

TITLE: In search of El DORAdo

Howy Jacobs is at the University of Tampere and a regular columnist for *EMBO reports*

In 2012, a group of scientists, editors and representatives of research organizations got together at the annual meeting of the American Society of Cell Biology and formulated a brief declaration of principles on how publication data (bibliometrics) should be used (or not) in evaluating research merits. This 'San Francisco Declaration on Research Assessment', or DORA (sfdora.org), has today been signed by more than 20,000 of us, and is widely considered the benchmark for best practice in research evaluation.

But has it really been implemented, and has it eliminated, as intended, the widespread misuse of citation metrics, most notoriously journal impact factors, in making decisions on hiring, promotions, research grant awards and even on publication itself?

Several years ago, I participated in a broad evaluation of a European university. During our initial panel meeting I suggested that we should record a decision that we would conduct all of our work in line with DORA and adhere to its principles. There were nods of agreement around the table, but I soon realized that most of my fellow assessors had either paid little attention to what DORA is all about or hadn't heard of it at all.

One of the most common misunderstandings I have come across is that DORA is only about journal impact factors, and all that is needed to rectify the problem is 'a more perfect metric'. Whilst some bibliometric indicators are clearly worse than others, this misses the main point. The impact and significance of scientific research cannot be numerically measured at all; it can only be assessed by a broad range of largely subjective criteria, mostly the judgement of those in the same or a related field, as well as by stakeholders in the wider academic community or society beyond the ivory tower. Unlike the Leiden Manifesto that would replace the impact factor with just another set of multidimensional metrics (leidenmanifesto.org), DORA exhorts us to apply principles of merit and significance when we judge research findings and those who generated them.

Another error I have encountered is the idea that DORA is only about assessing individual scientists, whilst raw bibliometrics are perfectly fine for judging whole departments, institutes, universities or even countries. I have often argued against such assertions based on the exact wording of this, that or the other clause of the original declaration, as if it were a badly written insurance policy that we are happy to wriggle out of, if we possibly can. DORA was drafted by science professionals, not by constitutional lawyers or insurance brokers, as a set of aspirational guidelines for the scientific community.

It was recently brought to my attention that a well-known European university that should have known better had explicitly asked its academics, who were participating in a broad research evaluation exercise, to state the current journal impact factor associated with every article in their personal bibliography. In other words, the flagrant opposite of DORA recommendations. As far as I know, all who simply ignored this injunction suffered no thunderbolt from heaven or other punishment. But this is far from an isolated example: a recent survey by McKiernan *et al* (2019) found that 40% of research-intensive institutions in the USA and Canada still use journal impact factor as a criterion in review, promotion and tenure decisions. I recently saw a job advert for a professorship in a leading German university, in my own field of mitochondrial physiology, where applicants were asked similarly to indicate the impact factors associated with all of their publications. I did not apply.

The trend towards 'a broader definition of impact' may be thought of as an improvement, but it suffers from similar defects. I don't believe the significance of scientific work can be measured by the number of retweets, by the financial worth of start-up companies launched, by the length of one's lists of international research collaborators, or by how many letters of recommendation we wrote in support of our friends. All of these are ephemeral measures of fashion and self-promotion. Indeed, the very concept of excellence as measured volumetrically in any way whatsoever has been questioned as meaningless and increasingly devalued (Binswanger, 2013).

The reasons why raw bibliometrics and any other volumetrics are such an unreliable and unfair basis for decision-making and actually limit scientific progress have been endlessly discussed for many years and I do not need to rehash the arguments here. I am simply disappointed that, more than 8 years after DORA was launched, it has had such limited impact.

I don't pretend to know all the reasons why, but one of them must surely be that research assessment is rarely properly resourced and rewarded; instead, reducing everything to a few headline numbers is far less costly in time and resources. Properly judging the merits of a fellow scientist is an extremely complex and time-consuming task. It requires extensive reading of the candidate's published work. It demands delving into how the candidate's work is perceived by the wider community of scholars, not just the chief editors of a few elite journals. It ideally includes assessing the candidate's contributions to teaching, mentoring and reviewing. To be done properly it should at the very least involve interviewing the candidate themselves, as well as close associates and peers. It cannot be undertaken by collaborators, or former students or superiors of the candidate, because their careers are likely to be intertwined too closely to allow for the necessary independence. Doing all this is clearly a major undertaking, requiring extensive time, expertise and freedom from other tasks, not just a cursory reading on the plane.

I believe the only way this can change is if funders and institutions themselves realize that they are spending their limited resources very poorly, unless they provide space and incentives for reviewers to devote the time and energy needed to do their job properly. In the long run, awarding grants, fellowships or promotions to the wrong people is a far costlier waste.

The expense implicit in proper research assessment is an overhead cost of research. One way it could be incentivized is if institutions were offered a bonus in their audited overhead percentages, if they can document that their academics conduct reviews in accord with DORA and as laid down in a rigorous code of conduct within their normal work duties. As part of this, working hours devoted to such tasks should be reported in the way we are already required to report working hours spent on particular research projects.

Alternatively, funders and institutions could support academics directly, by realistic extramural payments that reflect the time needed to do reviewing tasks in ways that demonstrably evaluate the real content and significance of research outputs, the logical ingenuity of new research proposals and the intrinsic value of a person's career choices to human progress. Research assessment is a core element of academic work and should be compensated properly by those who rely on it.

Till then, it's hardly surprising that reviewers have little motivation and devote little time to doing more than merely scanning a publication list for papers in top journals, with little regard for what authors actually contributed, or how original is their thinking.

In summary, I am still searching for El DORAdo, the university, research organization or country, which not only advertises that it abides by DORA, but actually implements it fully, both to the letter and in spirit.

CONFLICT OF INTEREST

The author declares that he has no conflict of interest.

McKiernan EC et al (2019)

doi: 10.7554/eLife.47338

Binswanger M (2013) doi: 10.1007/978-3-319-00026-8_3