

CURRENT SCIENCE

Volume 109 Number 5

10 September 2015

GUEST EDITORIAL

Damming the metric tide

‘There is a tide in the affairs of men.
Which, taken at the flood, leads on to fortune;
Omitted, all the voyage of their life
Is bound in shallows and in miseries.
On such a full sea are we now afloat,
And we must take the current when it serves,
Or lose our ventures.’

Thus spoke Brutus in Shakespeare’s *Julius Caesar*, plotting the emperor’s downfall, but the words have an uncanny resonance today among working scientists, beset by a rising tide of metrics.

The primary appeal of a life in science is the chance to pursue one’s curiosity about the world. In most cases this involves taking a mathematical measure: the parameterization of the manifold phenomena of physical and human nature is our stock-in-trade. For the vast majority of scientists, numbers are our friends. But there is growing concern about the numbers used to ‘measure’ scientists. We seem increasingly to be swamped by them: citation counts, journal impact factors, *h*-indices, grant scores, university rankings, and – the newest kid on the block – altmetrics. Metrics are tagging almost every aspect of academic life, dominating the management of research, arbitrating the success or failure of careers, and raising concerns about research reproducibility and fraud¹. Who among us does not imagine themselves floating on a sea of numbers, or think that we have to swim with the current ‘or lose our ventures’?

And yet, as scientists, we cannot ignore the numbers. Indeed it is our duty to understand them and that is precisely what ‘The metric tide’², a new independent UK report of the role of metrics in research assessment and management has sought to do.

The report was commissioned in 2014 by the then Minister for Universities and Science, David Willetts, primarily to examine the potential contribution of metrics to the process of the research excellence framework (REF), the peer review-based evaluation of university-based research that is undertaken in the UK every six years by the Higher Education Funding Council for England (Hefce). The review was conducted by a diverse steering group – of which the present author was a member – with representatives from funding agencies, publishers, research managers and various academic disciplines, and chaired by James Wilsdon (University of Sussex, UK). The steering group conducted the review in a highly consultative

manner that involved an open call for evidence and numerous workshops and meetings and dug deep into its brief. As part of the assembled evidence base, steering group members Paul Wouters and Mike Thelwall compiled an extensive review of the current literature on bibliometrics and scientometrics; and Hefce, performed a correlation analysis on the outcomes of the 2014 REF to determine whether any of the most commonly used metrics could have been used to predict the result.

After a year of deliberation, the final report was published on 9 July 2015 (alongside the literature review and correlation analysis). There is a handy executive summary that captures the main messages, but I would urge people to make time to read the report itself.

The report recognizes that the culture of quantitative, target-driven performance management that has grown up within commercial companies has also seeped into the public sector and is now pervasive in research management within universities. The acquisition of quantitative intelligence by universities enables them to account for public spending – a healthy practice in an open society – and to compete in a global market for researchers, funding and students. But the increasing use of metrics has been accompanied by a great deal of uneasiness among academics. Numerical indicators may provide some useful information for university leaders and research managers, since they are seen as a way to reduce the complexity decision-making between competing priorities to terms that are more objective and comprehensible. However, warns the report, in many cases metrics offer the allure but not the substance of objectivity. They need always to be understood in context and used intelligently.

The review found that there was no case for shifting to a REF process that would be based entirely on metrics – not least because it was clear from the correlation analysis that metrics were unable to accurately predict the judgements of the assessment panels that surveyed the UK research base in 2014. The core difficulty is that metrics lack the coverage or trustworthiness to supplant judgement on the quality of research activity in all its forms and disciplines. Peer review, for all its flaws, remains the gold standard since, in the view of the steering group, ‘no set of numbers, however broad, is likely to be able to capture the multifaceted and nuanced judgements on the quality of research outputs that the REF process currently provides’.

This applies not just to the REF. At all levels of research evaluation – individual, departmental, institutional, national and international – the review determined that metrics should only be used to *support* the judgements of experts or peers. That is not to reject metrics out of hand. There is recognition that metrics can inform human evaluations. They may be particularly useful in helping to challenge the biases inherent in peer review – for example, in testing assumptions about the influence diversity (of disciplines and researchers) and gender on research performance. But the message is clear: the use of metrics needs to be responsible.

That idea of responsible metrics is at the heart of the report and is fleshed out in five dimensions. Metrics should be robust – as accurate and appropriate as possible; humble – in playing a supporting rather than a leading role in evaluation; transparent – their acquisition and use should be open to interrogation; diverse – they should account for variation by field and researcher; and reflexive – open to modification so that unintended effects can be mitigated.

These various dimensions of responsible metrics are used to structure the 20 recommendations in the report, which are both principled and practical. They enshrine a sensitivity to the power of language, recommending first that the term ‘indicator’ be used in preference to ‘metric’, since quantitation of research outputs is usually no more than a proxy for our rather ill-defined notions of quality.

The recommendations also charge all stakeholders – universities, funders, publishers and researchers themselves – with the task of developing clear statements of principle on how they will make use, if any, of quantitative indicators in research assessment. These provisions are particularly important in addressing the need for ongoing dialogue between these groups to determine shared values and a shared understanding of what is meant by quality research. That dialogue is also an important part of the process of balancing research activity within the diverse portfolio of activities that constitute academic life.

The recommendations also implicitly recognize that the success of that dialogue is dependent on the transparency of the information used to assemble indicators – those being judged should have access to any underlying data used for assessment – and therefore pose a direct challenge to the commercial products that are used by universities (for example, the impact factors generated by Thomson-Reuters or the various systems of university ranking).

On a more practical level, the report calls for improvements in the efficiency and interoperability of the infrastructure used to gather information on research inputs and outputs. This should reduce the cost and increase the reliability of the quantitative information used to assemble indicators of research performance. Only through such steps is there any likelihood of reducing the burden of assessment – the REF is widely viewed as excessively onerous within the UK academia. Though the report sees no immediate prospect of moving the REF to a metrics-

only exercise, it leaves the door open for critical exploration of the added value that quantitative data might bring to its evaluations of university research.

Looking to the future, the report advocates the establishment of a Forum for Responsible Metrics to enable various stakeholders to work through issues raised by implementation of the report and, on a mischievous closing note, has set up a website (www.ResponsibleMetrics.org) to encourage sharing of good practice, but also to seek nominations for an annual Bad Metrics award for ‘the most egregious example of inappropriate use of quantitative indicators in research management’.

Since its publication in early July, reaction to the report has been broadly positive, which is a testament to the independence and robustness of the review process. Its introduction of the concept of responsible metrics may not break new ground, but it carries forward ideas that are embedded in the San Francisco Declaration on Research Assessment (Dora)³ and the Leiden Manifesto⁴, which have advocated corrective measures to avoid the misuse of performance metrics, and does so with a weight of evidence and official sanction that will not be easily dismissed.

Nevertheless, reports on their own do not solve problems. Fine arguments do not always rub up well against the blunter realities of university life. Meera Sabaratnam (School of Oriental and African Studies, University of London) welcomed the publication of the report, but sounded cautionary note in drawing attention to the structural problems within academia⁵ – not least the widespread use of metrics by time-poor academics and research managers, who frequently have to make assessments outside their fields of knowledge, and the powerful hold that the rankings culture has over universities in an increasingly competitive global market.

We cannot avoid the business of research evaluation – indeed, it would be irresponsible to do so. But we have to remember that it is not ultimately about numbers, even in science. Lives are at stake here, in terms of careers, achievements and human fulfilment. We cannot hope to turn the tide – the quantification of the world will only increase as our technology develops – but we might yet find a boat in which to navigate its currents with more control. I hope ‘The metric tide’ might enable more academics to articulate a vision of academic life that is rich, rewarding and meaningful.

1. Casadevall, A. and Fang, F. C., *Infect. Immunol.*, 2012, **80**; doi:10.1128/IAI.06183-11.

2. <http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/Title-104463.en.html>

3. <http://www.ascb.org/dora/>

4. Hicks, D. *et al.*, *Nature*, 2015, **520**, 429–431; doi:10.1038/520429a.

5. <http://thedisorderofthings.com/2015/07/09/can-metrics-be-used-responsibly-why-structural-conditions-push-against-this/>

Stephen Curry

Imperial College, London, UK
e-mail: s.curry@imperial.ac.uk