

ORCID or how to build a unique identifier for scientists in 10 easy steps

Martin Fenner, Gobbledygook

January 3, 2010

ORCID stands for Open Researcher and Contributor ID and was announced in early December. This blog post tries to summarize some of the problems that have to be solved to develop a unique identifier for scientists.

1. Identify the problem

The Researcher Identification Primer by the Gen2Phen Knowledge Center lists some of the problems that a unique identifier for scientists tries to solve, including

- Disambiguation of author names in the scientific literature and establishing/validating relationships between authors and publications.
- A solid foundation for permitting and tracking online scientific contributions, such as database submissions, scientific blogging, and community curation efforts.
- Knowledge discovery applications using some or all of the above components.

The Gen2Phen workshop co-organized by Gudmundur Thorisson in May 2009 discussed these issues in much more detail. One of several articles talking about the problem of disambiguation of author names (especially Asian author names) appeared in Nature News in February 2008. A December 2009 Nature editorial emphasized that a unique identifier for researchers will be especially valuable to track scientific contributions that are not related to authoring a paper. Phil Bourne and J. Lynne Fink also wrote about this in PLoS Computational Biology in December 2008: I Am Not a Scientist, I Am a Number. A number of tools have tried to solve this problem, but it is not possible to link the researcher identities in the many systems.

2. Define what you want to accomplish

Geoff Bilder gave a very good introduction to the problem at Science Online London in August and STM Innovations in December. Both talks were similar, but the latter is available as video and PDF. He emphasized that ORCID is about Knowledge Discovery and not Access Control, and explained the terminology for

subject, identifier, profile, persona and credential. Access Control is a related problem that is sometimes mixed in, but there is no requirement that a unique researcher identifier also has to provide secure access via whatever mechanism (Open ID is one solution to that problem).

3. Win support of stakeholders

Founding members of the ORCID initiative can be found on the ORCID homepage and include publishers, funders, universities, organizations and software companies. A number of important stakeholders are already part of the initiative, support by more funders (besides the Wellcome Trust) and software companies (particularly those that build reference managers or social networking sites for scientists) would be great. Probably the biggest name not on the list is the U.S. National Library of Medicine that runs the PubMed database of biomedical literature (the ORCID members Wellcome Trust and British Library are involved in UK PubMed Central).

4. Make decisions about the general design of the system

Some of the design decisions obviously are not set in stone at this stage. One continuing discussion is centralized vs. federated, and it looks like ORCID will be a centralized system similar to the DOI. Geoff Bilder has some good arguments for a centralized system. Another recurring theme is how much control an individual researcher has over his ORCID record. Although external assertion from publishers or funders will certainly be part of ORCID, the individual researcher will have an important role, not only because of privacy concerns, but also because this is the easiest way to fix errors that even the best automated algorithms for author assignment will produce. And it looks as if ORCID will be an extensible system that will for example allow publishers or social networking sites to add functionality they require. The discussion at the STM Innovations meeting in early December touches some of these issues and is recorded as video (after the talk by David Kochalko).

5. Pick a name

The name Open Researcher and Contributor ID (ORCID) is obviously a combination of ResearcherID (Thomson Reuters) and Contributor ID (CrossRef). I would have preferred a simpler name, but I guess we have to get used to ORCID.

6. Build on available tools

ORCID will be based on the ResearcherID software from Thomson Reuters. From what I've seen, the Open ID system will not be a central part of ORCID. But ORCID certainly will be designed to work together with Open ID and other authentication mechanisms. I don't know what Elsevier and the Scopus Author ID will contribute to ORCID.

7. Form an independent organization

In order to be adopted widely, ORCID must be run by an independent organization, and not by a single publisher, software company, research organization or funder. With the experience of running the DOI system to identify digital objects such as scientific papers, CrossRef would be one obvious candidate, but the ORCID founding members have yet to decide on that.

8. Secure financing

Starting and maintaining ORCID will obviously cost money. In my little survey about author identifiers back in April 2009, the opinions were split about who should pay for this. Journal publishers and database maintainers (referring to such databases as PubMed, Scopus or Web of Science) were the two most common answers. ORCID will make it easier for funding agencies to evaluate scientists and they might therefore also contribute to the system. Individual researchers hopefully will not have to pay for any of this, but their input in time is obviously required.

9. Promote ORCID

A Nature editorial in December was a good start to promote ORCID to a wider audience. A unique identifier for scientists will only become accepted if widely used. That's why it is important that publishers and funders quickly adopt this service. Software companies that build interesting tools around ORCID are also critical, e.g. integration of ORCID into manuscript submission systems (including the use of ORCID for the peer reviewers) and social networking sites (including of course Nature Network). My experience with the DOI for papers (e.g. the limited support in PubMed) tells me that adoption of ORCID will be a long process.

10. Involve individual researchers

Individual researchers currently have no way to get directly involved in ORCID. But some level of involvement is critical for an author identifier to work. The best place is currently probably the LinkedIn Group Unique Identifiers for Researchers started by Cameron Neylon. But I hope we soon see ORCID discussions on Nature Network and other social networking sites. The best place on Nature Network to discuss ORCID is currently probably the Scientific Researchers and Web 2.0: Social Not Working? Forum.