Dear DORA

Six months have passed since the publication of the San Francisco Declaration on Research Assessment (DORA). DORA recommended that the evaluation of research portfolios of individuals, institutions or even entire countries should be based on the real content of research, rather than the arbitrary use of impact factor (IF) points. The original signatories included the editors of some of the most prestigious academic journals, directors of top research institutes, as well as many Nobel laureates. Thousands of scientists have subsequently added their voices. DORA can no longer be portrayed as a statement of righteous indignation from an elite pressure group, but has become an expression of the collective view of the worlds' scientists.

Writing as an editor, one of the most astonishing things about the way a journal's IF is perceived is that most scientists consider it essentially immutable. I recently asked the members of my own laboratory to write down what they thought the IF is, without consulting any source. Only about half of them were aware that it had anything to do with citations, and less than a handful were close to the truth. This may account for its widespread misuse. Authors assume that the IF of a journal automatically guarantees their paper a specific level of exposure and recognition, as if it were a badge of quality delivered by an august body of heavenly sages.

The number of citations of individual papers published in any given journal typically follows a highly skewed distribution. A few papers attract many citations, whilst the vast majority are cited infrequently. Citation distributions also vary between journals, between subject areas and over time. If IF were calculated as a median it would look quite different. For assessing individual works, a journal's IF is simply meaningless.

DORA set out to eradicate the tyranny of the impact factor, enjoining funding agencies and institutions to ignore it as a mark of quality, and instead focus on the actual scientific value of published work. The problem is then how to recognize and measure it.

One unfortunate trend that I have noticed is that scientists and institutions increasingly decorate their CVs with actual numbers of citations of each published paper, along with aggregate metrics such as the *h*-index. But we need to ask whether personalized publication metrics are any more reliable an indicator of scientific quality than the cruder system they replace. This was certainly not the intention of the authors of DORA.

Where an article is published is still a major factor influencing how often it is cited. Authors still prefer to cite papers in 'good journals', even if very similar findings were reported elsewhere at the same time. Individual citation metrics are therefore distorted in similar ways as journal-based metrics. But let's imagine that we abolished all journals and instead operated a system whereby academic papers that passed peer-review were published in a single public repository. Would individual citations then be a reliable indicator of quality? A paper can be cited for a variety of reasons. It could report a trail-blazing discovery, or a useful new technique. But it might also just arouse controversy and subsequently prove erroneous. The reputation of the senior author remains important. The field trusts work reported by a distinguished investigator with a stellar track record more than that of a newbie in a far-away country of which little is known. A paper that is frequently cited may indicate originality, but it may also point to the opposite, that is, that the author has just jumped on a bandwagon. It may reflect an easy choice of topic, where

much can be learned by obvious and straightforward methods, rather than a painstaking analysis using techniques few can understand. Or it may demonstrate careful attention to serendipitous findings, leading to a new paradigm. And then we have the unsolved problem of the number of co-authors and their order. Is 7th authorship on a paper cited 1,000 times worth more than first authorship on a paper cited 70 times?

Rather than design ever more elaborate ways of objectifying such criteria, we need to follow the spirit of DORA and go back to more traditional means of peer-review. Assessment exercises should be based on a detailed reading of the papers of the person or unit being evaluated, perhaps accompanied by a written or oral justification of their main points and significance.

Even if we excluded all numerical measures of scientific output, we would still face the problem of divergent opinions. An evaluation panel split down the middle on a given portfolio will tend to trash it, even if half the panel thought it is the best science they ever read. Reviewers from Britain, where 80% is a top mark in an exam, may under-rate the best projects. The significance of research in niche areas might be invisible to most outsiders. Conversely, assessors drawn from too narrow a research area may fail to see the broader relevance.

At some point we are going to need a daughter of DORA; an agreed but flexible set of guidelines whereby scientific output can be more meaningfully judged, even if such assessment is laborious and imperfect. In the end, combining many subjective opinions is likely to prove safer and fairer than any supposedly objectified system.

Howy Jacobs

EMBO reports (2013) 14, 947; doi:10.1038/embor.2013.157