Information Services & Use 34 (2014) 93–96
DOI 10.3233/ISU-140736
IOS Press

Persistent identifiers can improve provenance and attribution and encourage sharing of research results

Laurel L. Haak
ORCID, 10411 Motor City Drive, Suite 750, Bethesda, MD 20817, USA
Tel.: +1 301 922 9062; E-mail: L.Haak@orcid.org

Abstract. ORCID is an organization that provides a registry of unique and persistent identifiers for researchers. In the publishing community, ORCID identifiers are used for authors and reviewers and enable persistent and trusted digital connections between researchers and their work. This article provides an overview of the uptake of ORCID identifiers and how they are being used to clarify author attribution and improve provenance.

Keywords: ORCID, persistent identifiers: open science, data attribution, authorship, attribution

Open science is all about sharing. With access to findings, data, and knowledgeable researchers, we can build upon today's research and create new knowledge and new opportunities. One of the largest incentives to sharing is attribution. Especially in research, where we aren't in it for the remuneration, ensuring that individuals get appropriate credit for their contributions underpins open science.

In place are processes and data systems to support authorship and presentation of works and to support awards and track citations. However, anyone who uses these systems knows that they are incomplete. Authorship acknowledgement has been based solely on a person's name, which may be shared with many others, misspelled, change over time, or be expressed in a culturally specific manner. Co-authorship is growing, but in many publication systems co-authors are not verified. "Author" is increasingly an archaic term: more and more people want to know what it is a person has contributed to a project and not just that they have authored a paper about the findings. Then, there are questions about the organizations that supported the work. What equipment or facilities were used? Where was someone employed and was there a public–private partnership? Who funded the work? And, what do we know about the data? How can we enable access to the data?

Clearly there is a fundamental need for naming protocols and methods to link together names of contributors, contributions, and supporting organizations and facilities. With these naming protocols in place we can start to develop authoritative attribution processes. And when I know I will get credit for my work, I am more likely to share it.

ORCID is working specifically on person naming protocols. ORCID provides a registry where individuals may obtain a unique and persistent identifier. Individuals may use that identifier in a number of research and scholarly workflows.

A researcher may include their ORCID identifier when they write a data management plan, deposit a dataset into a repository, or when they access a dataset for analysis purposes. This has come about because of the work over the last year by DataCite and the repository community in building ORCID

fields into metadata data schemas and integrating ORCID into data workflows [2–4]. Figshare takes this a step farther and provides a formatted reference to use to cite the data set [8], and in collaboration with the metrics community, has started to provide use and re-use information for datasets [14]. Together, these components provide machine-readable metadata that supports sharing, access, and rapid feedback on use and re-use.

Journal articles are a time-tested venue for sharing research results. ORCID identifiers are being integrated into the metadata schemas and submission workflows to ensure accurate attribution [9]. In addition, publishers are starting to ask reviewers for their ORCID identifier – and use ORCID linkages to qualify reviewers. We have been working with CASRAI to define a metadata standard for citing a peer review activity [10], and when the specification is completed later this summer, will be working to support the acknowledgement of these activities that are so important for ensuring quality and accountability of published research findings.

We have found some interesting challenges – and opportunities – for persistent identifiers in these data and article publication processes. One is validating co-authors. This process is largely managed by the corresponding author, who provides the co-author names, affiliations and email addresses. The co-author list can change during the review process, and in some cases authors are listed on an article without being contacted for their approval. This becomes an issue when the results of a paper are contested, and leads into another challenge, that of clearly identifying author roles and contributions.

The recent effort by Allen et al. [1] indicates that implementing a workflow for identifying author roles is practical, and a standard field set may be available for publishers to integrate in the near term, following a consultation with NISO and CASRAI [6]. Further, tools are being developed to link a person to their specific data contribution and insert it into a publication, making it possible to demonstrate a person's contribution but also to provide an actionable data element that supports interaction within the context of a journal article. Now, imagine combining the times that paper is cited with the times a chart or graph was downloaded and/or cited? That combines access and incentives to share data, and could provide a major assist to the open science movement.

Facilities managers are contacting ORCID. They are interested in tying together funded facilities use proposals with the data sets created and articles written by the researchers using the equipment. At present this is a largely manual task dependent on authors acknowledging the facility in the dataset or article, often in a free-text note buried in a manuscript. Here there is an opportunity to use persistent identifiers to link the contributor (ORCID) with their data or article (DOI) with a unique identifier for the facility. This will require an adjustment in what metadata are collected at the time of manuscript submission or data deposit, and journal editors can help to socialize this change.

How such linkages with organization identifiers can potentially transform the ways in which research is attributed is exemplified in two existing use cases: graduate student validation and research policy. Organizations seeking to validate an educational degree would usually ask for a transcript and a copy of a person's diploma. Now, we are seeing universities integrate identifiers into the thesis and dissertation defense process [5]. The thesis itself is assigned a DOI, the student includes their ORCID identifier when submitting the thesis, and the awarding institution posts the thesis metadata with their Ringgold organization identifier to the student's ORCID record. That binds together the student, the organization, and the thesis in a validated electronic record that the student can easily share with the human resources system at their next career stop.

Research funders are also interested in learning about connections between researchers, works, and organizations – and in particular the organizations that fund the research. Understanding these connections helps the funder evaluate the impact of their funding programs. Determining these connections is a

manual process, meaning that evaluations are usually performed sparingly with several years' hindsight and results often come in long after it is possible to adjust program features to best serve the community. We are starting to see funders integrate ORCID identifiers into grant application workflows, grantee profile records, and over time funders have indicated a willingness to pull information from ORCID records to support components of the post-award reporting process [7,11]. Universities are discussing ways to integrate ORCID identifiers into the grant workflow process, providing a more seamless process for the applicant and means to track applications, involved researcher, awards and research outputs – a key goal of the UMetrics project [12]. These linkages support data-driven research policy, reduce the time researchers spend responding to evaluation queries, and free up time to work on the project itself rather than the paperwork that currently engulfs grants.

These examples show how combining persistent identifiers for the researcher, their contributions, and the related facilities and universities can enable us to determine provenance. For each persistent identifier linked to an ORCID record, there is explanatory metadata, information on source, and whether the connection is a self-claim or third-party claim [13]. This information is available through the ORCID user interface and through our APIs, so anyone consuming the information can make a trust determination based on their particular needs. As we work with the community to roll out workflows to round-trip information from the point of publication/award/deposit to an individual's ORCID record, with improved provenance will come improved attribution, incentive to share a wide variety of research results.

The technology to support this vision is fairly straightforward. The identifiers exist. Numerous demonstrations of the capability and benefits of linkages between identifiers exist. Now, the community needs to come together in a concerted effort to integrate the identifiers into systems and workflows to allow researchers to share, and all of us to benefit.

References

[1] L. Allen, J. Scott, A. Brand, M. Hlava and M. Altman, Publishing: Credit where credit is due, *Nature* **508**(7496) (2014), 312–313.

[2] R. Bryant, Connecting research datasets and researchers, ORCID Blog Post, 17 June, 2013, available at: http://orcid.org/blog/2013/06/17/connecting-research-datasets-and-researchers.

[3] R. Bryant, ORCID, repositories, and researchers, ORCID Blog Post, 11 March 2014, available at: http://orcid.org/blog/2014/03/10/orcid-repositories-and-researchers.

[4] R. Bryant, ORCID supports interoperable exchange of datasets, ORCID Blog Post, 27 September 2013, available at: http://orcid.org/blog/2013/09/24/orcid-supports-interoperable-exchange-datasets.

[5] R. Bryant, Linking ORCID identifiers to theses and dissertations, ORCID Blog Post, 30 March 2014, available at: https://orcid.org/blog/2014/03/20/theses-dissertations-orcid.

[6] T.A. Carpenter, When a scholar is one among 500, what does it mean to be an author?, *The Scholarly Kitchen*, 22 April 2014, available at: http://scholarlykitchen.sspnet.org/2014/04/22/when-a-scholar-is-one-among-500-what-does-it-mean-to-be-an-author/.

[7] P. de Castro, Building pioneering functionality around ORCID integration: FCT and Portugal, GrandIR Blog, 18 February 2014, available at: http://grandirblog.blogspot.com/2014/02/building-pioneering-functionality.html.

[8] figshare ORCID integration, Blog post 86, available at: http://figshare.com/blog/figshare_ORCID_integration/86.

[9] L. Haak, Suggested practices for collection and display of ORCID iDs in publishing workflows, ORCID Blog Post, 8 August 2013, available at: http://orcid.org/blog/2013/08/08/suggested-practices-collection-and-display-orcid-ids-publishing-workflows.

[10] L. Haak, ORCID and CASRAI: Acknowledging peer review activities, ORCID Blog Post, 8 April 2014, available at: http://orcid.org/blog/2014/04/08/orcid-and-casrai-acknowledging-peer-review-activities.

[11] L. Haak, Link your ORCID record to your funding, ORCID Blog Post, 19 February 2014, available at: http://orcid.org/blog/2014/02/19/link-your-orcid-record-your-funding.

[12] L. Haak, Science of science policy and the need for data standards and persistent identifiers, ORCID Blog Post, 30 September 2013, available at: http://orcid.org/blog/2013/09/30/science-science-policy-need-data-standards-and-persistent-identifiers.

[13] L. Paglione, Plans for enabling multiple assertions in the ORCID Registry, ORCID Blog Post, 2 June 2014, available at: http://orcid.org/blog/2014/06/02/working-group-recommendations-multiple-assertions.
[14] H. Piwowar and J. Priem, Link your figshare and ImpactStory accounts, ImpactStory Blog, 2 December 2013, available at: http://blog.impactstory.org/link-your-figshare-and-impactstory-accounts/.