



Assess the real cost of research assessment

The Research Excellence Framework keeps UK science sharp, but the process is overly burdensome for institutions, says Peter M. Atkinson.

Twas the week before Christmas, and all through the United Kingdom, scientists were waiting nervously to see how many glittering prizes the government would stuff into their stockings. Those prizes — the results of the Research Excellence Framework (REF) exercise, to be announced on 18 December — will go some way towards determining which researchers in UK universities have a happy New Year.

The scale and importance of this assessment of publicly funded research is unique to the United Kingdom. Run every five years or so, the REF system grades the quality of research in dozens of fields across more than 100 institutions, and allocates government grant money accordingly. The winners enjoy high-quality ratings for their academic departments and the guarantee of a hefty chunk of cash to support their research. A poor rating can see a department starved of money or even closed down.

The government argues that this regular scrutiny has helped to consolidate the United Kingdom's place as a global scientific superpower. And an institution with an excellent rating in physics, say, or chemistry can use it to attract staff and students. But the REF comes at a heavy cost — the amount of time and work it takes institutions and staff to prepare submissions.

Work is already under way to prepare for the next exercise, expected to run in 2020. All involved should also start to think about how to do it differently, to keep the good points but minimize the workload.

Perhaps the largest burden for institutions is that of choosing which researchers will represent each subject in the assessment. Although it is departments and disciplines that are ultimately graded, their grades are based mainly on the outputs of individuals who work in them. But there is a tension here. Funding is per head, so of two equally rated departments, the one that submits the work of more researchers receives more money. But as the number of scientists included goes up, the overall quality of the research submitted goes down — even the very best departments have a limited number of truly world-leading researchers.

A chemistry department of 60 researchers, for example, can agonize over whether to submit the research of 50 or 40 of them. To make the decision, it will do its own assessment of the quality of each scientist's work, then rank the results and try to calculate where to draw the line between who is submitted and who is not. The department must not only grade the research of its own scientists, but also grade it according to how it thinks the REF will do so. The department must also consider where

departments at rival institutions are likely to draw their own lines. But, of course, there is rarely any information on a competitor's strategy. So game theory comes into play, but with few data to drive decision-making.

In my own research, I have found that such judgements are imprecise and vary to a large degree. Why? Because uncertainty is always present. Researchers asked to rate the quality of a colleague's work, from 0 to 10, for example, will rarely come up with the same score, and this uncertainty makes internal selection all the harder. Where does this leave the REF? Although the overall strategic effect of the exercise has been positive for the quality of UK science, the amount of effort it requires of institutions deserves a rethink.

More of the process could be automated, using 'big data' and bibliometric and machine-learning approaches. To reduce the workload on institutions — most of which already subscribe to systems that capture the real-time information needed — the REF should assess the outputs of all eligible staff, removing much of the selection burden. A machine cannot yet judge the quality of research output, but there are surrogates. For many subjects, bibliometric analysis can leverage the peer-review process that already occurs through publication, as well as the peer assessment implicit in citation data. (An independent review of the use of such metrics in a future REF was launched this year.)

The REF includes other subjective judgements of quality, including — for the first time this year — the socio-economic impact of research.

These impact reports are written specifically for the REF and so add considerable effort to the process. And it is arguably harder for the REF to judge and compare quality in this area. There is no guarantee, for example, that a spin-off company that generates 200 jobs and £20 million (US\$31 million) in investments will be judged to have more impact than a spin-off that generates 20 jobs and £2 million in investments. Automation is not possible here, but there is room for greater standardization of the dimensions by which impact is assessed and the criteria against which quality is judged.

As institutional access to big data increases and technology improves, it makes sense to use all the data available to inform judgements. An obvious benefit is that the REF could be updated annually on the basis of an electronic snapshot. These changes would not make the REF perfect, but it is not perfect now. They would, however, reduce its burden and allow institutions to focus on research. ■

Peter M. Atkinson is professor of geography at the University of Southampton, UK.
e-mail: p.m.atkinson@soton.ac.uk

THE
AMOUNT OF
EFFORT
THAT THE REF
REQUIRES OF
INSTITUTIONS
DESERVES A
RETHINK.

➔ NATURE.COM
Discuss this article
online at:
go.nature.com/occnga