

Essay

Open Access in Context: Connecting Authors, Publications and Workflows Using ORCID Identifiers

Josh Brown ^{1,*}, Tom Demeranville ² and Alice Meadows ³

¹ ORCID EU, 24 Paddock Gardens, Lymington SO41 9ES, UK

² ORCID EU, 86 Catherine Way, Bath BA17PA, UK; t.demeranville@orcid-eu.org

³ ORCID Inc., Brookline, MA 02445, USA; a.meadows@orcid.org

* Correspondence: j.brown@orcid-eu.org; Tel.: +44-7490-431-937

Academic Editor: Mary Anne Baynes

Received: 1 April 2016; Accepted: 2 September 2016; Published: 26 September 2016

Abstract: As scholarly communications became digital, Open Access and, more broadly, open research, emerged among the most exciting possibilities of the academic Web. However, these possibilities have been constrained by phenomena carried over from the print age. Information resources dwell in discrete silos. It is difficult to connect authors and others unambiguously to specific outputs, despite advances in algorithmic matching. Connecting funding information, datasets, and other essential research information to individuals and their work is still done manually at great expense in time and effort. Given that one of the greatest benefits of the modern web is the rich array of links between digital objects and related resources that it enables, this is a significant failure. The ability to connect, discover, and access resources is the underpinning premise of open research, so tools to enable this, themselves open, are vital. The increasing adoption of resolvable, persistent identifiers for people, digital objects, and research information offers a means of providing these missing connections. This article describes some of the ways that identifiers can help to unlock the potential of open research, focusing on the Open Researcher and Contributor Identifier (ORCID), a person identifier that also serves to link other identifiers.

Keywords: identifiers; ORCID; research information; open access; web technologies; open research

1. Introduction

The evolution of the academic web has provided researchers and the interested public with a radical set of possibilities. The cost of distribution and copying of resources has decreased at the same time as the volume of accessible information has soared. This was one of the inspirations for the Open Access movement [1]. The fundamental web technology of linking has enabled a degree of richness and interconnectedness in resources that was previously unimaginable. Many initiatives and products take advantage of these possibilities. The idea that a researcher can access an article, anywhere in the world, click on a resolvable link to articles cited (or to the data being reported on, or to digitized source material), and access them seamlessly is very attractive, and deceptively simple.

Resources are all too often hidden behind barriers. Links decay. ‘Reference Rot’ has become a significant issue [2]. However, a more fundamental challenge to opening up the processes and fruits of global research is that all too often links are not there, or are ambiguous. Two fundamental units in the scholarly endeavor are the researcher and the article. Even connecting these has proved challenging. The variations in naming conventions across journals, countries, cultures and alphabets are huge. One author could be referred to in her career as Jane Doe, Jane Ellen Doe, J. Doe, J.E. Doe, J.E. Doe-Pilkington, Dr. Jane Doe, Prof. Jane Doe and many other variations besides. Factor in all the other researchers with the same, or similar, names, and a simple search for works by a colleague can quickly become a hugely complicated discovery exercise. Correctly identifying a publication can be almost as

challenging, as it may appear in multiple formats, on multiple systems and platforms (the publisher's own platform, aggregator and library systems, and more), and even in multiple versions, including the author accepted manuscript (AAM) and version of record (VOR).

If we cannot unambiguously connect two such well-established units, how can we tackle new or emerging outputs in the research process, such as software or data visualization, or recognize an expanded range of roles and contributions, such as peer review and data curation? Correct attribution not only makes research more efficient, it is essential for openness.

Open Access (OA) is a special case in point. If a researcher publishes an article in an OA journal, there is often an audit and reporting trail implied. A funder may have a policy that requires outputs from funded projects to be OA [3]. An article processing charge (APC) may have been paid. The institution employing the author will want to track the publishing outputs and preferences of its staff. There may be an OA requirement for eligibility for research evaluation [4]. There might be a data management policy that requires deposit into a data center for underpinning datasets or open data. This could come from a funder or from a publisher [5,6]. The staff time involved in collecting, managing, and analyzing all of this information, and preparing it for formal reporting can be extremely significant, especially at global research institutions with substantial volumes of output.

In light of this, the challenge is clear. To fully deliver the potential benefits of open, digital scholarship, automatic, reliable, resolvable connections must be made systematically between researchers, their employment, their publications and other research outputs, their research activities, and the funding that supports it all. Truly open research is also transparent, which requires a mesh of information to surround each output or action. Attribution lies at the core of this transparency, as it forms a crucial link in the chain of relationships that underpin research.

In this context, ORCID (Open Researcher and Contributor Identifier) plays a vital role in establishing that chain [7]. ORCID maintains a central registry of unique persistent identifiers (ORCID iDs) for researchers and others engaged in research, scholarship, and innovation. These identifiers are openly available under a Creative Commons Zero license [8]. Researchers can register for an identifier at no cost, via a simple online registration process. This identifier can then be used when they apply for funding, when they submit a manuscript, or in internal research information systems. It provides an unambiguous, resolvable link to the individual to whom a research activity should be attributed. The registry also enables connections to other persistent identifiers (PIDs), whether for organizations or for research resources. These connections are curated and controlled by the researcher, and can be updated by the creators of identifiers and metadata, such as publishers, with the researcher's permission [9]. By connecting ORCID iDs with other PIDs for authoritative sources of data about the resource or activity, ORCID acts as a bridge, connecting researchers to information about their works.

These connections provide the beginnings of the rich, detailed context that openness both implies and requires. ORCID works with researchers and hundreds of organizations around the world to develop these connections, to help realize the benefits of 21st century communication technologies for the global research community [10]. This article sets out some of the ways that these partnerships have already improved the flow of research information, and describes some of the next steps that could be taken to further bolster openness and to improve the understanding of our progress towards Open Access.

2. Improving Attribution in Existing Workflows

ORCID is not alone in providing identifiers for researchers, and all published authors have multiple identifiers [11]. This is an inevitable consequence of publishing as many are automatically generated and assigned. Identifiers can be specific to a single vendor, country, discipline, institution, funder or publisher, or can emerge from wider community initiatives. Some focus on the needs of the library community or institutions, others may serve the researcher. National approaches also exist, or have existed in the past, such as DAI [12] in the Netherlands and JISC Names [13] in the UK.

The type and scope of these identities vary depending not just on the use-case they were designed to meet, but also on the use-cases they have evolved to address. Three of these are described below.

ISNI (International Standard Name Identifier) is the ISO certified global standard number for identifying contributors to creative works and those active in their distribution, including researchers, inventors, writers, artists, visual creators, performers, producers, publishers, aggregators, and more [14]. ISNI identifiers are semi-automatically derived from library catalogues and other trusted sources using algorithms and human intervention for quality control. They are not editable by the entities they reference, although users can suggest changes to existing profiles and report duplicates. Institutions that are members can submit data for matching and ISNI creation. ISNI requires at least two sources of matching information before an ID is created.

ResearcherID is a unique identifier assigned by Thomson Reuters that enables researchers to manage their publication lists, track their times cited counts and h-index, identify potential collaborators, and avoid author misidentification [15]. ResearcherID identifiers are closely integrated with other Thomson Reuters products, such as Web of Science. Like ORCID, ResearcherIDs are created and managed by users, and the two systems interoperate to enable the creation of bidirectional links.

Scopus Author Identifier assigns a unique number to groups of documents written by the same author via an algorithm that matches authorship based on certain criteria [16]. It is a service provided by Elsevier and integrated with their other products such as Mendeley. If a document cannot be confidently matched with an author identifier, it is grouped separately. In this case, you may see more than one entry for the same author. Scopus profiles cannot be amended by users but there is a process for author feedback.

These and many other identifiers are already embedded in many of the workflows that make up the research process. In this section, we describe some of the ways that interactions between these identifiers are improving the accuracy, currency, and reliability of research information.

2.1. Connecting Researchers and Their Organizational Affiliations

For researchers, reporting their employment or educational history is a central part of establishing their professional credentials. For the organizations employing or training researchers, the accuracy of these claimed affiliations is vital for the organization's understanding of its research portfolio, the balance of its activities and their impact, and also for its reputation. The claim that researcher X is affiliated with Organization Y is a fundamental component of research information, and the importance of its veracity cannot be overstated.

Organizational names suffer from similar ambiguities to those described for researcher names in the introduction to this article. Researchers typing organization names into free text fields on forms, or as part of a block of text in an article manuscript may express these names in varying ways (for example, M.I.T., MIT, Massachusetts Institute of Technology, etc.). This means that web or other searches for affiliations require using many variant terms, and may still not be exhaustive. While current identifiers for organizations tend to cover a defined subset of the vast range of organizations existing at any one time, there are a cluster of identifiers that are especially pertinent to research organizations; these are used in the ORCID registry to connect organizations to people.

At the simplest, data entry level, researchers can enter their own employment affiliation into their ORCID record. A dropdown list of names is provided based on characters entered, from which researchers can select the correct organization. Once this is done, the name and associated organization identifiers are linked to the metadata associated with that person's ORCID iD [17]. This means that the researcher is now connected to a canonically expressed, disambiguated organization name, underpinned by widely used identifiers.

In this scenario, the affiliation is asserted by the researcher. While this is no problem for the overwhelming majority of cases, in which trustworthy individuals act in good faith, an additional step—in which the claimed affiliation is validated—may be required for those seeking to re-use or publish this information. However, this additional step is unnecessary when employers or other

organizations make their own, validated claim (“assertion”) of affiliation with an individual, which can be done by using the ORCID API to add information to ORCID records through a simple process.

First, the researcher logs in to their institutional system, where their organization requests permission to connect to the researcher’s ORCID record and add affiliation information. If the researcher grants permission, the final step is for the organization to update the ORCID record with details of the affiliation. This may involve role information, or time bounds for the relationship (such as “October 2011 to the present”), but will always involve the unambiguous expression of the organization name and associated identifiers. This information can then be displayed in the web interface and in the metadata behind it, with the source shown as the asserting organization.

This employer assertion of affiliation provides an additional degree of trustworthiness in the connection between an individual and an organization. This is essential if the data are to be re-used, for example, by a publisher in adding affiliation information to an article or by a funder in assessing a grant application. It also offers a more robust underpinning for other potential uses of this information, which will be discussed later.

The value of these affiliations for OA are clear. Individuals are often the most visible connection between grant funding, an organization, a project, and a publication. By ensuring that these affiliations are openly and consistently expressed in a trusted way, the data can be embedded in publications and other outputs. Policy compliance, progress towards greater openness, and changes in citation level or other impacts can all be reliably tracked at the project or organizational level. For researchers, it can reduce the time spent re-keying information into multiple forms, reducing bureaucratic overhead and minimizing the risk of errors.

2.2. Automatic Updates for Newly Published Outputs

At the time of writing, the ORCID registry contains records linking individual researchers to more than 5,700,000 unique Digital Object Identifiers (DOIs) [18]. The DOI has long been established as a de facto standard for electronic academic publications, originally for journal articles, and now also used for book chapters, datasets, and other outputs. Connecting ORCID iDs for the authors of the outputs to those outputs’ DOIs is an obvious use case for improving the accuracy of publications records, citation counts, and attribution. Given that the workflows already exist for authors to register for an ORCID iD, for publishers to attach ORCID iDs to articles [19], and for DOI registration agencies to provide DOIs for those articles, the logical next step is to add one more piece to the workflow and connect the DOI and the ORCID iD.

Crossref is the major DOI registration agency for academic journal publishing [20]. It has established a workflow, in partnership with ORCID and academic publishers, enabling automatic updating (Auto-Update) of researchers’ ORCID records when a new article is published [21]. The process is the culmination of significant collaboration in the sector, and requires minimal effort from researchers. Automatic updating has also been set up by DataCite for research datasets [22]. Since the workflow is very similar, we will discuss the process as it currently operates for articles; for datasets, the first step involves a data center, rather than a scholarly publisher.

Auto-Update works simply by adding an additional step to the existing manuscript submission workflow that is already in place for many scholarly publishers. During manuscript submission, the publisher requests the author’s ORCID iD, which they provide by signing in to their ORCID account and granting permission for the publisher to read their ORCID record. When the article is accepted, the publisher then embeds the author’s ORCID iD in the article and associated metadata, which is sent to Crossref to mint a new DOI for the published article. Crossref detects the ORCID iD in the metadata, and updates the ORCID record with the metadata it has received from the publisher. If the author has not already granted permission, Crossref sends a message to their ORCID account notifying them that the article is published and the metadata is available.

Once this process has taken place, any system linked to the ORCID registry can be updated with the publication information, including institutional repositories, research information management

systems, and more. Since the update contains the DOI, the system can then pull a complete metadata record from the authoritative source for the data, populating the researcher's profile with the new metadata. This can be achieved, usually within 24 h of the DOI being minted, without the researcher having to lift a finger.

The ramifications for open research of these enhanced workflows are significant. Organizations tracking their research outputs can receive notifications of new outputs within a day of them being published. This accelerates the flow of new information and reduces the time currently spent on manual searches and data entry. A simple check by resolving the DOI can verify any rights metadata accompanying the DOI, enabling the organization to ensure that articles are indeed open and that they can be recorded and reported as compliant with institutional or funder policies. For datasets, the DOI points to a stable, accessible, curated home for the resource for the future, and indicates that action has been taken at a disciplinary or national level to preserve and share the data.

2.3. Supporting Transparency and Recognition for Reviews

Peer review, for publications, datasets, or funding applications, is fundamental to research. It is one of the ways in which researchers contribute to the general health of research, and is a form of disciplinary "good citizenship". As such, it should be recognized and rewarded as a first class research activity. To recognize peer review in this way, it must be made more transparent. This transparency is a clear step forward for open research, irrespective of whether or not the review itself is open.

In 2014, a community partnership, led by ORCID, F1000 Research and the standards body Consortia Advancing Standards in Research Administration Information (CASRAI) established a working group to define a set of metadata to describe peer review activities, with the aim of exposing it in their information systems [23–25]. This metadata, which includes identifiers for the reviewer, sponsoring organization, and the review itself, can be sent to the ORCID registry by a publisher or platform, enabling peer review activities to be included in an ORCID record, alongside the researcher's other outputs and professional contributions. From there, it can be sent to research management systems or funder reporting systems.

This functionality has now been rolled out by several organizations, including F1000 and the American Geophysical Union, and will be available to an increasing number of journals via implementations in systems such as Aries Editorial Manager, eJournalPress, and Publons [26–29]. A full citation for an open review can be shared or, for double-blind peer review, the journal can choose simply to acknowledge that the individual has performed reviews.

By using identifiers to expose peer review in this way, as a community we encourage a new level of openness in the system of peer review. By publishing review activities, we can get a fuller sense of the breadth of an individual's contributions to research. Reviewers unambiguously identified can be checked against author lists or known associations, improving the trustworthiness of both the review and the article. By adding yet more transparency to the research system, we can better understand the processes that shape articles, and support more innovation in scholarly communications, such as post-publication peer review.

3. Further Possibilities for Identifiers and Open Access

The developments described above set out some of the ways that identifiers are already serving as a tool to enhance the openness of research, and to make specific workflows around OA and related activities more efficient. In the future, they may also serve as a foundation for further steps in the evolution of OA workflows. As these new processes become embedded in scholarly communications practice, they will be the enablers of further gains in efficiency, transparency, and trust: the cornerstones of the Open movement.

3.1. Managing Article Processing Charges More Efficiently

One example would be to use identifiers to improve the management of Article Processing Charges (APCs) levied by many OA journals to cover publication costs. There are numerous challenges at present in managing the payment of APCs at an organizational level, but tools exist to overcome or ameliorate these.

By linking researchers to their employing organizations (as described in Section 2.1. above) and to the organizations that provide their research funding [30], using funder identifier systems such as Crossref Funding Data or the Global Research Identifier Database (GRID), eligibility for APC subsidies can be readily established [31,32]. The ability to delineate affiliations by time is crucial in this regard, as researchers may have moved on to pastures new while the products of their work at a previous employer were still working their way through the publication process. Tracking these progressions is a vital component in both career recording and the audit trail for APC expenditure. ORCID already has the ability to connect individuals, employers, and funding data, and will continue to enhance its support for other identifiers.

The auto-update functionality described in Section 2.2. above can already be used to streamline reporting processes, as noted, but could also serve as a valuable tool for auditing APC expenditure. Given that OA publishing represents a major shift for publisher business models, from charging for access to content to providing a publication service, the ability to generate close to real-time reports of publication activity will be a valuable indicator both for publishers and organizations.

Finally, the fact that these flows are being built from the ground up to be open and transparent, and rely extensively on APIs and other machine interfaces, creates exciting possibilities for other new services to be built on them. For example, the data they provide, and the flow of information they create, could be readily augmented from other sources to develop new kinds of analysis or new understandings of research activity.

3.2. Understanding OA as Both Producer and Consumer

Another potential benefit for OA from the increased use and integration of identifiers is the opportunity for enhanced understanding of publishing output. It would be hypothetically possible, given sufficiently robust uptake of ORCID iDs by researchers and publishers, to use the ORCID iD to connect articles to people and funding, assess the balance of subscription versus OA articles for any publisher or subscription portfolio, and tie this to the APCs paid at the organizational level. This would enable each organization to assess their contribution to OA progress.

This intelligence would also provide a valuable resource for governments, such as that of the Netherlands, which have set ambitious targets for OA publications [33]. For library consortia, it would provide a way to track the balance of subscription content versus OA year on year, assessing the impact of the APCs they pay, and comparing it to the remaining subscriptions they purchase. Such a tool would also be invaluable for publishers providing content to these consortia, as the objective data would enable both sides to lay the ghost of 'double-dipping' to rest [34].

4. Discussion

It is clear that the pursuit of Open Access relies on improvements and changes to many of the processes that underpin the research endeavor. It creates new uses for existing research information, and generates demand for new kinds of information. At the same time, existing research information and scholarly communications systems are showing signs of strain. Analysis and reporting is labor-intensive and slow. The administrative burden on researchers—and research managers and administrators—is too high. The transition to a digital research world is by no means complete or effective.

Improving our use of machine-to-machine communication and the automatic transmission of information across systems is vital if we are to improve on this situation. Identifiers play a central role

in this improvement, and the power they have to streamline workflows and generate better quality information has been amply demonstrated by the examples of identifier integration provided here.

That said, identifiers are just some of the tools that exist to both improve and expand the services that research and researchers depend upon. Increasing openness itself generates new possibilities. Open research is not just open to the people of the present, seeking to use, re-use, and better understand our research achievements. It is open to the future, creating possibilities we haven't even thought of yet. So, as we build our infrastructures for open research and work to continue to create new possibilities, we must be mindful not to foreclose the future.

Conflicts of Interest: The authors are employed by ORCID Inc. (Alice Meadows) and ORCID EU (Josh Brown and Tom Demeranville). ORCID Inc. provides some of the services and tools described in this article. ORCID EU exists to promote the adoption and integration of ORCID identifiers in European research infrastructures and systems.

References

1. Budapest Open Access Initiative. Available online: <http://www.budapestopenaccessinitiative.org/> (accessed on 31 March 2016).
2. Burnhill, P.; Mewissen, M.; Wincewicz, R. Reference rot in scholarly statement: Threat and remedy. *Insights* **2015**, *28*, 55–61. [CrossRef]
3. Registry of Open Access Repository Mandates and Policies (ROARMAP). Available online: <http://roarmap.eprints.org/> (accessed on 31 March 2016).
4. Higher Education Funding Council for England Open Access Research Policy. Available online: <http://www.hefce.ac.uk/rsrch/oa/Policy/> (accessed on 31 March 2016).
5. Engineering and Physical Sciences Research Council Policy Framework on Research Data Access Expectations. Available online: <https://www.epsrc.ac.uk/about/standards/researchdata/expectations/> (accessed on 31 March 2016).
6. Public Library of Science Data Availability Policy. Available online: <http://journals.plos.org/plosone/s/data-availability> (accessed on 31 March 2016).
7. Open Researcher and Contributor Identifier. Available online: <http://orcid.org/> (accessed on 31 March 2016).
8. Creative Commons CC0 License. Available online: <https://wiki.creativecommons.org/wiki/CC0> (accessed on 31 March 2016).
9. ORCID Privacy Policy. Available online: <https://orcid.org/privacy-policy> (accessed on 31 March 2016).
10. ORCID Integration Chart. Available online: <https://orcid.org/organizations/integrators/integration-chart> (accessed on 31 March 2016).
11. Aryani, A.; Barton, A.J.; Brase, J.; Brown, J.; Demeranville, T.; Herterich, P.; McAvoy, L.; Paglione, L.; Ruiz, S.; Thorisson, G.; et al. Workflow for Interoperability. 2015. Available online: https://figshare.com/articles/D4_2_Workflow_for_interoperability/1373669/1 (accessed on 31 March 2016).
12. Digital Author Identifier. Available online: <https://wiki.surfnet.nl/display/standards/DAI> (accessed on 29 June 2016).
13. Names: Developing a Name Authority Service to Provide Unique Identifiers for Authors. Available online: <https://www.jisc.ac.uk/rd/projects/names> (accessed on 29 June 2016).
14. International Standard Name Identifier (ISO 27729). Available online: <http://isni.org> (accessed on 29 June 2016).
15. ResearcherID. Available online: <http://www.researcherid.com/> (accessed on 29 June 2016).
16. Scopus Preview. Available online: <https://www.scopus.com/search/form/authorFreeLookup.uri> (accessed on 29 June 2016).
17. Organization Identifiers in ORCID (Ringgold Identifiers). Available online: <http://members.orcid.org/api/organizations-orcid-ringgold-identifiers> (accessed on 31 March 2016).
18. ORCID Statistics. Available online: <https://orcid.org/statistics> (accessed on 31 March 2016).
19. Requiring ORCID in Publication Workflows: Open Letter. Available online: <http://orcid.org/content/requiring-orcid-publication-workflows-open-letter> (accessed on 31 March 2016).
20. Crossref. Available online: <http://www.crossref.org/> (accessed on 31 March 2016).
21. Auto-update Has Arrived. Available online: <http://orcid.org/blog/2015/10/26/auto-update-has-arrived-orcid-records-move-next-level> (accessed on 31 March 2016).
22. DataCite. Available online: <https://www.datacite.org/> (accessed on 31 March 2016).

23. Faculty of 1000. Available online: <http://f1000.com/> (accessed on 31 March 2016).
24. Consortia Advancing Standards in Research Administration Information (CASRAI). Available online: http://casrai.org/Main_Page (accessed on 31 March 2016).
25. ORCID Peer Review Early Adopter Program. Available online: <http://orcid.org/content/peer-review-early-adopter-program> (accessed on 31 March 2016).
26. Hanson, B.; Lawrence, R.; Meadows, A.; Paglione, L. Early adopters of ORCID functionality enabling recognition of peer review: Two brief case studies. *Learn. Publ.* **2016**, *29*, 60–63. [[CrossRef](#)]
27. Aries Editorial Manager. Available online: <http://www.ariessys.com/software/editorial-manager/> (accessed on 31 March 2016).
28. EJournalPress. Available online: <http://www.ejournalpress.com/> (accessed on 31 March 2016).
29. Publons Partners with Orcid to Give More Credit for Peer Review. Available online: <https://orcid.org/blog/2015/10/12/publons-partners-orcid-give-more-credit-peer-review> (accessed on 31 March 2016).
30. Updating Awardee ORCID Records. Available online: <http://members.orcid.org/funder-workflow#award> (accessed on 31 March 2016).
31. Funding Data. Available online: <http://www.crossref.org/fundingdata/> (accessed on 31 March 2016).
32. Global Research Identifier Database. Available online: <https://www.grid.ac/> (accessed on 31 March 2016).
33. Dutch Government Policy on Open Access to Academic Publications. Available online: <https://www.government.nl/documents/parliamentary-documents/2014/01/21/open-access-to-publications> (accessed on 31 March 2016).
34. Australasian Open Access Strategy Group. Blog Post: “Addressing the Double Dipping Charge”. Available online: <http://aoasg.org.au/news-updates/blog-summary/addressing-the-double-dipping-charge/> (accessed on 31 March 2016).



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).