

Connecting researchers and research – investing in infrastructure to bring efficiencies to research. ORCID as an enabler of better research analytics

Liz Allen

l.allen@wellcome.ac.uk

Abstract There has been a boom in the emergence of research analytical platforms and tools to support research and researcher impact assessment and evaluation. However, different analytical platforms can contain different cuts of the same or similar data and can yield different conclusions to the same question. Stakeholders in the research ecosystem need simple ways to link the inputs, products and outputs of research to the researchers and research institutions associated with them – and ways that do not add to researcher administrative burdens. Launched in 2013, ORCID (Open Researcher and Contributor ID), is a not-for-profit, cross-sectoral collaboration, committed to building a global registry of unique and persistent researcher identifiers. ORCID effectively uses researcher IDs to connect and share research-related information across different systems via links and application programming interfaces.

ORCID has the potential to shed new light on the factors that influence researcher and research productivity and impact and enable analysis of research-related knowledge, information and people flows.

Keywords: ORCID; research identifiers; research evaluation; impact; career tracking

In the past five years there has been a boom in the emergence of research analytical platforms and tools to support research and researcher assessment and evaluation. This boom has been driven by the greater availability of research-related information on the web, open access research mandates encouraging researchers to share the results of their research, the appetite for information to support research evaluation and national research assessment exercises, and, practically, developments in text-mining and searching capabilities. For research evaluators — like myself — these are exciting times indeed.

However, the plethora of new tools and platforms emerging in such a short space of time can be overwhelming and confusing. Different analytical platforms can contain different cuts of the same or similar data and can yield different conclusions to the same question. Furthermore, they can make a considerable dent in an already stretched research evaluation budget. Setting aside the not insubstantial challenge of knowing which data are the important indicators of research quality and longer term impact - and I will come to this later - it is therefore fundamental that we don't let sexy analytics, infographics and seductive league tables and rankings distract us from the task in hand. Stakeholders in the research ecosystem who are trying to make sense of how research works and fund it more efficiently and effectively, *need denominators*

as their starting point; they need simple ways to link the inputs, products and outputs of research to the researchers and research institutions associated with them. Launched in 2013, ORCID (Open Researcher and Contributor ID; <http://orcid.org/>), a not-for-profit, cross sector collaboration, committed to building a global registry of unique and persistent researcher identifiers, was designed to address precisely this.

ORCID is effectively a piece of 'plumbing' intended to be laid across the research ecosystem. By creating unique and persistent IDs, ORCID provides researchers and contributors with a way to link themselves to their associated research works in publisher workflows, in manuscript submissions systems, in funder grant application and reporting systems and in research institutions' information management systems. Through a simple ID, information can be connected and shared, pushed and pulled across different systems via links and application programming interfaces. A range of agencies across the research ecosystem, including publishers, higher education institutions and research funders like the Wellcome Trust, have now started to integrate ORCID into their workflows, recognising the pivotal role that ORCID can play in bringing efficiencies to their business processes.

Reducing researcher burdens remains a key goal for ORCID, to protect researchers' time to do research. Via the use of an ID and system connections, it is easy to imagine how researcher time is freed up from manually entering the same data into different systems, multiple times and for slightly different purposes, during their career¹. Tight data privacy controls associated with an individual's ORCID record ensure that researchers remain in control of what data are linked to them and can be shared and with whom. And researchers seem very interested; to date over 1.3 million individuals have registered with an ORCID around the world — most without having any mandate to sign up.

For research funders and higher education institutions, information on the research works associated with those they support could be automatically harvested and tracked over time. Many agencies, in the attempt not to bother researchers, gather data on research works directly (peer reviewed publications being the prime example) via a range of bibliographic websites (eg PubMed, EPMC, Web of Science, Scopus). However, anyone who has ever tried to do this will recognise the challenge of identifying the correct individual if, for example, that person has a common surname or has changed his or her name, or has published using various combinations of his or her initials. If researchers had a unique identifier that was used consistently over time and across different platforms and systems, the challenge of disambiguating researchers could disappear. Easier identification of individuals would bring other benefits, including supporting peer review selection and faculty appraisal and recruitment, allowing learners and students to find individuals with whom they may want to collaborate, and supporting research and researcher evaluation and career tracking.

During 2013 and 2014, grants from the Sloan Foundation and JISC enabled pilot projects to test the value of ORCID integrations in university and research institution systems in

the USA and the United Kingdom respectively^{2,3}. ORCID was used in a variety of test cases across institutions, including supporting post-graduate career tracking, monitoring of open access publication mandates, and, staff and recruitment strategy. Evidenced by the continued involvement of the universities and HEIS with ORCID post pilot, both were deemed to be successful.

Using a simple and persistent ID as a piece of the research infrastructure across platforms also allows for flexibility and choice in use of analytical platforms and approach. Some funders and higher education institutions have large budgets to invest in building data and analytical systems to allow them greater insight into their funding investments and portfolios; others do not. If all that is required for system connectivity and to enable data harvesting and sharing are a few common fields and a unique ID, then the barriers to entry into the space of information and system inter-operability are considerably reduced. And it is easy to see the attraction of using ORCID to support national research assessment exercises, where different organisations are required to draw together a range of data and information from their own established and legacy systems that may or may not be stored in formats that are specified as part of a national submission. ORCID is already mandated by research funders in Portugal and Denmark, and other national research councils are attracted by the potential of ORCID. A strength of ORCID is that it is technology and platform agnostic, it is not a platform; it is a connector. It does not exclude participation; it opens doors.

One of the doors I would like to see open a bit wider is the one to the world in which we know how best to fund research to maximise funder investments, deliver impact and support the next generation of researchers. A persistent challenge for funding agencies is to understand exactly how their funding is making a difference; how best should they use their funds to maximize impact and be efficient? Can we better predict who or what is going to deliver the best outcomes with our funds? Can we pick the researchers who are most likely to become the future research leaders?

Research evaluators and data scientists are excited by the potential of ORCID to provide the denominators and shed light on the factors that influence researcher and research productivity and impact in all its guises, through its potential to create a 'researcher-centric view of science' and enable analysis of research-related knowledge, information and people flows⁴. Through this sort of analysis, we can start to understand some of the factors affecting how research progresses, the time taken to create an impact and what might be the pivotal catalysts that help things to move forward. And then we can make real inroads into understanding how research really works and how to do it better and better.

References

- 1 Anstey, A. (2014) How can we be certain who authors really are? Why ORCID is important to the *British Journal of Dermatology*. *British Journal of Dermatology*, 171, 4, 679–680. DOI: 10.1111/bjd.13381
- 2 See <http://orcid.org/content/adoption-and-integration-program>
- 3 See <http://orcidpilot.jiscinvolve.org/wp/>
- 4 Butler, D. (2012) Scientists: your number is up. *Nature*, 485, 564, doi:10.1038/485564a