

One ID to rule them all

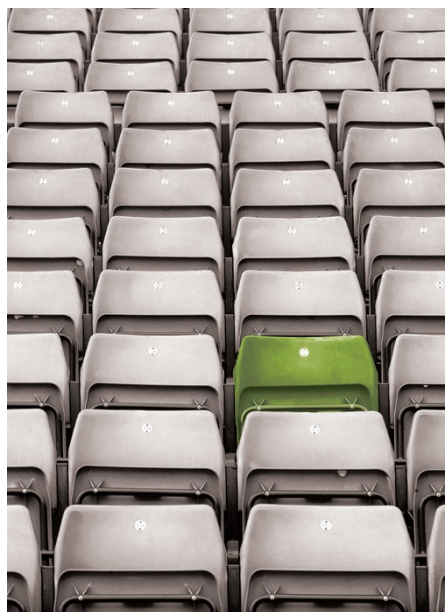
Some Springer Nature journals, including *Nature Physics*, are mandating Open Researcher and Contributor IDs (ORCID) for the corresponding authors of accepted papers. We provide some context to this initiative.

Being a scientist in the twenty-first century is not easy. On top of doing excellent research, you have to communicate your results through various outlets, from traditional journal articles to mainstream media, not to mention conference presentations, datasets, code and public outreach activities. The impact of all of these outputs is quantified and ultimately used to make important funding decisions. This drive for quantification is not unique to science, but sadly one of the defining trends of our age: from Amazon star ratings to the number of retweets of a news article, we are now more than ever influenced by numbers. And despite general discontent with this system, there is no sign of change, simply because, so far, no other realistic alternative has been found. Setting aside the question of whether scientific progress should be rated with a number, if we are to continue with this approach, we'd better get the number right. One way to do this requires widespread uptake of the ORCID system, which provides unique author identification and tallies a broad range of different research outputs.

The first problem in getting the number right is ensuring you have all the relevant data — aside from just peer-reviewed journal articles. Despite recent progress in extending digital object identifiers (DOIs) to non-traditional research outputs like datasets, figures and code, there is still relatively limited uptake from the scientific community and the culture of properly citing these items has yet to develop. Conference contributions are even more problematic, as it's unclear whether they should be tracked solely through proceedings or also through presentation slides made available in a repository. There is a clear need for a unique identifier that can pull conference contributions together with additional metadata, but no solution has yet emerged.

Another issue is which metrics to use for these non-traditional research outputs. The number of citations as an indicator of an article's impact has been studied for decades and, for better or worse, is the accepted norm. But it is not yet clear which metrics can be used for data or software and how these can be validated.

A problem that has plagued scientometrics for decades is the



RTIMAGES / ALAMY STOCK PHOTO

disambiguation of author names. Because, with few exceptions, we cannot clearly assign all the articles X. Wang has authored in the past ten years. Even if we restrict ourselves to the physicist X. Wang working in condensed matter, we cannot fully trust the association between the author's name and their scientific production. And the high mobility of researchers in a global environment does not make things easier. As a consequence, the actual numbers of active researchers in a specific area are not known precisely and for this reason any attempt to make comparisons between fields is not meaningful.

These are only a few of the problems, but they could be solved by a unique researcher identifier linking together all publications, datasets, grants and affiliation data. Such a tool would allow the researcher to connect to funders, publishers and other information providers, institutions, collaborators, conference organizers and potential employers — solving the disambiguation issue, even across different languages. This is essentially what ORCID does, so it is easy to see why funders, librarians and information providers are keen to mandate it.

ORCID started in 2012 as a non-profit, community-driven effort, which sets it apart from previous attempts like Thomson Reuters ResearcherID. It now has

well over three million registrants and over 650 member organizations. Following last year's open endorsement from organizations such as the Royal Society, eLife, PLoS and IEEE, there was a dramatic uptake in new ORCID registrations, which looks set to continue with the Springer Nature mandate and similar initiatives from other leading publishers. More organizations, including funders, academic institutions and libraries, are likely to follow with similar initiatives.

But for ORCID to live up to its full potential, there are hurdles to leap. First, the collection and validation of data need to be improved. Mandates from publishers and funders are boosting the number of new ORCID registrants, but many profiles remain empty and duplicate IDs may flourish. One solution would be to improve researchers' awareness of the actual value of their ORCID so that they put effort into curating their profiles. In this respect, proposed initiatives such as referee rewarding schemes, pushed by some publishers, or making the ORCID into a universal login to manuscript submission systems and conference registrations could provide serious incentives. Another solution is some type of manual curation. This has worked out well in discipline-specific information resources such as INSPIRE, whose success story is told by Bernard Hecker on page 523 of this issue. But it remains to be seen whether this approach is really scalable for ORCID.

Furthermore, the seamless integration between ORCID and various information providers is not easy to implement. Databases such as Crossref, Scopus and Web of Science; community resources and services such as INSPIRE, ADS and arXiv; publisher platforms, data and code repositories; and commercial services such as Google Scholar and ResearcherID all need to be able to suggest updates to ORCID and ORCID needs to be able to feed information back. Moreover, one needs to associate pre-ORCID era publications to active accounts.

ORCID faces a number of technical implementation challenges, but there is no fundamental reason why it should not live up to its promises and deliver the first fully centralized integration of all research data. Whether it will set us on course to truly improve our evaluation system is yet another question. □