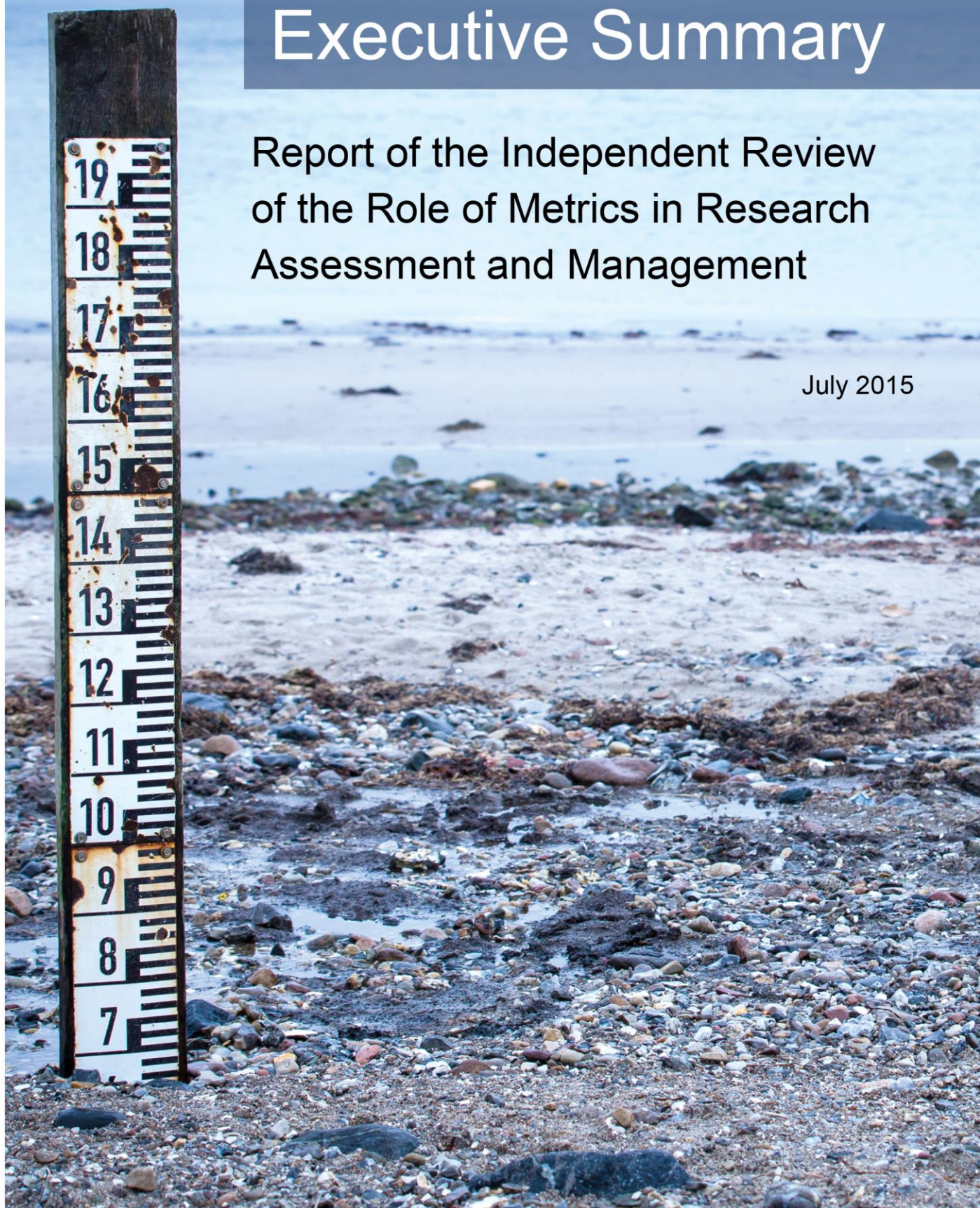


The Metric Tide

Executive Summary

Report of the Independent Review
of the Role of Metrics in Research
Assessment and Management

July 2015



Foreword

Metrics evoke a mixed reaction from the research community. A commitment to using data and evidence to inform decisions makes many of us sympathetic, even enthusiastic, about the prospect of granular, real-time analysis of our own activities. If we as a sector can't take full advantage of the possibilities of big data, then who can?



Yet we only have to look around us, at the blunt use of metrics such as journal impact factors, h-indices and grant income targets to be reminded of the pitfalls. Some of the most precious qualities of academic culture resist simple quantification, and individual indicators can struggle to do justice to the richness and plurality of our research. Too often, poorly designed evaluation criteria are “dominating minds, distorting behaviour and determining careers.”¹ At their worst, metrics can contribute to what Rowan Williams, the former Archbishop of Canterbury, calls a “new barbarity” in our universities.² The tragic case of Stefan Grimm, whose suicide in September 2014 led Imperial College to launch a review of its use of performance metrics, is a jolting reminder that what's at stake in these debates is more than just the design of effective management systems.³ Metrics hold real power: they are constitutive of values, identities and livelihoods.

How to exercise that power to positive ends is the focus of this report. Based on fifteen months of evidence-gathering, analysis and consultation, we propose here a framework for responsible metrics, and make a series of targeted recommendations. Together these are designed to ensure that indicators and underlying data infrastructure develop in ways that support the diverse qualities and impacts of UK research. Looking to the future, we show how responsible metrics can be applied in research management, by funders, and in the next cycle of the Research Excellence Framework.

The metric tide is certainly rising. Unlike King Canute, we have the agency and opportunity – and in this report, a serious body of evidence – to influence how it washes through higher education and research. Let me end on a note of personal thanks to my steering group colleagues, to the team at HEFCE, and to all those across the community who have contributed to our deliberations.

A handwritten signature in black ink, appearing to read 'J Wilsdon'.

James Wilsdon, Chair

¹ Lawrence, P.A. (2007) ‘The mismeasurement of science’. *Current Biology* Vol.17, Issue 15, pR583–R585.

² Annual Lecture to the Council for the Defence of British Universities, January 2015.

³ <https://www.timeshighereducation.co.uk/news/stefan-grimms-death-leads-imperial-to-review-performance-metrics/2019381.article>. Retrieved 22 June 2015.

Steering group and secretariat

The review was chaired by James Wilsdon FAcSS, Professor of Science and Democracy at the Science Policy Research Unit (SPRU), University of Sussex (orcid.org/0000-0002-5395-5949; @jameswilsdon).

Professor Wilsdon was supported by an independent steering group with the following members:

- Dr Liz Allen, Head of Evaluation, Wellcome Trust (orcid.org/0000-0002-9298-3168; @allen_liz);
- Dr Eleonora Belfiore, Associate Professor in Cultural Policy, Centre for Cultural Policy Studies, University of Warwick (orcid.org/0000-0001-7825-4615; @elebelfiore);
- Sir Philip Campbell, Editor-in-Chief, Nature (orcid.org/0000-0002-8917-1740; @NatureNews);
- Professor Stephen Curry, Department of Life Sciences, Imperial College London (orcid.org/0000-0002-0552-8870; @Stephen_Curry);
- Dr Steven Hill, Head of Research Policy, HEFCE (orcid.org/0000-0003-1799-1915; @stevenhill);
- Professor Richard Jones FRS, Pro-Vice-Chancellor for Research and Innovation, University of Sheffield (orcid.org/0000-0001-5400-6369; @RichardALJones) (representing the Royal Society);
- Professor Roger Kain FBA, Dean and Chief Executive, School of Advanced Study, University of London (orcid.org/0000-0003-1971-7338; @kain_SAS) (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 (orcid.org/0000-0003-4094-3719; @SimonRKerridge);
- Professor Mike Thelwall, Statistical Cybermetrics Research Group, University of Wolverhampton (orcid.org/0000-0001-6065-205X; @mikethelwall);
- Jane Tinkler, Social Science Adviser at the Parliamentary Office of Science and Technology (orcid.org/0000-0002-5306-3940; @janetinkler);
- Dr Ian Viney, MRC Director of Strategic Evaluation and Impact, Medical Research Council head office, London (orcid.org/0000-0002-9943-4989, @MRCEval);
- Paul Wouters, Professor of Scientometrics & Director, Centre for Science and Technology Studies (CWTS), Leiden University (orcid.org/0000-0002-4324-5732, @paulwouters).

The following members of HEFCE's Research Policy team provided the secretariat for the Steering Group and supported the review process throughout: Jude Hill, Ben Johnson, Alex Herbert, Kate Turton, Tamsin Rott and Sophie Melton Bradley. Members of HEFCE's Analytical Services Directorate, Hannah White and Mark Gittoes also contributed, particularly to the REF2014 correlation exercise. Vicky Jones from the REF team also provided advice.

The Metric Tide: Executive Summary

This document provides the main findings and recommendations of the Independent Review of the Role of Metrics in Research Assessment and Management, as set out in full in the main report, *The Metric Tide*. The review was chaired by Professor James Wilsdon, supported by an independent and multidisciplinary group of experts in scientometrics, research funding, research policy, publishing, university management and research administration. The full report is available on the HEFCE website at www.hefce.ac.uk/rsrch/metrics.

Scope of the review

This review has gone beyond earlier studies to take a deeper look at potential uses and limitations of research metrics and indicators. It has explored the use of metrics across different disciplines, and assessed their potential contribution to the development of research excellence and impact. It has analysed their role in processes of research assessment, including the next cycle of the Research Excellence Framework (REF). It has considered the changing ways in which universities are using quantitative indicators in their management systems, and the growing power of league tables and rankings. And it has considered the negative or unintended effects of metrics on various aspects of research culture.

The full report starts by tracing the history of metrics in research management and assessment, in the UK and internationally. It looks at the applicability of metrics within different research cultures, compares the peer review system with metric-based alternatives, and considers what balance might be struck between the two. It charts the development of research management systems within institutions, and examines the effects of the growing use of quantitative indicators on different aspects of research culture, including performance

management, equality, diversity, interdisciplinarity, and the ‘gaming’ of assessment systems. The review looks at how different funders are using quantitative indicators, and considers their potential role in research and innovation policy. Finally, it examines the role that metrics played in REF2014, and outlines scenarios for their contribution to future exercises.

The review has drawn on a diverse evidence base to develop its findings and conclusions. These include: a formal call for evidence; a comprehensive review of the literature (published as Supplementary Report I); and extensive consultation with stakeholders at focus groups, workshops, and via traditional and new media.

The review has also drawn on HEFCE’s recent evaluations of REF2014, and commissioned its own detailed analysis of the correlation between REF2014 scores and a basket of metrics (published as Supplementary Report II).

Headline findings

There are powerful currents whipping up the metric tide. These include growing pressures for audit and evaluation of public spending on higher education and research; demands by policymakers for more strategic intelligence on research quality and impact; the need for institutions to manage and develop their strategies for research; competition within and between institutions for prestige, students, staff and resources; and increases in the availability of real-time ‘big data’ on research uptake, and the capacity of tools for analysing them.

Across the research community, the description, production and consumption of ‘metrics’ remains contested and open to misunderstandings. In a positive sense, wider use of quantitative indicators, and the emergence of alternative metrics for societal impact, could support the transition to a more open, accountable and outward-facing research system. But placing too much emphasis on narrow, poorly-designed indicators – such as journal impact factors (JIFs) – can have negative consequences, as reflected by the 2013 San Francisco Declaration on Research Assessment (DORA), which now has over 570 organisational and 12,300 individual signatories.⁴ Responses to this review reflect these possibilities and pitfalls. The majority of those who submitted evidence, or engaged in other ways, are sceptical about moves to increase the role of metrics in research management. However, a significant minority are more supportive of the use of metrics, particularly if appropriate care is

exercised in their design and application, and the data infrastructure can be improved.

Peer review, despite its flaws and limitations, continues to command widespread support across disciplines. Metrics should support, not supplant, expert judgement. Peer review is not perfect, but it is the least worst form of academic governance we have, and should remain the primary basis for assessing research papers, proposals and individuals, and for national assessment exercises like the REF. However, carefully selected and applied quantitative indicators can be a useful complement to other forms of evaluation and decision-making. One size is unlikely to fit all: a mature research system needs a variable geometry of expert judgement, quantitative and qualitative indicators. Research assessment needs to be undertaken with due regard for context and disciplinary diversity. Academic quality is highly context-specific, and it is sensible to think in terms of research *qualities*, rather than striving for a single definition or measure of quality.

Inappropriate indicators create perverse incentives. There is legitimate concern that some quantitative indicators can be gamed, or can lead to unintended consequences; journal impact factors and citation counts are two prominent examples. These consequences need to be identified, acknowledged and addressed. Linked to this, there is a need for greater transparency in the construction and use of indicators, particularly for university rankings and league tables. Those involved in research assessment and management should behave responsibly, considering and

⁴ www.ascb.org/dora. As of July 2015, only three UK universities are DORA signatories: Manchester, Sussex and UCL.

pre-empting negative consequences wherever possible, particularly in terms of equality and diversity.

Indicators can only meet their potential if they are underpinned by an open and interoperable data infrastructure. How underlying data are

collected and processed – and the extent to which they remain open to interrogation – is crucial.

Without the right identifiers, standards and semantics, we risk developing metrics that are not contextually robust or properly understood. The systems used by higher education institutions (HEIs), funders and publishers need to interoperate better, and definitions of research-related concepts need to be harmonised.

Information about research – particularly about funding inputs – remains fragmented. Unique identifiers for individuals and research works will gradually improve the robustness of metrics and reduce administrative burden.

At present, further use of quantitative indicators in research assessment and management cannot be relied on to reduce costs or administrative burden. Unless existing processes, such as peer review, are reduced as additional metrics are added, there will be an overall increase in burden. However, as the underlying data infrastructure is improved and metrics become more robust and trusted by the community, it is likely that the additional burden of collecting and assessing metrics could be outweighed by the reduction of peer review effort in some areas – and indeed by other uses for the data. Evidence of a robust relationship between newer metrics and research quality remains very limited, and more experimentation is needed. Indicators such as patent citations and clinical guideline citations may have

potential in some fields for quantifying impact and progression.

Our correlation analysis of the REF2014 results at output-by-author level

(Supplementary Report II) has shown that individual metrics give significantly different outcomes from the REF peer review process, and therefore cannot provide a like-for-like replacement for REF peer review. Publication

year was a significant factor in the calculation of correlation with REF scores, with all but two metrics showing significant decreases in correlation for more recent outputs. There is large variation in the coverage of metrics across the REF submission, with particular issues with coverage in units of assessment (UOAs) in REF Main Panel D (mainly arts & humanities). There is also evidence to suggest statistically significant differences in the correlation with REF scores for early-career researchers and women in a small number of UOAs.

Within the REF, it is not currently feasible to assess the quality of UOAs using quantitative indicators alone. In REF2014, while some

indicators (citation counts, and supporting text to highlight significance or quality in other ways) were supplied to some panels to help inform their judgements, caution needs to be exercised when considering all disciplines with existing bibliographic databases. Even if technical problems of coverage and bias can be overcome, 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.

Similarly, for the impact component of the REF, it is not currently feasible to use quantitative indicators in place of narrative impact case studies, or the impact template.

There is a danger 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. For an exercise like the REF, where HEIs are competing for funds, defining impact through quantitative indicators is likely to constrain thinking around which impact stories have greatest currency and should be submitted, potentially constraining the diversity of the UK's research base. **For the environment component of the REF, there is scope to enhance the use of quantitative data in the next assessment cycle**, provided they are used with sufficient context to enable their interpretation.

There is a need for more research on research.

The study of research systems – sometimes called the ‘science of science policy’ – is poorly funded in the UK. The evidence to address the questions that we have been exploring throughout this review remains too limited; but the questions being asked by funders and HEIs – *‘What should we fund?’ ‘How best should we fund?’ ‘Who should we hire/promote/invest in?’* – are far from new and can only become more pressing. More investment is needed as part of a coordinated UK effort to improve the evidence base in this area. Linked to this, there is potential for the scientometrics community to play a more strategic role in informing how quantitative indicators are used across the research system, and by policymakers.

Responsible metrics

In recent years, the concept of ‘responsible research and innovation’ (RRI) has gained currency as a framework for research governance. Building on this, we propose the notion of **responsible metrics** as a way of framing appropriate uses of quantitative indicators in the governance, management and assessment of research. Responsible metrics can be understood in terms of the following dimensions:

- **Robustness:** basing metrics on the best possible data in terms of accuracy and scope;
- **Humility:** recognising that quantitative evaluation should support – but not supplant – qualitative, expert assessment;
- **Transparency:** keeping data **collection** and analytical processes open and transparent, so that those being evaluated can test and verify the results;
- **Diversity:** accounting for **variation** by field, and using a variety of indicators to support diversity across the research system;
- **Reflexivity:** recognising **systemic** and potential effects of indicators and updating them in response.

Recommendations

This review has identified 20 specific recommendations for further work and action by stakeholders across the UK research system. These recommendations draw on the evidence gathered and set out in the report, and should be seen as part of broader attempts to strengthen research governance, management and assessment, which have been gathering momentum, and where the UK is well positioned to play a leading role internationally. The recommendations are listed below, with targeted recipients in brackets.

Supporting the effective leadership, governance and management of research cultures

1 The research community should develop a more sophisticated and nuanced approach to the contribution and limitations of quantitative indicators. Greater care with language and terminology is needed. The term ‘metrics’ is often unhelpful; the preferred term ‘indicators’ reflects a recognition that data may lack specific relevance, even if they are useful overall. *(HEIs, funders, managers, researchers)*

2 At an institutional level, HEI leaders should develop a clear statement of principles on their approach to research management and assessment, including the role of quantitative indicators. On the basis of these principles, they should carefully select quantitative indicators that are appropriate to their institutional aims and context. Where institutions are making use of league tables and ranking measures, they should explain why they are using these as a means to achieve particular ends. Where possible, alternative

indicators that support equality and diversity should be identified and included. Clear communication of the rationale for selecting particular indicators, and how they will be used as a management tool, is paramount. As part of this process, HEIs should consider signing up to DORA, or drawing on its principles and tailoring them to their institutional contexts. *(Heads of institutions, heads of research, HEI governors)*

3 Research managers and administrators should champion these principles and the use of responsible metrics within their institutions. They should pay due attention to the equality and diversity implications of research assessment choices; engage with external experts such as those at the Equality Challenge Unit; help to facilitate a more open and transparent data infrastructure; advocate the use of unique identifiers such as ORCID iDs; work with funders and publishers on data interoperability; explore indicators for aspects of research that they wish to assess rather than using existing indicators because they are readily available; advise senior leaders on metrics that are meaningful for their institutional or departmental context; and exchange best practice through sector bodies such as ARMA. *(Managers, research administrators, ARMA)*

4 HR managers and recruitment or promotion panels in HEIs should be explicit about the criteria used for academic appointment and promotion decisions. These criteria should be founded in expert judgement and may reflect both the academic quality of outputs and wider contributions to policy, industry or society. Judgements may sometimes usefully be guided by metrics, if they are relevant to the criteria in question and used responsibly; article-level citation metrics, for instance, might be useful indicators of academic impact, as long as they are interpreted in the light of disciplinary norms and with due regard to their limitations. Journal-level metrics, such as the JIF,

should not be used. *(HR managers, recruitment and promotion panels, UUK)*

5 Individual researchers should be mindful of the limitations of particular indicators in the way they present their own CVs and evaluate the work of colleagues. When standard indicators are inadequate, individual researchers should look for a range of data sources to document and support claims about the impact of their work. *(All researchers)*

6 Like HEIs, research funders should develop their own context-specific principles for the use of quantitative indicators in research assessment and management and ensure that these are well communicated, easy to locate and understand. They should pursue approaches to data collection that are transparent, accessible, and allow for greater interoperability across a diversity of platforms. *(UK HE Funding Bodies, Research Councils, other research funders)*

7 Data providers, analysts and producers of university rankings and league tables should strive for greater transparency and interoperability between different measurement systems. Some, such as the Times Higher Education (THE) university rankings, have taken commendable steps to be more open about their choice of indicators and the weightings given to these, but other rankings remain ‘black-boxed’. *(Data providers, analysts and producers of university rankings and league tables)*

8 Publishers should reduce emphasis on journal impact factors as a promotional tool, and only use them in the context of a variety of journal-based metrics that provide a richer view of performance. As suggested by DORA, this broader indicator set could include 5-year impact factor, EigenFactor, SCImago, editorial and publication times. Publishers, with the aid of Committee on Publication Ethics (COPE), should encourage responsible authorship

practices and the provision of more detailed information about the specific contributions of each author. Publishers should also make available a range of article-level metrics to encourage a shift toward assessment based on the academic quality of an article rather than JIFs. *(Publishers)*

Improving the data infrastructure that supports research information management

9 There is a need for greater transparency and openness in research data infrastructure. A set of principles should be developed for technologies, practices and cultures that can support open, trustworthy research information management. These principles should be adopted by funders, data providers, administrators and researchers as a foundation for further work. *(UK HE Funding Bodies, RCUK, Jisc, data providers, managers, administrators)*

10 The UK research system should take full advantage of ORCID as its preferred system of unique identifiers. ORCID iDs should be mandatory for all researchers in the next REF. Funders and HEIs should utilise ORCID for grant applications, management and reporting platforms, and the benefits of ORCID need to be better communicated to researchers. *(HEIs, UK HE Funding Bodies, funders, managers, UUK, HESA)*

11 Identifiers are also needed for institutions, and the most likely candidate for a global solution is the ISNI, which already has good coverage of publishers, funders and research organisations. The use of ISNIs should therefore be extended to cover all institutions referenced in future REF submissions, and used more widely in internal HEI and funder management processes. One component of the solution will be to map the various organisational identifier systems against ISNI to allow the various existing systems to interoperate. *(UK HE Funding Bodies, HEIs, funders, publishers, UUK, HESA)*

12 Publishers should mandate ORCID iDs and ISNIs and funder grant references for article submission, and retain this metadata throughout the publication lifecycle. This will facilitate exchange of information on research activity, and help deliver data and metrics at minimal burden to researchers and administrators. *(Publishers and data providers)*

13 The use of digital object identifiers (DOIs) should be extended to cover all research outputs. This should include all outputs submitted to a future REF for which DOIs are suitable, and DOIs should also be more widely adopted in internal HEI and research funder processes. DOIs already predominate in the journal publishing sphere – they should be extended to cover other outputs where no identifier system exists, such as book chapters and datasets. *(UK HE Funding Bodies, HEIs, funders, UUK)*

14 Further investment in research information infrastructure is required. Funders and Jisc should explore opportunities for additional strategic investments, particularly to improve the interoperability of research management systems. *(HM Treasury, BIS, RCUK, UK HE Funding Bodies, Jisc, ARMA)*

Increasing the usefulness of existing data and information sources

15 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. *(HEFCE, RCUK, HEIs, Jisc, HESA, ARMA)*

16 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. *(BIS, other government departments, UK HE Funding Bodies, RCUK)*

Using metrics in the next REF

17 For the next REF cycle, we make some specific recommendations to HEFCE and the other HE Funding Bodies, as follows. *(UK HE Funding Bodies)*

- a. **In assessing outputs, we recommend that quantitative data – particularly around published outputs – continue to have a place in informing peer review judgements of research quality.** This approach has been used successfully in REF2014, and we recommend that it be continued and enhanced in future exercises.
- b. **In assessing impact, we recommend that HEFCE and the UK HE Funding Bodies build on the analysis of the impact case studies from REF2014 to develop clear guidelines for the use of quantitative indicators in future impact case studies.** While not being prescriptive, these guidelines should provide suggested data to evidence specific types of impact. They should include standards for the collection of metadata to ensure the characteristics of the research being described are captured systematically; for example, by using consistent monetary units.
- c. **In assessing the research environment, we recommend that there is scope for enhancing the use of quantitative data, but that these data need to be provided with sufficient context to enable their interpretation.** At a minimum this needs 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.

Coordinating activity and building evidence

18 The UK research community needs a mechanism to carry forward the agenda set out in this report. We propose the establishment of a Forum for Responsible Metrics, which would bring together research funders, HEIs and their representative bodies, publishers, data providers and others to work on issues of data standards, interoperability, openness and transparency. UK HE Funding Bodies, UUK and Jisc should coordinate this forum, drawing in support and expertise from other funders and sector bodies as appropriate. The forum should have preparations for the future REF within its remit, but should also look more broadly at the use of metrics in HEI management and by other funders. This forum might also seek to coordinate UK responses to the many initiatives in this area across Europe and internationally – and those that may yet emerge – around research metrics, standards and data infrastructure. It can ensure that the UK system stays ahead of the curve and continues to make real progress on this issue, supporting research in the most intelligent and coordinated way, influencing debates in Europe

and the standards that other countries will eventually follow. (*UK HE Funding Bodies, UUK, Jisc, ARMA*)

19 Research funders need to increase investment in the science of science policy. There is a need for greater research and innovation in this area, to develop and apply insights from computing, statistics, social science and economics to better understand the relationship between research, its qualities and wider impacts. (*Research funders*)

20 One positive aspect of this review has been the debate it has generated. As a legacy initiative, the steering group is setting up a blog (www.ResponsibleMetrics.org) as a forum for ongoing discussion of the issues raised by this report. The site will celebrate responsible practices, but also name and shame bad practices when they occur. Researchers will be encouraged to send in examples of good or bad design and application of metrics across the research system. Adapting the approach taken by the Literary Review’s “Bad Sex in Fiction” award, every year we will award a “Bad Metric” prize to the most egregious example of an inappropriate use of quantitative indicators in research management. (*Review steering group*)

The full report is available on the HEFCE website, together with two supplementary reports (a literature review and a correlation analysis of REF2014 scores and metrics).

www.hefce.ac.uk/rsrch/metrics/

Join the conversation on Twitter: **#HEFCEmetrics**