A personal take on science and society

World view

Stop misusing data when hiring academics

By Cameron Neylon

Social-media impact, university rankings and online-aggregator metrics are broken — better ways to assess researchers are needed.

Look me up in InCites — a tool made by US research-analytics company Clarivate — and you'll find a biochemist who hasn't published all that much recently. Google Scholar, Google's tool for searching academic publications, shows more of my work in the past few years on research evaluation and open research. Like everyone else, I prefer the larger numbers — and Google Scholar provides them — but they come with more errors, owing to how Google's algorithms crawl and index author lists. My problems are relatively minor; one colleague has tens of citations according to InCites, and tens of thousands according to scrapers such as Google Scholar.

Researchers, especially those early in their careers, need to know how to marshal sources of evidence, such as these publication counts, to make their case to hiring and promotion committees. These tools are treated as trusted sources, despite the fact that they can give very different values. Other data, such as retweets or 'likes' on online videos, are sometimes used as a proxy for societal impact, but the relevance of these is even more questionable when used inappropriately.

This is a serious and deeply ironic problem across the scientific enterprise. As researchers, we are used to using partial, imperfect, incomplete data to make decisions and draw conclusions. Those imperfections are smoothed out through statistical processes, error calculations and good research practice. But best practices are often not applied in the evaluation of the publishing records of researchers: where are the error bars on a tenure decision, university ranking or grant application?

Policy, hiring, funding and promotion decisions are being built from this shaky evidence. If these pieces of evidence were research data, their collection, description, analysis and interpretation would never pass peer review.

This problem spans institutions as well as individuals and disciplines. My colleague Karl Huang, an open-knowledge researcher at Curtin University in Perth, Australia, and I investigated the data underlying university rankings (C.-K. K. Huang *et al. Quant. Sci. Stud.* **1** 445–478; 2020). We created a simple citation-based ranking of 155 universities, and fed it data from each of three sources: Web of Science, Scopus and Microsoft Academic, all of which are tools for searching publication records. Three universities shifted more than 110 places, and 45 moved more than 20, when the data source changed.

It shouldn't be a surprise that different sources and ranking approaches give different results. But we continue to

> **" Best practices are often not applied in the evaluation of publishing records: where are the error bars on a tenure decision?"**

Cameron Neylon is professor of research communications at Curtin University in Perth, Australia, and co-lead of the Curtin Open Knowledge Initiative.
e-mail: cameron.neylon@curtin.edu.au

ignore the differences — and make policy, funding and career decisions as though any individual metric can provide an answer. So we're getting many crucial decisions wrong — at the individual and institutional levels.

What needs to change? The policy landscape has shifted in the past decade. The Agreement on Reforming Research Assessment, published last week (see go.nature.com/3pmwd), answers the Paris Call on Research Assessment to evaluate research on "intrinsic merits and impact, rather than on the number of publications and where they are published, promoting qualitative judgement provided by peers, supported by a responsible use of quantitative indicators". In other words, its writers are sick of out-of-context numbers, too. It follows the 2013 San Francisco Declaration on Research Assessment and the 2015 Leiden Manifesto, which both called for similar policy shifts.

To make these calls effective, academia needs a cultural change in terms of the evidence that is used to evaluate research output. This will happen only when the entire enterprise demands higher standards. We should tell the stories behind our work and success more qualitatively — with more meaningful words and fewer meaningless numbers. This would better respect the variety of disciplines and the many ways that researchers make an impact.

Senior researchers should be critically evaluating the quality of evidence presented when judging job and grant applicants or conducting departmental reviews. And we should support early-career researchers by creating guidelines and training to help them to prepare the best possible cases for advancement.

It is unfair but inescapable that much of the work will fall on the shoulders of early and mid-career researchers, for whom evaluations are most crucial. They have a choice of whether to provide more rigorous and complete evidence in their applications or just the usual numbers. But this is an opportunity to reshape the stories of their research, while making the assessment of their work fairer.

These changes are already occurring. The policy landscape is shifting following the Agreement on Reforming Research Assessment and similar initiatives in many countries. Increasingly, evaluations for grants, promotions and jobs require qualitative cases supported by quantitative evidence. More senior scholars are asking for higher standards of evidence for research assessment. And I am increasingly seeing junior scholars making sophisticated, rigorous and diverse cases for the value of their research to promotion panels or in grant evaluations.

Real change will occur only when those being evaluated are prepared to show the real value and impact of their research, beyond citation counts, retweets or h-indices. I might like the larger numbers, but I'd prefer to work in a world fed by informative ones.