try to bridge the identifier space and the scholarly communication space. Sharing some knowledge in this column is one attempt to do this.

Background

The world of persistent identifiers is vast and diverse. The wider landscape is already well covered elsewhere.[1] My goal in writing this column is to focus on describing a few types of open identifiers that scholarly communication librarians are likely to encounter in their work, share useful things about these identifiers, and provide tips for leveraging their potential.

There are a few overarching concepts that I want to emphasize in this article:

- Identifiers can have many purposes
- Identifiers are not an end in and of themselves
- Identifiers work better together
- Identifiers work better when they are open
- You do not need to know everything about identifiers

These concepts underscore that identifiers are not homogenous, that the purpose of identifiers goes beyond mere identification, that the power of identifiers comes from connecting

Maria Gould is product manager/research data specialist at the California Digital Library, email: maria.gould@ucop.edu.

them in and through open metadata and open scholarly infrastructure, and that one of the benefits of identifiers is that they can operate behind the scenes without our intervention.

Common identifiers in scholarly communication

What we might think of as the "built environment" of the scholarly communication landscape comprises a set of fundamental components that enable scholarship to be created, disseminated, and accessed. Of course, scholarly communication entails much more than this, but for the purpose of this column, my focus is on these core infrastructural functions.

Persistent identifiers have been developed to uniquely represent, provide long-term access to, and connect many of these components. They identify such entities as research contributors, outputs, organizations and facilities, instruments and materials, publishers and repositories, and funders and awards. We might think of these entities in terms of three general categories: people, places, and things.

Persistent identifiers for these components are valuable for several reasons: they facilitate disambiguation, they enable discovery and tracking of research, and they establish connections that help us to understand the contexts and relationships in which research is being produced and consumed. As discussed in more detail below, these identifiers are created and managed in different ways, perform different functions, and may be part of services and infrastructure that entail varying costs and degrees of openness. In this article, my focus is on open identifiers (available for free or as part of openly available infrastructure).

While the identifying function that they provide is important, identifiers do much more. An identifier on its own can actually do very little. A standalone string of numbers and characters will not convey any meaning or insights about the object it represents. It is the metadata associated with the identifier, and the connections that identifiers can enable, that make them meaningful and powerful.

Understanding identifiers in context

Let's take a deeper dive into a few of these identifiers for people, places, and things to illustrate how they work and why they are important.

ORCID is a good place to start.[2] An ORCID iD is a unique identifier for researchers and research contributors. ORCID iDs help to disambiguate individuals with the same name or similar names and provide a stable reference point for names that might change over time. This is important to ensure that researchers receive credit for their contributions and for these contributions to be discoverable and citable. ORCID also helps to streamline research workflows and save researchers' time by minimizing the amount of information that needs to be entered multiple times across systems.[3]

Published research outputs are typically associated with DOIs, or digital object identifiers. DOIs can be registered for many content types, including journal articles, monographs, preprints, dissertations, and datasets. Crossref and DataCite are the primary registration agencies for DOIs.[4] DOIs are stable reference points that remain the same even if the published location of a research output changes. This helps to maintain long-term access to the scholarly record. DOIs can also be registered for entities that go beyond publications. For example, Crossref also supports DOIs for funding organizations and for grants, and DataCite supports DOIs for data management plans.

While ORCID iDs for people and DOIs for outputs have been relatively well established for a number of years now, there has not been an equivalently developed open identifier for research organizations. This changed with the recent launch of the Research Organization Registry (ROR).[5] ROR IDs uniquely identify organizations even as they change their names and go through other transformations over time. This makes it easier to track research outputs by institution.

ORCID iDs for researchers, DOIs for research outputs, and ROR IDs for research organizations collectively exemplify the utility of persistent identifiers for scholarly communications more generally. Given the complexity of information that populates the scholarly communications landscape, persistent identifiers help to disambiguate this information, facilitate more efficient discovery and tracking, and enable long-term access.

These and other identifiers are most powerful when they are used together and when they include interoperable metadata.[6] A ROR ID on its own cannot do much besides identify an institution, but when a ROR ID is included in metadata for a DOI, it can be easier to find published research associated with the corresponding institution. When a ROR ID is part of the metadata included in a researcher's ORCID record, institutions can more easily identify their researchers and track their research outputs.

Untangling complexities

The examples being discussed here—ORCID iDs, DOIs, and ROR IDs—also reflect some dimensions and nuances of persistent identifiers that might not be obvious or well understood, resulting in confusion or lack of clarity.

One point of confusion that I have observed is about how identifiers are created or obtained, and by whom. This is understandable because not all identifiers are the same in this regard.

For example, any individual researcher can register for an ORCID iD and retain control over the information displayed in their ORCID records.[7] The process of creating a DOI is different. DOIs must be registered via a registration agency, which requires being an institutional member of Crossref or DataCite, or they must be publishing work via a platform that automatically registers DOIs, such as Zenodo or most journal publishing systems. DOI registrations also require following policy requirements and best practices for the types of objects that can be associated with a DOI and the metadata used to describe these objects. ROR IDs are not created by individuals or institutions themselves, but instead added to the ROR registry through a community-based curation process.[8]

Another point of confusion about identifiers has to do with costs. The ORCID, DOI, and ROR examples represent some important nuances and distinctions in this area as well. ORCID, Crossref, DataCite, and ROR all make their metadata openly available via APIs and public data files. While an individual can obtain an ORCID iD at no cost, ORCID has institutional membership options that offer enhanced access to ORCID services. Crossref and DataCite's global membership bodies consist largely of publishers, repositories, and libraries. In the case of an individual wanting or needing to obtain a DOI for a research output, this would typically happen by publishing with an existing Crossref or DataCite member (such as a journal or repository) or by self-depositing work on a platform that registers DOIs. Lastly, ROR is not a membership organization and does not charge fees for the creation or use of ROR IDs.

In all these cases, the question of costs is more complex than meets the eye. Even if open scholarly infrastructure is free to access and use, it still costs money to build and maintain.

This is an opportunity for libraries to make strategic investments in infrastructure that can disseminate open access to knowledge at scale.[9] The Principles of Open Scholarly Infrastructure (POSI) provide guidance for research stakeholders and infrastructure providers about building and supporting open and sustainable scholarly infrastructure.[10]

A third and common misconception about identifiers is that identification is the end goal or only goal. Of course, identification is important and necessary, but it is also essential to think about identifiers as enablers of connections. These connections happen through open and interoperable metadata. When a research output is published and a DOI is registered for this output, the metadata associated with the DOI can include other identifiers that make it possible to realize a wealth of insights about scholarship and make administrative tasks more efficient. ORCID iDs in DOI metadata allow for auto-population of research works in an ORCID record, obviating the need for a researcher to enter this information manually. Metadata in a DOI about related works can enable discovery of preprints that preceded an article, datasets cited in the article, or translated versions that can be picked up by a broader audience. Including ROR IDs for affiliations in DOI metadata can allow research administrators and funders to track research associated with a specific institution or funding body. The implementation of identifiers in scholarly infrastructure depends on this rich metadata to power connections and insights, and when this metadata is open and interoperable, it makes scholarly communications more sustainable and resilient.[11]

Putting identifiers into practice

Scholarly communication librarians can benefit from a general understanding of persistent identifiers and how they can be implemented in institutional contexts. However, becoming an expert in every type of persistent identifier is not necessary or worthwhile, as in many cases identifiers can operate invisibly in the background to streamline workflows and connect information.

Scholarly communication librarians can play a key role in supporting the implementation of open and interoperable persistent identifiers.[12] One way to lead by example is to sign up for an ORCID iD if you don't already have one. A deeper way to engage with identifiers could mean encouraging your institution to join a membership organization to register identifiers. Advocating for identifiers can also involve paying closer attention to the types of identifiers available in the tools and services that are being purchased and licensed, and prioritizing options that allow for open, reusable, and interoperable metadata.

Persistent identifiers and the infrastructure around them can be an exciting opportunity for developing new technical skills and engaging professionally in open initiatives. This could mean learning how to work with scholarly APIs, navigating open indexing platforms like OpenAlex,[13] or joining community groups to discuss open infrastructure projects. I might have missed these opportunities in my prior role, but it may not be too late for you! ⁊⁊

Notes

1. Some general overviews of common persistent identifiers for scholarly communication include Herbert Van de Sompel, Robert Sanderson, Harihar Shankar, and Martin Klein, "Persistent Identifiers for Scholarly Assets and the Web: The Need for an Unambiguous Mapping," *International Journal of Digital Curation* 9, no. 1 (July 23, 2014): 331–42, https://doi

.org/10.2218/ijdc.v9i1.320; Alice Meadows, Laurel L. Haak, and Josh Brown, "Persistent Identifiers: The Building Blocks of the Research Information Infrastructure," *Insights the UKSG Journal* 32 (March 13, 2019): 9, https://doi.org/10.1629/uksg.457; Frances Madden, René van Horik, Stephanie van de Sandt, Artemis Lavasa, and Helena Cousijn, "Guides to Choosing Persistent Identifiers," May 28, 2020, https://doi.org/10.5281/zenodo.3862655; Maria Gould and Maria Praetzellis, "Open Persistent Identifiers: The Building Blocks of Sustainable Scholarly Infrastructure," *Research Library Issues* 302 (2021): 5–18, https://doi.org/10.29242/rli.301.2.

2. ORCID homepage, https://orcid.org/.

3. Laurel L. Haak, Martin Fenner, Laura Paglione, Ed Pentz, and Howard Ratner, "ORCID: A System to Uniquely Identify Researchers," *Learned Publishing* 25, no. 4 (October 1, 2012): 259–64, https://doi.org/10.1087/20120404.

4. Crossref homepage, https://www.crossref.org/; DataCite hompage, https://datacite.org/.

5. Maria Gould, "Hear Us ROR: Announcing Our First Prototype and Next Steps," *ROR Blog*, February 2, 2019, https://ror.org/blog/2019-02-10-announcing-first-ror-prototype/.

6. Helena Cousijn et al., "Connected Research: The Potential of the PID Graph," *Patterns* 2, no. 1 (2021): 100180, https://doi.org/10.1016/j.patter.2020.100180.

7. "ORCID + Researchers," https://info.orcid.org/researchers/.

8. "ROR Updates," Github repository, https://github.com/ror-community/ror-updates.

9. Maria Gould and John Chodacki, "Pathways to Open Access: Open Infrastructure and CDL," *Pathways to Open Access* (blog), Office of Scholarly Communication, University of California, August 18, 2022, https://osc.universityofcalifornia.edu/2022/08/pathways-to-oa-open-infrastructure/.

10. Geoffrey Bilder, Jennifer Lin, and Cameron Neylon, "The Principles of Open Scholarly Infrastructure," 2020, https://doi.org/10.24343/C34W2H.

11. Helena Cousijn, Ginny Hendricks, and Alice Meadows, "Why Openness Makes Research Infrastructure Resilient," *Learned Publishing* 34, no. 1 (2021): 71–75, https://doi.org/10.1002/leap.1361.

12. For more specific guidance on how institutional stakeholders can take advantage of persistent identifiers, see John Chodacki Cynthia Hudson-Vitale, Natalie Meyers, Jennifer Muilenburg, Maria Praetzellis, Kacy Redd, Judy Ruttenberg, Katie Steen, Joel Cutcher-Gershenfeld, and Maria Gould, "Implementing Effective Data Practices: Stakeholder Recommendations for Collaborative Research Support," Association of Research Libraries, September 23, 2020, https://doi.org/10.29242/report.effectivedatapractices2020.

13. OpenAlex homepage, https://openalex.org/.