Lord Stern's review of the Research Excellence Framework - response form

The call for evidence is available at: www.gov.uk/government/consultations/research-excellence-framework-review-call-for-evidence

The closing date for responses is **Thursday 24 March 2016**.

Please return completed forms to:

Hannah Ledger
Research Strategy Unit
Department for Business, Innovation and Skills
1 Victoria Street
London
SW1H 0ET

Email: REFreview@bis.gsi.gov.uk



Comments:

Click here to enter text.

# Questions

Name: Professor James Wilsdon

Email: j.wilsdon@sheffield.ac.uk

Address: Faculty of Social Sciences, 2nd Floor, Edgar Allan House, 241 Glossop Road
University of Sheffield, Sheffield S10 2GW

Name of Organisation (if applicable): Independent Review of the Role of Metrics in Research Assessment and Management

Please check the box that best describes you as a respondent to this consultation

| | **Respondent type** |
|---|---|
| ☐ | Alternative higher education provider (with designated courses) |
| ☐ | Alternative higher education provider (no designated courses) |
| ☐ | Awarding organisation |
| ☐ | Business/Employer |
| ☐ | Central government |
| ☐ | Charity or social enterprise |
| ☐ | Further Education College |
| ☐ | Higher Education Institution |
| ☐ | Individual (Please describe any particular relevant interest; teaching staff, student, etc.) |
| ☐ | Legal representative |
| ☐ | Local Government |
| ☐ | Professional Body |
| ☐ | Representative Body |
| ☐ | Research Council |
| ☐ | Trade union or staff association |
| ☒ | Other (please describe) |

If you selected 'Other,' please give details

We are responding here collectively as members of the Independent Review of the Role of Metrics in Research Assessment and Management (which was set up by HEFCE at the request of David Willetts, when he was Minister for Universities and Science). The views expressed do not represent the positions of the organisations or employers of group members, but draw on our original report *The Metric Tide* (July 2015), and subsequent deliberations in the context of the Stern Review. Our report is available here: http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/Independentresearch/2015/The,Metric,Tide/2015_metric_tide.pdf

The membership of the group is shown below.

- James Wilsdon, Professor of Research Policy, University of Sheffield (Chair);
- Dr Liz Allen, Former Head of Evaluation, Wellcome Trust (now F1000);
- Dr Eleonora Belfiore, Associate Professor in Cultural Policy, Centre for Cultural Policy Studies, University of Warwick;
- Sir Philip Campbell, Editor-in-Chief, Nature;
- Professor Stephen Curry, Department of Life Sciences, Imperial College London;
- Dr Steven Hill, Head of Research Policy, HEFCE;
- Professor Richard Jones FRS, Pro-Vice-Chancellor for Research and Innovation, University of Sheffield (representing the Royal Society);
- Dr Chonnettia Jones, Head of Insight and Analysis, Strategy, Wellcome Trust;
- Professor Roger Kain FBA, Dean and Chief Executive, School of Advanced Study, University of London (representing the British Academy);
- Dr Simon Kerridge, Director of Research Services, University of Kent, and Chair of the Board of the Association of Research Managers and Administrators;
- Professor Mike Thelwall, Statistical Cybermetrics Research Group, University of Wolverhampton;
- Jane Tinkler, Social Science Adviser at the Parliamentary Office of Science and Technology;
- Dr Ian Viney, MRC Director of Strategic Evaluation and Impact, Medical Research Council head office, London;
- Paul Wouters, Professor of Scientometrics & Director, Centre for Science and Technology Studies (CWTS), Leiden University.

**NB We have chosen to respond to questions 1, 4, 5 & 6 as these relate most directly to issues that we analysed in The Metric Tide.**

# Section 1

**The primary purpose of the REF is to inform the allocation of quality-related research funding (QR).**

1. **What changes to existing processes could more efficiently or more accurately assess the outputs, impacts and contexts of research in order to allocate QR? Should the definition of impact be broadened or refined? Is there scope for more or different use of metrics in any areas?**

Debates about the REF suffer from conflicting interpretations of its purpose, and the Stern Review presents an important opportunity to clarify this question. While the consultation is right to say that the allocation of quality-related funding is *an* important purpose, it is not *the* primary purpose. Rather, as defined by HEFCE, the exercise has three purposes:

1. ***Allocation****:* "Our higher education funding bodies use the assessment outcomes to inform the selective allocation of their grant for research to the institutions which they fund";

2. ***Accountability****:* "The assessment provides accountability for public investment in research and produces evidence of the benefits of this investment."

3. ***Benchmarking****:* "The assessment outcomes provide benchmarking information and establish reputational yardsticks, for use within the higher education (HE) sector and for public information."

In our view, as the exercise has evolved through successive cycles, it has accrued two additional functions:

4. ***Influencing research cultures and behaviours****.* Initially the RAE incentivized productivity. But in the most recent cycle, it has been the biggest driver of a serious shift towards embracing and valuing impact through the system. The next cycle is likely to be used to nudge change through the system in other areas, through its requirements over open access and (perhaps) the use of unique identifiers like ORCID IDs. Given the reach of the REF into all corners of university research it is arguably the most efficient and effective way of introducing changes of this kind.

5. ***Performance management within HEIs***. This is the most controversial purpose, but there is no denying the fact that across the university system, the REF is now used as a *de facto* management framework for research activities. Whether one thinks this is good or bad, RAE/REF has progressively been internalised and institutionalised as a convenient framework for HEI leaders and managers to monitor performance and make strategic (and sometimes unpopular or controversial) management decisions.

Many of the debates we have as a sector about the REF result from people talking, literally, at cross-purposes. If you insist that the exercise is all – or primarily – about QR allocation, then your approach to the design of the exercise will be different to someone who sees a value in three, or all five, of these purposes. You might, for example, argue that metrics more readily provide an answer to a narrowly-defined algorithmic challenge of allocation. Similarly, in the debate over costs and burden, the range of purposes you include will fundamentally alter your cost-benefit equation.

As a method for allocating research funding, the REF is already highly efficient, particularly when compared to other approaches e.g. research council grant processes (for which "full-cost" accounting, taking into account the costs of proposal development, simply isn't available). As the Technopolis study of REF 2014 has noted, a significant proportion of the increased costs in 2014 were associated with the addition of the impact case studies. These costs may fall next time around as HEIs will be more experienced in producing the case studies, but there is no metric-based alternative that can be applied here.

No matter how the RAE/REF had been designed or configured, HEIs have continued to pursue strategies to maximise their outcomes. As a result, the formal and informal functions of the RAE/REF have become deeply embedded in the HE system.

The success of the RAE/REF in improving research productivity and performance (and now in measuring impact) is directly related to the development of the UK's research system: institutional strategies have co-evolved with the assessment process. This makes it hard to 'decouple' the role of the REF from the wider performance of the UK system. Other countries have national systems for research assessment, which differ from the REF but also work well. In the UK or elsewhere, these systems cannot be understood separately from their cultural, institutional and political context.

One of the arguments we make in *The Metric Tide* is that, even if you radically simplify or metricise the REF, it won't remove the need for functions 4 and 5 within the UK's system. A lot of the time, effort and energy that currently goes into managing the REF as we know it would simply be diverted into optimising performance against university rankings, or for other funders, using slightly different criteria (and no doubt, requiring the purchase of costly bibliometrics, altmetrics and related analytical services to underpin institutional strategies).

The positive and negative effects of HEI maximising strategies must remain a key consideration in the design of the next exercise, particularly when thinking about the scope for increasing or changing the role of metrics.

A great deal of evidence and analysis of the potential and the pitfalls of using metrics in the assessment of outputs, impacts and environment is provided in *The Metric Tide*. We hope that the Stern Review is able to draw productively on some of this work.

In summary, when our review looked at whether metrics could be used to capture the impacts of research, both we (and the separate review by Digital Science/Kings College London) found that robust metrics simply do not exist for most forms of impact. Our concern would be that a metric-based approach would mean that the concept of impact might narrow and become too specifically defined by the ready availability of indicators for some types of impact and not for others.

Similar constrictions would occur for research outputs in a metric-based system, where (chiefly because of problems of calibration, coverage, robustness and transparency) certain types of research would become inappropriately elevated over others. When HEIs responded to such a situation through their maximising strategies, this could have a profoundly constricting effect on research performance and diversity. The effects on women, BME and early career researchers, and researchers in arts and humanities disciplines, would be particularly problematic.

Coupled with the notable lack of correlation between REF peer review outcomes and metrics outcomes, our review therefore concluded that it is not yet possible to supplant peer review with metrics in the assessment of outputs or impacts. We did, however, argue that the role of quantitative data could be enhanced in the next exercise, and we proposed a Forum for Responsible Metrics be established by HEFCE, Research Councils, Jisc and others to develop guidelines for data collection, standardisation and interoperability.

For the research environment component of the REF, we also believe that there is significant scope to increase the use of metrics, as long as such data were appropriately contextualised. At a minimum this would need to include information on the total size of the UOA to which the data refer. In some cases, the collection of data specifically relating to staff submitted to the exercise may be preferable, albeit more costly. In addition, data on the structure and use of digital information systems to support research (or research and teaching) may be crucial to further develop excellent research environments.

It is worth noting that other aspects of the current design of the REF are perceived to constrain research in certain ways. For instance, the requirement that all submitted staff have four research outputs may have led to HEIs pursuing maximising strategies on research output volume, which would favour certain types of research output (e.g. journal articles) over others (e.g. books). In thinking about how the system might be reconfigured, it will be crucial to think about how such constrictions can be tackled or alleviated. It is not clear that metrics are the answer, nor that HEI responses to an enhanced role for metrics in the system would not itself prove costly. Indeed, as our review notes, adding metrics-based assessment to peer review would very likely increase the overall costs of the exercise.

## Section 2

**4. What data should REF collect to be of greater support to Government and research funders in driving research excellence and productivity?**

The information that the REF produces is currently widely used in HEI management and in wider funding and policy contexts. However, as our review noted, there is scope to increase its use further. Our report recommended that HEFCE, funders, HEIs and Jisc should explore how to leverage data held in existing platforms to support the REF process, and vice versa. Further debate is also required about the merits of local collection within HEIs and data collection at the national level. BIS should identify ways of linking data gathered from research-related platforms (including Gateway to Research, Researchfish and the REF) more directly to policy processes in BIS and other departments, especially around foresight, horizon scanning and research prioritisation. There is a role here for the new High Level Landscape Mapping Group, which is being chaired by Jane Elliott, Chief Executive of the ESRC.

At the moment, it is not possible to integrate some of the information produced by and for the REF with data and information from other sources. In *The Metric Tide*, we noted the specific example of impact case studies, which could be made more useful if impact could be tracked back to any project funding e.g. from research councils, and combined with or draw upon information held in Researchfish. We would see value in the inclusion of grant funding information, and unique identifiers for researchers (ORCID), organisations (e.g. ISNI) and outputs (DOI) cited in case studies – this would enable impact to be more effectively tracked and combined with other data and evidence.

To underpin this, we see a significant need to improve the interoperability and interoperation of research information management systems, with further investment needed to improve standards and transparency. The current research information landscape is severely fragmented, with HEIs running a variety of different and incompatible software systems to manage and track their research. These systems are typically expensive and proprietary, providing “metrics” that are not demonstrably transparent or robust. We see issues in HEIs continuing to rely to a significant extent on these systems in their planning and management of research, including for the next REF.

## Section 3

5. **How might the REF be further refined or used by Government to incentivise constructive and creative behaviours such as promoting interdisciplinary research, collaboration between universities, and/or collaboration between universities and other public or private sector bodies?**

As already noted, the REF has significant effects on behaviours, largely due to efforts by HEIs to optimise their own outcome in the exercise. We consider that there are a number of areas where the REF has already promoted positive behavioural change: in

the increase in the productivity of researchers, in the generation of impact from research, and more recently in the pursuit of open access to research publications.

To improve the interoperability and management of research information, our report recommended that the REF mandate the adoption of ORCID identifiers by submitted staff. We believe this will be a key step towards increasing the uptake of ORCIDs en route to a more efficient and effective research management landscape in the UK.

The REF is clear that the “reputation” of the venue in which research outputs are published will have no bearing on how outputs are assessed, but we would note that there remain some doubts about this in the HE community. Anecdotal evidence suggests that the widely-discredited Journal Impact Factor is still widely used by HEI managers to determine selection of outputs for the REF, with consequent damaging effects on research culture. We consider this behaviour to be irresponsible, but the fact it goes on suggests deeper issues relating to academic culture, prestige and performance incentives that run deeper than just REF. To address this, we recommended that HEIs and funders both develop context-specific principles for the use of quantitative indicators in the assessment and management of research, and that these principles be well communicated and made easy to locate and understand. Our report also encouraged HEIs to sign up to the San Francisco Declaration on Research Assessment (DORA), which was echoed in the recent report by Adam Tickell on open access. At present only three UK HEIs are signed up. There is clearly scope for improvement here.

## Section 4

**6. In your view how does the REF process influence, positively or negatively, the choices of individual researchers and / or higher education institutions? What are the reasons for this and what are the effects? How do such effects of the REF compare with effects of other drivers in the system (e.g. success for individuals in international career markets, or for universities in global rankings)? What suggestions would you have to restrict gaming the system?**

The best way to restrict gaming is to limit the extent to which HEIs are able or incentivised to strategically optimise their submissions in response to REF drivers. The best way to limit negative effects on individual researchers is to limit the extent to which HEIs are able or incentivised to select staff. Here we offer some design options for consideration by the Stern Review, without arguing for any method in particular:

a. **Submit everyone.** All research active staff are required to submit (say) four research outputs, with tailoring for circumstances as in REF2014. On the upside, this would remove the advantage that HEIs seek to gain by 'ranking' researchers through internal benchmarking and/or mock REF exercises. It would in theory eliminate any negative effect on individual researchers of being 'non-submitted'. On the downside, HEIs would be incentivised to move research-active but unproductive or less prominent research staff onto teaching-only contracts, which may have an equally negative effect on those concerned. Such an approach would also unfairly reward high-output disciplines and may unhelpfully incentivise increases in output volume at the expense of focus or quality elsewhere, as those HEIs with high-output researchers would have a wider pool of outputs to choose from. It would also significantly increase the cost of the exercise. It is also worth noting that this would have a variable impact on institutions (with some already far closer to 100% submission rates than others), and would create a lack of comparability with previous exercises.

b. **Submit everything.** All research outputs attributed to the HEI are submitted (with a connected number of impact case studies etc.). On the upside, this would remove the advantage that HEIs seek to gain by carefully selecting outputs for submission, including through 'mock REFs' and the like. It would encourage a focus on quality at an earlier stage in the publication cycle. On the downside, assessing all outputs would substantially increase the cost of the exercise (though this could be mitigated through a sampling approach as discussed below). In addition, a focus on quality over quantity could limit productivity and lead to good but not 'world-leading' research going unpublished.

c. **Decouple outputs from staff.** An HEI would be required to submit a fixed number of outputs determined by the overall number of research-active staff submitted to the exercise (potentially everybody, if combined with a. above). HEIs would be free to select those outputs without a per-researcher limit. On the upside, this approach would reduce (but not necessarily eliminate) the fixed 'in-or-out' nature of staff selection, by allowing the HEI to make more fine-grained judgments about numbers of outputs. It would also reduce the pressure on low-output-volume disciplines to produce more volume (which, some have argued, has encouraged moves towards

journal article publishing and away from book publishing). On the downside, this approach would not necessarily reduce gaming, as significant work would be needed to identify outputs from the fixed pool. It would also have the effect of 'rewarding' some staff more than others for what would be a higher individual contribution to the exercise (though in practice these higher contributions might reflect the reality in some departments).

As an alternative to the use of metrics, **sampling approaches** should be seriously considered for assessment of outputs in the next REF cycle. A number of dimensions might need to be considered to ensure that the sample of outputs is fair and representative and commands the confidence of the sector:

- *Equality and diversity*. Is the sample representative of the contributions of all groups (including the nine protected characteristics in the Equality Act 2010)? [Note this may require further and more granular information on staff circumstances than is currently required]
- *Career stage*. Is the sample representative of the contributions of academics at all career stages? [Note the same point about staff details applies here.]
- *Disciplinary coverage.* Is the sample representative of the mix of disciplines undertaken by researchers at that HEI?
- *Output type*. Is the sample representative of the breadth and variety of different types of research undertaken at the HEI?

We would not recommend any approach to bias the sample towards outputs that perform in particular ways according to particular metrics, particularly if the sample selected for peer review were perceived 'outliers' to some otherwise broadly metric-based assessment. The use of metrics to choose and assess outputs would not be appropriate or responsible, given the well-documented issues with the coverage, representativeness, diversity, transparency and robustness of current and popular research metrics, and the consequent negative effects of institutional gaming to satisfy the selection metrics.

Likewise, we would not recommend that the use of metrics to guide sampling approaches be limited to particular disciplines. A key strength of the REF is that it assesses all research from all disciplines equally and fairly. To do otherwise would harm the credibility of the exercise and would risk having adverse effects on interdisciplinary research.

Final thoughts

**The Review is keen to hear of creative ideas and insights and to be open in its approach.**

7. Are there additional issues you would like to bring to the attention of the Review?

   We would welcome the opportunity to engage directly with the Stern Review panel as their deliberations advance, to offer further advice on the opportunities and limitations of greater use of metrics and quantitative data within the REF, based on the evidence we gathered and analysis we carried out for *The Metric Tide*.

Thank you for taking the time to let us have your views. We do not intend to acknowledge receipt of individual responses unless you tick the box below.

Please acknowledge this reply ☐

**IND/16/1a**