PERSISTENT, UNIQUE IDENTIFIERS FOR AUTHORS – ORCID AND SMALLER PUBLISHERS

Gudmundur A. Thorisson

Abstract

in October 15th the ORCID initiative (http://orcid.org) launched its long-awaited central registry service for scholarly authors and contributors. The new service enables researchers to obtain a unique, persistent personal identifier and to maintain a centralized record of their published works, grants and other scholarly activities. At the time of writing, over 30,000 users have signed on with the service in the first two months and several 3rd party online services are already linked to it, with many more to come in the next year.

In this article I briefly outline the background to this important initiative and the identification problems it was created to solve, and conclude with remarks on what all this means for small, independent journal operations like ScieCom.

Building a community

ORCID - short for Open Researcher and Contributor ID - is a global, open, community-based non-profit organization created to tackle challenges relating to name ambiguity in scholarly communication. Established in 2009, ORCID differs fundamentally from previous initiatives which have either i) failed or ii) been successful only in specific countries or disciplines[1] in that the project is backed by broad, diverse community of stakeholder organizations in research. This community includes commercial and non-profit publishers, academic institutions, research libraries, funding agencies and many other organizations and individuals. In short, everyone who is constantly dealing with mistaken author or reviewer identity, missing papers, shared names and a thousand other identification-related problems in the scholarly research domain.

Linking researchers with their research

The organization's slogan - "Connecting research and researchers" - captures the ambition of the project well. The over-arching aim is to not merely provide unique, persistent identifiers, but also to create and maintain the necessary supporting informatics services and tools (aka *identification infrastructure*) that make it possible for organizations to embed identifiers in their information workflows and computer systems and use them to track researchers and link them to their research activities. Currently this is often very problematic for organizations, largely due to the inherent non-uniqueness of person names and the resulting ambiguity in assigning authorship across the rapidly growing body of published literature.

In their recent coverage of ORCID, Nature[2] cited the classic case of Y. Wang who appears to have authored nearly 4,000 papers in the year 2011 alone. A related and entertaining story of name confusion is that of two academics in China who share a common family name and first initial[1], are both physics researchers, and work *in the same university department*, no doubt causing no end of trouble for departmental administrators. Cases like this will trip up even the most sophisticated data mining and disambiguation algorithm approaches (see e.g. ref **3**). The take-home message is that the "author name problem" is big enough and complex enough that it will likely never be solved with automated methods alone.

Scholars in identity crisis

On the other side of the table, as individual researchers we ourselves also often have to deal with identification-related problems. Some of these are rather obvious; one would expect that the numerous scholars named Y. Wang must have their hands full keeping track of the works they have authored, and ensuring that they are accurately represented, for example, on the Internet and not confused with their colleagues of the same name.

---

[1] Many Chinese names have identical spelling after transliteration from the Chinese logographic writing system to the Western Latin alphabet.

Mistaken identity can have non-trivial - and even serious - consequences for scholars. For example, Melissa Terras, a scientist working in the humanities, reported a case[4] where an error in a publisher's computer system caused her to be listed as the author of a decidedly non-scientific book on Tarot symbolism. The error was propagated and amplified across the Internet and caused Melissa numerous troubles which took a long time to work out. Another, less obvious example is mis-identification in the selection of expert reviewers for manuscripts submitted to peer-reviewed journals.

Solving existing problems - creating new opportunities

ORCID is now putting the necessary infrastructure in place to enable the community to start seriously tackling these identification challenges. On one hand, the new tools will help with retroactively sorting out the current mess, i.e. un-tangling authorship for works that are already published. On the other hand, and more importantly, it will be possible to prospectively address the myriad identification-related problems in scholarly publishing workflows, grant management, institutional research management and other settings going forward. These were, after all, the primary drivers for creating ORCID in the first place.

This evolution is already taking place at rapid pace. For example, Scopus, one of the two major literature indexing services, has built and launched a wizard-based tool[2] which enables an ORCID user to seamlessly populate his/her profile with publication lists retrieved from the Scopus system. Another ORCID-integrated service is ImpactStory[3] which builds an impact analysis report of a scholar's research outputs. The analysis goes beyond traditional citation-based measurements, using as input broad evidence of use as diverse as online views and downloads, social media sharing, commenting and bookmarks, and WikiPedia mentions[4].

I and many of my colleagues are especially excited about ORCID's potential to serve an enabling platform to support the creation of new and innovative tools and services, such as ImpactStory. A major area of opportunity is support for the modern-day "digital scholar" - that is, infrastructure that enables researchers to be linked with, and receive credit for, a broad range of so-called *non-traditional* research outputs or knowledge contributions, including (to name a few) research datasets[5], presentation slides, source code for scientific software, curation of biomedical databases, contributions to WikiPedia articles and much more.

The operative word is *open*

Openness is a key element in everything that ORCID does and this is a major reason why the initiative has garnered so much backing. Most of the organization's ten principles[6] feature openness of one meaning or another. Several deal with openness in governance, organization membership, data and more. For end users (i.e. researchers), the most important principle is the one that states that *anybody* who so wishes can create and manage their ORCID identifier and corresponding profile in the system, free of charge.

In the two months since launch, over 30,000 users have already jumped in and registered.
This is a good start, given that these early adopters can use their IDs with only a small number of ORCID-enabled 3rd party services at the moment. But the long-term scope of the project is international and trans-disciplinary, and the total number of scholarly authors worldwide may be much as two orders of magnitude larger than this (no one knows how many, it goes without saying). So how can ORCID attract the interest of millions of scholars worldwide and get them to register?

Can ORCID become the new black?

Excited as I am about the potential for new tools & services, in reality such developments will take some time to appear, mature and be adopted (or fail otherwise). They are therefore not likely to be significant in driving early ORCID adoption by researchers. Amongst those who have been following and working in this space, there is general consensus that traditional publishing is where the early action will take place.

---

[2] http://orcid.scopusfeedback.com
[3] http://impactstory.org
[4] This field of study is commonly referred to as *alternative metrics*, see http://altmetrics.org

[5] Research data are increasingly published in online repositories such as Dryad, see http://datadryad.org and ref. **5**
[6] http://about.orcid.org/about/what-is-orcid/our-principles

This prediction is based on the simple fact that most researchers routinely come into contact with publishers when they need to publish their work. For many authors, the time when they submit a manuscript to (say) a journal for peer-review is probably one of the very few key events in their busy[7] academic life when they are likely to be receptive to the concept of author identifiers. Therefore, this is the best time to promote ORCID and highlight the benefits of registering (e.g. that they don't need fill out an author registration form for the umpteenth time). Put another way, *unpublished content* is where both active scholarly authors *and* their publishers (as key stakeholder groups) have the most incentives from adopting ORCID early on, compared to other major publishing-focused use cases (e.g. works published by deceased or otherwise inactive authors)[6].

Following this line of reasoning, a great deal of emphasis has been placed on facilitating integration between the central ORCID service and software used by publishers. Integration means not only embedding of ORCID identifiers in workflows (e.g. ask authors to supply their ID when they submit their manuscript) but also displaying them in author lists on article web pages and full text PDFs, and so on. Several major commercial and non-profit publishers, as well as makers of manuscript tracking systems used by many publishers, are already working on integrating their systems. We can expect to see many of those come online sometime in the first half of 2013.

ORCID and small-scale publishers

What does this mean for smaller, independent scholarly journals, especially those on a shoestring (or even zero) budget? In particular, how can e-journal outfits like ScieCom, their authors and their readers benefit from this emerging new technical infrastructure? I can recommend as a general background reading a recent paper[7] authored by the ORCID leadership which outlines the main issues and key benefits from integration to publishers, repositories and other organizations. Here I want to highlight a pair of issues which I consider to be of key relevance to smaller players in the publishing space.

First there is the technical obstacle. Connecting to ORCID programmatically via the application programming interface (API) requires certain modifications to the software used to run an e-journal. Journals which run on commercial, closed platforms are tied to whatever functionality is "in the box", and so will not be able to connect until the software vendor gets around implements the required ORCID integration.

Many smaller journals, on the other hand, run on free, open source software[8]. This means that, in principle at least, it is perfectly possibly for each journal to implement ORCID functionality by simply modifying the source code as needed. But the technical expertise required for this is likely beyond most individuals or groups running a small journals, and so most of them are likely stuck in the same boat as journals using commercial solutions. The good news is that the majority of these journals (including ScieCom) are powered by a single platform - Open Journal Systems (OJS)[9] - which is used by thousands of groups worldwide to disseminate knowledge on an incredibly diverse range of topics. The dominance of OJS should greatly simplify the task of bringing ORCID to this community. That is, the required extra functionality can be implemented just once in the OJS platform, and subsequently reused by all the OJS-based journals the next time they upgrade their system.

Second, there is the orthogonal problem of cost. Certain parts of the ORCID API can be used by individuals and organizations free of charge to search and retrieve profile data. But inn order to get the kind of full integration that a journal would need, the journal (or single multi-journal publisher outfit) must have access to the full member API. This is where ORCID's business plan for becoming financially sustainable comes in: organizations who benefit from integration (e.g. by saving costs) will be charged annual membership fees, and those fees will pay for ongoing costs of operating the service.

A membership fee based model in itself is not in itself a bad thing: after all, somehow the bills must be paid to keep the service running. However, for various reasons the membership fee structure[10] that ORCID has started with is inflexible and very biased in favour of larger publishers and institutions with large budgets

---

[7] Another key event is submission of grant proposals to funding agencies

[8] See the Open Source Initiative (OSI) website: http://opensource.org

[9] http://pkp.sfu.ca/ojs

[10] http://about.orcid.org/about/membership

who are expected to be early heavy users of the system. The consequence of this is a financial barrier to participation for smaller e-publishing outfits with a limited or no budget who would not be able to pay the annual fee. The good news is that ORCID is now in the process of revising the current model and expects to introduce additional membership plan options in 2013 that will better suit this category of "small integrators".

Conclusions

I have focused here on the publishing-focused use cases for ORCID and prospects for uptake of the service amongst scholarly authors and publishers. For smaller journals, one of the two key factors - improve the membership fee structure - is something that ORCID can influence to facilitate broad adoption. The other one – support in open source software tools - is a task for the journals and software developers themselves to take on, ideally in concert with and as part of the ORCID developer community which is gradually taking form. See the developer portal at http://dev.orcid.org if you are interested in getting involved.

I want to mention another important route to ORCID adoption: namely introduction and integration at the national level. This is a strategy that will not work in larger countries (USA, UK, Germany) because of their sheer size and diversity in research infrastructure. There is, however, substantial interest in going this route in smaller countries with sufficiently advanced research information infrastructure, including some of the Scandinavian & Baltic nations. For further reading on this topic, I suggest the article by Adrian Price elsewhere in this issue of ScieCom where he reports on a plan now in preparation for adopting ORCID nationally in Denmark.

These are interesting times. ORCID now takes its first steps as an organization and as an emerging key piece of scholarly communication infrastructure. At the time of writing, the new registry service is limited in functionality and is experiencing some early growing pains, but wrinkles are constantly being ironed out with the help of a growing and actively participating community, as evidenced by the feedback gather via http://support.orcid.org. I invite you to join us.

References

1. Fenner, M. **Author Identifier Overview**. LIBREAS 24–29 (2011). Available at: http://edoc.hu-berlin.de/docviews/abstract.php?lang=&id=37867

2. Butler, D. **Scientists: your number is up.** Nature 485, 564–564 (2012). http://dx.doi.org/10.1038/485564a

3. Smalheiser, N. & Torvik, V. **Author Name Disambiguation**. *In*: Annual Review of Information Science and Technology (ARIST) 43, 287–313 (2009). Available at: http://arrowsmith.psych.uic.edu/arrowsmith_uic/tutorial/ARIST_preprint.pdf

4. Terras, M. **What's in a name? Academic Identity in the metadata age, or, I didn't see #tarotgate coming**. Retrieved 14/12/2012. Available at: http://melissaterras.blogspot.co.uk/2012/10/whats-in-name-academic-identity-in.html

5. Vision, T. J. **Open Data and the Social Contract of Scientific Publishing**. BioScience 60, 330–330 (2010). http://dx.doi.org/10.1525/bio.2010.60.5.2

6. Fenner, M., Gómez, C. G. & Thorisson, G. A. **Collective Action for the Open Researcher & Contributor ID (ORCID)**. Serials: The Journal for the Serials Community 24, 277–279 (2011). http://dx.doi.org/10.1629/24277

7. Haak, L. L., Fenner, M., Paglione, L., Pentz, E. & Ratner, H. **ORCID: a system to uniquely identify researchers.** Learned Publishing 25, 259–264 (2012). http://dx.doi.org/10.1087/20120404

Gudmundur A. Thorisson Faculty of Life and Environmental Sciences, University of Iceland and ORCID

Gudmundur has been involved with ORCID in the past 3 years and as of October he is employed part time by the organization via the EC-funded ODIN project (http://odin-project.eu). His interests include, among other things, technical outreach and building & fostering a vibrant ORCID developer community.

E-mail: gthorisson@gmail.com
Personal website: http://gthorisson.name
ORCID profile: http://orcid.org/0000-0001-5635-1860