Learned Publishing, 25: 259–264
doi:10.1087/20120404

ORCID: a system to uniquely identify researchers

Laurel L. HAAK[a], Martin FENNER[a], Laura PAGLIONE[a], Ed PENTZ[a,b] and Howard RATNER[a,c]

[a]*ORCID Inc.*
[b]*CrossRef*
[c]*Nature Publishing Group*

ABSTRACT. The Open Researcher & Contributor ID (ORCID) registry presents a unique opportunity to solve the problem of author name ambiguity. At its core the value of the ORCID registry is that it crosses disciplines, organizations, and countries, linking ORCID with both existing identifier schemes as well as publications and other research activities. By supporting linkages across multiple datasets – clinical trials, publications, patents, datasets – such a registry becomes a switchboard for researchers and publishers alike in managing the dissemination of research findings. We describe use cases for embedding ORCID identifiers in manuscript submission workflows, prior work searches, manuscript citations, and repository deposition. We make recommendations for storing and displaying ORCID identifiers in publication metadata to include ORCID identifiers, with CrossRef integration as a specific example. Finally, we provide an overview of ORCID membership and integration tools and resources.


Laurel L. Haak


Martin Fenner


Laura Paglione


Ed Pentz


Howard Ratner

The ORCID initiative

ORCID (Open Researcher & Contributor ID)[1] is an international, interdisciplinary, open and not-for-profit organization created to solve the researcher name ambiguity problem for the benefit of all stakeholders, including research institutions, funding organizations, publishers, and researchers themselves. The core mission of ORCID is to provide a registry of persistent unique identifiers for researchers and scholars.

Working with stakeholders to embed these identifiers in research workflows, including manuscript submission, will support timely and complete attribution by automating the contributor–research linkage. In turn, the ORCID registry can serve an important role in supporting efforts in the publishing community including conflict-of-interest reporting and author role acknowledgement.

While several author identifier initiatives exist already, they are limited by organization, discipline, or geographic region – or they are part of a proprietary system. However, researchers increasingly work across disciplines and institutions, and are geographically mobile. ORCID is designed for the researcher community: the organization works across all of these boundaries to provide a registry for individuals or their organizations to create identifiers and manage ORCID records. In its first phase, ORCID will provide a self-claim system that allows individuals fine-grained control of privacy settings, as well as data exchange with grant and manuscript submission systems and other identification systems such as Scopus,[2] RePEc,[3] ResearcherID,[4] and VIVO.[5]

Use of an identifier must have clear benefits, otherwise it will not be adopted. In discussions with many stakeholders it became clear that simply providing a unique

identifier would not be enough, and that the ORCID identifier needed to be integrated into research workflows and linked to information on research activities such as publications, grants, patents, and datasets.

ORCID has taken the stance that the use of an identifier should first and foremost reduce the reporting burden for researchers, both in the immediate task of filling out basic information on forms, as well as in longer-term progress reporting. A good example for this is the manuscript submission process, for which much of the identifying information requested from authors (name, affiliation, email address) could be retrieved from the ORCID registry. System-to-system authentication also provides a researcher with the option to create a 'trusted' relationship with a publisher, so that when a manuscript is accepted the publisher can update the researcher's ORCID record with publication metadata. Other systems can then use the data from the ORCID record to maintain, for example, university profile databases and local digital research repositories, or to support grant progress reporting at funding agencies.

ORCID also recognizes that, first and foremost, individuals own their record

ORCID also recognizes that, first and foremost, individuals own their record. A central principle of the ORCID initiative is that researchers control the defined privacy settings of their own ORCID record data. Individual record holders can control what information is displayed publicly, what is shared with trusted partners, and who those trusted partners are. Furthermore, ORCID does not collect sensitive information. The only information required to register for an identifier is name and email address, and only the ORCID identifier is always publicly available. All other information in the registry can be marked as non-public.

Pain points for publishers

Similar to researchers, publishers are affected by name ambiguity in direct and indirect ways. On the direct path are author databases, which for many journals are collections of duplicate records requiring a substantial investment to disambiguate and manage. This in turn has an impact on the ability to understand an author's history, perform accurate name-based searches, and find and manage reviewers. Another pain point is the processing of citation metadata, which without author identifiers requires manual disambiguation to match authors and articles.

Perhaps a little less directly, publishers face issues with authorship roles on a daily basis. Who should be listed as an author? How can an author role be appropriately acknowledged? How can publishers discern author responsibility? Linked to this is conflict-of-interest reporting. Who needs to report what, and in what context? Clearly, there is a role for a central registry that crosses disciplines, work places, sectors, and national boundaries. By supporting linkages across multiple datasets – clinical trials, publications, patents, datasets – such a registry becomes a switchboard for researchers and publishers alike in managing the dissemination of research findings.

Use cases

Unique author identifiers are the only way we can address these issues. To be effective identifiers will need to be incorporated in several publisher workflows. Perhaps the most straightforward is the *manuscript submission system*, where an ORCID identifier can be collected from the corresponding author at the time of submission. Associating a unique identifier with an author reduces duplicate author accounts and enables publishers to provide a more accurate representation of an author's prior work and citations. If additional data are linked with the identifier, association with the author profile may assist the author in filling out the submission form, including affiliation information. ORCID identifiers for co-authors are preferably collected after the manuscript is accepted for publication. In this scenario journals could ask all co-authors to enter their ORCID identifiers via authentication with the ORCID service, together with the information on author contributions and potential conflicts of interest. At this stage authors could also agree to a trusted party relationship with the publisher for subsequent ORCID record updates.

In addition to manuscript submission, another use case for publishers is *searching for prior work*, to support vetting of reviewers, creation of author profiles for authors, or to assist in the processing of citations, not all of which might be using DOIs or other unique identifiers. Searching by name is fraught, in particular because authors and/or the journals they publish in use multiple name forms for the same person. Any search needs to incorporate these multiple forms. An identifier linked with an author name would eliminate this ambiguity, and allow not just for multiple variants of the same name but also continuity along a career for those authors who have changed their names. Many publishers have started to build value-added services for their authors, and professional societies that are also publishers are linking author and membership information. Publishers are also connecting author and reviewer databases. ORCID identifiers will greatly facilitate the creation and maintenance of these author profiles, in particular the disambiguation of authors and the linking to external information.

Including ORCID identifiers in *manuscript citations* would greatly improve citation accuracy with regards to author names, which in turn would assist authors and their institutions in managing publication lists.

Many authors are asked to deposit their accepted manuscripts into *institutional and/or discipline-specific repositories*. In some cases, the publisher handles this process; in other cases, it is the author's responsibility to deposit, or a library's task to find articles produced by university faculty. An ORCID identifier will facilitate local deposition, in particular if done by the library on behalf of their researchers. For those repositories linked to research funding, an identifier is a key component of efforts to link funding with research outputs, and would assist in a number of research reporting efforts currently underway.

Other use cases for publishers include managing login credentials, cross-journal transfer, improving the collection of disclosure information from authors, and management of open access fees and permissions.

Best practices

For any identifier to be effective, it must be widely adopted by the research community – not only among individuals but also at the touch points where research findings are disseminated: publication submission, dataset deposition, research grant and contract applications, faculty and staff profiles, patent applications, etc.

How an identifier is collected at these touch points is critical. Providing a field for a researcher to type in their identifier may create more problems than it solves, in particular due to typographical errors, lack of validation, and inability to update with current data. A more effective method of collection would employ authentication and storage in an existing touch point 'profile' such as the author profile, where the identifier could be reused each time the author submits a manuscript. This process would support disambiguation and maintenance of author databases, and could also be used to support unified sign-on, another method to reduce the likelihood of duplicate profile creation.

for any identifier to be effective, it must be widely adopted by the research community

For longer-term benefits to be realized, publishers will need not only to collect the ORCID identifier but also to store it (and the ID type) with the paper's metadata, deposit it with the paper in various systems, and determine how best to display that metadata in versions of the manuscript. ORCID uses a semantically opaque identifier[6] – meaning that is not possible to deduce the name or other identifying information from the identifier. It is therefore important to provide a visual display of the ORCID in association with the author name. Best practices for publishers and other organizations on how to include and display ORCID identifiers are still evolving, but based on discussions with publishers and repository managers we recommend including ORCID identifiers at least in the following scenarios:

1. as a footnote or in-line in the HTML and PDF versions of published manuscripts;
2. in the article metadata used on journal websites;
3. in the article metadata sent to CrossRef and bibliographic databases such as PubMed;

4. in downloadable reference lists using the RIS, BibTeX or Endnote format.

Similar recommendations apply to the use and display of ORCID identifiers in research datasets and grant applications. The ultimate goal for ORCID is to have the ORCID unique identifier used whenever a scholarly contribution is made or reported. This not only includes journal articles and books, but also conference abstracts, research datasets, scientific presentations, and other scholarly contributions. Scholarly publishers will play a central role in the adoption of the ORCID identifier.

Publishers use a number of different DTDs. Ideally, the ORCID identifier should be associated with the author name field, so that each author has an identifier and identifier type specified in a <IDtype> format. Some organizations have already made recommendations on how to include ORCID identifiers in the documents they provide. The DataCite Metadata schema[7] describes the core metadata properties for datasets using DataCite DOIs and includes fields for the ORCID identifier (<nameIdentifier>, <nameIdentifierScheme>). The Journal Article Tag Suite (JATS),[8] a NISO draft standard for standardized markup for journal articles based on the NLM DTDs developed at the National Library of Medicine, uses the <contrib> element to identify authors and other contributors to a work and allows the <ext-link> element to store identifier and identifier type inside of it. The @ext-link-type attribute can be used to give an indication of the type of resource to which the external link points. The JATS tag set does not constrain the values of this attribute, and ordinary text is acceptable. Thus, to specify ORCID, a value of <ext-link ext-link-type="orcid"> would be used.

with the launch of ORCID, researchers will be able to retrospectively search CrossRef metadata for their own publications

ORCID/CrossRef integration

From the start, the CrossRef and ORCID systems have been envisioned as complementary infrastructures for uniquely identifying researchers and enabling researchers to connect with their publications. CrossRef's current system for recording author names makes no effort to normalize or disambiguate names. This makes author-name based queries on the system impossible to do with any meaningful reliability. CrossRef members would like to see ORCIDs deposited along with CrossRef author metadata to improve the querying system. CrossRef would need to make minimal changes in its data structures and API extensions to support ORCID deposits and publishers would need to ensure that ORCIDs were gathered at manuscript submission. By submitting ORCIDs to CrossRef with publication metadata, the publisher would be verifying the linkage between author and publication necessary for updating an author's ORCID record.

With the launch of ORCID, researchers will be able to retrospectively search CrossRef metadata for their own publications and add them to their ORCID record. Once ORCIDs start being deposited with CrossRef metadata, CrossRef will be in a position to help ORCID 'push' updates to researchers every time a new publication is deposited with ORCIDs into CrossRef. At launch, researchers also will be able to link their ORCID record to other external IDs, and thereby import past publication and information into their ORCID record. At a later date, ORCID will support importing of additional information types through this linking method.

These kinds of 'pushed' updates from trusted sources such as CrossRef will eventually play an important role in the ORCID system. In essence, a CrossRef pushed update which is then claimed by the researcher indicates two parties (the publisher and the researcher) agree that the publication DOI and the ORCID are related to each other. Contrast this with the use-case of a researcher manually importing or entering a publication into their ORCID record. In this latter case, the 'publisher' has played no role in ensuring that that particular ORCID was associated with the respective DOI. In short, this would be a 'pure self claim' as opposed to a 'verified' claim. In addition to the claims by researchers and publishers, claims by other partners – such as the institution of the author or the funder who paid for the research – will further increase the trust in the claims made in

the ORCID registry. This community review and validation process also offers a level of assurance that the data in ORCID are accurate. More details on how ORCID will deal with different claims about the same piece of work can be found in a white paper.[9] Another important benefit of 'pushed' updates is in ensuring that ORCID records are current, which in turn will increase the value of ORCID to all stakeholders. This functionality will be available once publishers start to collect ORCID identifiers and deposit them along with CrossRef metadata.

Integrating with the ORCID service

ORCID is an open initiative; individuals may create, share, and maintain an ORCID record free of charge. ORCID software is made available under an MIT Open Source license. The public data in the ORCID Registry is searchable without a license, and ORCID will provide an annual public data file for free under a Creative Commons Zero waiver. ORCID is sustained by organizational memberships. For the fee, member organizations receive more frequently updated data, authenticated access to the ORCID registry, are registered as a trusted organization, and may create or update ORCID records on behalf of employees or students.[10]

ORCID currently provides two application programming interface (API) classes for the community to use. The *Tier 1* service can be used by individuals and organizations to query and retrieve public data. The Tier 1 Query API may be used without any registration or configuration. The *Tier 2* service is intended for third parties who need to query and retrieve limited access data, update or add new record data, and require production-level integration. Authentication and authorization follows the OAuth 2 standard. Tier 2 is available to member organizations and requires registration of the client application as an OAuth 'consumer'. The technical design of the ORCID service has been described in more detail elsewhere.[11]

The ORCID developer's portal (http://dev.orcid.org) provides a number of resources to developers in the research community. These include use cases, API classes, a test server, and help documents. These resources are updated as feedback from the community is received. Currently, a number of publishers are testing ORCID integration using these resources, including Nature Publishing Group, Springer, and Wiley; professional society publishers including the American Association for Cancer Research and the Association for Computing Machinery, and service providers including Aries, eJournal Press, Highwire Press, and Scholar One. The goal is to have ORCID integration in the first manuscript tracking systems implemented when the ORCID registry launches in October 2012, alongside ORCID records created by early adopter academic institutions. This will kick-start the use of ORCID identifiers by the research community.

Summary

By creating a registry for researchers and working with stakeholders to link digital research documents and other contributions to this registry, ORCID aims to provide a high-fidelity solution to the name ambiguity problem in scholarly communication. Benefits include reduced reporting workload, improved attribution, and a better understanding of knowledge flows to support research, collaboration, and evaluation. This vision is only possible if all of the stakeholders work together. This paper details steps for publishers to integrate ORCIDs into manuscript submission systems, and provides recommendations for specifying and displaying ORCID metadata. In addition to publishers, ORCID is working with research organizations, funders, and researchers to detail use cases and requirements ensure that the system is responsive to their needs.

benefits include reduced reporting workload, improved attribution, and a better understanding of knowledge flows

References

1. *Open Researcher & Contributor ID (ORCID)*, http://about.orcid.org (accessed 27 July 2012).
2. *Scopus*, http://www.info.sciverse.com/scopus/about (accessed 27 July 2012).
3. *Research Papers in Economics (RePEc)*, http://repec.org/ (accessed 27 July 2012).
4. *ResearcherID*, http://www.researcherid.com/ (accessed 27 July 2012).

5. VIVO | enabling national networking of scientists, http://vivoweb.org (accessed 27 July 2012).
6. *Structure of the ORCID Identifier.* http://dev.orcid.org/structure-orcid-identifier (accessed 16 August 2012).
7. DataCite Metadata Schema for the Publication and Citation of Research Data. 2011. doi:10.5438/0005
8. Journal Publishing Tag Set 2011. http://jats.nlm.nih.gov/publishing/ (accessed 25 July 2012).
9. Bilder, G. Disambiguation without de-duplication: modeling authority and trust in the ORCID system, http://about.orcid.org/content/disambiguation-without-de-duplication-modeling-authority-and-trust-orcid-system> (accessed 27 July 2012).
10. *ORCID membership and subscriptions*, http://about.orcid.org/membership (accessed 27 July 2012).
11. Wilson, B. and Fenner, M. 2012. Open Researcher & Contributer ID (ORCID): solving the name ambiguity problem. *Educause Review*, 47: 54–55.

Laurel L. HAAK, Laura PAGLIONE,
ORCID, Inc.
Email: l.haak@orcid.org

Martin FENNER
Board member ORCID

Ed PENTZ
Board member ORCID and CrossRef

Howard RATNER
Chair ORCID Board and Nature Publishing Group