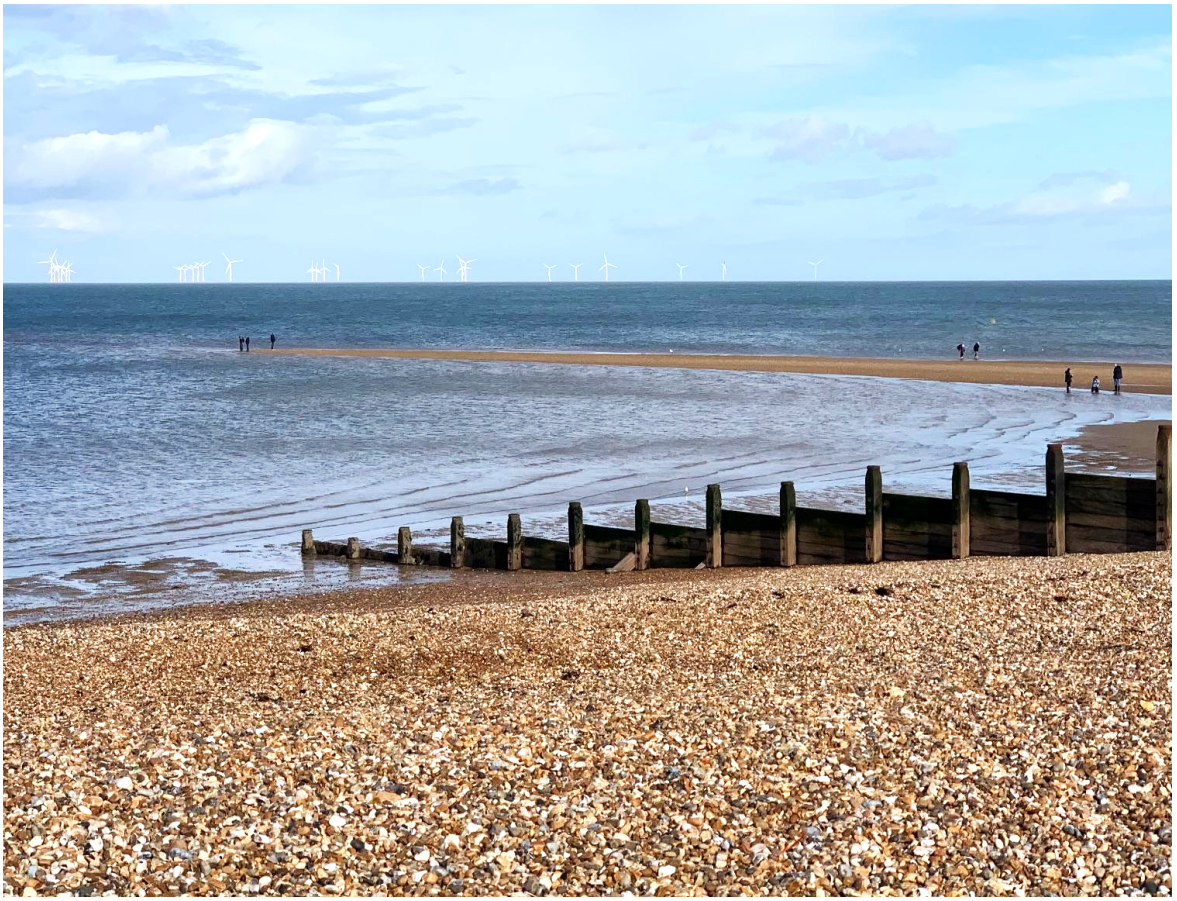

HARNESSING THE METRIC TIDE:

indicators, infrastructures and priorities for responsible research assessment in the UK

Stephen Curry, Elizabeth Gadd and James Wilsdon

Report of *The Metric Tide Revisited* panel
December 2022

For citation:
Curry, S., Gadd, E & Wilsdon, J. (2022). *Harnessing the Metric Tide: indicators, infrastructures & priorities for UK responsible research assessment. Report of The Metric Tide Revisited panel*, December 2022. ISBN 978-1-7397102-1-7. https://doi.org/10.6084/m9.figshare.21701624

Cover photograph:
Whitstable beach by Stephen Curry

Contents

ABOUT THIS REVIEW

The review was led by a three-person expert panel:

Stephen Curry: Professor of Structural Biology and Assistant Provost for equality, diversity and inclusion at Imperial College London, and Chair of the Declaration on Research Assessment steering committee. He was a steering group member and co-author of *The Metric Tide* in 2015. https://orcid.org/0000-0002-0552-8870

Dr Elizabeth Gadd: Research Policy Manager, Loughborough University, Chair of the International Network of Research Management Societies (INORMS) Research Evaluation Group and Vice-Chair of the Coalition on Advancing Research Assessment. https://orcid.org/0000-0003-4509-7785

James Wilsdon: founding Director of the Research on Research Institute (RoRI) and (from January 2023) Professor of Research Policy in the Department of Science, Technology, Engineering and Public Policy (STEaPP) at University College London (UCL), He was review Chair of *The Metric Tide* in 2015 and is a member of the UK Forum for Responsible Research Metrics (FRRM). https://orcid.org/0000-0002-5395-5949

The panel's work was supported by the Future Research Assessment Programme (FRAP) team at Research England—in particular **Noémie Aubert Bonn,** who lead-authored the literature review (included here at Appendix A),

Terms of reference

The Future Research Assessment Programme (FRAP) is led by the four UK higher education funding bodies. As a contribution to FRAP, in May 2022, our panel was asked to lead a brief review of the role of metrics in research management and assessment. Over a six-month period, the review has taken a short, sharp, evidence-informed look at existing and potential uses of metrics against four tightly-defined objectives:

1) To revisit the conclusions and recommendations of the last detailed review of these questions, *The Metric Tide* (2015), and assess progress against these;

2) To consider whether developments over recent years in the infrastructures, methodologies and uses of research metrics negate or change any of those 2015 conclusions, or suggest additional priorities;

3) To look afresh at the role of metrics in any future Research Excellence Framework (REF) and consider whether design changes now under consideration as part of the FRAP suggest similar or different conclusions to those reached in 2015;

4) To offer updated advice to the UK's higher education and research funding bodies on the most effective ways of supporting and incentivising responsible research assessment and responsible uses of metrics.

Review methodology and timetable

The panel was keen to hear and encourage contributions from the wider research community and held a series of **three roundtables** to invite formal inputs from experts and key stakeholder groups in July 2022 **(see Appendix C).** These captured advice and input from experts and stakeholder groups. Each roundtable included 20 to 25 live panellists and they were open to anyone to attend and contribute. The first roundtable brought together research managers, library and information professionals and university planners; the second, university leaders, researchers, and representatives from sector bodies and learned societies; and the third, experts in scientometrics, research policy and evaluation, plus providers of research data and infrastructure. Recordings, chat transcripts, and poll results from these seminars were used to inform the panel's conclusions. The panel also liaised extensively with the FRAP International Advisory Group, FRAP Programme Board and UK Forum for Responsible Research Metrics (FRRM)[1] for informal advice and review.

Acknowledgements

Our thanks go to Professor David Sweeney CBE, former Executive Chair of Research England, for commissioning us to undertake this rapid review, and to his successor, Professor Dame Jessica Corner, for her encouragement and support. We are also indebted to a wider circle of Research England colleagues–including Steven Hill, Catriona Firth, Noemie Aubert-Bonn and Natascha Klein–who have provided practical and intellectual support to the review, and provided an informal sounding board for our emerging ideas. Noemie merits particular thanks for her dedicated and thoughtful approach to the literature review that informed our work (see Appendix B). We received invaluable feedback on early versions of the report from members of the Future Research Assessment Programme (FRAP) Board, the UK Forum for Responsible Research Metrics, and from Sir Peter Gluckman and his colleagues on FRAP's International Advisory Group. Cameron Neylon, Helen Young and Jackie Njoroge read and strengthened initial drafts. Ludo Waltman and Mike Thelwall reviewed and improved the literature review. For additional advice, evidence and data, we are grateful to Juan Pablo Pardo-Guerra at the University of California, San Diego; Victoria Moody and Christopher Brown at Jisc; Cassidy Sugimoto at Georgia Institute of Technology; Jonathan Adams at Clarivate; Andrew Plume at Elsevier ICSR; Daniel Hook and Juergen Wastl at Digital Science. Finally we would like to express warm thanks to the speakers and attendees at the three roundtables we held in July 2022, and at the launch event on 12 December 2022, and to all those across the community who have stimulated our thinking and enriched our deliberations. We invite and encourage feedback on this review via email at: responsiblemetrics@gmail.com. Errors or omissions remain ours alone.

Stephen Curry, Elizabeth Gadd and James Wilsdon, December 2022

[1] Research on Research Institute. *RoRI PEER REVIEW project*. https://researchonresearch.org/projects#!/tab/273951116-5

SUMMARY AND HEADLINE RECOMMENDATIONS

This report was commissioned by the joint UK higher education (HE) funding bodies as part of the Future Research Assessment Programme (FRAP). It revisits the findings of the 2015 review *The Metric Tide* and provides a rapid and independent review of the use of indicators in research management and assessment. Three roundtable events held during the summer of 2022 brought together research managers, librarians, researchers, academies, institutional leaders, scientometricians and metric vendors to gather insights from the community on current thinking on the use of metrics in research assessment.

While this review feeds into the larger FRAP process, the authors have taken full advantage of their independence and sought to stimulate informed and robust discussion about the options and opportunities of future REF exercises. The report should be read in that spirit: as an input to ongoing FRAP deliberations, rather than a reflection of their likely or eventual conclusions.

The report is written in three sections. Section 1 plots the development of the responsible research assessment agenda since 2015 with a focus on the impact of *The Metric Tide* review and progress against its recommendations. Section 2 revisits the potential use of metrics and indicators in any future REF exercise, and proposes an increased uptake of 'data for good'. Section 3 considers opportunities to further support the roll-out of responsible research assessment policies and practices across the UK HE sector. Appendices include an overview of progress against the recommendations of *The Metric Tide* and a literature review.

We make **ten recommendations** targeted at different actors in the UK research system:

1: Put principles into practice.

If they are not already doing so, UK institutions and other stakeholders should participate in the growing global movement to implement **responsible research assessment (RRA)**. In particular, we strongly encourage participation in the recently formed **Coalition for Advancing Research Assessment (CoARA),** which builds on DORA, the Leiden Manifesto, *The Metric Tide*, the work of INORMS and others to provide guidance and mutual learning towards the implementation of responsible research assessment practices.[2] This is a route for institutions to share lessons from earlier implementation work, or to access support and encouragement for efforts to develop good practice – and, crucially, to build the trust of their research communities that change is happening.

2: Evaluate with the evaluated.

Higher education institutions (HEIs), research funders, and other stakeholders involved in research evaluation should enable and incentivise the co-design and co-interpretation of research assessments with research-active and research-enabling staff.

[2] The Declaration on Research Assessment (DORA) recently issued this helpful clarification on how it relates to CoARA and the new Agreement: https://sfdora.org/2022/09/21/dora-and-the-new-agreement-on-reforming-research-assessment/

3: Redefine responsible metrics.

In light of evolving RRA debates, we propose some simple yet important refinements to the definition of responsible metrics offered in *The Metric Tide*.

Box 1: A revised framework for responsible metrics (new additions in **bold**):

Responsible metrics are founded on the principle that the qualities of research reside both in the outputs and impacts of research work, and in the way it is conducted. They have the following dimensions:

- **Robustness:** basing metrics on the best possible data in terms of accuracy and scope;
- **Humility:** recognising that quantitative **indicators should not supplant qualitative, expert assessment, but should be used where appropriate to strengthen or complement peer review**;
- **Transparency:** opening up data collection and analytical processes, so those being evaluated are **included in the design of the evaluations** and can test and verify the results;
- **Diversity:** accounting for variation by field, and using a range of indicators to reflect and support a plurality of research, of **research and research-enabling**[3] **staff characteristics**, and researcher career paths across the system;
- **Reflexivity:** recognising and anticipating the systemic and potential effects of indicators, and updating them in response.

4: Revitalise the UK Forum.

The UK Forum for Responsible Research Metrics (FRRM) should receive a renewed and expanded mandate by Universities UK, UKRI and other UK research funders as the **UK Forum for Responsible Research Assessment (FoRRA)**. It should be effectively resourced to undertake and provide consultation, deliberation, advocacy and evidence-informed advice, and should provide a UK focal point for engagement with international initiatives, and third-party evaluators. FoRRA should complete work to develop a set of principles for responsible research information management. It should also invite representatives of the Higher Education Strategic Planners Association (HESPA) and Universities Human Resources (UHR) to join, with a view to improving best practice exchange between institutional stakeholders.

5: Avoid all-metric approaches to REF.

Whatever the ultimate conclusions of the FRAP process on REF purposes and design, it is unlikely that an all-metric approach will deliver what the research community, government and stakeholders need from the exercise. This pertains with particular force to the assessment of research impacts, where (despite some developments with potential) available indicators or infrastructure cannot approximate the richness of the current case study format. There may be

[3] We define research enabling staff to include all those who make a direct contribution to the research process but would not describe themselves as researchers. For more on this issue, see *The Hidden REF* https://hidden-ref.org/about/.

more opportunities to use metrics, in combination with qualitative modes of assessment, if the level at which assessment is applied moves higher than individual Units of Assessment (UoAs) (e.g., to main panels or entire institutions) in future cycles of the REF.

6: Reform the REF over two cycles.
Redesigning the REF is inevitably complex: the framework has co-evolved over four decades with the system it is assessing. Reforms must be approached carefully but not at the expense of ambition. We propose a phased reform of REF over two cycles. A clear direction of travel should be set, with specific milestones, so that the sector knows what to prepare for, and which aspects of culture, process and data infrastructure require investment and improvement. This radical yet gradual approach could harness the potential of indicators in responsible research assessment, while minimising negative or unforeseen effects.

7: Simplify the purposes of REF.
Given that RRA principles state that a clearly defined purpose should frame approaches to assessment, the UK research funding bodies should, in consultation with the wider community, agree on a simplified statement of REF purposes. Alongside this, we propose renaming the REF—for example as the 'Research Qualities Framework (RQF)'—in order to replace the contested and ill-defined term 'excellence' with an alternative that reflects plural dimensions of research quality and impact, and encompasses processes, cultures and behaviours.

8: Enhance environment statements.
Within future REF cycles, there should be greater weight overall on research environments. Institutional environment statements should be restructured to reflect additional dimensions of research culture, and to draw responsibly on data, indicators and other evidence. Given the challenge of undertaking size-independent assessments of environment (due to reliance on the physical aspects of a university's research endeavour, including wealth, age and history), we propose replacing 'environment statements' with 'people and culture statements'. This would help to capture important aspects of research activity that can be assessed in a size-independent way.

9: Use data for good.
The UK research funding bodies should commit to co-designing a set of value-led indicators for a healthy research ecosystem with community stakeholders. Bearing in mind the need for data infrastructures that are, as far as practicable, open and interoperable, these 'data for good' could then enhance existing HESA data and be used as an input to the people and culture aspects of future assessments. Such data should adhere to new principles for responsible research information management (Recommendation 4), especially around openness and transparency, and funding bodies should commit to only using services that meet these principles.

10: Rethink university rankings.

HEIs should be encouraged to take a more responsible approach to their engagement with, and promotion of, university league tables, as many league table providers continue to promote and intensify harmful incentives in research culture from outside the academic community, while resisting moves towards responsible metrics. This may include becoming a signatory to the INORMS *More Than Our Rank*[4] initiative. The House of Commons Science and Technology Committee should also initiate an inquiry into the effects of league tables on UK research culture.

Box 2: Harnessing the Metric Tide: 10 headline recommendations

1: Put principles into practice.
2: Evaluate with the evaluated.
3: Redefine responsible metrics.
4: Revitalise the UK Forum.
5: Avoid all-metric approaches to REF.
6: Reform the REF over two cycles.
7: Simplify the purposes of REF.
8: Enhance environment statements.
9: Use data for good.
10: Rethink university rankings.

[4] INORMS. *More Than Our Rank*. https://inorms.net/more-than-our-rank/

FOREWORD BY SIR PETER GLUCKMAN

"Not everything that can be counted counts, and not everything that counts can be counted"

This well-known aphorism often but wrongly attributed to Einstein remains central to ongoing discussions and debate about research assessment. Words like quality, impact and excellence can have very different meanings to different stakeholders and yet these words find their way into various forms of metric and indicator used within the academic endeavour.

Research assessment can and should have very different meanings in different contexts. It can apply to an individual seeking a fellowship, job, promotion or tenure; to a team seeking funding; to review of a publication or other outputs; or to the evaluation of the institution conducting research. One size will not fit all.

The principles that define science, broadly speaking, (acknowledging that research is a larger term) require that the products of its research endeavours are open to scrutiny by peers and its claims are only acceptable in the face of such scrutiny. As a result, scholarly knowledge evolves and is not constant. This makes the business of evaluating research outputs central to the enterprise. But there are many different types of output both within and across disciplines and this adds complexity. And the research enterprise is evolving to include a greater diversity of contributors, more team based activity, and the emergence of new forms of cross disciplinary and transdisciplinary activity. Genuine transdisciplinarity may be best assessed by looking at non-traditional outputs. Meanwhile, the utilitarian expectation of funders for more actionable or exploitable knowledge continues to grow

Since the first bibliometric analyses appeared in the 1980s, they have become increasingly central to the academic exercise, and much more sophisticated. Scholars, their institutions, their funders and governments have all turned to metrics as a tool, even if their limitations have been increasingly obvious. The so-called gold standard against which research quality might be measured (at least in some contexts), namely peer review itself, is under pressure and merits critical rethinking. But academic and policy stakeholders, alongside the publishing industry, have created a veritable complexity of overt and covert incentives within research. The result is mixed, with outcomes that are not always desirable. But funders require and deserve some accountability, and in the UK as in some other countries this has led to dual funding systems for higher education, where one stream is based on some form of assessment of research quality – no matter how contentious it may be.

In this context, when *The Metric Tide* was published in 2015, it became a vital document in opening up these vexed questions for examination. It was critically valuable in informing the design of REF 2021, the most recent exercise by which 'research excellence' has been assessed by the UK's higher education funding bodies.

The REF itself has multiple objectives – some related to funding allocation and accountability, others related to using its data and incentives for strategic evolution within the higher education sector.

Now that the REF 2021 exercise is complete and as the UK system reflects on what comes next, it is timely to revisit the 2015 report.

In those seven years, much has happened. Research assessment has become a richer field of academic inquiry. The limitations of some forms of metrics (e.g. H indices, impact factors, citation rates) have become very apparent. Yet new or enhanced metrics that take a broader perspective on evaluating research may become possible through the use of big data techniques.

As the UK government reviews the purposes and approaches to research assessment for the anticipated next cycle, the role and place of research assessment and metrics remains debated and can generate strong opinions. This short review makes a timely and valuable contribution to the ongoing discussion about how the REF might evolve. How will concepts such as quality, impact and the broader research environment be considered? To what extent do metrics help? To what extent do they embed complex processes and induce unhelpful behaviours in areas such as institutional ranking?

The authors of this review are to be congratulated for what has been a rapid reconsideration. As an attendee at the third of the workshops that informed the report, I was impressed by the quality of the discussion that ensued. I hope this report assists in ensuring that the research enterprise within and beyond the UK continues to support activities that advance knowledge frontiers and meet the needs of all societies, the peoples that live within them and our home—planet earth.

Sir Peter Gluckman is President of the International Science Council & Chair of the FRAP International Advisory Group

1: FROM METRICS TO RESPONSIBLE RESEARCH ASSESSMENT (RRA)

1.1 Tidal flow: the rise of responsible research assessment (RRA)

The ebb and flow of research assessment practices exert powerful and contested forces that are reshaping the research landscape. Research assessment is seen as a necessary process for the effective allocation of jobs, promotions and funding, but its scale and reach has grown enormously since the 1980s. This is evident not only within research institutions and funding organisations, but also at the national level, where through exercises like the Research Excellence Framework (REF)[5] governments seek accountability and justification for public investment.

Over the years, the steady surge in research assessment and the expansion of the REF has raised questions about the robustness, burden and cost of the task, and its impacts on individual researchers and their organisational and disciplinary cultures. These concerns have been exacerbated by the influx of research metrics, or indicators, which seek to bring a quantitative scaffold to complex qualitative judgements.

The academic critique of the use of metrics in research assessment stimulated by these concerns has grown alongside them. It largely serves an academic audience but ever so often spills into the policy sphere – either by activism or invitation. One such invitation, *The Metric Tide* report was the work of an independent review group convened at the behest of the then Minister for Universities and Science, David Willetts.

Though it had the REF clearly in view, the review group spent fifteen months examining the uses and limitations of metrics in research assessment, their use across different disciplines, their potential for tracking research quality and impact, and the ways in which they were influencing research culture, management, and issues of equality and diversity within universities and institutes.

The Metric Tide review made twenty specific recommendations, which are appraised in more detail below (and in **Appendix A**). In brief, while the review held to the view that metrics could not be used on their own to assess the quality of Units of Assessment (UoAs) in the REF, it found there was scope to use them to support the expert judgement of peer reviewers. Such usage needed to remain mindful of the risks of perverse incentives, the review recommended, and to be as open as possible to enhance trustworthiness and interoperability.

[5] REF 2021. *What is the REF?* https://www.ref.ac.uk/about-the-ref/what-is-the-ref/

The advance of responsible research assessment

The report attracted significant attention nationally and internationally. In 2015, it was a timely addition to a swelling chorus of voices warning against the narrowing conceptions of research quality that had accompanied the uncritical use of bibliometric data in research assessment. *The Metric Tide* was preceded by the 2013 San Francisco Declaration on Research Assessment (DORA[6]), an initiative from a group of editors and academic publishers that drew attention to the unintended effects of the growing obsession with journal impact factors. It also coincided with the 2015 Leiden Manifesto,[7] a move by the bibliometric and scientometric community to take greater responsibility for metrics by distilling best practice in metrics-based research assessment to meet the need for greater accountability and transparency from both indicators and evaluators. In harmony with both DORA and the Leiden Manifesto, *The Metric Tide* was the first contribution to articulate the notion of *responsible metrics*, insisting that they be robust, humble, transparent, diverse and reflexive (see 1.5 below).

Since 2015, calls for reform of research assessment, in particular for more responsible uses of metrics, have grown louder and more international. The interactions between research assessment and other shifts in the academic landscape, such as the drive to open scholarship; concerns over research integrity; and overdue attention being paid to issues of equality, diversity, bullying and harassment, have become more apparent and urgent.

The task ahead

What hasn't changed is the struggle to marry the task of research assessment to its most important purposes not just in ways that are responsible and avoid unintended effects, but that are aligned with intersecting movements to support more fruitful, inclusive and positive research cultures. Yet more work is needed to articulate how assessment can identify, incentivise and reward diverse qualities in the outputs, impacts and processes of research.

It is timely to reflect on *The Metric Tide* in light of these developments. As with the original review, although our primary purpose here is to inform and improve future REF exercises, we also hope to contribute to ongoing debates and reforms that reach far beyond the UK.

While the UK reflects on the future of its national research assessment framework through the FRAP process, related discussions are underway in other research-intensive countries. In August 2022, Australia's government halted the current round of its Excellence in Research for Australia (ERA) assessment pending a root-and-branch review, which some speculate will lead to it being scrapped.[8] Around the same time, New Zealand's equivalent exercise, the Performance-Based Research Fund (PBRF), was postponed owing to concerns about the effects of Covid-19, sparking renewed calls for its abolition.[9] Heated debates are ongoing in

[6] *San Francisco Declaration on Research Assessment*. https://sfdora.org/read/
[7] Hicks, D., Wouters, P., Waltman, L., Rijcke, S. d. & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature* 520, 429-431, doi:10.1038/520429a
[8] See recent coverage here: https://www.timeshighereducation.com/news/australia-postpones-research-assessment-exercise; https://www.timeshighereducation.com/news/australian-research-assessment-exercise-has-achieved-purpose
[9] For more on New Zealand's PBRF: https://www.timeshighereducation.com/news/new-zealand-research-assessment-delayed-again

Spain[10] and Italy[11] about the need for reforms to research assessment systems that are perceived by critics as too narrow or restrictive. Other countries, including Sweden and Norway, are considering adopting new or expanded assessment frameworks. From a policy perspective, one dilemma in these discussions is that, despite varied methodologies and approaches across national assessment frameworks, there is little compelling evidence to support the effectiveness of one approach to design and implementation over another. Yet as a recent study by Clarivate-ISI argues, there is stronger evidence for macro benefits to comparative research performance that flow from having a national assessment framework.[12]

In this report, we review the conclusions and recommendations of *The Metric Tide* to assess progress in the light of moves towards more holistic approaches to research assessment (Chapter 1); to offer recommendations for the role of metrics as part of RRA in a future REF (Chapter 2); and to highlight broader ways of supporting and incentivising responsible research assessment (Chapter 3).

1.2 Turning tides: RRA in the wider context of research cultures

The executive summary of *The Metric Tide* is at pains to outline the breadth and depth of its examination of the absorption of metrics not just into processes of research assessment, but into the very marrow of scholarly endeavour:

> *"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."*

Though it remains an important landmark, that report now sits in a landscape that has undergone significant shifts. There has been considerable ingestion by UK institutions of the philosophies underpinning DORA and the Leiden Manifesto over the last seven years. This has been driven in part by the transformation of DORA[13] from a static declaration to an active initiative since 2017, which has provided advocacy and support for the development and uptake of new approaches to research assessment. Over a similar period, the co-authors of the Leiden Manifesto[14] who are associated with the Leiden Centre for Science and Technology Studies[15] have been active advocates for reform, both through their own work

[10] https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87158; https://blogs.lse.ac.uk/impactofsocialsciences/2022/11/02/reforming-research-assessment-in-spain-requires-greater-university-autonomy/
[11] https://www.timeshighereducation.com/opinion/italian-miracle-illusion; https://researchprofessionalnews.com/rr-news-europe-italy-2022-5-further-controversy-hits-italy-s-promotion-evaluation-process/
[12] Adams, J., Beardsley, R., Bornmann, L., Grant, J., Szomszor, M. and Williams, K. (2022) *Research Assessment: Origins, evolution, outcomes. What have assessment initiatives achieved?* Clarivate Institute for Scientific Information (ISI). https://clarivate.com/lp/research-assessment-origins-evolutions-outcomes/
[13] *San Francisco Declaration on Research Assessment*. https://sfdora.org/read/
[14] Hicks, D., Wouters, P., Waltman, L., Rijcke, S. d. & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature* 520, 429-431, doi:10.1038/520429a
[15] *Centre for Science and Technology Studies*. https://www.cwts.nl/

(e.g., developing principles for the responsible use of university rankings[16] in 2017) and work with DORA (e.g., to develop new Tools for Advancing Research Assessment[17]) and more recently through the Research on Research Institute (RoRI)[18] on a range of projects.

Further attention has been drawn to research assessment by the publication of the Hong Kong Principles[19] in 2018, which emphasise the importance of considering research integrity when assessing researchers (a baton now being picked up by the newly formed UK Committee on Research Integrity[20]), and by INORMS's critical evaluation[21] of the metrics and methodologies of university ranking.

Policy shifts

Amid this growth of statements and initiatives, there have also been notable policy shifts, particularly by funders. Both the Wellcome Trust and UKRI are implementing policies and processes that enact and incentivise more responsible approaches to research assessment. These are moves that give practical form to the awareness that 'assessment shapes culture'.[22]

These moves are happening not just in the UK; witness, for example, the arrival of FOLEC,[23] the Latin American Forum on Scientific Evaluation, and the Dutch National Recognition and Rewards Programme.[24] Moreover, the international agenda for research assessment reform is increasingly propelled by push for open access and open science. This is evident, for example, in Plan S,[25] in the French Plan for Open Science,[26] in the recent UNESCO recommendations on Open Science,[27] and related initiatives by the G7 governments[28]—all of which are built on the presumption that opening up research practices must proceed hand-in-hand with reform of an incentive structure that is too closely tied to publication metrics.

From principles to practice

A welcome shift from declarations of principles to the harder work of implementation was noted in the RoRI working paper[29] discussed at the Global Research Council meeting in 2020. This phase of the reform movement is becoming increasingly embedded and is supported by

[16] Waltman, L., Wouters, P. & Eck, N. J. v. (2017). Ten rules for ranking universities. *Research Professional* https://www.researchresearch.com/news/article/?articleId=1368350
[17] DORA. *Project TARA*. https://sfdora.org/project-tara/
[18] *Research on Research Institute*. https://researchonresearch.org/
[19] Moher, D. *et al.* (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biology* 18, e3000737, doi:10.1371/journal.pbio.3000737
[20] UKRI. (2022). *UK Committee on Research Integrity*. https://www.ukri.org/what-we-offer/supporting-healthy-research-and-innovation-culture/research-integrity/uk-committee-on-research-integrity/
[21] Gadd, E., Holmes, R. & Shearer, J. (2021). Developing a Method for Evaluating Global University Rankings. *Scholarly Assessment Reports* 3, 2, doi:10.29024/sar.31
[22] UKRI publishes new report on responsible research assessment. (2021) *UKRI* (28 May). https://www.ukri.org/news/ukri-publishes-new-report-on-responsible-research-assessment/
[23] CLACSO. *FOLEC*. https://www.clacso.org/en/folec/
[24] *Recognition and Rewards*. https://recognitionrewards.nl/
[25] Plan S. *Plan S principles*. https://www.coalition-s.org/plan_s_principles/
[26] Ministère de l'enseignement supérieur et de la recherche. *Second French Plan for Open Science*. https://www.ouvrirlascience.fr/second-national-plan-for-open-science/
[27] UNESCO. (2021). *UNESCO Recommendation on Open Science*. SC-PCB-SPP/2021/OS/UROS, Paris. https://en.unesco.org/science-sustainable-future/open-science/recommendation
[28] *G7 2021 Research Compact*. (2021). https://www.gov.uk/government/publications/g7-2021-research-compact/g7-2021-research-compact
[29] Curry, S. *et al.* (2020). *The changing role of funders in responsible research assessment: progress, obstacles and the way ahead (RoRI Working Paper No.3)*. 10.6084/m9.figshare.13227914.v2 Research on Research Institute. https://doi.org/10.6084/m9.figshare.13227914.v2

the work of many stakeholders. One of the most notable recent developments is the formation of a European-led Coalition on Advancing Research Assessment (CoARA).[30] The coalition is made up of research organisations that have committed to work together to innovate and experiment with new approaches to research assessment and to implement and evaluate reformed practices within five years.

RECOMMENDATION 1: *Put principles into practice.*
If they are not already doing so, UK institutions and other stakeholders should participate in the growing global movement to implement responsible research assessment (RRA). In particular, we strongly encourage participation in the recently formed Coalition for Advancing Research Assessment (CoARA), which builds on DORA, the Leiden Manifesto, *The Metric Tide*, the work of INORMS and others to provide guidance and mutual learning towards the implementation of responsible research assessment practices.[31] This is a route for institutions to share lessons from earlier implementation work, or to access support and encouragement for efforts to develop good practice – and, crucially, to build the trust of their research communities that change is happening.

These are certainly interesting times for research assessment reform but a reality check is in order. As noted in our review of the recommendations in *The Metric Tide,* public statements by institutions do not always align with what happens in practice. Reports in 2021 of the misuse of research metrics in a redundancy process at one UK university, raised concerns[32] about its attitude to its obligations as a DORA signatory and questions about the most effective way to hold institutions to account.[33] In the Netherlands, bold moves to abandon the use of journal impact factors[34] in research assessment encountered vocal resistance[35] from a section of the professoriate. Dealing with such issues is a delicate matter for voluntary initiatives like DORA, but one that needs to be faced squarely by funding organisations that have powerful leverage over the institutions they fund.

Growing scholarly critique

In parallel with these political and practical developments, the scholarly and professional critique of research assessment practices has continued apace. The discourse on the impacts of assessment and the use of metrics, which formed the spine of *The Metric Tide* report, has become richer and more extensive. A survey of the literature is included in Appendix B; here we will only touch briefly on some of the more salient themes to have emerged.

In his book *The Quantified Scholar*,[36] an important recent contribution to these debates, Juan Pablo Pardo-Guerra has analysed how the norms and expectations generated by the

[30] European Commission. (2022). *Reforming research assessment: The Agreement is now final*. https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/reforming-research-assessment-agreement-now-final-2022-07-20_en
[31] See this helpful clarification from DORA on how it relates to the new Agreement: https://sfdora.org/2022/09/21/dora-and-the-new-agreement-on-reforming-research-assessment/
[32] Responsible research assessment faces the acid test. (2021). *Nature* 595, 471-472, doi:10.1038/d41586-021-01991-z
[33] Curry, S. (2021). How should Dora be enforced? *Research Professional News* (8 September). https://www.researchprofessionalnews.com/rr-news-uk-views-of-the-uk-2021-9-how-should-dora-be-enforced/
[34] Woolston, C. (2021). Impact factor abandoned by Dutch university in hiring and promotion decisions. *Nature* 595, 462, doi:10.1038/d41586-021-01759-5
[35] Chawla, D. S. (2021). Scientists at odds on Utrecht University reforms to hiring and promotion criteria. *Nature index* (9 August). https://www.nature.com/nature-index/news-blog/scientists-argue-over-use-of-impact-factors-for-evaluating-research
[36] Pardo-Guerra, J. P. (2022). *The Quantified Scholar*. Columbia University Press.

quantification of 'excellence' in the REF operate 'upstream' in the research process, and are internalised by academics themselves. His analyses suggest that by narrowing the parameters of what is considered a worthwhile contribution, research assessment exercises like the REF risk stifling originality, leading to disciplinary homogenisation. His findings, which focus on the social sciences, echo earlier work on how "thinking with indicators"[37] perturbs the design and conduct of research programmes in the life sciences.

The utility of the common but ill-defined notion of 'excellence' has also come under greater scrutiny from Moore and colleagues.[38] Its lack of conceptual clarity[39] grants the term cover to sustain problematic issues of hyper-competition, reproducibility, fraud and homophily. Like Pardo-Guerra, who argues for 'evaluative diversity' to address the epistemic and moral deficiencies of much of current research assessment practice, Moore et al, call for the diverse features of excellence to be better recognised and defined. That call is powerfully reinforced by critiques that have highlighted the biases inherent in the concept of excellence,[40] which sustains 'epistemic injustice',[41] by feeding through to unequal acknowledgements of the contributions of women[42] or indeed anyone who isn't a white, able-bodied, heterosexual man.[43]

The link between research assessment and research culture

Deepening concerns about the need to explicitly consider equity, diversity and inclusion (EDI) within the construction of our research culture are not confined to the scholarly literature. Within the UK, they have become increasingly embedded within the policies and practices of universities and funders. In part this reflects the impact of benchmarking schemes such as Athena Swan[44] and the Race Equality Charter[45] (both of which demand thorough quantitative and qualitative analyses). But there is also much greater awareness of the legacies of historical exclusion, the scourge of sexual harassment[46] and other forms of ill-treatment, and how these problems are sustained by a culture of evaluation that is too fixated on quantitative aspects of research productivity. DORA has highlighted the complex intersections[47] between the research assessment reform, the drive for open scholarship, and the need to create a more equitable and inclusive academy: these challenges need to be tackled together.

The original *Metric Tide* report recommended the development and adoption of indicators that support equality and diversity as a counterweight to these problematic impacts of assessment. Some progress in this direction has been made now that institutions can opt to

[37] Müller, R. & de Rijcke, S. (2017). Thinking with indicators. Exploring the epistemic impacts of academic performance indicators in the life sciences. *Research Evaluation* 26, 157-168, doi:10.1093/reseval/rvx023
[38] Moore, S., Neylon, C., Paul Eve, M., Paul O'Donnell, D. & Pattinson, D. (2017). "Excellence R Us": university research and the fetishisation of excellence. *Palgrave Communications* 3, 16105, doi:10.1057/palcomms.2016.105
[39] Hatch, A. (2019). To fix research assessment, swap slogans for definitions. *Nature* 576, 9, doi:10.1038/d41586-019-03696-w
[40] van den Brink, M. & Benschop, Y. (2011). Gender practices in the construction of academic excellence: Sheep with five legs. *Organization* 19, 507-524, doi:10.1177/1350508411414293
[41] Clavero, S. & Galligan, Y. (2021). Delivering gender justice in academia through gender equality plans? Normative and practical challenges. *Gender, Work & Organization* 28, 1115-1132, doi:https://doi.org/10.1111/gwao.12658
[42] Ross, M. B. *et al.* (2022). Women are credited less in science than men. *Nature* 608, 135-145, doi:10.1038/s41586-022-04966-w
[43] Cech, E. A. The intersectional privilege of white able-bodied heterosexual men in STEM. *Science Advances* 8, eabo1558, doi:10.1126/sciadv.abo1558
[44] Advance HE. *Athena Swan Charter*. https://www.advance-he.ac.uk/equality-charters/athena-swan-charter
[45] Advance HE. *Race Equality Charter*. https://www.advance-he.ac.uk/equality-charters/race-equality-charter/members
[46] Universities UK. (2021). *Changing the culture: two years on*. https://www.universitiesuk.ac.uk/what-we-do/policy-and-research/publications/changing-culture-two-years
[47] The intersections between DORA, open scholarship, and equity. (2020). *DORA* (18 August). https://sfdora.org/2020/08/18/the-intersections-between-dora-open-scholarship-and-equity/

include EDI data within their REF environment statements; as part of the REF institutions must also submit an equality impact assessment to demonstrate fair treatment of all staff.

However, the need to build a research culture that is positive and equitable is, if anything, more urgent today than in 2015. Given UKRI's endorsement[48] of the government's R&D People and Culture Strategy,[49] which aims "to create a more inclusive, dynamic, productive and sustainable research and development sector in the UK", we need to reconsider whether the conceptions of 'excellence' that are transmitted via the REF remain too narrow. We are firmly of the view that assessment of research excellence within the REF (and elsewhere) should more explicitly consider the plural qualities of the research process alongside the qualities of the output and outcomes (a point emphasised in *The Metric Tide*).

Innovations supporting research assessment

The capacity for shifting focus in this way has been enhanced over recent years by a number of innovations in publishing and in research metrics. It remains to be seen how some of these innovations will play out – but they give rise to opportunities for creating richer and more diverse accounts of contributions to research. We cannot hope to paint a complete picture but note with interest the enormous growth of preprints – boosted significantly by the Covid-19 pandemic – in fields beyond the domains of physics and mathematics that have long been served by the arXiv,[50] and the arrival of Overton,[51] a digital platform that identifies links between academic papers and policy documents.

Increased capacity for more quantitative and more transparent assessments of research contributions and research culture can be seen in the Contributor Roles Taxonomy (CRediT[52]) developed to define specific contributions to research papers; in the Initiative for Open Citations,[53] which has succeeded in persuading thousands of publishers to make citation lists from academic paper freely available; and in the pioneering inclusion of metrics for gender equality and adoption of open access in the Leiden Ranking[54] of universities produced by the CWTS. This latter ranking innovation is an important example of Sugimoto and Larivière's proposal[55] to use indicators as a "force for social good". A similar motivation lies behind the development of university rankings that attempt to to track university sustainability[56] or contributions to the UN Sustainable Development Goals,[57] though the proxy indicators used are more indirect and therefore more difficult to interpret meaningfully.[58]

---

[48] UK Research and Innovation. (2021). *UKRI welcomes government's R&D People and Culture Strategy*. https://www.ukri.org/news/ukri-welcomes-governments-rd-people-and-culture-strategy/
[49] Department for Business, E., & Industrial Strategy,. (2021). *R&D People and Culture Strategy*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1004685/r_d-people-culture-strategy.pdf
[50] *arXiv*. https://arxiv.org/
[51] *Overton*. https://www.overton.io/
[52] *CRediT*. https://credit.niso.org/
[53] *Initiative for Open Citation*. https://i4oc.org/
[54] CWTS. *Leiden Ranking*. https://www.leidenranking.com/
[55] Larivière, V. & Sugimoto, C. (2019). Indicators for social good. *CWTS* (15 May). https://www.cwts.nl/blog?article=n-r2w2c4
[56] *UI Green Metric*. https://greenmetric.ui.ac.id/
[57] Times Higher Education. *Impact Rankings 2022*. https://www.timeshighereducation.com/impactrankings
[58] Curry, S. (2020). The still unsustainable goal of university ranking. *Reciprocal Space* (26 April). http://occamstypewriter.org/scurry/2020/04/26/still-unsustainable-university-rankings/

Reducing the assessment burden

While the proliferation of these and other new technologies enable richer and more capacious descriptions of research 'excellence' that might be deployed to rebalance the unintended effects of assessment processes focused on outputs, great care must also be taken not to add unduly to the administrative loading of evaluative exercises like the REF. The growing burden of the REF has preoccupied policy makers for over a decade. It was the motivation behind the 2009 pilot exercise,[59] which concluded that the REF could not easily be streamlined through a greater reliance on bibliometrics, and a prominent consideration for *The Metric Tide* review five years later. Reducing the REF's burden was also a core focus of the 2016 Stern Review,[60] but initial indications[61] are that the resulting reforms resulted in little, if any, progress towards this in REF 2021.[62]

Although the REF did not fall within the remit of the 2022 Tickell review of research bureaucracy,[63] the principles and recommendations that it developed resonate with those of *The Metric Tide* and seem likely to feed usefully into the FRAP. These include emphasis on transparency, on the integration of research information systems, on clearer and more expansive conceptions of research excellence, and – most germane to the discussion at hand – on ensuring that processes embody fairness and support commitments to EDI. A similar philosophy lies behind the ongoing Concordat Review,[64] which is attempting to reduce bureaucracy by harmonising the various frameworks for enhancing the culture and environment in which UK research takes place.

The REF, or indeed any research assessment process, should work with the grain of progressive norms with the research community. The importance and difficulty of achieving this cannot be underestimated. The task is to arrive at ways of combining quantitative data with qualitative judgements that are meaningful, proportionate, and have a clarity of purpose; they must drive up the qualities of the research process, its outputs and impacts, while seeking to build a culture that is inclusive, positive and people-centred. There is no prospect of arriving at an optimal solution given the complexity of the systems being evaluated and the political contestation of shifting norms. Nevertheless, the emergence in recent years of more holistic and values-based conceptions of what a healthy research ecosystem could look like offers hope that real progress can and should be made.

1.3 Tidal monitoring: revisiting the 2015 recommendations

To get a sense of the impact of the *Metric Tide* report we have reviewed and commented on the progress made in respect of each of the 20 recommendations (see Appendix A). Although

[59] REF 2014. *Bibliometrics pilot exercise*. https://www.ref.ac.uk/2014/about/background/bibliometrics/
[60] Stern, N. (2016). *Independent report - Research Excellence Framework review*. Department for Business, Energy & Industrial Strategy. https://www.gov.uk/government/publications/research-excellence-framework-review
[61] Firth, C. (2021). What has the real-time REF review taught us about future research assessment? *Jisc Blog* (2 December). https://www.jisc.ac.uk/blog/what-has-the-real-time-ref-review-taught-us-about-future-research-assessment-02-dec-2021
[62] Manville, C. et al. (2021). *Understanding perceptions of the Research Excellence Framework among UK researchers: The Real-Time REF Review*. doi:10.7249/RRA1278-1, RAND Corporation,, Santa Monica, CA:. https://doi.org/10.7249/RRA1278-1
[63]Department for Business Energy & Industrial Strategy, UK Research and Innovation & Kwarteng, K. (2022). Government recommends cutting unnecessary bureaucracy in research sector. Gov.uk (28 July). https://www.gov.uk/government/news/government-recommends-cutting-unnecessary-bureaucracy-in-research-sector
[64] Research concordats and agreements review - Phase 2. (2022). *Universities UK* (19 October). https://www.universitiesuk.ac.uk/what-we-do/policy-and-research/publications/research-concordats-and-agreements

the recommendations were not formulated as an action plan, we have attempted to RAG-rate each one.

Broadly speaking, there is clear progress. Most recommendations are rated amber, while some are green and only two (16 and 20) are considered not to have made any headway.

Recommendations 1-8: Engagement with the responsible metrics agenda

Recommendations 1-8 were targeted at key stakeholders within the research community including HEI leaders, research managers, HR managers, individual researchers, funders, data providers and publishers, and aimed to support effective leadership, governance and management of research culture. In each case we see evidence of much greater awareness of the importance of responsible metrics and moves to use that awareness to re-shape assessment processes. For example, there has been a significant growth in public commitments by Higher Education Institutions (HEIs), funders and publishers to DORA, the Leiden Manifesto or research assessment policies that align with their calls for more responsible approaches to metrics; members of the UK Association of Research Managers and Administrators (ARMA) have played a leading role in international efforts in critiquing metrics and creating frameworks for value-led approaches to evaluation; and some data providers have taken concrete steps to address deficiencies in aggregate metrics such as impact factors as part of efforts to foster more responsible usage.

The responsible metrics agenda has therefore achieved considerable gravitational pull among stakeholders, but we still lack clear evidence that it has become the default. Concerns remain that public commitments may be largely performative. We heard from our roundtables, for example, that greater awareness of responsible metrics had driven some poor practices behind closed committee room doors. Greater openness around actual practices is needed to provide the verification that engenders trust.

Recommendations 9-16: Improving data infrastructures

Recommendations 9-16 aimed to improve the data infrastructure that supports research information management. Here too there has been progress, particularly on the uptake of identifiers such as ORCIDs[65] for researchers, ROR IDs[66] for institutions, and DOIs[67] for documents and other outputs. The adoption of such tools should foster greater interoperability and reduce the burden of information-gathering.

Greater interoperability will benefit from the adoption of common standards and development of open data infrastructures. Although some very significant steps forward have been taken (*e.g.* the successful Initiative for Open Citations[68]), achieving these goals is complicated by the multiplicity of stakeholders (e.g., UKRI, HEIs, funders, HESA, UUK) and the existence of a market for the sale of closed research data. There is powerful and articulate advocacy for open research infrastructure within the academic community, which draws strength from the open scholarship movement. However, what is less clear is whether there is sufficient vision and leadership among funders on the national or international stage to make this a

[65] *ORCID*. https://orcid.org/
[66] *ROR*. https://ror.org/
[67] *doi*. https://www.doi.org/
[68] *Initiative for Open Citation*. https://i4oc.org/

sustainable reality. This is a question that obviously takes us beyond the REF, which has already been a significant factor in driving open access by stimulating the growth and use of institutional repositories.

Recommendation 17: Metrics and the REF

The three components of Recommendation 17, which focused on the use of metrics in the REF have been largely achieved. First, quantitative data (typically article citation counts) are provided to sub-panels that request them; and second, guidelines have been produced – and adopted – regarding the use of metrics in impact case studies and in environment statements (the two components that in addition to research outputs, are scored in the REF[69]). Third, while specific quantitative indicators that can support environment statements have been described,[70] no such prescription was made for impact case studies. The logic behind these decisions is that metrics could facilitate the comparison of research environments in different institutions, but risk inhibiting the presentation of the full richness and variety of examples of impact beyond the academy.

Recommendations 18-20: Coordinating responsible metrics

Finally, Recommendations 18-20 addressed the need for greater coordination and evidence-building to embed responsible metrics practices. Two of these have given rise to new bodies, the UK Forum for Responsible Research Metrics (FRRM[71]) and the Research on Research Institute (RoRI[72]), which works with an international consortium of research funders and scholarly communications organisations, and has a mission to accelerate transformational research on research systems, cultures and decision-making. The third, a proposal for a 'bad metrics' prize, failed to capture the imagination of the research community.

The FRRM, formed in 2016, has provided a valuable locus to bring together key stakeholders and helped with the practical realisation of recommendations that touched most closely on the REF. That said, it has achieved only limited visibility and traction within the UK research community and could play a more prominent role in the sector-wide dialogue needed to underpin reform of research assessment.

1.4 Tide marks: contributions by the UK system to the RRA agenda

As noted in the previous subsection, *The Metric Tide* report had several concrete impacts on the conduct of the REF in 2021. These were mediated in part through the recommendations of Lord Stern's independent review of the REF,[73] conducted in 2016. This confirmed the *Metric Tide's* view on the primacy of peer review in research assessment while allowing for qualitative assessments of research outputs and environment to be supported by judicious use of key metrics.

---

[69] REF 2021. *What is the REF?* https://www.ref.ac.uk/about-the-ref/what-is-the-ref/

[70] *Guidance for institutions on environment indicators*. https://www.ref.ac.uk/media/1019/guidance-on-environment-indicators.pdf

[71] The UK Forum for Responsible Research Metrics. (2022). *Universities UK* (30 September). https://www.universitiesuk.ac.uk/topics/research-and-innovation/uk-forum-responsible-research-metrics

[72] *Research on Research Institute*. https://researchonresearch.org/

[73] Stern, N. (2016). *Independent report - Research Excellence Framework review*. Department for Business, Energy & Industrial Strategy. https://www.gov.uk/government/publications/research-excellence-framework-review

The Peer Review Tide

While acknowledging the widespread view of peer review as a 'gold standard' in research assessment, *The Metric Tide* nevertheless provided a critical assessment of its strengths and weaknesses. Since its publication, the scrutiny of peer review has only intensified. Commentators and scholars have observed[74] limitations regarding its quality, reproducibility, transparency, equity, inclusion, efficiency and the lack of incentives to participate. It should be noted that much of this evidence focuses on journal peer review, rather than the review of outputs for quality ratings such as those required by the REF. Experiments with open, portable and double-blind peer review seek to address some peer review concerns. However, there are questions about the long-term sustainability of the current volume of peer review within the academy.

Roundtable participants noted inconsistencies regarding concerns about quantitative assessment approaches when qualitative approaches were equally (if distinctively) problematic. Professor Mike Thelwall observed that whilst artificial intelligence (AI) assessment mechanisms were accused of being 'black box' and unable to replicate peer review, peer review is often also opaque and difficult to reproduce.

Part of this opacity relates to the *implicit* use of metrics in peer review, as revealed by studies demonstrating a high level of correlation between peer review and journal metrics that are likely due to the unwitting or unspoken use of this information in peer judgements. For example, Professor Alan Dix, a member of the REF 2014 Computer Science and Informatics sub-panel has documented his experience[75] of comparing citation metrics with REF peer review scores. The exercise suggested peer review was subject to latent bias against applied sub-disciplines and, by extension, women who were more likely to work in those areas.

Further research to provide a better understanding of peer review is ongoing, but the judicious and transparent use of appropriate metrics to not only support peer review decisions but to triangulate and sense-check those decisions would seem to be prudent given what we already know.

Ongoing influence of the *Metric Tide*

Mindful of the growing burden of the REF, Lord Stern also echoed the report's call for metrics to be open and standardised to make them interoperable between agencies so as to reduce the cost of accessing and reusing such information. Although full interoperability remains to be achieved (see Appendix 1), citation data provided by Clarivate for REF 2021 was open to interrogation[76] by the institutions being assessed.

The Metric Tide report has also had impacts beyond the REF, as discussed briefly above. A survey of UK institutions conducted as part of a 2018 FRRM review[77] of progress toward the responsible use of metrics found that a clear majority of respondents (63 out of 96) agreed with the framework outlined in *The Metric Tide*. Indeed it is not difficult to find examples of

[74] Research on Research Institute. *RoRI PEER REVIEW project*. https://researchonresearch.org/projects#!/tab/273951116-5

[75] *REF2014 Citation Analysis*. https://alandix.com/ref2014/

[76] Clarivate Analytics will provide citation data during REF 2021. (2018). *REF 2021* https://www.ref.ac.uk/guidance-and-criteria-on-submissions/news/clarivate-analytics-will-provide-citation-data-during-ref-2021/

[77] UK Forum for Responsible Research Metrics. *UK Progress towards the use of metrics responsibly: Three years on from The Metric Tide report* (2018). http://dera.ioe.ac.uk/id/eprint/31945

universities[78] that explicitly reference the report in public commitments to good research assessment practice. An international survey this year by the LIS-Bibliometrics forum noted a decline since 2018 in the use of *the Metric Tide* as a cited influence on institutional principles for research assessment. However, this likely reflects the growth and activity of organisations behind other initiatives in this domain which espouse similar approaches, rather than any loss of faith in the idea of responsible metrics. Notably, the report continues to accrue about 150 citations per year.[79]

Other 'soft power' influences of the report are also detectable; for example, in the UK government's 2018 Research and Innovation Strategy,[80] which affirms its commitment to greater openness in research and to improving both research governance and evaluation; and in the more recent UK Research and Development Roadmap,[81] which specifically references the need to tackle "the problematic uses of metrics in research". This marks a notable shift of awareness and commitment since the Brown era, when the REF reform efforts were focused more narrowly on the potential of metrics to reduce the burden of the exercise.

With others, *The Metric Tide* has also helped the UK system to be more prominent in the global movement for research assessment reform. A notable example of this is a working paper prepared by RoRI[82] in 2020 for the Global Research Council which highlighted the influential role that funders can play and developed the concept of responsible metrics as a component of a broader framework of responsible research assessment.

1.5 Tidal swell: expanding the scope and potential of responsible metrics

Having briefly reviewed how the recommendations of *The Metric Tide* have stood the test of time since 2015 and surveyed the changing scholarly landscape, which is reshaping thinking on the integration of quantitative and qualitative information in research assessment, we turn finally to one of the main ideas introduced by the report: the notion of **responsible metrics.** These were originally conceived to have the dimensions of robustness, humility, transparency, diversity and reflexivity. In our view, these dimensions remain largely fit for purpose. However, as discussed above, and as articulated in a 2020 RoRI working paper,[83] the focus on responsible metrics has now been folded into the broader framework of **responsible research assessment (RRA).** This can be defined as *"an umbrella term for approaches to assessment which incentivise, reflect and reward the plural characteristics of high-quality research, in support of diverse and inclusive research cultures".*

---

[78] University of Bath. *Principles of research assessment and management*. https://www.bath.ac.uk/corporate-information/principles-of-research-assessment-and-management/

[79] Google Scholar. *The Metric Tide: report of the independent review of the role of metrics in research assessment and management. 2015*. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=DQlScvcAAAAJ&citation_for_view=DQlScvcAAAAJ:ZuybSZzF8UAC

[80] *UK International Research and Innovation Strategy*. (2019). Department for Business, Energy & Industrial Strategy. https://www.gov.uk/government/publications/uk-international-research-and-innovation-strategy

[81] *UK Research and Development Roadmap*. (2020). Department for Business, Energy & Industrial Strategy. https://www.gov.uk/government/publications/uk-research-and-development-roadmap

[82] Curry, S. *et al.* (2020). *The changing role of funders in responsible research assessment: progress, obstacles and the way ahead (RoRI Working Paper No.3)*. 10.6084/m9.figshare.13227914.v2 Research on Research Institute. https://doi.org/10.6084/m9.figshare.13227914.v2

[83] Curry, S. *et al.* (2020). *The changing role of funders in responsible research assessment: progress, obstacles and the way ahead (RoRI Working Paper No.3)*. 10.6084/m9.figshare.13227914.v2 Research on Research Institute. https://doi.org/10.6084/m9.figshare.13227914.v2

To reflect this shift, we propose a series of amendments to the defined properties of responsible metrics. First, to acknowledge the movement towards an enlarged conception of open research and to more explicitly recognise the need for responsible metrics to support the diversification of the research community, we propose amendments to the definitions of Transparency and Diversity respectively. In particular, we now include the principles of co-design and co-interpretation of research assessments as important facets of transparency.

Second, we want to take this opportunity to clarify in the dimension of humility the context in which metrics might be used to support qualitative assessments.

Third, and most importantly, we propose the addition of an underpinning tenet to capture the consensus that 'assessment shapes culture' and that assessment of research qualities should therefore intentionally take account of research processes as well as outputs and impacts. How a research output is achieved matters as much as what is achieved or the impact it has.

RECOMMENDATION 2: *Evaluate with the evaluated.*
HEIs, research funders, and other stakeholders involved in research evaluation should enable and incentivise the co-design and co-interpretation of research assessments with research-active and research-enabling staff.

RECOMMENDATION 3: *Redefine Responsible Metrics.*
In light of evolving RRA debates, we propose some simple but important refinements to the definition of responsible metrics offered in *The Metric Tide*.

Box 1: A revised framework for responsible metrics (additions in **bold**):

Responsible metrics are founded on the principle that the qualities of research reside both in the outputs and impacts of research work, and in the way it is conducted. They have the following dimensions:

- **Robustness:** basing metrics on the best possible data in terms of accuracy and scope;
- **Humility:** recognising that quantitative **indicators should not supplant qualitative, expert assessment, but should be used where appropriate to strengthen or complement peer review**;
- **Transparency:** opening up data collection and analytical processes, so those being evaluated are **included in the design of the evaluations** and can test and verify the results;
- **Diversity:** accounting for variation by field, and using a range of indicators to reflect and support a plurality of research, of **research and research-enabling**[84] **staff characteristics**, and researcher career paths across the system;
- **Reflexivity:** recognising and anticipating the systemic and potential effects of indicators, and updating them in response.

[84] We define research enabling staff to include all those who make a direct contribution to the research process but would not describe themselves as researchers. For more on this issue, see *The Hidden REF* https://hidden-ref.org/about/.

Principles and definitions are important ingredients in progressing research assessment but what matters is how they are implemented in practice. Broadly speaking, the principles articulated in *The Metric Tide* and the recommendations arising from its analysis of the wider cultural and managerial impacts of research metrics have helped to propel some significant changes. However, as we have seen, it remains difficult to make a full assessment of how effectively its recommendations and values have been adopted, while the need to consider culture and environment as integral to research quality has become a much more central concern. To provide greater impetus and support for opening up pathways to real change, we envisage a reinvigorated role for the UK Forum for Responsible Research Metrics:

RECOMMENDATION 4: *Revitalise the UK Forum.*
The UK Forum for Responsible Research Metrics (FRRM) should receive a renewed and expanded mandate by Universities UK, UKRI and the other UK research funders as the UK Forum for Responsible Research Assessment (FoRRA). It should be effectively resourced to undertake and provide consultation, deliberation, advocacy and evidence-informed advice, and should provide a UK focal point for engagement with international initiatives, and third-party evaluators. FoRRA should complete work to develop a set of principles for responsible research information management (See Section 3.1). It should also invite representatives of the Higher Education Strategic Planners Association (HESPA) and Universities Human Resources (UHR) to join, with a view to improving best practice exchange between institutional stakeholders (See Section 3.3).

Metrics and bibliometrics

The Metric Tide report was careful not to conflate the term metrics with bibliometrics. However, questions as to the legitimate place of bibliometrics in research assessment were undoubtedly one of the main reasons the metric tide report was commissioned. Indeed, the supporting analyses focussed entirely on the potential of bibliometrics to replicate peer review scores. There has been extensive discussion around the appropriate place of bibliometrics in research evaluation practices before and since the publication of *The Metric Tide*, including the question as to whether they have any place at all.

As previously reported, much of this discourse addresses the limitations of bibliometrics to fairly assess those in certain demographic groups, especially where small sample sizes are used (e.g., the comparative evaluation of individual researchers); concerns about the coverage of the databases used to run the analyses[85] are also frequently reported.[86] Initiatives such as the Citation Typing Ontology[87] have sought to expose the multiple possible meanings of an individual citation, thereby challenging the view that all citations can be seen as positive. An increasing number of tools such as Semantic Scholar[88] and Scite[89] are now

[85] Bramer, W. M., Rethlefsen, M. L., Kleijnen, J. & Franco, O. H. (2017). Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Systematic Reviews* 6, 245, doi:10.1186/s13643-017-0644-y
[86] Bardiau, M. & Dony, C. (2022). *"Counting Back": What kind of bibliodiversity does the Impact Factor brand reflect? A case study of IF journals included in the 2021 Journal Citations Report*. https://zenodo.org/record/7193256#.Y0fKUkzP2Ul
[87] Peroni, S. & Shotton, D. (2012). FaBiO and CiTO: ontologies for describing bibliographic resources and citations. *Journal of Web Semantics* 17, 33-43, doi:10.1016/j.websem.2012.08.001
[88] *Semantic Scholar*. https://www.semanticscholar.org/
[89] *scite*. https://www.scite.ai

using machine learning and semantic web technologies to identify citation sentiments; while these provide more nuanced citation analyses, questions about their reliability remain.[90]

Responsible use guidance such as that provided by the Metric Tide report and the Leiden Manifesto address some of the risks of bibliometric analysis. However, when it comes to reporting on bibliometric analyses, any understanding of their limitations are, at best, only acknowledged in the methodology, or, at worst, ignored altogether. It is rare, particularly in HEIs, for the limitations of bibliometrics to be considered so flawed as to lead to the cessation of bibliometric analysis altogether, despite this being the more rigorous course of action in some circumstances. Guidance on this issue from a renewed Forum for Responsible Research Assessment would be very welcome.

We turn now to consider in more detail the scope for diversifying and enhancing the quantitative information used to evaluate the processes and output of research – particularly in the UK through the Research Excellence Framework.

[90] Gadd, E. (2020). AI-based citation evaluation tools: good, bad or ugly? *The Bibliomagician* (23 July). https://thebibliomagician.wordpress.com/2020/07/23/ai-based-citation-evaluation-tools-good-bad-or-ugly/

2: DATA FOR GOOD: THE FUTURE OF UK RESEARCH ASSESSMENT

2.1 Radical yet phased reform

No-one now aged under 60 has worked in a UK higher education institution that wasn't shaped to some extent by the cycles of national research assessment that have spun through that system since the first "research selectivity exercise" in 1986. Over four decades and eight cycles, this assessment process—operating since 2014 as the Research Excellence Framework (REF)—has become a highly sophisticated "evaluation machine", to adopt a term coined by the political scientist Peter Dahler-Larsen.[91]

As with any complex machinery, the designers and users of the REF regularly look to the potential of new processes and upgrades to enhance, reboot or streamline its operations. The REF is simultaneously admired, being seen by some as a leading assessment system to emulate; relied upon as a robust, fair and accountable basis on which to determine the annual allocation of around £2 billion of quality-related funding; and contested, as a source of bureaucracy, competition and conformity (an "instrument of terror" by one recent account).[92]

The UK was already completing its second research assessment cycle when Tim Berners-Lee invented the World Wide Web. In the years since, advances in ICTs, data science, scientometrics and related fields have transformed the possibilities and practices of measurement and management—and research assessment has co-evolved alongside. Many regard applications of machine learning and artificial intelligence (AI) as the latest example of "general purpose technologies" with the capacity to boost productivity and transform working practices across entire economies, including HE and research, with specific calls by some to build these technologies into future iterations of the REF.[93]

So it is welcome that within the Future Research Assessment Programme (FRAP), the four UK HE funding bodies are looking afresh at the potential contribution of metrics, and at newer approaches using machine learning and AI—in a balanced way that sets these questions in the context of responsible use and responsible research assessment.

[91] Dahler-Larsen, P. (2012). *The Evaluation Society.* Stanford, CA: Stanford University Press.
[92] McIntyre, F. (2021). REF feared as 'instrument of terror' by researchers. *Research Professional* (24 March). https://www.researchprofessional.com/0/rr/news/uk/ref-2014/2021/3/REF-feared-as--instrument-of-terror--by-researchers.html
[93] Balbuena, L. D. (2018). The UK Research Excellence Framework and the Matthew effect: Insights from machine learning. *PLoS ONE* 13, e0207919, doi:10.1371/journal.pone.0207919

Indicators are not enough

One thing that has not changed since the publication of the *Metric Tide* report is the acceptance that 'not everything that counts can be counted', and that even where we can 'count things that count', this cannot supplant peer review. Our data infrastructures are improved but have some way to go, not least in terms of interoperability. Moreover, our data sources are not yet comprehensive and, in any case, do not yield indicators that can be used without the supporting context provided by expert commentary. As such, it is unlikely that an all-metric approach will deliver what the research community, government and other stakeholders need from the next REF exercise. This pertains with particular force to the assessment of research impacts, where, despite some developments with potential, the available data infrastructure cannot approximate the richness of the current case study format. That said, there may yet be more opportunities to use indicators, in combination with qualitative modes of assessment, if the design and level at which assessment is applied changes in future cycles of the REF**.**

RECOMMENDATION 5: ***Avoid an all-metric approach to REF.***
Whatever the ultimate conclusions of the FRAP process on REF purposes and design, it is unlikely that an all-metric approach will deliver what the research community, government and stakeholders need from the exercise. This pertains with particular force to the assessment of research impacts, where (despite some developments with potential) available indicators or infrastructure cannot approximate the richness of the current case study format. There may be more opportunities to use metrics, in combination with qualitative modes of assessment, if the level at which assessment is applied moves higher than individual Units of Assessment (UoAs) (e.g., to main panels or entire institutions) in future cycles of the REF.

In future cycles, we would like to see the REF evolve further to:

1. Reward progress towards better research cultures;
2. Be a more formative process (and less of a retrospective audit);
3. Reduce competition and increase collaboration and collegiality;
4. Become less burdensome on researchers and institutions.

To properly harness the potential of metrics and indicators to enrich RRA practices within the REF, we need to combine care with ambition. We would like the FRAP to explore options for phased reform that engages the sector in a process of radical change. We recognise that any reforms of the REF can be complicated, and the sector requires predictability and time to adapt to new approaches. Piloting and trialling of changes, initially on a smaller scale, is also important to identify effects and unforeseen consequences, given the scale, reach and financial importance of the REF, and the lengths to which institutions often go to achieve even marginal gains in 'performance'.

RECOMMENDATION 6: ***Reform the REF over two cycles.***
Redesigning the REF is inevitably complex: the framework has co-evolved over four decades with the system it is assessing. Reforms must be approached carefully but not at the expense of ambition. We propose a phased reform of REF over two cycles. A clear

direction of travel should be set, with specific milestones, so that the sector knows what to prepare for, and which aspects of culture, process and data infrastructure require investment and improvement. This radical yet gradual approach could harness the potential of indicators in responsible research assessment, while minimising negative or unforeseen effects.

Date	Exercise	Coordinating body	Key features
1986	Research Selectivity Exercise	Universities Grants Committee	37 cost-centres; 4-part questionnaire on research income, expenditure, planning priorities & output
1989	Research Selectivity Exercise	Universities Funding Council	152 units of assessment; 70 peer review panels; 2 outputs per member of staff
1992	Research Assessment Exercise (RAE)	HEFCE	HEIs select which staff to submit; 5-point scale; 2800 submissions to 72 UoAs; introduction of census date
1996	Research Assessment Exercise (RAE)	HEFCE	Up to four outputs per researcher; 69 UoAs
2001	Research Assessment Exercise (RAE)	HEFCE	2600 submissions to 69 units of assessment; 5 umbrella groups of panel chairs for consistency
2008	Research Assessment Exercise (RAE)	HEFCE	67 sub-panels under 15 main panels; results presented as quality profiles
2014	Research Excellence Framework (REF)	HEFCE	4 main panels; 36 sub-panels; introduction of 20% impact element
2021	Research Excellence Framework (REF)	UKRI (Research England + devolved funding councils)	All staff with significant responsibility for research included. Impact 25% weighting. Flexible number of outputs.

Figure 1: The eight cycles of UK research assessment since 1986

2.2 REF purposes

Before anyone can answer the question of whether and how metrics should be used in any future REF, there is a need to revisit and clarify what the REF is for. Its stated purposes have multiplied in recent years—and some now want to add further purposes, such as incentivizing improvements in research culture.

Many would consider the REF's primary purpose to be the **selective allocation of quality-related (QR) funding** for research. In addition, the UK funding bodies have over several REF cycles emphasised two additional purposes (REF 2017): to provide **accountability for public investment** in research by producing evidence of the benefits of this investment; and **to provide benchmarking information** and establish reputational yardsticks, for use within the HE sector and for public information.

To these three agreed purposes, the Stern Review[94] proposed adding a further three:

- It provides a rich evidence base **to inform strategic decisions about national priorities** across science, social science, engineering, medicine and arts and humanities research.
- It can create a **strong performance incentive** for universities and for individual academics.
- It can be used by universities and other bodies **to inform strategic decisions and resource allocation.**

Can a single assessment framework deliver on six or seven simultaneous objectives? What hierarchy of importance should be applied to these? What happens when they are in tension? These are fundamental questions that the FRAP process needs to answer. Only then, can we have a sensible discussion about design options, and the role of metrics and indicators as a methodological option to deliver these. The choice of purposes of the REF also fundamentally shapes any cost-benefit-burden analysis of the exercise.

Considering the purposes of REF

A precise reformulation of these purposes is beyond the scope of this review. Nevertheless we wish to provoke discussion of these critical issues. We therefore propose starting with a focus on the core purposes of accountability and adaptation. Accountability: to evaluate the diverse outcomes, qualities and impacts of research in order to demonstrate accountability to government and wider society (on the basis of which the allocation of QR funding can take place); and adaptation: to positively influence and incentivize improvements in research cultures, capacities and capabilities. This new second purpose would compel a shift towards formative assessment.

The primary purpose here is accountability – and the potential then for effective allocation of funding is a product of that accountability. As such, allocation no longer needs to be elevated to a purpose in its own right. It is worth noting that when one correlates the REF 2021 QR allocations[95] with REF-submitted FTE[96] it produces a correlation coefficient of 0.96. The REF 2014 allocations followed a similar pattern for both QR and the Scottish Research Excellence Grant (REG) funding[97]. One option would therefore be to settle the matter of allocation *entitlement* via FTE volume and for access to that funding allocation to only be granted to HEIs after participating in a research quality assurance mechanism which was contextualised and more formative.

Similarly, we would argue that benchmarking at the unit of assessment level should no longer be a stated purpose of REF, as this is methodologically dubious and encourages some of the more problematic aspects of institutional and strategic responses to the REF[98] (Wilsdon,

[94] Stern, N. (2016). *Independent report - Research Excellence Framework review*. Department for Business, Energy & Industrial Strategy. https://www.gov.uk/government/publications/research-excellence-framework-review

[95] Research England. (2021). *QR funding: supporting information for 2021 to 2022*. UKRI. https://www.ukri.org/publications/quality-related-research-qr-funding-supporting-information-for-2021-to-2022/

[96] REF 2021. *Submissions data*. https://ref.ac.uk/results-analysis/submissions-data/

[97] Correlating total grants for research and innovation AY 2020-2021 with the Category A FTE submitted to REF 2014.

[98] Wilsdon, J. (2022). What the FRAP happens next? Four priorities for reforming the REF. *Times Higher Education* (26 May). https://www.timeshighereducation.com/opinion/what-frap-happens-next-four-priorities-reforming-ref

2022). Research management in HEIs can still benefit from an enhanced evidence base to inform internal strategy and prioritisation, especially on aspects of research culture. However, any such data, and related metrics and indicators, should be independently useful to HEIs and not generated solely for the purposes of assessment. From a simplified set of purposes, the UK research funding bodies should then clarify the outcomes that the REF is intended to support**.** Examples of these might include:

- High-quality research;
- Real world impact;
- High-quality training, mentorship and institutional leadership;
- A diverse, engaged, content and motivated population of researchers and research enabling staff.

Following on from a revised and clearer set of purposes and outcomes, we would propose a renaming of the REF to remove the problematic term 'excellence' (see section 1.2). When even the chief executive of UKRI acknowledges that "no one knows what research excellence means",[99] we need to replace this word with an alternative that offers greater clarity and reflects plural dimensions of research quality and impact, while also encompassing processes, cultures and behaviours. One suggestion is the 'Research Qualities Framework (RQF)'.

RECOMMENDATION 7: *Simplify the purposes of REF.*
Given that RRA principles state that a clearly defined purpose should frame approaches to assessment, the UK research funding bodies should, in consultation with the wider community, agree on a simplified statement of REF purposes. Alongside this, we propose renaming the REF—for example as the 'Research Qualities Framework (RQF)'—in order to replace the contested and ill-defined term 'excellence' with an alternative that reflects plural dimensions of research quality and impact, and encompasses processes, cultures and behaviours.

2.3 REF design and levels of assessment

Once the REF's purposes have been simplified, and its ideal outcomes specified, the next step is to consider design choices for the structure of the assessment process, and the levels in the system at which the assessment is applied (for example, entire institutions; Main Panel clusters of disciplines; individual Units of Assessment.). Only when these choices have been made is it possible to properly consider different methodological options, including the uses of metrics and indicators. Of course, at greater levels of aggregation (e.g., institution-level), some indicators would become more viable. (***See Recommendation 5: Avoid all-metric approaches to REF.)***

We must again emphasise that precise recommendations on the design and level of assessment are beyond the scope of this review. However, given that any decisions around the responsible use of metrics are intimately bound up with these design choices, we offer the following observations for consideration as part of the wider FRAP deliberations.

[99] Wilsdon, J. (2022). What the FRAP happens next? Four priorities for reforming the REF. *Times Higher Education* (26 May). https://www.timeshighereducation.com/opinion/what-frap-happens-next-four-priorities-reforming-ref

There have been calls to broaden the sector's view as to who has "significant responsibility for research" (using REF terminology) and what constitutes a high quality research contribution. Initiatives such as the Hidden REF sought to celebrate a wider range of individuals and contributions to the UK research endeavour. The BEIS R&D People and Culture Strategy promotes "broaden[ing] the range of experiences and accomplishments that are recognised" through the introduction of the Résumé for Researchers. It also champions recognising "the importance of leadership and management skills" through the next assessment exercise. The need to look beyond input and output measures and to consider process measures - *how* research is done - is seen to be of particular importance.

RECOMMENDATION 8: *Enhance environment statements.*

Within future REF cycles, there should be greater weight overall on research environments. Institutional environment statements should be restructured to reflect additional dimensions of research culture, and to draw responsibly on data, indicators and other evidence. Given the challenge of undertaking size-independent assessments of environment (due to reliance on the physical aspects of a university's research endeavour, including wealth, age and history), we propose replacing 'environment statements' with 'people and culture statements'. This would help to capture important aspects of research activity that can be assessed in a size-independent way.

An expanded emphasis on the research environment with a particular focus on research culture would allow for more sophisticated use of a range of data and indicators – much of which is already available. Debates often focus on the negative consequences of over-reliance on certain indicators, but other indicators can of course also be used in positive ways to measure and incentivise culture change.

Box 3 below illustrates a few examples of positive indicators that could be included in the environment and culture section of future REF exercises. Some of these were suggested by our roundtable participants in response to the question, "what do we value about UK research that might be *legitimately and usefully* quantified either for monitoring or assessment?". There needs to be careful consideration of which of these indicators accords with the principles of responsible metrics, which are already gathered (for example, for Athena Swan), which provide meaningful information, which might not be wholly related to an institution's research activity, and which might create additional burdens or unproductive incentives for gaming.

Box 3: DATA FOR GOOD

- Gender and ethnicity pay gaps amongst research staff;
- Percentage of research staff on short term contracts;
- Measures of research staff wellbeing and contentment in workplace surveys
- Volume of teamwork; collaborations; co-produced research (with users);
- Open research indicators;
- Policy impacts e.g. via citations in policy literatures;
- Peer review activities;
- Citizenship contributions (from workload models);
- Measures of support for EDI;
- Career pathways;
- Research leadership (through surveys);
- Research leave taken;
- Effective measures for dealing with bullying and harassment.

RECOMMENDATION 9: *Use data for good.*
The UK research funding bodies should commit to co-designing a set of value-led indicators for a healthy research ecosystem with community stakeholders. Bearing in mind the need for data infrastructures that are, as far as practicable, open and interoperable, these 'data for good' could then enhance existing HESA data and be used as an input to the people and culture aspects of future assessments. Such data should adhere to new principles for responsible research information management (Recommendation 4), especially around openness and transparency, and funding bodies should commit to only using services that meet these principles.

Greater availability of data for monitoring

Regardless of the design of the next REF, it would be of benefit to UK HE institutions if a wider range of data and indicators were available to them to plot their progress towards improving their research culture and environment. Indeed, an overarching principle for any metrics and indicators selected for the next REF should be that they also provide value to the assessed institutions independently of any REF exercise.

Once a set of useful indicators had been identified, it would be then one small step, were environment statements a continued feature of the next exercise, to make them more structured (addressing specific aspects of the institutional environment and culture) and buttressed by such indicators, and to give them more weight in the overall scoring. It has been noted[100] that assessment of research environments will always favour larger, wealthier, more established research-intensive HEIs. A refocusing of this element of the assessment to give greater consideration to size-independent matters of research culture (e.g., gender pay gap, leadership surveys) may give all research-active institutions an opportunity to succeed.

[100] Rogers, C. (2022). REF 2021: Rules give more to those that have most. *Research Professional News* (18 May). https://www.researchprofessionalnews.com/rr-news-uk-views-of-the-uk-2022-5-ref-2021-rules-give-more-to-those-that-have-most

Box 4: OPTIONS FOR CONSIDERATION BY FRAP

To reinforce moves to more responsible research assessment, the FRAP Board may also wish to consider the following design changes for the next REF:

- Require the submission of fewer outputs from a cross-section of staff, including outputs where the contributions of research-enabling staff are listed;
- Undertake a trial in selected UoAs of output assessment through the introduction of "scholarly case studies" (e.g. selected narratives that explain the academic impact of a body of work produced by a cohort of research-active and research-enabling staff), supported where appropriate by responsible metrics;
- For the environment section, require compulsory indicators on select elements (e.g., RRA practices, staff satisfaction) within structured submissions;
- Retain impact case studies in their current form.

For the REF after next, further reforms which could be considered by the FRAP Board include:

- Streamlining the exercise by moving from ~34 UoAs to larger-scale assessment under four Main Panels (A, B, C, D).
- Assess these Main Panel clusters through a dashboard of data; formal submissions; self-assessment; external sampling of outputs; and on-site panel visits.

3: STRENGTHENING RESPONSIBLE RESEARCH ASSESSMENT

3.1 Responsible data infrastructures, services and indicators

The generation of reliable and comparable metrics and indicators depends on the collection, management, sharing, analysis and presentation of data in ways that meet stakeholder expectations and build community confidence. Roundtable participants noted a range of concerns about the current data infrastructure. These included, for example, the loss of control over both the data they generate in the course of their research (e.g., abstracts and citation linkages) and the data they provide (often repeatedly, via different systems) that describes their research (e.g., ResearchFish submissions, CRediT information supplied to journals). A related set of anxieties surrounded the ingest of data provided freely by the research community into poorly-designed, commercially-owned systems, and the formation of these data into indicators in which the community has little trust, that are then sold back to HE institutions at significant cost.

Many responsible research evaluation principles, including those of the original Metric Tide report, demand the openness and transparency of data used for evaluation purposes, especially towards those being evaluated. Whilst some institutions seek to adhere to such principles for internal evaluations, they are less frequently met by external evaluation tools or exercises. This includes the REF itself, which (in the absence of a suitable non-commercial alternative) uses commercially-provided bibliometric data to support peer review. Individual REF peer review scores are also not made publicly available.

Open bibliographic data infrastructure

In recent years, positive advances have been made towards an open bibliographic data infrastructure. The Initiative for Open Citations (I4OC) recently announced that 100% of Crossref publishers now make their cited references openly available. The Initiative for Open Abstracts (I4OA) is also now underway. OpenAlex[101], building on Crossref and the former Microsoft Academic Graph, provides access to hundreds of millions of entities (works, authors, venues, institutions and concepts) using well-established Permanent Identifiers (including Digital Object Identifiers (DOI), Open Researcher and Contributor IDs (ORCID), Linking International Standard Serial Numbers (ISSN-L), Research Organisation Register IDs (ROR) and Wikidata IDs). The combination of these sources, and others, provides a strong foundation for community-developed infrastructure initiatives to enable access to research outputs and

[101] *OpenAlex*. https://openalex.org/

analysis of their usage. Such initiatives include Lens.org[102], a patent and scholarly data search, mapping and analysis tool; the Curtain Open Knowledge Initiative (COKI[103]), which provides institution- and country-level open access monitoring; and many other citation-based discovery and evaluation tools.

Many of the pieces are now in place to form a functioning and open research management data infrastructure, but the task is not yet complete. Small additional investments, for example, in making ORCID a mandatory element of an institution's HESA submission, encouraging publishers to consistently surface ROR IDs to Crossref, and an exploration of the use of Research Activity IDs[104] (RAIDs) to aggregate data at project rather than institutional level, could all reap significant additional benefits. These might include better visibility of collaborative multidisciplinary research projects, and improved discovery.

Principle-led data infrastructures

However, investments alone will not be sufficient to create usable infrastructure. We also need to put in place principles, based on values agreed by the community, to guide not just the construction of research information systems, tools and services, but how they are to be governed, developed, accessed and maintained. These principles should ensure that all the data submitted to, and surfaced by, such services adheres to community best practice expectations. They are essential to build trust in any resulting metrics, indicators and evaluation approaches. This was a recommendation of the *Metric Tide* report and work was started – but not finished – on developing a set of principles for the responsible management of research information management data. In the meantime, the development of a set of Principles for Open Scholarly Infrastructure[105] by Lin, Bilder and Neylon and the Invest In Open Infrastructures[106] (IOI) work will provide a good basis for completing this important work.

Once agreed principles are in place, it will be important to incentivise the uptake of data and services that adhere to them. This could be achieved through ensuring any future REF is underpinned only by services that meet the principles, and that the exercise rewards those institutions that use them to guide their internal research information discovery, monitoring and assessment practices. ***(See Recommendation 4: Revitalise the UK Forum and Recommendation 9: Use data for good.)***

Expanding our data collection

The data captured via the bibliographic record is not the only thing we care about in the research ecosystem. In the UK, the Higher Education Statistics Agency (HESA) data already provide a solid base for understanding and monitoring the HE landscape. Data is collected on staff, students, graduates, finances, estates management, and the HE Business & Community Interaction (HEBCI) datasets that inform the Knowledge Exchange Framework (KEF). The provision by HEIs of some elements of the HESA data return is optional. The Higher Education

[102] *Lens.org*. https://www.lens.org/

[103] *Curtin Open Knowledge Initiative*. https://open.coki.ac/

[104] *Research Activity Identifier*. https://www.raid.org.au/

[105] *The Principles of Open Scholarly Infrastructure*. https://openscholarlyinfrastructure.org/

[106] *Invest in Open Infrastructure*. https://investinopen.org/

Strategic Planners Association (HESPA) is one of a number of sector bodies working with the higher education data insight group[107] to enhance understanding of data on higher education.

As we seek to reconsider what we value about the UK research endeavour and therefore what we may wish to monitor and evaluate, it would seem sensible to expand our existing data collection to include a broader range of measures and, where appropriate, to mandate reporting of those elements of institutional HESA submissions that are currently optional. For example, the submission of data on professional services staff (including many technicians) is currently optional, but could support better visibility of the contribution of research-enabling staff.

The HEBCI data already forms the most significant source for the KEF, requiring relatively little additional input from HEIs. Were the research management data collection to expand to cover a wider range of aspects of the research endeavour, it could be possible to see a mechanism similar to the KEF dashboards used for research monitoring and benchmarking purposes whilst noting that these are currently only used by HEIs in England. The KEF dashboards have been praised for their responsible data visualisation, by using profiles rather than rankings; clusters to compare like institutions; and by combining qualitative and quantitative data as appropriate to the aspect of KE under scrutiny ***(See Recommendation 9: Use data for good.)***

3.2 From principles to cultures and practices

One of the challenges for institutions seeking to take more responsible approaches to their internal use of research metrics is the fact that they continue to be assessed by third parties using approaches that do not adhere to the same principles. Roundtable participants shared concerns about the poor use of metrics by professional bodies accrediting degree programmes; by the measures used to assess institutions for Research Degree Awarding Powers; and by national and international league tables. The UK government has itself recently bestowed questionable legitimacy on the global university rankings by using them as a proxy by which to identify 'High Potential Individual'[108] visa candidates.

Suppliers of research assessment products sometimes use methods and indicators that do not support responsible evaluations. For example, the UK university that controversially and inappropriately planned to use SciVal's Field-Weighted Citation Impact (FWCI) indicator to identify individuals for redundancy[109] is unlikely to have done so if the indicator was not readily available in the tool for researcher-level evaluations.

Outside the UK, despite international momentum towards more responsible research assessment, there is still widespread use of Journal Impact Factors and other reductive indicators to assess candidates for funding and jobs.

---

[107] HESPA. *Higher Education Data Insight Group (HEDIG)*. https://hespa.ac.uk/advocacy/hedig/

[108] Gov.uk. *High Potential Individual (HPI) visa*. https://www.gov.uk/high-potential-individual-visa

[109] Else, H. (2021). Row erupts over university's use of research metrics in job-cut decisions. *Nature* 592, 19, doi:10.1038/d41586-021-00793-7

The influence of third-party assessments

Campbell's Law tells us that we get what we measure. In the face of constant exposure to poor third-party assessment practices, it is difficult for institutions to fully embrace responsible assessment approaches as they absorb and cascade the measures used to assess them. For example, despite the recommendation of the 2015 *Metric Tide* report for institutions to think more critically about their engagement with league tables, the financial and reputational benefits that result from a good ranking position may leave institutions with little choice but to seek to improve on these measures despite them having no value to the institution beyond the league table and potentially undermining EDI work.[110] Uncritical internal dissemination of analyses of institutional and departmental scores in the latest rankings sends a questionable message to staff about the value placed on crude and distorting proxies. Action taken on behalf of the UK HE sector to make visible and engage with some of these unhelpful external drivers would be very welcome. ***(See Recommendation 4: Revitalise the UK Forum.)***

Living up to our principles

In addition to some form of collective engagement with third-party assessors and providers of assessment tools, there is also an opportunity to further support individual institutions to live up to their own responsible research assessment principles. Adherence to principles for the responsible use of metrics is now encoded in the Research England Terms & Conditions and many UK HEIs have either signed DORA or developed their own responsible metrics policy. However, taking such a step is often the beginning rather than the end of an institution's responsible research assessment journey. In some cases, public statements by institutions do not always align with what happens in practice.

Ensuring that responsible assessment principles are fully embedded throughout an institution is a challenging process, particularly for senior leaders charged with establishing KPIs, promotion, recruitment and redundancy criteria in an ever more competitive international HE landscape. Demands from senior leaders for simple answers to complex questions are seen in some quarters to be the root of much poor evaluation practice. Research-enabling professionals, for example, have noted the difficulties of dealing with requests for metrics from senior managers which they felt were not in line with their institutional policy. Research-active staff have noted similar challenges when on the receiving end of assessments that did not meet the standards set. Senior leaders can themselves feel ill-equipped to deal with breaches of their policy. (***See Recommendation 1: Put principles into Practice & Recommendation 4: Revitalise the UK Forum.***)

Ensuring that senior managers with academic or professional services responsibilities have sufficient understanding to confidently exercise ownership of and commitment to their institution's responsible research assessment principles is key if these are to be effective. The provision of mechanisms to support institutions in this regard would be beneficial. One approach might be to ensure senior leaders draw on appropriate evaluation expertise when designing assessments. Similarly, running equality impact assessments (EIAs) on any evaluation designs (especially where they are being used for allocation purposes) could

[110] University Wankings. (2021). in *Socially Responsible Higher Education* 67-79 (Brill). doi:10.1163/9789004459076_006 isbn:9789004459076 https://doi.org/10.1163/9789004459076_006

identify concerns at an early stage. One option being developed by the INORMS Research Evaluation Group to resolve the conundrum posed by league tables universities face is the *More Than Our Rank*[111] initiative; this provides institutions with a straightforward way to acknowledge the limitations of the university rankings and to describe their qualities and achievements on their own terms.

RECOMMENDATION 10: *Rethinking the rankings*.
HEIs should be encouraged to take a more responsible approach to their engagement with, and promotion of, university league tables, as many league table providers continue to promote and intensify harmful incentives in research culture from outside the academic community, while resisting moves towards responsible metrics. This may include becoming a signatory to the INORMS *More Than Our Rank*[112] initiative. The House of Commons Science and Technology Committee should initiate an inquiry into the effects of league tables on UK research culture.

3.3 Measuring what matters with the people who matter

Since the publication of the 2015 *Metric Tide* report the importance of involving evaluated communities in the design and interpretation of the evaluation mechanisms used to assess them has come to the fore. The International Development Research Centre's (IDRC) Research Quality Plus (RQ+)[113] method, the Centre for Science & Technology Studies (CWTS) Leiden's Evaluative Inquiry[114] approach, DORA's SPACE rubric,[115] and the INORMS Research Evaluation Group's SCOPE framework[116] all operate under the principle that evaluators should 'evaluate *with* the evaluated'. Recent research assessment reforms such as the Dutch Recognition & Rewards Programme[117] and the EC Agreement on Reforming Research Assessment[118] have also been highly consultative. We have incorporated this philosophy into our proposals for creating responsible data infrastructures and services in section 3.1 above.

Participants at roundtable events repeatedly called for the voices of research-active and research-enabling staff to be heard in the design of research assessments to mitigate unintended consequences and maximise buy-in and a sense of ownership. Indeed, it has been suggested that the sector's high regard for peer review may have its foundation in a desire for agency over the assessment process, rather than a belief in the superiority of peer review *per se*. Co-designed indicators, properly contextualised, may provide stakeholders with the same agency without the need to rely unnecessarily on peer review. Involvement in the design of indicators may also better equip stakeholders to interpret them. ***(See Recommendation 2: Evaluate with the evaluated.)***

---

[111] INORMS. *More Than Our Rank*. https://inorms.net/more-than-our-rank/
[112] INORMS. *More Than Our Rank*. https://inorms.net/more-than-our-rank/
[113] Research Quality Plus. (2018). *IDRC* (14 June). https://www.idrc.ca/en/rqplus
[114] Holtrop, T. (29 November 2018). The evaluative inquiry: a new approach to research evaluation. http://blogs.lse.ac.uk/impactofsocialsciences/2018/11/29/the-evaluative-inquiry-a-new-approach-to-research-evaluation/
[115] Hatch, A. & Schmidt, R. (2021). *SPACE to evolve academic assessment: A rubric for analyzing institutional conditions and progress indicators*. 10.5281/zenodo.4927605, Zenodo. https://doi.org/10.5281/zenodo.4927605
[116] INORMS Research Evaluation Group. (2021). *The SCOPE Framework*. Emerald Publishing. https://inorms.net/wp-content/uploads/2022/03/21655-scope-guide-v10.pdf
[117] *Recognition and Rewards*. https://recognitionrewards.nl/
[118] CoARA. (2022). *Agreement on Reforming Research Assessment*. https://coara.eu/agreement/the-agreement-full-text/

Two groups were identified through the roundtables as having a greater role to play in the development of institutional research assessment approaches going forward. These were Human Resources (HR) and Strategic Planning staff. HR staff may question the relevance of their roles in relation to research assessment, but as the institutional owners of recruitment, redundancy, promotion and probation processes and EDI ambitions, their support for responsible research evaluation is critical. Similarly, whilst research managers and strategic planners may oversee responsible metrics policies and oversee research management data collection, institutional planners are the owners of institutional KPIs, league table submissions, and the visualisation and presentation of a much broader range of HE management data (including HESA). As such, they have an important role to play in actualising institutions' responsible research evaluation policies. ***(See Recommendation 4: Revitalise the UK Forum.)***

Community-led evaluation design

There are technical aspects to co-design that need to be considered. For research managers, one of the current limitations of HESA data is the time delay between submitting their own data and having access to that of other institutions for benchmarking. Roundtable participants spoke positively about the Snowball Metrics[119] initiative, which promised clearly defined indicators calculated to set 'recipes' with the option to share early data with other HEIs on a 'I'll show you mine if you show me yours' basis. Snowball stalled due to technical difficulties with the data exchange software and concerns about the small range of institutions involved in the development of the indicators. A formal piece of stakeholder-led work (perhaps using an evaluation framework such as SCOPE[120]) should be undertaken to agree what a healthy university research sector looks like, what the indicators of a healthy research ecosystem might be, and how we might monitor (and visualise) those things in timely ways that do not risk impairing the health of the sector. In this way, if additional measures are brought into the design agreed by the FRAP, they should be of value to the institution beyond their contribution to REF preparations. ***(See Recommendation 9: Use data for good.)***

What might such indicators look like? The Kain report[121] on use of metrics in the environment statement and Research England's guidance on environment indicators[122] provide a good starting point for a set of additional indicators. These included the percentage of staff on permanent contracts, gender, ethnicity and disability profiles of staff, and signatory status on a range of concordats and charters. Further possibilities are provided in Box 2 above. Ultimately, it is important that the design and selection of new indicators should be a community-owned exercise conducted within a formal evaluation framework that sense-checks the outcomes for any unintended consequences and discriminatory effects. To avoid the danger of adding unduly to the burden of the REF, development of new indicators should focus on the most important KPIs.

---

[119] *Snowball Metrics: Standardized research metrics – By the sector for the sector*. https://snowballmetrics.com/

[120] INORMS Research Evaluation Group. (2021). *The SCOPE Framework*. Emerald Publishing. https://inorms.net/wp-content/uploads/2022/03/21655-scope-guide-v10.pdf

[121] UK Forum for Responsible Research Metrics. (2018). *UK Progress towards the use of metrics responsibly: Three years on from The Metric Tide report*. http://dera.ioe.ac.uk/id/eprint/31945

[122] *Guidance for institutions on environment indicators*. https://www.ref.ac.uk/media/1019/guidance-on-environment-indicators.pdf

4: CONCLUSION AND NEXT STEPS

This review is one contribution to a mixed-method, multi-stage process of evaluation and reflection on the future of the REF and research assessment more broadly. That process – the Future Research Assessment Programme (FRAP) – will continue for several months yet.

As with the original *Metric Tide* review, we are also speaking here to a broader sweep of audiences and agendas. Responsible research assessment is a global conversation, and in addition to DORA, the Leiden Manifesto and other key initiatives, we now have the important contribution that CoARA is expected to make. The UK research community has played a crucial role in these international movements over the past decade and it is important that it continues to do so.

Looking back at the recommendations of *The* M*etric Tide*, it is encouraging to see how much progress there has been. But significant work remains. And the contours of the conversation have shifted questions of research culture to the centre. In 2015, we talked a lot about the move from quality and impact to diverse qualities and impacts. Now our ten new recommendations emphasise the importance of *how* research is done and impacts are generated, and the wider range of contributors that make this possible. As Professor Dame Ottoline Leyser, Chief Executive of UKRI, reminded us last year:

> *"There are hardly any domains of research and innovation where it is possible to make progress as a lone genius…Research and innovation need technicians, administrators, project managers, librarians, archivists, IT specialists and communication experts, to name but a few….We need to build a truly inclusive system that values and nurtures a much wider range of careers and career paths."*[123]

Some may feel that a fresh review of metrics was unnecessary, but a healthy research culture is one that is committed to reflecting critically on its own arguments and practices. It is one that invests progressively in open data, infrastructures and expertise. And it is one that systematically turns the tools of research back on itself: as we work collectively, if unevenly, towards a more evidence-informed research system.

[123] Leyser, O. (2021). Research's 'lone genius' image is unhelpful. *UKRI* (4 February). https://www.ukri.org/blog/researchs-lone-genius-image-is-unhelpful/

These are all necessary but not sufficient ingredients. Progress on this agenda also requires leadership from those institutions and individuals who set the terms of trade in UK research. These leaders need to rise above short-term horizons to make the case for the kinds of research cultures that so many of us want to see. The FRAP process will provide those leaders with the evidence that they need. The choice of how and when to act is theirs.

Appendix A: Progress of recommendations from *The Metric Tide* (2015)

No.	Recommendation	Comment
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)	We detect greater awareness of how quantitative indicators should and shouldn't be used, and of the proxy nature of many metrics. In the roundtable discussions, concern was expressed that greater awareness had driven some problematic uses of metrics 'underground'. This could be addressed by institutions providing public statements on how quantitative indicators will be used in assessment and clear pathways for deviant practice to be challenged.
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)	Over 90 universities and research institutes in the UK have signed DORA or openly adopted responsible research assessment policies. This represents significant progress (stimulated in part by the Wellcome Trust requiring funded institutions to adopt responsible research assessment principles aligned with DORA, the Leiden Manifesto or equivalent initiatives). However, it is often difficult to find publicly available evidence that these policies or commitments are being implemented robustly – or that sufficient progress has been made to developing indicators to support to the quality of the research *process* (*e.g.* research integrity, mentorship, equitable treatment of staff and students; addressing bullying and harassment). Moreover, we see little evidence that institutions using league tables have been open about their reasons for doing so or their awareness of the limitations of university ranking methods.
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 [ECU (now part of Advance HE)]; 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)	ARMA has worked hard to support the roll-out of responsible metrics within institutions by offering webinars and hosting a dedicated Research Evaluation Special Interest Group. In their term as Chair of the INORMS Council they instituted the INORMS Research Evaluation Group. This was the birthplace of the value-led SCOPE Framework for responsible research evaluation which has been adopted by a number of institutions, the guide 'Five Arguments to Persuade HE Leaders to Evaluate Responsibly', and the University Ranker Ratings. ARMA continues to support the INORMS REG beyond their term of office. Research support professionals have supported the former Consortium on Advancing Standards in Research Administration Information (CASRAI) including the development of the CRediT taxonomy. They are also strong advocates for ORCID IDs, work with CRIS suppliers to support data interoperability, and led much of the work on the Snowball initiative.

No.	Recommendation	Comment
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)	We have seen progress in this area linked to the uptake of DORA and other responsible research assessment practices. As with recommendation 2, evidence of impact is difficult to ascertain. In line with the responsible metric principle of transparency, there is scope to use the REF to do a national audit of practice in this area; e.g. require a public link to the university recruitment and promotion criteria and processes as part of a structured environment statement. Staff involved in assessment should be given the time to do it to a high standard. Publication of criteria and processes should enable internal and external challenge (and foster a culture of trust); it would likely also stimulate the uptake of good practice. Cautions against the use of aggregate metrics such as the JIF should extend to the H-index and the scoring mechanisms of university ranking).
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)	We detect greater awareness in some quarters, but it is impossible to assess quantitatively. The work of UKRI and others on the use of narrative or evidence-based CVs should help researchers explore alternative ways of presenting their own achievements and assessing others. It would not be appropriate for national funders to take action to enforce researcher behaviour. However, institutions have a role to play by ensuring that researchers have a voice in developing new assessment processes. The FRRM could support institutions in doing so.
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)	There has been notable progress in this area, not just within the UK (e.g. the Wellcome Trust policy noted above, and the adoption of narrative CVs by UKRI and other funders), but internationally. Nevertheless there remains scope for achieving greater consistency in responsible research assessment practices between UK funders, which, as noted by the recent Tickell review, has the potential for reducing research bureaucracy. The funding organisations responsible for the REF, the largest UK research assessment exercise, should take responsibility for achieving this.
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)	There have been some useful moves towards greater transparency (e.g. in allowing institutions to verify citation data provided by Clarivate to REF panels). We also note more openness from Clarivate around responsible practice (including the publication of their 'Profiles, not metrics' paper and their recent decision to reduce the JIF to 1 decimal place) and the commitments made by Elsevier, another data provider, to DORA and the Leiden Manifesto. Nevertheless, there is concern that the transfer of university data to commercial companies remains a barrier to interoperability. Moreover, if anything, we have seen disengagement of university rankers from informed critique.

No.	Recommendation	Comment
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)	There is some evidence of progress in this area (e.g., adoption of citation distributions by a number of journals, made easier since these are now provided by Clarivate). A growing number of publishers (including major companies such as Springer-Nature, Elsevier, Wiley, Taylor & Francis) have made public commitments to DORA. As with research institutions, we would encourage publishers to make it clear to authors and readers exactly how these commitments are enacted.
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)	In parallel with the momentum behind Open Access, and Open Scholarship more broadly, there have been moves to create a more open infrastructure for the data used to monitor research. These include practical steps such as the extraordinarily successful Initiative for Open CItations, and articulation of Principles for Open Scholarly Infrastructure. Although there are community-led efforts to create open infrastructures for research and research data (including Open Alex and the Invest in Open Infrastructure initiative), the landscape is complicated by the fact that much research intelligence is sold by private companies. Work to develop a set of principles to underpin open and trustworthy research information management was initiated by the FfRRM and others, but has stalled. This report recommends that this work is reinstated, perhaps at a higher level, to ensure community expectations are met in the design, purchase and implementation of infrastructures and services that support research information management infrastructure. We see this as necessary to ensure the transparency demanded by the responsible use of metrics.
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)	The use of ORCIDs in REF 2021 was optional rather than mandatory, however, we believe this recommendation still has merit since it aligns with the drive to reduce bureaucratic burden for institutions and researchers. We understand there is evidence from Australia that mandating ORCIDs as part of their ERA assessment exercise was effective at increasing uptake. It may be that submitting ORCIDs as part of the UK HESA staff submission may be a better way of encouraging uptake. This should be explored.

No.	Recommendation	Comment
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)	The open, stakeholder-governed Research Organisations Registry ID has become the 'priority Permanent Identifier (PID)' following a consultation as part of the UK PIDs for OA project. The use of ROR IDs is now commonplace in community-developed data services (e.g., OpenAlex). However ROR IDs do not currently cover commercial organisations with which UK research organisations might work. Bridging identifiers that map between different organisational IDs (including ISNIs) are therefore important. The more systematic uptake of ROR IDs, for example by publishers disclosing data to Crossref, would lead to a better connected data infrastructure.
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)	It is not entirely clear how many publishers have mandated ORCID IDs or other PIDs and in many cases funders have more leverage in this space than publishers. (Many funders mandate an acknowledgement of their funding source on publications, for example). However, it is now generally agreed that a better way of progressing the adoption of PIDs is through consensus and advocacy rather than mandating behaviours. The UK Research Identifiers National Coordination Committee (RINCC) has proposed the establishment of a UK PID Support Network which could take on this role. It is recommended that this work is prioritised and supported. (UKRI, Jisc)
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)	To fully include, describe and connect the wide variety of output that results from research activity, and to reduce duplicate data entry, all outputs require PIDs. However, this does not necessarily need to be the DOI. Indeed there are good reasons why this might not be appropriate. Not all outputs are digital, for example; work is currently ongoing to consider identifiers for Practice-based Research. Some outputs already have well-established PIDs, such as books (IBSNs) and journals (ISSN-Ls). It is therefore recommended that work to support the development of PIDs that describe a wider range of outputs is prioritised and funded. (UKRI, Jisc)

No.	Recommendation	Comment
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)	This recommendation was somewhat vague so it is difficult to comment accurately as to whether it was achieved. Certainly, Jisc continues to support a programme of research information management infrastructure activity and ARMA members have supported work in this space also (e.g., CRediT and Snowball). A particular challenge is the identification of a clear locus of responsibility for championing the prioritisation of (and therefore investment in) research information infrastructure developments. This leads to some infrastructure work being funded on a project basis affecting its sustainability. It is recommended that the mechanisms by which research information infrastructure is governed and realised are reviewed to ensure they are working effectively. A sector-owned approach to creating sustainable and persistent research information management infrastructure should be supported to ensure the level of standards, integrity and equity is developed and maintained over the long term. BEIS, Jisc, ARMA, the Forum for Responsible Research Metrics and research funders should continue to develop this approach in support of the recommendations of Professor Adam Tickell's report on reducing research bureaucracy.
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)	Jisc's new research and innovation strategy focuses on how the data produced through the processes of research management could be used on a greater scale for greater efficiency and equity. It recognises that expanded approaches to sustainable, longitudinal data assets are required. The Open Data about Research project (ODAR) is in development and work is ongoing to ensure high quality, standards-based and machine actionable metadata is appropriate to adequately describe target research datasets and to mint DataCite DOIs. We note that ResearchFish data has been used just 25 times in UKRI publications between 2016-2020, and we are unconvinced that the benefits of this platform outweigh the costs.
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)	We see little evidence from BEIS that the information gathered via the REF is used to inform strategic planning and question whether this should remain as a declared purpose of the REF. It may be more agile to develop more focused or time-limited analytical methods to use the information gathered in the REF for policy or strategy development at national level.

No	**Recommendation**	**Comment**
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.	Done, though only about one third of the UoAs requested quantitative data. In REF 2021.
	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.	Done, in a necessarily 'light touch' and flexible manner. We agree with the recommendations that emerged from the FRRM that overly prescriptive demands for quantitative indicators in impact case studies might put at risk the capture of the fullness and richness of UK research impact.
	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.	Done. The FRRM developed a useful set of principles for deciding on the use of quantitative indicators in REF environment statements which were adopted in REF 2021. Nevertheless we see opportunities for facilitating more quantitative comparisons between institutional research environments by the use of more structured statements (perhaps modelled on narrative/evidence-based CVs) and the integration of cultural indicators gathered from other benchmarking schemes (*e.g.* Athena Swan, Race Equality Charter).

No	**Recommendation**	**Comment**
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)	The FRRM has provided a valuable forum for discussion and for steering implementation of some of the recommendations of *The Metric Tide* (e.g the three-year review in 2019; guidance on indicators for environment statements and impact case studies). However, we believe there is scope for the forum to be more proactive and more visible in its work to support the adoption of responsible research assessment practices in the UK.
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)	This recommendation has been achieved through the establishment of the Research on Research Institute in 2019. It has quickly established a national and international profile and placed itself on a sustainable footing having secured international funding and incorporated itself as a nonprofit social enterprise. Now in 'phase 2' of its operations, the institute's next wave of projects will span four themes: • experiments with evaluation; • infrastructures and data sharing; • research priorities and portfolios; and • impacts, indicators and culture change
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)	This recommendation has not captured the imagination of the researcher community. We would in any case prefer to be positive and celebrate examples of best practice, in a similar vein to the case studies collected by DORA. We recommend that universities extol the quality of their RRA practices as part of the REF Environment statements. This will further contribute to building a repository of good practice.

Appendix B: Literature Review

Compiled and written by: Noémie Aubert Bonn (Research England)

Reviewed and edited by: Stephen Curry, Elizabeth Gadd, Mike Thelwall, Ludo Waltman and James Wilsdon.

Contents

Upon its original publication, *The Metric Tide* was accompanied by a literature review which examined the roles of scientometrics, peer review, and alternative metrics in measuring and assessing research.[124] In a similar vein, this appendix accompanies the report '*Harnessing the Metric Tide: indicators, infrastructures and priorities for responsible research assessment in the UK'* to provide an overview of important developments that have changed scientometrics and shaped the discussion on research assessment in the seven years since *The Metric Tide* was published.

This literature review is divided into three sections. In the first, we look at recent advances in scientometric databases and in metrics and alternative metrics. In the second, we consider more broadly international developments, going through key statements and documents that have shaped the discussion on research assessment since 2015. Finally, we give an overview of common recommendations for responsible research assessment included in the statements and documents described in section 2, and showcase examples of these recommendations in research settings.

SECTION 1 – ADVANCES IN SCIENTOMETRIC INDICATORS SINCE 2015

1.1 Advances in research databases

Academic search engines and bibliographic databases have continued to evolve and grow, reaching greater coverage of scientific outputs than ever before, with many of them covering over a hundred million records and a billion cited references. Beyond their coverage, bibliographic databases also evolve and adapt to the growing diversity of research outputs, building means to capture datasets, computer code, grant data, patents, and policy documents, for example. The publishing landscape is also changing, with new elements being added as relevant research outputs. For instance, the growing prevalence of preprints, accentuated by the COVID-19 pandemic, changed the needs for metrics capture[125] and encouraged databases to include preprints in their records. Scopus included preprints in early 2021.[126] Several new databases have also become available, and while we cannot cover all of them, we want to mention a few which have assumed a dominant role in the scientometric community.

Dimensions[127] is one of these new databases that is now commonly used by the scientific community. Dimensions was released in 2018 and rapidly became widely used in scientometrics and meta-research. Dimensions is a product of Digital Science, the same company that now owns Altmetric.com. In addition to publications, Dimensions also covers a large grants database and linkage data as well as a variety of additional outputs such as

[124] Wilsdon, J. et al. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. doi:10.13140/RG.2.1.4929.1363, HEFCE. https://re.ukri.org/documents/hefce-documents/metric-tide-2015-pdf/

[125] Watson. (2022). Rise of the preprint: how rapid data sharing during COVID-19 has changed science forever. *Nature Medicine* 28, 2-5, doi:10.1038/s41591-021-01654-6

[126] McCullough. (2021). Preprints are now in Scopus! *SCOPUS* https://blog.scopus.com/posts/preprints-are-now-in-scopus

[127] Dimensions. https://www.dimensions.ai

patents, clinical trials, attention indicators, and policy documents.[128] While specific services in Dimensions need to be paid for, for example large API downloads, the database provides free access to non-commercial users and has therefore become widely used among researchers.

Another new dataset of bibliographic references is COCI,[129] the OpenCitations Index of Crossref[130] open DOI-to-DOI citations. COCI harvests open data from Crossref. Initially, only a very small proportion of the citations in Crossref were open. Yet, with the push and advocacy from the Initiative for Open Citations (I4OC) and other players, publishers started to open their citations. In June 2022 Crossref changed its user terms to ensure that "all Metadata and Identifiers registered with Crossref are made available for reuse without restriction through (but not limited to) public APIs and search interfaces."[131]

Microsoft Academic is another example. Released in 2016 as a follow up to Microsoft Academic Search, a database retired in 2012, Microsoft Academic was found to provide the largest citation coverage after Google Scholar.[132] It had the advantage of not limiting its content to journals, but rather used AI and web crawlers to populate its content while also enabling open and free downloads of large volumes of data through its API. Despite these capabilities, the database was retired at the end of 2021.[133] Data from the Microsoft Academic database continues to be used by other databases (see below).[134]

OpenAlex[135] can be seen as a successor of Microsoft Academic. It is a promising database that is likely to become increasingly important in the next few years. OpenAlex collects scholarly entities (i.e., 'works', 'authors', 'venues', 'institutions', and 'concepts') and links them together to create a huge web of connections. The data is available via an API and a full database snapshot, and a website with a graphical user interface is also being developed. Much like COCI, OpenAlex was developed with an open science mindset: it is completely free, openly accessible to anyone, and based on an open-source codebase. In fact, the entire database is available under a Creative Commons CC0 licence. The data is largely derived from the Microsoft Academic database, complemented by data from Crossref, Open Researcher and Contributor ID[136] (ORCID), Research Organisation Registry[137] (ROR), the Directory of Open Access Journals[138] (DOAJ) and other data repositories. OpenAlex is one of the most comprehensive open scientometric databases.

---

[128] Herzog, Hook and Konkiel. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies* 1, 387-395, doi:10.1162/qss_a_00020

[129] OpenCitations. COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. https://opencitations.net/index/coci

[130] *Crossref*. https://www.crossref.org/

[131] Hendricks, Rittman and Bartell. (2022). Amendments to membership terms to open reference distribution and include UK jurisdiction. *CrossRef Blog* (4 April). https://www.crossref.org/blog/amendments-to-membership-terms-to-open-reference-distribution-and-include-uk-jurisdiction/

[132] Martín-Martín, Thelwall, Orduna-Malea and Delgado López-Cózar. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics* 126, 871-906, doi:10.1007/s11192-020-03690-4; Harzing and Alakangas. (2017). Microsoft Academic: is the phoenix getting wings? *Scientometrics* 110, 371-383, doi:10.1007/s11192-016-2185-x

[133] Next Steps for Microsoft Academic – Expanding into New Horizons. (2021). *Microsoft Academic* (May 4). https://www.microsoft.com/en-us/research/project/academic/articles/microsoft-academic-to-expand-horizons-with-community-driven-approach/

[134] Tay, Martín-Martín and Hug. (2021). Goodbye, Microsoft Academic – Hello, open research infrastructure? *LSE Impact Blog* (27 May). https://blogs.lse.ac.uk/impactofsocialsciences/2021/05/27/goodbye-microsoft-academic-hello-open-research-infrastructure/

[135] OpenAlex. https://openalex.org/

[136] ORCID. https://orcid.org/

[137] ROR. https://ror.org/

[138] Directory of Open Access. https://doaj.org

Another interesting and relatively new data source is Overton[139], which provides a vast searchable index of policy documents, guidelines, think tank publications, and working papers. In other words, Overton captures the link between research, people, and policies, enabling users to track the evolution of ideas from research or think tanks to governments and legislations.

Together, the broadening of existing scientometric databases and the creation of new ones has enabled novel research and harvesting possibilities. For instance, bibliographic databases now enable the collection of data that extends far beyond scientific publications. Books, preprints, grants, patents, datasets, software, and code are increasingly discoverable, enabling users to grasp a much richer picture of the scientific landscape. The increasing openness of the data also reshapes the types of uses that can be made, enabling users with different levels of resources to access and verify data. Yet, the multiplicity and enduring limits of such databases are not without challenges. For example, coverage often differs between the databases, and all databases suffer from gaps and inaccuracies in their data. The choice of a database can thus have an impact on the way science is measured and represented. Knowing that some disciplines are underrepresented in specific databases,[140] building a strong understanding of the limits of each database and reflecting on the right database to use continues to be crucial for adequate use. Finally, although the more comprehensive data enables a much richer understanding of scientific knowledge production than it did some years ago, it still has some lacunae in capturing grey literature such as policy documents and reports without Digital Object Identifiers[141] (DOIs), non-English publications, and implementation of research results in practice. To provide value, the increasing reliance on metrics needs to remain aware of such limits, and conversely, research organisations need to keep investing in improving the availability of high-quality data on their outputs.

1.2 Advances in data and indicators

Some elements of the data and indicators landscape also changed since the release of *The Metric Tide* in 2015.

First, there have been changes in the popularity and use of some indicators. For example, Altmetric.com,[142] which was discussed in *The Metric Tide* as an emerging source of alternative indicators, has acquired substantial popularity in the past few years and is now commonly displayed in journals and preprint servers from many publishers and research societies, including John Wiley & Sons, Taylor and Francis, The JAMA Network, and Springer Nature.

Conversely, there has been a decreased focus on the Journal Impact Factor (JIF) with a growing number of publishers and individual journals signing the San Francisco Declaration on Research Assessment (DORA[143]). As a result, a few journals entirely removed the JIF from

[139] Overton. https://www.overton.io/
[140] Martín-Martín, Thelwall, Orduna-Malea and Delgado López-Cózar. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics* 126, 871-906, doi:10.1007/s11192-020-03690-4
[141] doi. https://www.doi.org/
[142] Altmetric. https://www.altmetric.com/
[143] *San Francisco Declaration on Research Assessment*. https://sfdora.org/read/

their websites and communications.[144] Most publishers still advertise the JIF on their website, but it is now generally accompanied by additional indicators such as the Source Normalized Impact per Paper[145] (SNIP) or the SCImago Journal Rank[146] (SJR). A proposal to accompany the JIF with a graph of its distribution[147] has also been adopted by Clarivate as a way to contextualise the indicator and better inform users of its meaning.[148] Consequently, although it is still ubiquitous, the JIF is less often presented in isolation and more often contextualised with other indicators as was recommended in *The Metric Tide*. Along the same lines, in 2020, the Ministry of Science and Technology and the Ministry of Education in China announced that they would change research evaluations by moving away from bibliometric indicators based on the Web Of Science database, notably the Science Citation Index (SCI) and the Journal Impact Factor.[149] This decision was motivated in part by a lack of scientific communication in Chinese outlets in the early days of the COVID-19 pandemic and a realisation that the focus on international publications may have delayed the Chinese response to the pandemic.[150] While the concrete details on how this reform should be implemented are still largely missing, the move away from SCI indicators and JIFs from a country who "is now the largest contributor of research papers worldwide"[151] could have important repercussions for the future role and popularity of such metrics.

New data and indicators have also been created or explored for use in research assessment.

Journal-level indicators. CiteScore,[152] for example, is a journal-level indicator that was proposed by Elsevier in late 2016. Like the JIF, CiteScore provides information about the average citation performance of journals, but with the difference that its data can be freely accessed.[153] While the increased openness and transparency of CiteScore has been welcomed, the addition of yet another journal-level indicator to the research assessment landscape also raised concerns, especially at a time when journals were starting to respond to criticisms of the JIF. Despite these worries, CiteScore is now used and displayed by several large-scale publishers including Elsevier, Emerald, Frontiers, Hindawi, Inderscience, MDPI, SAGE, Taylor & Francis, and Walter de Gruyter.

Article-level indicators. At the article-level, PlumX provides a notable set of article level altmetrics.[154] Launched in 2015, it has grown in visibility since its purchase by Elsevier in 2017. PlumX provides an overview of citations (including Scopus citations and clinical or policy

[144] Casadevall, Bertuzzi, Buchmeier, Davis, Drake, Fang, Gilbert, Goldman, Imperiale, Matsumura, McAdam, Pasetti, Sandri-Goldin, Silhavy, Rice, Young and Shenk. (2016). ASM Journals Eliminate Impact Factor Information from Journal Websites. *mSystems* 1, e00088-00016, doi:10.1128/mSystems.00088-16; PLOS. PLOS and DORA. https://plos.org/publish/dora/

[145] SNIP. https://journalinsights.elsevier.com/journals/0969-806X/snip

[146] Scimago Journal & Country Rank. https://www.scimagojr.com

[147] Callaway. (2016). Beat it, impact factor! Publishing elite turns against controversial metric. *Nature* 535, 210-211, doi:10.1038/nature.2016.20224; Larivière, Kiermer, MacCallum, McNutt, Patterson, Pulverer, Swaminathan, Taylor and Curry. (2016). A simple proposal for the publication of journal citation distributions. *bioRxiv*, doi:10.1101/062109

[148] Refresh of the Journal Citation Reports data. (2018). *Clarivate Blog* (17 October). https://clarivate.com/blog/refresh-of-the-journal-citation-reports-data/

[149] China's research-evaluation revamp should not mean fewer international collaborations. (2020). *Nature* 579, doi:10.1038/d41586-020-00625-0; Zhang and Sivertsen. (2020). The New Research Assessment Reform in China and Its Implementation. *Scholarly Assessment Reports* 2, 3, doi:10.29024/sar.15

[150] Shu, Liu and Larivière. (2022). China's Research Evaluation Reform: What are the Consequences for Global Science? *Minerva* 60, 329-347, doi:10.1007/s11024-022-09468-7

[151] Ibid p. 338

[152] Elsevier. CiteScore: a new metric to help you track journal performance and make decisions. https://www.elsevier.com/connect/editors-update/citescore-a-new-metric-to-help-you-choose-the-right-journal

[153] Teixeira da Silva and Memon. (2017). CiteScore: A cite for sore eyes, or a valuable, transparent metric? *Scientometrics* 111, 553-556, doi:10.1007/s11192-017-2250-0

[154] Plum Analytics. About PlumX Metrics. https://plumanalytics.com/learn/about-metrics/

citations), usage (including clicks, downloads, views, library holdings, and video plays), captures (including bookmarks, code forks, 'favourites', readers, and watchers), mentions (for example on blog posts, comments, reviews, Wikipedia references, or news media), and social media activity such as shares, likes, comments, or tweets, making it similar to the Altmetric.com approach, but with slightly different coverage in certain sources (e.g., less capture of Twitter in favour of a greater capture of Facebook and Mendeley[155]). PlumX is now the dominant altmetric on all Elsevier products, including Elsevier journals, Scopus, PURE, and ScienceDirect. Crossref Event Data[156] may offer an open alternative to commercial services such as PlumX and Altmetric.com, but its adoption still seems very limited. Another new service is *scite.*[157] It offers citation statistics that distinguish between citations classified as 'contrasting', 'supporting' and 'mentioning'.

Author-level data and indicators. New author-level data and indicators have also become available or gained importance in the past few years. Author profiles on ORCID are a prime example that contain a variety of personal data (e.g., employment, publications, funding, and peer-review). ORCID IDs enable the disambiguation of authors (e.g., authors with the same name) and can therefore improve the reliability of indicators based on author characteristics. One of the recommendations from *The Metric Tide* and of the Expert Group on Altmetrics commissioned by the European Commission in 2016 was in fact to make ORCID accounts mandatory for all European researchers in order to improve the usefulness and the reliability of data.[158] Several research institutions now instruct research staff to create an ORCID and some publishers – including Open Research Europe, the publishing platform of the European Commission – only allow manuscript submissions through an ORCID account. In the UK, Wellcome Trust and NIHR now mandate ORCID identifiers from corresponding applicants, while the REF strongly encourages them and expects to require them in future exercises.[159] Several other funders worldwide are moving in this direction.[160] Although ORCID does not provide indicators directly, it allows other organisations to do so. For instance, ORCID data is used by ImpactStory[161] – an open-source web-based software that enables authors to create author profiles that showcase, among other things, the openness and reuse of their research outputs.

Increasingly, narrative CVs are also being adopted by funders and research institutions around the world. Some examples include the Résumé for Research and Innovation (R4RI) at UKRI, SciCV at the Swiss National Science Foundation's (SNSF), the Individual Narrative Profile from the Luxembourg National Research Funds, and many others. We will discuss narrative CVs further in Section 3, but in short, they are CV formats that include a narrative element, enabling individuals to explain the impact of their research in their own words. For

[155] Karmakar, Banshal and Singh. (2021). A large-scale comparison of coverage and mentions captured by the two altmetric aggregators: Altmetric.com and PlumX. *Scientometrics* 126, 4465-4489, doi:10.1007/s11192-021-03941-y

[156] Crossref. Event Data. https://www.crossref.org/services/event-data/

[157] scite. https://www.scite.ai

[158] European Commission, Directorate-General for Research and Innovation, Peters, Frodeman, Wilsdon, Bar-Ilan, Lex and Wouters. (2017). *Next-generation metrics : responsible metrics and evaluation for open science*. (Publications Office). doi:10.2777/337729 https://data.europa.eu/doi/10.2777/337729

[159] *REF 2021 Decisions on staff and outputs*. (2017 (Updated 2018)). Bristol, England. https://www.ref.ac.uk/media/1034/ref-2017_04-decisions-updated-11042018.pdf; NIHR. Generic supporting information for applicants. https://www.nihr.ac.uk/documents/generic-supporting-information-for-applicants/28196#orcid; Wellcome. Open Researcher and Contributor ID (ORCID). https://wellcome.org/grant-funding/open-researcher-and-contributor-id-orcid

[160] ORCID. Funders' ORCID Policies. https://info.orcid.org/funders-orcid-policies/

[161] Impactstory. https://profiles.impactstory.org

this reason, narrative CVs provide a different way to capture elements that are not necessarily captured by metrics. At the moment, these CVs generally remain with funders or research institutions and are not openly accessible, but research on the elements captured in these CVs could open new venues about the elements of impact that are important to researchers. Ongoing research from the Luxembourg National Research Funds[162] as well as from a collaboration between the Swiss National Science Foundation (SNCF) and the Leiden Center for Science and Technology Studies (CWTS)[163] provide insights on the overall reception and use of these CVs, but the actual content and possible extraction of information and indicators is still not streamlined.

While these areas provide new venues for individual indicators, other indicators have also been retired. For example, the ResearchGate Score which was released in 2015 and highly criticised for its lack of transparency[164] was removed in March 2022, with an open letter stating that the indicator did not fit the criteria of a good indicator.[165]

Granular indicators. The seeds for new indicators at a more granular level have also started to appear. For example, many journals now ask authors to describe their contributions to the papers published. A standard author contribution taxonomy called CRediT (Contributor Role Taxonomy) is now implemented across an increasing number of journals, enabling analyses of individual contributions, roles, and team dynamics. So far this has only been done for scholarly analysis rather than assessment.[166]

Datasets are also starting to be included in many of the major databases. This is aided by the addition of Digital Object Identifiers[167] (DOIs) on datasets from data repositories such as Zenodo, Figshare, and the Open Science Framework, which opened the door to analyses linking datasets to publications and other outputs.[168] DOIs are also increasingly added to other types of material including policy reports and scientific presentations, aligning with the recommendations from *The Metric Tide*. In parallel, there are more and more institutional, disciplinary, and general data repositories that support data sharing and Google Dataset Search to help find them.

With the increasing digitisation of science, many more measurements of otherwise invisible contributions are becoming possible. These include Research Activity Identifiers[169] (RAiDs), which provide unique identifiers that can be used to link a variety of elements attached to a research project such as people (including researchers but also support staff and students) publications, instruments, and institutions involved. RAiDs were introduced by the Australian

[162] Luxembourg National Research Fund. (2022). *Narrative CV: Implementation and feedback results*. https://www.fnr.lu/narrative-cv-implementation-and-feedback-results/

[163] Singh Chawla. (2022). Swiss funder unveils new CV format to make grant evaluation fairer. *Nature* 606, 1033-1034, doi:10.1038/d41586-022-01599-x; Strinzel, Kaltenbrunner, van der Weijden, von Arx and Hill. (2022). SciCV, the Swiss National Science Foundation's new CV format. *bioRxiv*, doi:10.1101/2022.03.16.484596

[164] Kraker, Jordan and Lex. (2015). The ResearchGate Score: a good example of a bad metric. *LSE Impact Blog* (9 December). https://blogs.lse.ac.uk/impactofsocialsciences/2015/12/09/the-researchgate-score-a-good-example-of-a-bad-metric/

[165] Why we're removing the RG Score (and what's next). (2022). *ResearchGate Updates* (29 March). https://www.researchgate.net/researchgate-updates/removing-the-rg-score

[166] Larivière, Pontille and Sugimoto. (2021). Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). *Quantitative Science Studies* 2, 111-128, doi:10.1162/qss_a_00097

[167] doi. https://www.doi.org/

[168] European Commission, Directorate-General for Research and Innovation, Peters, Frodeman, Wilsdon, Bar-Ilan, Lex and Wouters. (2017). *Next-generation metrics : responsible metrics and evaluation for open science*. (Publications Office). doi:10.2777/337729 https://data.europa.eu/doi/10.2777/337729

[169] Research Activity Identifier. https://www.raid.org.au

Research Data Commons and are in active use in several Australian organisations.[170] Other identifiers that have risen since 2015 include identifiers for research organisations such as the Research Organization Registry (ROR) and the Global Research Identifier Database (GRID), the latter of which has recently been discontinued and merged with ROR. Peer review is also becoming more visible through open and transparent peer-review models which can attribute DOIs to peer-reviews, through post-publication peer-review sites such as PubPeer,[171] PREreview[172] and ScienceOpen,[173] and through linkable peer-review acknowledgements in ORCID and Web of Science Author Records (the latter recently acquired and integrated records previously captured by Publons). While the use of these options in research assessment still seems rare, they can offer new ways to capture the dynamics of knowledge creation.

SECTION 2 – INTERNATIONAL DEVELOPMENTS ON RESEARCH ASSESSMENT: FROM EARLY ENDORSEMENT TO MOMENTUM FOR CHANGE

Beyond metrics, the landscape and ongoing discussions on research assessment have progressed substantially since the publication of *The Metric Tide* in 2015. An abundance of new position papers, policy documents, and initiatives from a variety of stakeholders have been released, demonstrating a keen interest in responsible research assessment and a strong momentum for change. Statements and public announcements are useful for situating awareness of different regions and stakeholders and indicating how this awareness has come about. In this section, we provide a short overview of important changes that played a role in shaping the landscape of research assessment.

2.1 DORA: signatories and growing support

Although the 2012 Council of Canadian Academies (CAC) report[174] may be one of the first policy documents to draw explicit attention to responsible research assessment, the San Francisco Declaration on Research Assessment[175] (DORA), published a year later, is without doubt one of the most well-known and influential position statements on the topic. Published in May 2013, DORA played a crucial role in raising awareness of the responsible use of research metrics. At the time of publication of *The Metric Tide* in 2015, DORA had gathered signatures from 570 organisations (e.g., universities, research funders, publishers). Since then, the number of organisation signatories has increased five-fold, with over 2600 organisations now having signed the declaration, 242 of which are based in the UK.

On a global scale, DORA also had an important impact on research assessment practice, particularly since it transformed itself in 2018 into a campaigning initiative backed by

[170] Manista. (2020). There's A PID For That, Part 2: Projects. *Jisc scholarly communications* (13 October). https://scholarlycommunications.jiscinvolve.org/wp/2020/10/13/theres-a-pid-for-that-part-2-projects/
[171] PubPeer. https://pubpeer.com/
[172] PREReview. https://prereview.org/
[173] ScienceOpen.com. https://www.scienceopen.com/
[174] The Council of Canadian Academies. (2012). *Informing Research Choices: Indicators and Judgment*. Report No. 978-1-926558-42-4, Ottawa, Canada.
[175] American Society for Cell Biology. (2013). *San Francisco Declaration on Research Assessment*. https://sfdora.org/read/

international funding that collaborates widely to develop and disseminate practical guidance.[176] Arguably, it has set the scene for several national, organisational, and institutional position statements, changes, and reforms. A timeline highlighting a selection of international research assessment statements is presented in Figure 1.

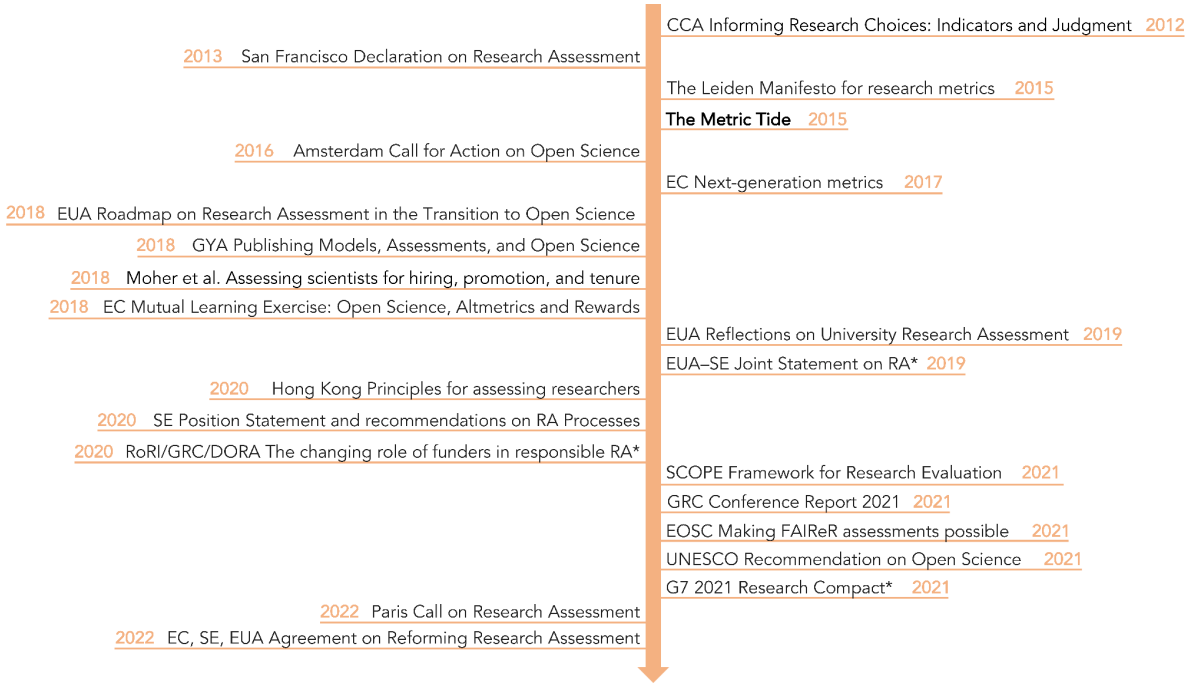

Figure 1. Selection of international statements and recommendations on research assessment.
*Abbreviations: CAC: Council of Canadian Academies; RA: Research Assessment; EUA: European University Association; GYA: Global Young Academy; SE: Science Europe; RoRI: Research on Research Institute; GRC: Global Research Council; EOSC: European Open Science Cloud; EC: European Commission. * Not included in the content summary of Table 1 given the breadth of scope or lack of explicit recommendations*

2.2 Responsible research assessment enter discussions on good science

In the years immediately following the publication of DORA, two important publications further added to the responsible research assessment agenda. Firstly, the 'Research: increasing value, reducing waste' series published in The Lancet in early 2014 accentuated the role of rewards and incentives in the problem of non-reproducible science. Secondly, the report 'The culture of scientific research in the UK' published by the Nuffield Council of Bioethics in late 2014 further confirmed the perception that the exaggerated focus on high JIF publication in hiring, tenure, and promotion was damaging for science.[177] The report went on to discuss the role of the Research Excellence Framework (REF) as a driver of the pressure to publish in high impact journals. These reports, published around the time of the REF 2014 results, certainly helped move the research assessment agenda forward.

[176] Curry. (2018). Let's move beyond the rhetoric: it's time to change how we judge research. *Nature* 554, 147, doi:10.1038/d41586-018-01642-w
[177] Nuffield Council of Bioethics. (2014). *The culture of scientific research in the UK*. https://nuffieldbioethics.org/publications/the-culture-of-scientific-research

2.3 Metrics pushback: The Leiden Manifesto and The Metric Tide

At about the same period, the research community was becoming more attuned to some of the negative consequences of the metrification of research assessments. In the European Union for example, a new Composite Indicator for Scientific and Technological Research Excellence was introduced in 2013 as part of the Europe 2020 strategy. The indicator was created as the EU aimed to boost investment in research and make Europe an attractive research area, allowing to measure and compare the research performance of different EU countries. The composite was constructed through a feasibility study entitled 'ERA monitoring' that was tasked to measure (i) progress in the construction and integration of a European Research Area, (ii) structural change towards a more knowledge-intensive economy in Europe, and (iii) research excellence, including the vitality of the research environment and the quality of research outputs in both basic and applied research.[178] The resulting indicator was calculated from four dimensions: highly cited publications, top rankings in the Leiden Ranking and the Scimago Institutional Ranking, numbers of patent applications, and number of ERC grants.[179] This new indicator raised concerns from researchers since it reframed excellence as a measurable concept which focused mostly on output and shifted away from research investments, capacity building, and research processes.[180] The metrification of high-level research assessment rapidly transitioned into the criteria used to assess researchers within research institutions. Consequently, metrics that had been devised for the purpose of studying science and high-level research performance started being used inappropriately to assess individual researchers.

It is against this increasingly metrics-focused background that the Leiden Manifesto for research metrics was released.[181] The Leiden Manifesto voiced a concern from scientometricians, social scientists, and research administrators and proposed a way of taking responsibility for the risk of misuse that the new indicators had introduced. The release of the Leiden Manifesto was followed by *The Metric Tide* only a few months later. Both documents raised awareness on the need to use research metrics responsibly and re-emphasised the need to think more broadly about research assessment.

2.4 Open Science as an accelerator for responsible research assessment

The growing interest in open science and its importance on research policy agendas in the past decade has also helped to motivate the drive towards responsible research assessment. In fact, looking at Figure 1, the six first statements to follow *The Metric Tide* all arose from actions aiming to foster Open Science.

The entanglement of research assessment and open science is complex and unresolved. Research assessment practices that depended on impact metrics such as the JIF were readily identified as a barrier to the establishment of open access journals. Although many legacy

[178] Vertesy and Tarantola. (2012). *Composite Indicators of Research Excellence*. Report No. JRC72592, Publications Office of the European Union, Luxembourg (Luxembourg). https://publications.jrc.ec.europa.eu/repository/handle/JRC72592

[179] Joint Research Centre, Institute for the Protection and Security of the Citizen, Vertesy, Hardeman, Saisana and Van Roy. (2014). *An analysis of national research systems (I) : a composite indicator for scientific and technological research excellence*. (Publications Office). doi:10.2788/95887 https://data.europa.eu/doi/10.2788/95887

[180] Sørensen, Bloch and Young. (2016). Excellence in the knowledge-based economy: from scientific to research excellence. *European Journal of Higher Education* 6, 217 - 236

[181] Hicks, Wouters, Waltman, Rijcke and Rafols. (2015). The Leiden Manifesto for research metrics. *Nature* 520, 429-431, doi:10.1038/520429a

publishers recognise that JIFs are often used inappropriately in research evaluations, they remain a prominent marketing tool and their allure for researchers, while dimmed, has by no means been extinguished. This explains the ongoing efforts on open access and open science policy-making (e.g., in Plan S (see below) and the UNESCO Recommendation on Open Science) to emphasise that a reform of research assessment is needed to ensure that such practices become embedded as the norm. Further impetus in this direction arises from moves to measure, recognise, and reward efforts towards open science.[182] Although this increased focus on measuring open science raised a worry about a potential distraction away from open science's initial values,[183] at least 13 of the core research assessment documents analysed in the next section support a better recognition for open science practices.

In addition, there is growing recognition that many of the metrics and indicators used to measure research performance failed to be open themselves; often calculated behind closed doors of commercial companies.[184] The lack of transparency and accessibility of the indicators used for assessing researchers is problematic because it imposes constraints on the autonomy of the research community to decide what kinds of metrics it does or does not want to use in research assessments, it restricts options for testing, verifying, and improving research indicators to suit the needs of the research community,[185] it enables undetected gaming and inappropriate research practices,[186] and it often requires substantial financial investments merely for obtaining access to the indicators.[187]

Widespread recognition of these drawbacks has stimulated efforts to make responsible research assessment a core aspect of global moves towards open science. In 2016, the Amsterdam Call for Action on Open Science defined goals relating to five core themes.[188] In the first and most important theme of 'Removing barriers to open science', the goal is to 'Change assessment, evaluation, and reward systems in science'. In the same year, as part of its Open Science agenda the European Commission asked an Expert Group on Altmetrics to "consider how indicators of research performance could be used effectively to enhance the strategic goals of the commission and the risks and opportunities that new forms of data pose to the research enterprise."[189] The Expert Group's report 'Next Generation Metrics: Responsible metrics and evaluation for open science' made twelve targeted recommendations urging policy makers, research funding organisations, and other

[182] Munafò, Nosek, Bishop, Button, Chambers, Percie Du Sert, Simonsohn, Wagenmakers, Ware and Ioannidis. (2017). A manifesto for reproducible science. *Nature Human Behaviour* 1, doi:10.1038/s41562-016-0021; Wilsdon, J. et al. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. doi:10.13140/RG.2.1.4929.1363, HEFCE. https://re.ukri.org/documents/hefce-documents/metric-tide-2015-pdf/

[183] Gadd. (2018). Measuring openness: should we be careful what we wish for? *The Bibliomagician* (21 August). https://thebibliomagician.wordpress.com/2018/08/21/measuring-openness-should-we-be-careful-what-we-wish-for/

[184] Sugimoto and Larivière. (2018). *Measuring research What everyone needs to know*. (Oxford University Press). isbn:978-0-19-064012-5 http://ebookcentral.proquest.com/lib/umontreal-ebooks/detail.action?docID=5160923; Wilsdon, J. et al. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. doi:10.13140/RG.2.1.4929.1363, HEFCE. https://re.ukri.org/documents/hefce-documents/metric-tide-2015-pdf/

[185] Wilsdon, J. et al. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. doi:10.13140/RG.2.1.4929.1363, HEFCE. https://re.ukri.org/documents/hefce-documents/metric-tide-2015-pdf/

[186] Gordon, Lin, Cave and Dandrea. (2015). The Question of Data Integrity in Article-Level Metrics. *PLoS Biology* 13, e1002161, doi:10.1371/journal.pbio.1002161

[187] Sugimoto and Larivière. (2018). *Measuring research What everyone needs to know*. (Oxford University Press). isbn:978-0-19-064012-5 http://ebookcentral.proquest.com/lib/umontreal-ebooks/detail.action?docID=5160923

[188] *Amsterdam Call for Action on Open Science*. (2016). https://www.openaccess.nl/sites/www.openaccess.nl/files/documenten/amsterdam-call-for-action-on-open-science.pdf

[189] Neylon. (2016). Submission to the European Commission Expert Group on Altmetrics. *Science in the Open - The online home of Cameron Neylon* (2 August). http://cameronneylon.net/blog/submission-to-the-european-commission-expert-group-on-altmetrics/

stakeholders to use research metrics responsibly for open science, particularly with regards to research assessment.[190]

The following years saw further work exploring the linked challenges of open science and research assessment. In 2017 the European University Association (EUA) organised a conference and issued a series of reports on research assessment as part of a series of activities to situate research institutions in the transition to open science. These include a 'Roadmap to Research Assessment in the Transition to Open Science' in 2018, and a report which surveyed European research institutions about their assessment practices 'Reflections on University Research Assessments: key concepts, issues and actors' in 2019.[191] Plan S, the European-wide initiative for open access publishing launched in 2018 and further strengthened the case that a reform of research assessment was needed to accompany a transition to open access. The 10th principle of Plan S explicitly states that "*The funders commit that when assessing research outputs during funding decisions they will value the intrinsic merit of the work and not consider the publication channel, its impact factor (or other journal metrics), or the publisher.*" The Global Young Academy (GYA) also joined in, with a conference and associated report 'Publishing Models, Assessments, and Open Science' published in 2018.[192] That same year, the European Commission published its report on 'Open Science, altmetrics and reward'.[193]

As a result, the discourse on open science has played a pivotal role in building awareness towards the need to reshape research assessment. At the same time, discussions on research assessment have expanded to embrace multiple dimensions of research and research culture, including for example, research integrity. In 2019, research assessment was the major theme discussed at the 6th World Conference on Research Integrity held in Hong Kong. The 'Hong Kong Principles for assessing researchers' which emerged from these discussions were then published in 2020.[194] For its part, DORA has drawn attention to the intersections between research assessment, open science and equity that need to be considered in any process of reform.[195]

2.5 From statements to coordination and action

These early actions and statements were essential in building an agenda for change and in raising awareness among different stakeholder groups. Yet greater collaboration and coordination is needed to move recommendations to large-scale action. This need for

[190] European Commission, Directorate-General for Research and Innovation, Peters, Frodeman, Wilsdon, Bar-Ilan, Lex and Wouters. (2017). *Next-generation metrics : responsible metrics and evaluation for open science*. (Publications Office). doi:10.2777/337729 https://data.europa.eu/doi/10.2777/337729

[191] European University Association. (2018). *EUA Roadmap on Research Assessment in the Transition to Open Science*. Brussels, Belgium. https://eua.eu/downloads/publications/eua-roadmap-on-research-assessment-in-the-transition-to-open-science_v20-08-2019.pdf; Saenen and Borell-Damián. (2019). *EUA Briefing - Reflections on University Research Assessment: Key concepts, issues and actors*. Brussels, Belgium. https://eua.eu/resources/publications/825:reflections-on-university-research-assessment-key-concepts,-issues-and-actors.html

[192] Global Young Academy. (2018). *Publishing models, assessments, and open science*. Halle, Germany. https://globalyoungacademy.net/wp-content/uploads/2018/10/APOS-Report-29.10.2018.pdf

[193] Miedema, Mayer, Holmberg and Leonelli. (2018). *Mutual Learning Exercise Open Science: Altmetrics and Rewards*. doi:10.2777/468970, Publications Office of the European Union. https://data.europa.eu/doi/10.2777/468970

[194] Moher, Bouter, Kleinert, Glasziou, Sham, Barbour, Coriat, Foeger and Dirnagl. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biology* 18, e3000737, doi:10.1371/journal.pbio.3000737

[195] The intersections between DORA, open scholarship, and equity. (2020). *DORA* (18 August). https://sfdora.org/2020/08/18/the-intersections-between-dora-open-scholarship-and-equity/

coordination was one of the core recommendations of *The Metric Tide*, and we are beginning to see some important movements on this front.

On a national level, funding organisations from several countries have joined forces to reshape research assessment on a broader scale. In the UK, major funders who signed DORA – for instance UKRI and Wellcome Trust – started requiring funded research institutions to uphold responsible research assessment principles,[196] a requirement that flips the first mover's disadvantage around. Given the strong influence of funders on the strategic actions of research universities, these decisions have the potential to generate broad scale changes in a multitude of UK institutions. In Norway, Universities Norway, which groups together 32 accredited universities and university colleges, also helped to build a national framework for research career assessments. Their report, issued in 2021, proposes a toolbox for recognition and rewards of academic careers.[197]

In the Netherlands, the Recognition and Rewards programme, developed from the position paper 'Room for everyone's talent' is another important example of national coordination.[198] The programme brought together a broad coalition of public knowledge institutions and funders to agree on a new assessment system to adopt across the country. The programme led to important changes to assessment practices in the Netherlands, but it also highlighted that there was still resistance for change and disagreement among researchers on the role of metrics in assessing researchers.[199] Part of this resistance came from a concern that changing assessment practices in the Netherlands may impact the compatibility with assessments in other countries with a leading role in science (for instance the US), thereby reducing the competitiveness of Dutch research and the mobility of Dutch researchers. This perceived first mover's disadvantage was accompanied by a defence of the JIF as an indicator of quality, and a fear that abandoning the JIF would "lead to randomness and a compromising of scientific quality",[200] although this perspective was also robustly disputed. The resistance also highlighted a few trends which, although not formally documented, were often described in the discussions that ensued. First, resistance seemed to be discipline-dependent, expressed more strongly in biomedical fields than in social sciences and humanities. Second, there also seemed to be generational differences, with critics on average being more senior than proponents of assessment reform. This last point was captured time and again in later broad discussions on reforming research assessment.

---

[196] Wellcome. (2021). Guidance for research organisations on how to implement responsible and fair approaches for research assessment. https://wellcome.org/grant-funding/guidance/research-organisations-how-implement-responsible-and-fair-approaches-research; Guidance for UKRI grant assessors (reviewers and board/panel members, etc.). https://www.ukri.org/wp-content/uploads/2020/10/UKRI-22102020-Final-DORA-statement-external.pdf

[197] Universities Norway. (2021). *NOR-CAM - A toolbox for recognition and rewards in academic careers*. Oslo. https://www.uhr.no/en/_f/p3/i86e9ec84-3b3d-48ce-8167-bbae0f507ce8/nor-cam-a-tool-box-for-assessment-and-rewards.pdf

[198] VSNU, NFU, KNAW, NWO and ZonMw. (2019). *Room for everyone's talent*. The Hague, The Netherlands. https://vsnu.nl/recognitionandrewards/wp-content/uploads/2019/11/Position-paper-Room-for-everyone's-talent.pdf

[199] Poot and Mulder. (2021). Banning journal impact factors is bad for Dutch science. *Times Higher Education* (3 August). https://www.timeshighereducation.com/opinion/banning-journal-impact-factors-bad-dutch-science; Nieuwe Erkennen en waarderen schaadt Nederlandse wetenschap. (2021). *ScienceGuide* (19 july). https://www.scienceguide.nl/2021/07/nieuwe-erkennen-en-waarderen-schaadt-nederlandse-wetenschap/; We moeten af van telzucht in de wetenschap. (2021). *ScienceGuide* (21 july). https://www.scienceguide.nl/2021/07/we-moeten-af-van-telzucht-in-de-wetenschap/; Chawla. (2021). Scientists at odds on Utrecht University reforms to hiring and promotion criteria. *Nature index* (9 August). https://www.nature.com/nature-index/news-blog/scientists-argue-over-use-of-impact-factors-for-evaluating-research

[200] Poot and Mulder. (2021). Banning journal impact factors is bad for Dutch science. *Times Higher Education* (3 August). https://www.timeshighereducation.com/opinion/banning-journal-impact-factors-bad-dutch-science.,

Beyond national initiatives, the Latin American Forum for Research Assessment (FOLEC) is an interesting example of international coordination. FOLEC provides a forum for discussion, exchange of information, and guidance on research assessment, enabling its members to agree on common principles and discuss implementation and best practices.[201] In Europe, a 2019 joint statement between EUA and Science Europe also brought together the commitments of research performing organisations and research funding organisations[202] for the European landscape. The statement proposed that the two organisations would commit to build a strong dialogue between their members to commit to key actions towards responsible research assessment. Similar international initiatives such the European Commission Open Science Policy Platform and Mutual Learning Exercise mentioned earlier further helped to bring awareness to responsible research assessment in European countries.

Changes are also appearing at a global scale. Over the past several years DORA has been instrumental in uniting voices, raising awareness, and enacting change towards responsible research assessment globally. Other initiatives are now starting to appear to address research assessment reform as a worldwide challenge.

In November 2020, the Global Research Council (GRC) held a five-day virtual conference entirely dedicated to responsible research assessment. Ahead of the conference, the Research on Research Institute (RoRI), CWTS-Leiden, the National Research Foundation of South Africa, and DORA joined forces to create a working paper summarising a selection of important initiatives on responsible research assessment and the results of a survey showcasing current practices among GRC members.[203] The event resulted in a call for action to all GRC members and the creation of a GRC working group that continues working and experimenting on Responsible Research Assessment.[204]

UNESCO and the G7 also recently highlighted the need for changes in research assessment. The *UNESCO Recommendation on Open Science* published in 2021 includes the need for "Encouraging responsible research and researcher evaluation and assessment practices, which incentivize quality science, recognizing the diversity of research outputs, activities and missions."[205] Similarly, the G7 introduced a reconsideration of responsible research assessment approaches as one of its actions to be followed by the Working Group on Open

[201] Latin American Forum for Research Assessment (FOLEC). (2020). Towards a Transformation of Scientific Research Assessment in Latin America and the Caribbean: Proposal for a Declaration of Principles. Latin American Council of Social Sciences (CLACSO). https://www.clacso.org/en/una-nueva-evaluacion-academica-para-una-ciencia-con-relevancia-social-2/; Latin American Forum for Research Assessment (FOLEC). (2020). Towards a Transformation of Scientific Research Assessment in Latin America and the Caribbean: Evaluating Scientivic Research Assessment. Latin American Council of Social Sciences (CLACSO). https://www.clacso.org/en/una-nueva-evaluacion-academica-para-una-ciencia-con-relevancia-social/; Latin American Forum for Research Assessment (FOLEC). (2020). Towards a Transformation of Scientific Research Assessment in Latin America and the Caribbean: Diagnosis and Proposals for a Regional Initiative. Latin American Council of Social Sciences (CLACSO). https://www.clacso.org/en/diagnostico-y-propuestas-para-una-iniciativa-regional/

[202] Science Europe and European University Association. (2019). *Joint Statement on Research Assessment*. doi:10.5281/zenodo.4925722. https://doi.org/10.5281/zenodo.4925722

[203] Curry, Rijcke, Hatch, Pillay, Weijden and Wilsdon. (2020). *The changing role of funders in responsible research assessment: progress, obstacles and the way ahead*. Research on Research Institute. https://rori.figshare.com/articles/report/The_changing_role_of_funders_in_responsible_research_assessment_progress_obstacles_and_the_way_ahead/13227914

[204] Global Research Council. (2021). *Responsible Research Assessment - Call to Action*. https://globalresearchcouncil.org/fileadmin//documents/GRC_Publications/RRA_Call_to_Action/RRA_Call_to_Action_English.pdf

[205] UNESCO. (2021). *UNESCO Recommendation on Open Science*. doi:SC-PCB-SPP/2021/OS/UROS, Paris. https://en.unesco.org/science-sustainable-future/open-science/recommendation

Science and the Working Group on the Security and Integrity of the Research Ecosystem.[206] Close on the heels of these initiatives, the Paris Open Science European Conference that took place in early 2022 also served as a rich discussion forum for research assessment. The *Paris Call on Research Assessment* presented during that conference called for the "creation of a coalition of research funding organisations, research performing organisations, and assessment authorities, willing and committed to reform the current research assessment system along commonly agreed objectives, principles and actions."[207]

Following this call, the *Agreement on Reforming Research Assessment*[208] was released in the summer of 2022. The Agreement was drafted as a joint effort between Science Europe, the European University Association, the European Commission and Karen Stroobants in close collaboration with a core group representing the diversity of the research community across Europe and in consultation with a large assembly of more than 350 organisations from over 40 countries. The agreement is the framework for the creation of a global coalition of stakeholders who commit to reform their research assessment procedures in line with the ten commitments laid out in the agreement, the Coalition for Advancing Research Assessment (CoARA).

From these commitments, the fourth core commitment addresses another point that relates to coordination, namely the need to abandon rankings of research organisations in research assessments in order to "*help avoid that metrics used by international rankings, which are inappropriate for assessing researchers, trickle down to research and researcher assessment.*" Most university rankings have been identified as having significant methodological challenges which lack rigour and transparency, indicators that are a poor proxy for the quality they claim to assess which are then weighted without any real justification into a single composite figure provided without error reporting. Ultimately they are accused of fostering competition rather than collaboration between research institutions.[209] The reliance of research institutions on rankings thus shapes research assessment and inhibits moves towards more responsible forms of assessment.

Despite these well-known issues, rankings continue to influence universities' "strategic, organisational, managerial, or faculty action"[210] and even influence more profound decisions such as visa and residency. The UK Government's High Potential Individual Visa scheme introduced in 2022 which offers visas to individuals who graduated from institutions which appeared in the top 50 rankings in two of three ranking tables (i.e., Times Higher Education World University Rankings, Quacquarelli Symonds World University Rankings and The Academic Ranking of World Universities) has raised a lot of controversy. The Netherlands also has a Highly Educated Migrant scheme which provides advantages to graduates of top 200 institutions on the Times Higher Education World University Rankings, Quacquarelli Symonds

[206] *G7 2021 Research Compact*. (2021). https://www.gov.uk/government/publications/g7-2021-research-compact/g7-2021-research-compact

[207]. Paris Call on Research Assessment. https://osec2022.eu/paris-call/

[208] CoARA. Agreement on Reforming Research Assessment. https://coara.eu/agreement/the-agreement-full-text/

[209] Curry. (2020). The still unsustainable goal of university ranking. *Reciprocal Space* (26 April). https://occamstypewriter.org/scurry/2020/04/26/still-unsustainable-university-rankings/; Gadd. (2020). University rankings need a rethink. *Nature* 587, doi:10.1038/d41586-020-03312-2; Gadd, Holmes and Shearer. (2021). Developing a Method for Evaluating Global University Rankings. *Scholarly Assessment Reports* 3, 2, doi:10.29024/sar.31

[210] Hazelkorn. (2019). The dubious practice of university rankings. https://elephantinthelab.org/the-accuracy-of-university-rankings-in-a-international-perspective/; Hazelkorn. (2007). The Impact of League Tables and Ranking Systems on Higher Education Decision Making. doi:doi:https://doi.org/10.1787/hemp-v19-art12-en

World University Rankings, or the Academic Ranking of World Universities. Recently, the initiative *More Than Our Rank*[211] was developed to highlight the limitations of the global university rankings and to provide institutions with a way to describe their qualities and achievements on their own terms.

SECTION 3 – COMMON RECOMMENDATIONS FOR RESPONSIBLE RESEARCH ASSESSMENT

There is no doubt that momentum for reform of research assessment has been building worldwide. Many organisations have taken a public position on the issue and issued recommendations for more responsible research assessment. Nevertheless, the plethora of recommendations can be challenging to implement and sometimes generate confusion. In this section, we explore a few of the trends and common themes that are presented in the statements and recommendation documents discussed above and we illustrate some of the implementation efforts that help deliver on their promises.

We extracted common topics from the recommendations and organised them around five main themes as follows:

- Responsible research indicators;
- Responsible assessment culture;
- Data infrastructure;
- Efficiency and coordination;
- Evidence building;

Table 1 summarises the content that we extracted in each document used. We limited Table 1 to documents that contain explicit recommendations, principles, actions, commitments or steps, and captured the information provided in these sections. Given the diversity in terminology and concepts used, we are conscious that Table 1 may not fully represent the elements covered in each document. Nevertheless, we believe it helps provide an overview of the types of themes raised in recommendations for responsible research assessment.

[211] INORMS. More Than Our Rank. https://inorms.net/more-than-our-rank/

		CCA 2012	DORA 2013	Hicks 2015	Wilsdon 2015	NL 2016	EC 2017	EUA 2018	GYA 2018	Moher 2018	EC 2018	EUA 2019	SE 2020	WCRI 2020	INORMS 2021	GRC 2021	EOSC 2021	UNESCO 2021	PARIS 2022	EC 2022
Responsible research indicators	Use research indicators responsibly	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Consider a broad range of research outputs and activities		✓	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
Responsible assessment culture	Show committment towards responsible research assessment				✓	✓	✓	✓			✓				✓	✓	✓	✓		✓
	Make research assessment processes transparent (e.g., using transparent and open indicators)	✓	✓	✓	✓	✓		✓				✓	✓		✓	✓	✓			✓
	Train assessors and foster diversity				✓				✓		✓		✓		✓	✓				✓
	Mitigate biases				✓								✓		✓	✓				✓
Data infrastructure	Enable reuse, verifiability, and interoperability	✓	✓	✓	✓	✓	✓		✓		✓		✓		✓	✓	✓	✓		✓
Efficiency and coordination	Consider efficiency of assessment processes				✓				✓		✓		✓		✓	✓	✓			
	Coordinate and engage in mutual learning				✓	✓		✓			✓	✓	✓			✓	✓	✓	✓	✓
Evidence building	Build evidence on research assessment	✓		✓	✓	✓	✓	✓		✓	✓		✓		✓	✓	✓	✓		✓
Target: University-level and higher-level assessment		•		•	•		•	•			•				•	•	•			
Target: Researcher-level assessment			•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Page numbers where only sections were used		xi - xvi			xv - xxii	6-7	16-17				108-119						15 - 25	27-29		

Table 1. Overview of themes covered in selected recommendation documents. *Themes were only captured in sections explicitly providing recommendations, principles, commitments, steps, or actions within each document. When only sections of these recommendations were used, for instance in broad documents on open science, we note the page numbers of the sub-sections categorised. The extent and depth with which the themes were presented differed between documents. Given differences in the terminology and concepts used, it is possible that themes were not always captured or were interpreted as belonging to different categories. Full reference lists are available in Box 1.*

Box 1. Statements and recommendations classified in Table 1

CCA 2012	The Council of Canadian Academies. (2012). *Informing Research Choices: Indicators and Judgment*. Report No. 978-1-926558-42-4, (Ottawa, Canada).
DORA 2013	American Society for Cell Biology. (2013). *San Francisco Declaration on Research Assessment*. https://sfdora.org/read/
Hicks 2015	Hicks, D., Wouters, P., Waltman, L., Rijcke, S. d. & Rafols, I. (2015) The Leiden Manifesto for research metrics. *Nature* 520, 429-431, doi:10.1038/520429a
Wilsdon 2015	Wilsdon, J. *et al.* (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. 10.13140/RG.2.1.4929.1363 (HEFCE).
NL 2016	*Amsterdam Call for Action on Open Science*. (2016). https://www.openaccess.nl/sites/www.openaccess.nl/files/documenten/amsterdam-call-for-action-on-open-science.pdf
EC 2017	European Commission, Directorate-General for Research and Innovation, Peters, I. *et al.* (2017) *Next-generation metrics : responsible metrics and evaluation for open science*. (Publications Office). doi:10.2777/337729 https://data.europa.eu/doi/10.2777/337729
EUA 2018	European University Association. (2018). *EUA Roadmap on Research Assessment in the Transition to Open Science*. (Brussels, Belgium). https://eua.eu/downloads/publications/eua-roadmap-on-research-assessment-in-the-transition-to-open-science_v20-08-2019.pdf
GYA 2018	Global Young Academy. (2018). *Publishing models, assessments, and open science*. (Halle, Germany). https://globalyoungacademy.net/wp-content/uploads/2018/10/APOS-Report-29.10.2018.pdf
Moher 2018	Moher, D. *et al.* (2018) Assessing scientists for hiring, promotion, and tenure. *PLoS Biology* 16, e2004089, doi:10.1371/journal.pbio.2004089
EC 2018	Miedema, F., Mayer, K., Holmberg, K. & Leonelli, S. (2018). *Mutual Learning Exercise - Open Science: Altmetrics and Rewards*. (European Commission). http://publications.europa.eu/resource/cellar/449cc187-693f-11e8-ab9c-01aa75ed71a1.0001.03/DOC_2
EUA 2019	Saenen, B. & Borell-Damián, L. (2019). *EUA Briefing - Reflections on University Research Assessment: Key concepts, issues and actors*. (Brussels, Belgium). https://eua.eu/resources/publications/825:reflections-on-university-research-assessment-key-concepts,-issues-and-actors.html
SE 2020	Science Europe. (2020). *Position Statement and Recommendations on Research Assessment Processes*. https://doi.org/10.5281/zenodo.4916155
WCRI 2020	Moher, D. *et al.* (2020) The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biology* 18, e3000737, doi:10.1371/journal.pbio.3000737
INORMS 2020	INORMS Research Evaluation Group. (2021). *The SCOPE Framework*. (Emerald Publishing). https://inorms.net/wp-content/uploads/2022/03/21655-scope-guide-v10.pdf
GRC 2021	Fraser, C. *et al.* (2021). *Responsible Research Assessment - Global Research Council (GRC) Conference Report 2021*. https://www.globalresearchcouncil.org/fileadmin/documents/GRC_Publications/GRC_RRA_Conference_Summary_Report.pdf
EOSC 2021	Mustajoki, H. et al. (2021). *Making FAIReR assessments possible*, doi:10.5281/zenodo.4701375. https://doi.org/10.5281/zenodo.4701375
UNESCO 2021	UNESCO. (2021). *UNESCO Recommendation on Open Science*. SC-PCB-SPP/2021/OS/UROS (Paris). https://en.unesco.org/science-sustainable-future/open-science/recommendation
PARIS 2022	OSEC. (2022). *Paris Call on Research Assessment*. https://osec2022.eu/paris-call/
EC 2022	CoARA. (2022). *Agreement on Reforming Research Assessment*. https://coara.eu/agreement/the-agreement-full-text/

3.1 Responsible research indicators

3.1.1 Use research indicators responsibly

The element most often addressed in recommendations on responsible research assessment is, perhaps not surprisingly, the responsible use of research indicators.

What do we mean by responsible research indicators? First and foremost, it relates to the need for indicators to align with the mission and values of the entity under evaluation and to be relevant to the context in which they are being used. For example, indicators that may be used to responsibly assess a researcher are not necessarily the same as those that can be used to assess an institution, a country, researchers from different disciplines, or a journal. This is in fact one of the main ideas behind DORA's first principle to stop using journal-level indicators to assess the quality of individual researchers or papers. The SCOPE framework provides useful guidance to help assessors to consider the context of evaluation and what is valued in the assessment before identifying appropriate qualitative or quantitative ways in which to evaluate.[212]

Another facet of responsible use relates to the need for research indicators to be clear about their limitations. The fact that the impact factor is given to three decimal places, for example, has long been criticised for creating a false impression of precision.[213] It is perhaps a sign of the growing acknowledgement of the importance of responsible research assessment that Clarivate has this year decided to reduce the JIF to one decimal place.[214]

Finally, statements relating to the responsible use of research indicators also mention that metrics and indicators should be used to support – not supplant – decisions made through peer review, and whenever possible to assess the content and merits of the work instead of relying on indicators. Examples of processes that facilitate the responsible use of research indicators can be found in the case studies presented in 'Reimagining academic assessment: stories of innovation and change' developed by DORA in collaboration with EUA and SPARC Europe.[215]

3.1.2 Consider a broad range of research outputs and activities

The indicators currently used in research assessment place a strong emphasis on research outputs – predominantly to published journal articles – as indicators of excellence. Most of the key statements mentioned above include at least one recommendation on widening the range of research activities considered by research assessment. For example, statements mention the need to consider open science practices such as open data and code sharing,[216]

[212] INORMS Research Evaluation Group. (2021). *The SCOPE Framework*. Emerald Publishing. https://inorms.net/wp-content/uploads/2022/03/21655-scope-guide-v10.pdf

[213] Gingras. (2014). *Bibliometrics and Research Evaluation: Uses and Abuses*. (The MIT Press). isbn:978-0-262-03512-5

[214] Clarivate Announces Changes to 2023 Journal Citation Report. (2022). *NISO Member News* (26 July). https://www.niso.org/niso-io/2022/07/clarivate-announces-changes-2023-journal-citation-report

[215] DORA. (2021). Reimagining academic assessment: stories of innovation and change. https://sfdora.org/dora-case-studies/

[216] Moher, Bouter, Kleinert, Glasziou, Sham, Barbour, Coriat, Foeger and Dirnagl. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biology* 18, e3000737, doi:10.1371/journal.pbio.3000737; European Commission, Directorate-General for Research and Innovation, Peters, Frodeman, Wilsdon, Bar-Ilan, Lex and Wouters. (2017). *Next-generation metrics : responsible metrics and evaluation for open science*. (Publications Office). doi:10.2777/337729 https://data.europa.eu/doi/10.2777/337729

peer-review activities,[217] and responsible practices such as complete and transparent reporting.[218] Changes are also starting to appear in assessment practices. For example, Cancer Research UK, an important funder of cancer research in the UK, reframed its definition of eligible achievements to include a broad array of possible outputs including datasets, preprints, or consensus statements.[219]

Along the same line, the implementation of Narrative CVs or Résumés for Researchers[220] requires researchers to list a variety of achievements which are not limited to research publications, giving them space to explain why specific achievements are important to them. Several research and funding institutions have been experimenting with narrative CVs, and UKRI announced its adoption of Résumé for Research and Innovation (R4RI) at the end of 2021. Despite the important mindset change needed to move away from traditional CV formats, narrative CVs raise broadly positive views. The Luxembourg National Research Funds, for instance, implemented narrative CVs in 2021 and surveyed assessors, panel members, and researchers to capture their impressions. Overall, they found a general acceptance of all stakeholders with most respondents feeling positive about the introduction of the narrative CVs.[221] Similarly, the Swiss National Science Foundation's (SNSF) SciCV – a new CV format that includes four narratives to help researchers describe their most important contributions to science – was recently evaluated in a survey with 123 grant applicants and 159 reviewers and the results suggest that "Survey respondents rated the narrative elements of the SciCV format as the most useful."[222]

The limited ability of current research assessment to capture impacts beyond academia has also been criticised in the past few years and was discussed in *The Metric Tide*. It may be argued that citation metrics provide a limited proxy for a form of impact within the scientific community. However, these metrics cannot capture social impact, translation of the research in practice, or the relevance of the research for policy decisions.[223] In an analysis of 864 review, promotion and tenure documents from 129 universities in the US and Canada, Alperin and colleagues found that public and community dimensions of research work are rarely explicitly linked to incentives and support structures and that "the public dimension of impact, in any form, is minimally addressed."[224] Such documents also generally place more emphasis on research outputs than on the societal aspects of the work.[225]

---

[217] Moher, Bouter, Kleinert, Glasziou, Sham, Barbour, Coriat, Foeger and Dirnagl. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biology* 18, e3000737, doi:10.1371/journal.pbio.3000737; Science Europe. (2020). *Position Statement and Recommendations on Research Assessment Processes*. https://doi.org/10.5281/zenodo.4916155,

[218] Moher, Bouter, Kleinert, Glasziou, Sham, Barbour, Coriat, Foeger and Dirnagl. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biology* 18, e3000737, doi:10.1371/journal.pbio.3000737; Moher, Naudet, Cristea, Miedema, Ioannidis and Goodman. (2018). Assessing scientists for hiring, promotion, and tenure. *PLoS Biology* 16, e2004089, doi:10.1371/journal.pbio.2004089

[219] Cancer Research UK. (2018). Improving how we evaluate research: how we're implementing DORA. https://www.cancerresearchuk.org/funding-for-researchers/research-features/2018-02-20-improving-research-evaluation-dora

[220] Royal Society. Résumé for Researchers. https://royalsociety.org/topics-policy/projects/research-culture/tools-for-support/resume-for-researchers/

[221] Luxembourg National Research Fund. (2022). *Narrative CV: Implementation and feedback results*. https://www.fnr.lu/narrative-cv-implementation-and-feedback-results/

[222] Singh Chawla. (2022). Swiss funder unveils new CV format to make grant evaluation fairer. *Nature* 606, 1033-1034, doi:10.1038/d41586-022-01599-x

[223] Lebel and McLean. (2018). A better measure of research from the global south. *Nature* 559, 23-26, doi:10.1038/d41586-018-05581-4

[224] Alperin, Muñoz Nieves, Schimanski, Fischman, Niles and McKiernan. (2019). How significant are the public dimensions of faculty work in review, promotion and tenure documents? *eLife* 8, e42254, doi:10.7554/eLife.42254

[225] Schimanski and Alperin. (2018). The evaluation of scholarship in academic promotion and tenure processes: Past, present, and future [version 1; peer review: 2 approved]. *F1000Research* 7, doi:10.12688/f1000research.16493.1

The REF led the way in enabling an in-depth and retrospective assessment of research impacts that go beyond typical research outputs. Though contributing significantly to the burden and cost of the exercise, which remains unique in the world,[226] the value of the information unearthed by the exercise is not in dispute. As a result, the weight on the impact component in the REF scoring calculation was increased from 20% in REF2014 to 25% in REF2021. Other tools, such as the Research Quality Plus (RQ+)[227] approach used by the Canadian International Development Research Centre (IDRC) can provide a framework by which societal impact can be measured.[228] The RQ+ approach is designed for use by funders in evaluating proposals for research, yet its principles and its different dimensions and stages of impact can provide valuable insight for assessing impact in research.

The Dutch Recognition and Reward Programme added another element to this broadening of research assessment by enabling researchers to decide on a particular area of focus such as research, education, impact, leadership, and patient care. At different stages in their careers, researchers are invited to choose a combination of areas in which they wish to invest their efforts and upon which their individual career assessments will focus.[229] The qualification portfolio at UMC Utrecht is another example of this diversification of profiles in practice.[230] Beyond benefiting equality, diversity, and inclusion, as we detail in section 3.2.4, embracing more diversity in the outputs, roles, and profiles of research-enabling staff is crucial in fostering a rich and sustainable research system.

3.2 Responsible assessment culture

3.2.1 Show commitment towards responsible research assessment

Many organisations have signalled their commitment to research assessment reform by signing DORA, but as DORA has frequently emphasised, such commitments should be made public, not just so that staff understand any changes assessment criteria, but also to empower them to challenge deviations from the principles of the declaration.[231] In the Agreement on Reforming Research Assessment, recommendations are framed as 'commitments' rather than 'principles.'[232] To enact their commitment, the agreement emphasises the need for signatories to raise awareness with different stakeholders and to invest resources so that commitments lead to genuine change.[233] Several statements reinstated the need for stakeholders to be explicit about their commitment to responsible research assessment (e.g., having it as one of their core missions or in a statement of principles). For instance, the report of the Expert

[226] Sutton. (2020). The increasing significance of impact within the Research Excellence Framework (REF). *Radiography* 26, S17-S19, doi:https://doi.org/10.1016/j.radi.2020.02.004

[227] Research Quality Plus. (2018). *IDRC* (14 June). https://www.idrc.ca/en/rqplus

[228] McLean, Ofir, Etherington, Acevedo and Feinstein. (2022). *Research Quality Plus - Evaluating Research Differently*. International Development Research Centre, Ottawa. https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/60945/IDL-60945.pdf?sequence=2

[229] VSNU, NFU, KNAW, NWO and ZonMw. (2019). *Room for everyone's talent*. The Hague, The Netherlands. https://vsnu.nl/recognitionandrewards/wp-content/uploads/2019/11/Position-paper-Room-for-everyone's-talent.pdf

[230] Benedictus, Miedema and Ferguson. (2016). Fewer numbers, better science. *Nature* 538, 453-455, doi:10.1038/538453a

[231] Hatch and Curry. (2020). Changing how we evaluate research is difficult, but not impossible. *eLife* 9, e58654, doi:10.7554/eLife.58654

[232] CoARA. Agreement on Reforming Research Assessment. https://coara.eu/agreement/the-agreement-full-text/

[233] European University Association. (2018). *EUA Roadmap on Research Assessment in the Transition to Open Science*. Brussels, Belgium. https://eua.eu/downloads/publications/eua-roadmap-on-research-assessment-in-the-transition-to-open-science_v20-08-2019.pdf; CoARA. Agreement on Reforming Research Assessment. https://coara.eu/agreement/the-agreement-full-text/

Group on altmetrics recommends that the European Commission encourages EU research institutions to create clear statements of principles on research assessment.[234]

3.2.2 Make research assessments processes transparent

The need for transparency in assessment processes was at the core of most statements reviewed. Using open data and open algorithms for indicator calculations was a common recommendation to improve this transparency. Statements also mentioned the importance of providing clear criteria upon which assessments are conducted, particularly emphasising the need for clear and unambiguous terminology. Broad and generic terms such as 'world-leading', 'excellent', 'impactful', 'significant' are often left undefined, creating an opportunity for personal biases and confusion.[235] Providing clear definitions of the terms used can help improve the clarity and the transparency of assessments. In practice, clearer definitions are starting to be added to statements and assessment documents. For example, the REF provides a definition of impact as "an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia". This definition is very similar to the one used in the position paper of the Recognition and Reward Programme 'Room for everybody's talent' which describes impact as "The contribution made by scientific research, in both the short and the long term, to changes in, or the development of, sectors of society and to challenges facing society. Such sectors of society include the economy, culture, public administration, and healthcare, while the challenges include such issues as climate change, immigration, quality of life, the environment, the rule of law, and security." (taken from KNAW[236]). Nonetheless, core concepts such as excellence are still rarely defined, at times meaning something akin to research integrity and reliability, at times leaning towards ground-breaking and internationally renowned, and at times coming closer to societal impact.[237]

3.2.3 Train assessors and foster diversity

The importance of ensuring that assessors are diverse and trained to understand assessment processes is also mentioned by several recommendations. For example, Science Europe's *Position Statement and Recommendations on Research Assessment Processes* states that everyone involved in assessment processes should be "trained and equipped to detect, monitor, and act on potential biases, discrimination, or unfair treatment."[238] The position statement also explains that assessors should be trained so they can understand the value of qualitative assessments and the limits of quantitative tools and metrics. In this regard, initiatives such as the Metrics Toolkit[239] can help to improve metrics literacy on a broader scale by highlighting the strength and limitations of well-known metrics. The *Responsible Research Assessment - Global Research Council (GRC) Conference Report 2021* mentions

[234] Wilsdon, Bar-Ilan, Frodeman, Peters and Wouters. (2017). *Next-generation metrics: Responsible metrics and evaluation for open science*. doi:10.2777/337729, European Commission. https://data.europa.eu/doi/10.2777/337729
[235] Hatch. (2019). To fix research assessment, swap slogans for definitions. *Nature* 576, 9, doi:10.1038/d41586-019-03696-w
[236] VSNU, NFU, KNAW, NWO and ZonMw. (2019). *Room for everyone's talent*. The Hague, The Netherlands. https://vsnu.nl/recognitionandrewards/wp-content/uploads/2019/11/Position-paper-Room-for-everyone's-talent.pdf
[237] Moore, Neylon, Paul Eve, Paul O'Donnell and Pattinson. (2017). "Excellence R Us": university research and the fetishisation of excellence. *Palgrave Communications* 3, 16105, doi:10.1057/palcomms.2016.105
[238] Science Europe. (2020). *Position Statement and Recommendations on Research Assessment Processes*. https://doi.org/10.5281/zenodo.4916155
[239] Metrics Toolkit https://www.metrics-toolkit.org/

that assessors should be trained but also repeatedly reminded of the principles of assessment methods to "ensure they do not fall back to using old methods."[240] Training and guidance is also important to ensure that assessors are comfortable with the assessment methods they use and that they understand what is expected from them. Both the *Mutual learning exercise - Open science: altmetrics and rewards* and the Global Young Academy *Publishing models, assessments, and open science* emphasise these points, although the latter focuses mostly on review of manuscripts rather than researcher assessment.[241] This recommendation has been taken up by several funders who introduced narrative CVs (see section 3.1.2) and now offer training and detailed guidance to help assessors navigate and assess this new format of CVs.

3.2.4 Mitigate biases

Many statements also require adherents to address biases in research assessment. In recent years, scientometric research has shown disparities between different groups of researchers, for example between men and women or between other over- and underrepresented groups. For example, research has shown that women are more likely to have a high volume of teaching and administrative assignments and to be involved with the technical and procedural aspects of research, making them less likely to be lead authors on research publications.[242] Studies on authorship position and contributions add to this issue, showing that women play a crucial role in scientific production but that their contribution often does not result in lead author positions.[243] Researchers from minority ethnic groups and specific geographic regions also face biases and unfair attribution[244] as well as unequal access to funding and opportunities.[245] During the COVID-19 pandemic, work disruptions exacerbated existing disparities, in particular towards women with children and black women.[246] Disparities are also exacerbated by the Matthew Effect, which further reduces the chances of already underrepresented groups to succeed in academia. Some studies have also shown that applications with male names tend to be perceived as more competent, more hireable, having

[240] Fraser, Nienaltowski, Goff, Firth, Sharman, Bright and Dias. (2021). *Responsible Research Assessment - Global Research Council (GRC) Conference Report 2021*. https://www.globalresearchcouncil.org/fileadmin/documents/GRC_Publications/GRC_RRA_Conference_Summary_Report.pdf

[241] Miedema, Mayer, Holmberg and Leonelli. (2018). *Mutual Learning Exercise Open Science: Altmetrics and Rewards*. doi:10.2777/468970, Publications Office of the European Union. https://data.europa.eu/doi/10.2777/468970; Global Young Academy. (2018). *Publishing models, assessments, and open science*. Halle, Germany. https://globalyoungacademy.net/wp-content/uploads/2018/10/APOS-Report-29.10.2018.pdf

[242] Macaluso, Lariviere, Sugimoto and Sugimoto. (2016). Is Science Built on the Shoulders of Women? A Study of Gender Differences in Contributorship. *Academic Medicine* 91, 1136-1142, doi:10.1097/ACM.0000000000001261; Astegiano, Sebastián-González and Castanho. Unravelling the gender productivity gap in science: a meta-analytical review. *Royal Society Open Science* 6, 181566, doi:10.1098/rsos.181566; Ross, Glennon, Murciano-Goroff, Berkes, Weinberg and Lane. (2022). Women are credited less in science than men. *Nature* 608, 135-145, doi:10.1038/s41586-022-04966-w

[243] Larivière, Ni, Gingras, Cronin and Sugimoto. (2013). Bibliometrics: global gender disparities in science. *Nature* 504, 211-213, doi:10.1038/504211a; Beaudry and Lariviere. (2016). Which gender gap? Factors affecting researchers' scientific impact in science and medicine. *Research Policy* 45, 1790-1817, doi:10.1016/j.respol.2016.05.009

[244] Rochmyaningsih. (2018). Showcase scientists from the global south. *Nature* 553, 251, doi:10.1038/d41586-018-00662-w; Powell. (2018). These labs are remarkably diverse — here's why they're winning at science. 558,, 19-22, doi:10.1038/d41586-018-05316-5

[245] Check Hayden. (2015). Racial bias continues to haunt NIH grants. *Nature* 527, 286-287, doi:10.1038/527286a

[246] Viglione. (2020). Are women publishing less during the pandemic? Here's what the data say. *Nature* 581, 365-366, doi:10.1038/d41586-020-01294-9

greater worth for mentoring, and meriting higher starting salary than identical applications with female names,[247] but there are also studies that show opposite effects.[248]

Several recommendations are made for minimising biases, often linking back to other recommendations. For instance, recognising a broad range of research activities and a diversity of skills (see section 3.1.2) can help to recognise and support researchers and research-enabling staff who specialise in activities and areas that are not necessarily output-driven. Along those lines, the European project ON-MERRIT (Observing and Negating Matthew Effects in Responsible Research & Innovation Transition), which investigated the potential of open and responsible research practices to worsen inequalities, recommended that funders and institutions must move beyond narrow quantitative indicators to value quality, openness, collaboration and responsibility in research, and that they should recognise the full range of academic tasks.[249] Including diverse assessor profiles on assessment committees[250] and training assessors to recognise and confront biases are also mentioned as ways to mitigate biases in research assessments (see section 3.2.3).

Many funders and research institutions have started implementing such measures. For example, the Health Research Board (HRB) Ireland mentioned that it will not only consider gender balance and possible biases in assessing grant applications, but that it will also make efforts to consider these biases in the drafting and implementation of HRB funding calls.[251] The League of European Research Universities (LERU) also issued detailed guidance to help its member institutions address implicit biases in academia, including in recruitment and career advancement processes.[252] This guidance was followed by a thorough position paper advocating for a more integrated approach to equality, diversity and inclusion at universities, including in the assessment of researchers.[253] On top of providing a solid evidence base explaining how biases impact research assessment, the position paper offers concrete 'recommended actions' to help research-intensive universities and other stakeholders address biases. These recommendations include a plea for greater transparency and granularity at all levels of research assessment, for instance by asking funders to publish data on the characteristics of applicants and reviewers; for conference organisers to actively seek speakers and panellists from under-represented groups; and for universities to develop policies to set what constitutes best practices and to train assessors so they become "conversant with the effects of bias in publishing and funding decisions".

[247] Moss-Racusin, Dovidio, Brescoll, Graham and Handelsman. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences* 109, 16474-16479, doi:10.1073/pnas.1211286109

[248] Williams and Ceci. (2015). National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences* 112, 5360-5365, doi:10.1073/pnas.1418878112

[249] Cole, Reichmann and Ross-Hellauer. (2022). *Global Thinking. ON-MERRIT recommendations for maximising equity in open and responsible research (1.0)*. doi:10.5281/zenodo.6276753, Zenodo. https://doi.org/10.5281/zenodo.6276753

[250] Science Europe. (2020). *Position Statement and Recommendations on Research Assessment Processes*. https://doi.org/10.5281/zenodo.4916155

[251] Health Research Board. (2019). *HRB Gender Policy*. https://www.hrb.ie/fileadmin/user_upload/HRB_Gender_Policy_Nov_2019.pdf

[252] Gvozdanović and Maes. (2018). *Implicit bias in academia: A challenge to the meritocratic principle and to women's careers - And what to do about it*. LERU. https://www.leru.org/publications/implicit-bias-in-academia-a-challenge-to-the-meritocratic-principle-and-to-womens-careers-and-what-to-do-about-it

[253] Buitendijk, Curry and Maes. (2019). *Equality, diversity and inclusion at universities: the power of a systemic approach*. LERU, Leuven, Belgium. https://www.leru.org/publications/equality-diversity-and-inclusion-at-universities

3.3 Data infrastructure

3.3.1 Enable reuse, verifiability, and interoperability

Several recommendations lean towards a need to improve the data infrastructure that supports research information management. In particular, a call for increasing the openness and the interoperability of research indicators is often highlighted as necessary to improve research assessment. This push toward open, transparent, and verifiable indicators is raised in the recommendations from DORA, *The Metric Tide*, and the Leiden Manifesto, among others (see Table 1).

The initiatives for Open Citations and Open Abstracts advocate for journals to make the bibliographic, citation, and abstract data of published articles openly available through Crossref. As a result of this, Crossref is becoming an increasingly rich source of data on research articles and other outputs.[254] The availability of data on open peer reviews of journal articles and preprints is also improving, thanks to infrastructures such as Crossref and Sciety. In addition, the use of persistent identifiers such as DOIs, ORCIDs, and ROR IDs is increasing. This enhances the interoperability between different data infrastructures. The OpenAIRE initiative launched by the European Commission in 2010 continues to be a core player in this area in more recent years. OpenAIRE created the platform Zenodo in 2013, and the platform continues to grow and to allow a whole array of output types to become citable and findable through DOIs. More recently, OpenAIRE contributed to the development of the European Open Science Cloud (EOSC). Together, these initiatives support a more open, reliable, and transparent metrics and indicators landscape.

3.4 Efficiency and coordination

3.4.1 Consider efficiency of assessment processes

The cost of assessing researchers, and the importance of doing so efficiently, is sometimes mentioned as a consideration. This is specified in the Science Europe Position Statement and Recommendations on Research Assessment Processes.[255] It is also addressed in the SCOPE framework, which calls on evaluators to “Evaluate only when necessary” but also to consider the cost-benefit ratio of any analysis as part of the ‘PROBE’ stage of SCOPE. Ghent University, which was one of the case studies in ‘Reimagining academic assessment: stories of innovation and change’, applied this principle to their researcher evaluation, reducing research assessment frequency from 2 or 3 years to five years in 2016. The Joint UK funding bodies always run post-hoc cost analyses of their national research assessment exercises. One such cost analysis is included in the Future Research Assessment Programme[256] (FRAP) to review the costs and benefits of REF 2021 in order to contribute towards policy development for future research assessment exercises.

[254] van Eck and Waltman. (2022). Crossref as a source of open bibliographic metadata. doi:10.31222/osf.io/smxe5
[255] Science Europe. (2020). *Position Statement and Recommendations on Research Assessment Processes*. https://doi.org/10.5281/zenodo.4916155
[256] Future Research Assessment Programme. https://www.jisc.ac.uk/future-research-assessment-programme

3.4.2 Coordinate and engage in mutual learning

Finally, as mentioned in section 2, the past few years have brought an increased realisation that stronger coordination between different initiatives for responsible research assessment is needed to move from statements to action. Several alliances are moving in this direction, including the Coalition for Advancing Research Assessment (CoARA[257]), the Global Research Council Working Group on Responsible Research Assessment[258] and targeted actions from G7 and UNESCO as discussed above.[259] In line with these alliances comes a growing recognition for the need to help one another implement better research assessment and engage in mutual learning. Many initiatives provide concrete examples of changes in research assessment. For instance, the Hong Kong Principles devoted much of its statement to examples of innovative actions to help users understand how they can implement responsible research practices.[260] DORA resources (see sfdora.org) and the case studies captured as part of the 'Reimagining academic assessment: stories of innovation and change' project also provide many detailed examples of what responsible research assessment can look like in practice.[261] The project 'Tools to Advance Research Assessment' (TARA), which is expected to deliver soon, will also contain a toolbox of such initiatives.

3.5 Evidence building

3.5.1 Build evidence on research assessment

The need to build evidence on research assessment processes is another crucial element raised by many recommendations for responsible research assessment (see Table 1) and something that was recommended by *The Metric Tide* report. Assessments have a tendency to "start with the data sources they have available to them" instead of determining what is really needed for a valuable assessment.[262] Yet if assessments are adapted to assess what they really value, new practices should be thoroughly studied to evidence the possible consequences and impacts that they may have both on the assessed and on research practices and environments.[263] This need for evidence has led to the creation of consortiums and research centres dedicated to understanding metrics and research systems to better inform research assessment. The Research on Research Institute (RoRI) is one example of a new wave of meta-research centres and networks that is active in evidence building on

[257] CoARA. Coalition for Advancing Research Assessment. https://coara.eu
[258] Global Research Council. (2021). *Responsible Research Assessment - Call to Action*. https://globalresearchcouncil.org/fileadmin//documents/GRC_Publications/RRA_Call_to_Action/RRA_Call_to_Action_English.pdf,
[259] *G7 2021 Research Compact*. (2021). https://www.gov.uk/government/publications/g7-2021-research-compact/g7-2021-research-compact; UNESCO. (2021). *UNESCO Recommendation on Open Science*. doi:SC-PCB-SPP/2021/OS/UROS, Paris. https://en.unesco.org/science-sustainable-future/open-science/recommendation
[260] Moher, Bouter, Kleinert, Glasziou, Sham, Barbour, Coriat, Foeger and Dirnagl. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biology* 18, e3000737, doi:10.1371/journal.pbio.3000737
[261] DORA. (2021). Reimagining academic assessment: stories of innovation and change. https://sfdora.org/dora-case-studies/
[262] INORMS Research Evaluation Group. (2021). *The SCOPE Framework*. Emerald Publishing. https://inorms.net/wp-content/uploads/2022/03/21655-scope-guide-v10.pdf
[263] Hicks, Wouters, Waltman, Rijcke and Rafols. (2015). The Leiden Manifesto for research metrics. *Nature* 520, 429-431, doi:10.1038/520429a; Wilsdon, Bar-Ilan, Frodeman, Peters and Wouters. (2017). *Next-generation metrics: Responsible metrics and evaluation for open science*. doi:10.2777/337729, European Commission. https://data.europa.eu/doi/10.2777/337729

research cultures, evaluation and decision-making.[264] Consortia such as the UK Forum for Responsible Research Metrics (FRRM), UK Reproducibility Network (UKRN) and the Future Research Assessment Programme (FRAP) are also active in gathering evidence and sharing good practices. Given the momentum, evidence is already surfacing, shaping our understanding of assessments. For example, several projects on narrative CVs have looked at the perspectives of assessors and assessed and at the implications that such assessments may have on the process.[265] Some universities such as UMC Utrecht are "undertaking an in-depth evaluation" of the impact that their changes in criteria for promotion and internal evaluation had on researchers.[266] Given that the final commitment of the Agreement on Reforming Research Assessment encourages signatories to "Evaluate practices, criteria and tools based on solid evidence and the state-of-the-art in research on research", and knowing that the Horizon Europe WIDERA programme from the European Commission is planning funds to support the Coalition for Advancing Research Assessment (CoARA) with cascading opportunities for implementation and implementing new assessment methods, it can be expected that empirical research on research assessment will continue to appear.

CONCLUDING REMARKS

This review complements the report *'Harnessing the Metric Tide: indicators, infrastructures and priorities for responsible research assessment in the UK'* by providing an overview of advances in scientometric databases, research indicators, and research assessment since 2015. A rich array of works, statements, and initiatives were published in this period, bringing research assessments within the top priorities of international research policy agendas. Although scrutiny around quantitative indicators has increased, metrics and indicators remain crucial to research assessments. Capturing new elements that can be used to create more diverse and inclusive indicators and understanding how indicators can be used responsibly to assess research activities are key to the future of assessments. Increasingly, research assessments also emphasise the critical role of qualitative approaches, for instance through a reaffirmation of the place for narrative elements and peer-review in improving research assessments.

Together, the works, statements, and initiatives issued since 2015 remind us of the need to continue moving towards responsible use of research indicators, both by using the right metrics in the right contexts and by diversifying the activities and outputs captured by the indicators used in assessing researchers. They also reaffirm the need for responsible assessment cultures in which assessment purposes are clear, methodologies are transparent, assessors are competent, and biases are mitigated. They also bring a recognition that data

[264] More information on RoRI's work here: https://researchonresearch.org/. Other examples include the International Centre for the Study of Research (ICSR) at Elsevier https://www.elsevier.com/icsr; the Association for Interdisciplinary Meta-Research and Open Science (AIMOS) https://aimos.community/; the Meta-Research Innovation Centre (METRICS) at Stanford https://metrics.stanford.edu/; and the work of the Center for Open Science https://www.cos.io/ in convening a series of Metascience meetings in 2019, 2021 and (forthcoming) in 2023, https://metascience.info/.

[265] Singh Chawla. (2022). Swiss funder unveils new CV format to make grant evaluation fairer. *Nature* 606, 1033-1034, doi:10.1038/d41586-022-01599-x; Luxembourg National Research Fund. (2022). *Narrative CV: Implementation and feedback results*. https://www.fnr.lu/narrative-cv-implementation-and-feedback-results/

[266] Hatch and Curry. (2020). Changing how we evaluate research is difficult, but not impossible. *eLife* 9, e58654, doi:10.7554/eLife.58654 Science Europe and European University Association. (2019). *Joint Statement on Research Assessment*. doi:10.5281/zenodo.4925722. https://doi.org/10.5281/zenodo.4925722

infrastructures should be open and reusable. In addition, they emphasise the importance of maximising efficiency in research assessments, both through cost-considerations and through increasing coordination between stakeholders to improve mutual learning and compatibility of new assessment principles. Finally, the works, statements, and initiatives issued since 2015 continue to advocate a need for evidence on research assessment and on the impact that new assessment methods may have on research and research communities.

A clear momentum for change is surfacing, and we can only anticipate that the coming years will be very rich in research, advances, and innovation for research assessment.

REFERENCES

Alperin, J. P. *et al.* (2019). How significant are the public dimensions of faculty work in review, promotion and tenure documents? *eLife* 8, e42254, doi:10.7554/eLife.42254

Altmetric. https://www.altmetric.com/

American Society for Cell Biology. (2013). *San Francisco Declaration on Research Assessment*. https://sfdora.org/read/

Amsterdam Call for Action on Open Science. (2016). https://www.openaccess.nl/sites/www.openaccess.nl/files/documenten/amsterdam-call-for-action-on-open-science.pdf

Astegiano, J., Sebastián-González, E. & Castanho, C. d. T. Unravelling the gender productivity gap in science: a meta-analytical review. *Royal Society Open Science* 6, 181566, doi:10.1098/rsos.181566

Beaudry, C. & Lariviere, V. (2016). Which gender gap? Factors affecting researchers' scientific impact in science and medicine. *Research Policy* 45, 1790-1817, doi:10.1016/j.respol.2016.05.009

Benedictus, R., Miedema, F. & Ferguson, M. W. J. (2016). Fewer numbers, better science. *Nature* 538, 453-455, doi:10.1038/538453a

Buitendijk, S., Curry, S. & Maes, K. (2019). *Equality, diversity and inclusion at universities: the power of a systemic approach*. LERU, Leuven, Belgium. https://www.leru.org/publications/equality-diversity-and-inclusion-at-universities

Callaway, E. (2016). Beat it, impact factor! Publishing elite turns against controversial metric. *Nature* 535, 210-211, doi:10.1038/nature.2016.20224

Cancer Research UK. (2018). Improving how we evaluate research: how we're implementing DORA. https://www.cancerresearchuk.org/funding-for-researchers/research-features/2018-02-20-improving-research-evaluation-dora

Casadevall, A. *et al.* (2016). ASM Journals Eliminate Impact Factor Information from Journal Websites. *mSystems* 1, e00088-00016, doi:10.1128/mSystems.00088-16

Chawla, D. S. (2021). Scientists at odds on Utrecht University reforms to hiring and promotion criteria. *Nature index* (9 August). https://www.nature.com/nature-index/news-blog/scientists-argue-over-use-of-impact-factors-for-evaluating-research

Check Hayden, E. (2015). Racial bias continues to haunt NIH grants. *Nature* 527, 286-287, doi:10.1038/527286a

China's research-evaluation revamp should not mean fewer international collaborations. (2020). *Nature* 579, doi:10.1038/d41586-020-00625-0

Clarivate Announces Changes to 2023 Journal Citation Report. (2022). *NISO Member News* (26 July). https://www.niso.org/niso-io/2022/07/clarivate-announces-changes-2023-journal-citation-report

CoARA. *Coalition for Advancing Research Assessment*. https://coara.eu

CoARA. (2022). *Agreement on Reforming Research Assessment*. https://coara.eu/agreement/the-agreement-full-text/

Cole, N. L., Reichmann, S. & Ross-Hellauer, T. (2022). *Global Thinking. ON-MERRIT recommendations for maximising equity in open and responsible research (1.0)*. doi:10.5281/zenodo.6276753, Zenodo. https://doi.org/10.5281/zenodo.6276753

Crossref. https://www.crossref.org/
Crossref. *Event Data*. https://www.crossref.org/services/event-data/
Curry, S. (2018). Let's move beyond the rhetoric: it's time to change how we judge research. *Nature* 554, 147, doi:10.1038/d41586-018-01642-w
Curry, S. (2020). The still unsustainable goal of university ranking. *Reciprocal Space* (26 April). https://occamstypewriter.org/scurry/2020/04/26/still-unsustainable-university-rankings/
Curry, S. *et al.* (2020). *The changing role of funders in responsible research assessment: progress, obstacles and the way ahead*. Research on Research Institute. https://rori.figshare.com/articles/report/The_changing_role_of_funders_in_responsible_research_assessment_progress_obstacles_and_the_way_ahead/13227914
Dimensions. https://www.dimensions.ai
Directory of Open Access. https://doaj.org
doi. https://www.doi.org/
DORA. (2021). *Reimagining academic assessment: stories of innovation and change*. https://sfdora.org/dora-case-studies/
Elsevier. *CiteScore: a new metric to help you track journal performance and make decisions*. https://www.elsevier.com/connect/editors-update/citescore-a-new-metric-to-help-you-choose-the-right-journal
European Commission *et al.* (2017). *Next-generation metrics : responsible metrics and evaluation for open science*. (Publications Office). doi:10.2777/337729 https://data.europa.eu/doi/10.2777/337729
European University Association. (2018). *EUA Roadmap on Research Assessment in the Transition to Open Science*. Brussels, Belgium. https://eua.eu/downloads/publications/eua-roadmap-on-research-assessment-in-the-transition-to-open-science_v20-08-2019.pdf
Fraser, C. *et al.* (2021). *Responsible Research Assessment - Global Research Council (GRC) Conference Report 2021*. https://www.globalresearchcouncil.org/fileadmin/documents/GRC_Publications/GRC_RRA_Conference_Summary_Report.pdf
Future Research Assessment Programme. https://www.jisc.ac.uk/future-research-assessment-programme
G7 2021 Research Compact. (2021). https://www.gov.uk/government/publications/g7-2021-research-compact/g7-2021-research-compact
Gadd, E. (2018). Measuring openness: should we be careful what we wish for? *The Bibliomagician* (21 August). https://thebibliomagician.wordpress.com/2018/08/21/measuring-openness-should-we-be-careful-what-we-wish-for/
Gadd, E. (2020). University rankings need a rethink. *Nature* 587, doi:10.1038/d41586-020-03312-2
Gadd, E., Holmes, R. & Shearer, J. (2021). Developing a Method for Evaluating Global University Rankings. *Scholarly Assessment Reports* 3, 2, doi:10.29024/sar.31
Gingras, Y. (2014). *Bibliometrics and Research Evaluation: Uses and Abuses*. (The MIT Press). isbn:978-0-262-03512-5
Global Research Council. (2021). *Responsible Research Assessment - Call to Action*. https://globalresearchcouncil.org/fileadmin//documents/GRC_Publications/RRA_Call_to_Action/RRA_Call_to_Action_English.pdf
Global Young Academy. (2018). *Publishing models, assessments, and open science*. Halle, Germany. https://globalyoungacademy.net/wp-content/uploads/2018/10/APOS-Report-29.10.2018.pdf
Gordon, G., Lin, J., Cave, R. & Dandrea, R. (2015). The Question of Data Integrity in Article-Level Metrics. *PLoS Biology* 13, e1002161, doi:10.1371/journal.pbio.1002161
Guidance for UKRI grant assessors (reviewers and board/panel members, etc.). https://www.ukri.org/wp-content/uploads/2020/10/UKRI-22102020-Final-DORA-statement-external.pdf
Gvozdanović, J. & Maes, K. (2018). *Implicit bias in academia: A challenge to the meritocratic principle and to women's careers - And what to do about it*. LERU. https://www.leru.org/publications/implicit-bias-in-academia-a-challenge-to-the-meritocratic-principle-and-to-womens-careers-and-what-to-do-about-it

Harzing, A.-W. & Alakangas, S. (2017). Microsoft Academic: is the phoenix getting wings? *Scientometrics* 110, 371-383, doi:10.1007/s11192-016-2185-x

Hatch, A. (2019). To fix research assessment, swap slogans for definitions. *Nature* 576, 9, doi:10.1038/d41586-019-03696-w

Hatch, A. & Curry, S. (2020). Changing how we evaluate research is difficult, but not impossible. *eLife* 9, e58654, doi:10.7554/eLife.58654

Hazelkorn, E. (2007). The Impact of League Tables and Ranking Systems on Higher Education Decision Making. doi:doi:https://doi.org/10.1787/hemp-v19-art12-en

Hazelkorn, E. (2019). The dubious practice of university rankings. https://elephantinthelab.org/the-accuracy-of-university-rankings-in-a-international-perspective/

Health Research Board. (2019). *HRB Gender Policy*. https://www.hrb.ie/fileadmin/user_upload/HRB_Gender_Policy_Nov_2019.pdf

Hendricks, G., Rittman, M. & Bartell, A. (2022). Amendments to membership terms to open reference distribution and include UK jurisdiction. *CrossRef Blog* (4 April). https://www.crossref.org/blog/amendments-to-membership-terms-to-open-reference-distribution-and-include-uk-jurisdiction/

Herzog, C., Hook, D. & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies* 1, 387-395, doi:10.1162/qss_a_00020

Hicks, D., Wouters, P., Waltman, L., Rijcke, S. d. & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature* 520, 429-431, doi:10.1038/520429a

Impactstory. https://profiles.impactstory.org

INORMS. *More Than Our Rank*. https://inorms.net/more-than-our-rank/

INORMS Research Evaluation Group. (2021). *The SCOPE Framework*. Emerald Publishing. https://inorms.net/wp-content/uploads/2022/03/21655-scope-guide-v10.pdf

The intersections between DORA, open scholarship, and equity. (2020). *DORA* (18 August). https://sfdora.org/2020/08/18/the-intersections-between-dora-open-scholarship-and-equity/

Joint Research Centre *et al.* (2014). *An analysis of national research systems (I) : a composite indicator for scientific and technological research excellence*. (Publications Office). doi:10.2788/95887 https://data.europa.eu/doi/10.2788/95887

Karmakar, M., Banshal, S. K. & Singh, V. K. (2021). A large-scale comparison of coverage and mentions captured by the two altmetric aggregators: Altmetric.com and PlumX. *Scientometrics* 126, 4465-4489, doi:10.1007/s11192-021-03941-y

Kraker, P., Jordan, K. & Lex, E. (2015). The ResearchGate Score: a good example of a bad metric. *LSE Impact Blog* (9 December). https://blogs.lse.ac.uk/impactofsocialsciences/2015/12/09/the-researchgate-score-a-good-example-of-a-bad-metric/

Larivière, V. *et al.* (2016). A simple proposal for the publication of journal citation distributions. *bioRxiv*, doi:10.1101/062109

Larivière, V., Ni, C., Gingras, Y., Cronin, B. & Sugimoto, C. R. (2013). Bibliometrics: global gender disparities in science. *Nature* 504, 211-213, doi:10.1038/504211a

Larivière, V., Pontille, D. & Sugimoto, C. R. (2021). Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). *Quantitative Science Studies* 2, 111-128, doi:10.1162/qss_a_00097

Latin American Forum for Research Assessment (FOLEC). (2020). *Towards a Transformation of Scientific Research Assessment in Latin America and the Caribbean: Diagnosis and Proposals for a Regional Initiative*. Latin American Council of Social Sciences (CLACSO). https://www.clacso.org/en/diagnostico-y-propuestas-para-una-iniciativa-regional/

Latin American Forum for Research Assessment (FOLEC). (2020). *Towards a Transformation of Scientific Research Assessment in Latin America and the Caribbean: Evaluating Scientivic Research Assessment*. Latin American Council of Social Sciences (CLACSO). https://www.clacso.org/en/una-nueva-evaluacion-academica-para-una-ciencia-con-relevancia-social/

Latin American Forum for Research Assessment (FOLEC). (2020). *Towards a Transformation of Scientific Research Assessment in Latin America and the Caribbean: Proposal for a Declaration of Principles*. Latin American Council of Social Sciences (CLACSO). https://www.clacso.org/en/una-nueva-evaluacion-academica-para-una-ciencia-con-relevancia-social-2/

Lebel, J. & McLean, R. (2018). A better measure of research from the global south. *Nature* 559, 23-26, doi:10.1038/d41586-018-05581-4

Luxembourg National Research Fund. (2022). *Narrative CV: Implementation and feedback results*. https://www.fnr.lu/narrative-cv-implementation-and-feedback-results/

Macaluso, B., Lariviere, V., Sugimoto, T. & Sugimoto, C. R. (2016). Is Science Built on the Shoulders of Women? A Study of Gender Differences in Contributorship. *Academic Medicine* 91, 1136-1142, doi:10.1097/ACM.0000000000001261

Manista, F. (2020). There's A PID For That, Part 2: Projects. *Jisc scholarly communications* (13 October). https://scholarlycommunications.jiscinvolve.org/wp/2020/10/13/theres-a-pid-for-that-part-2-projects/

Martín-Martín, A., Thelwall, M., Orduna-Malea, E. & Delgado López-Cózar, E. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics* 126, 871-906, doi:10.1007/s11192-020-03690-4

McCullough, R. (2021). Preprints are now in Scopus! *SCOPUS* https://blog.scopus.com/posts/preprints-are-now-in-scopus

McLean, R., Ofir, Z., Etherington, A., Acevedo, M. & Feinstein, O. (2022). *Research Quality Plus - Evaluating Research Differently*. International Development Research Centre, Ottawa. https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/60945/IDL-60945.pdf?sequence=2

Metrics Toolkit https://www.metrics-toolkit.org/

Miedema, F., Mayer, K., Holmberg, K. & Leonelli, S. (2018). *Mutual Learning Exercise Open Science: Altmetrics and Rewards*. doi:10.2777/468970, Publications Office of the European Union. https://data.europa.eu/doi/10.2777/468970

Moher, D. *et al.* (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biology* 18, e3000737, doi:10.1371/journal.pbio.3000737

Moher, D. *et al.* (2018). Assessing scientists for hiring, promotion, and tenure. *PLoS Biology* 16, e2004089, doi:10.1371/journal.pbio.2004089

Moore, S., Neylon, C., Paul Eve, M., Paul O'Donnell, D. & Pattinson, D. (2017). "Excellence R Us": university research and the fetishisation of excellence. *Palgrave Communications* 3, 16105, doi:10.1057/palcomms.2016.105

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J. & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences* 109, 16474-16479, doi:10.1073/pnas.1211286109

Munafò, M. R. *et al.* (2017). A manifesto for reproducible science. *Nature Human Behaviour* 1, doi:10.1038/s41562-016-0021

Mustajoki, H. et al. (2021). Making FAIReR assessments possible. doi:10.5281/zenodo.4701375. https://doi.org/10.5281/zenodo.4701375

Next Steps for Microsoft Academic – Expanding into New Horizons. (2021). *Microsoft Academic* (May 4). https://www.microsoft.com/en-us/research/project/academic/articles/microsoft-academic-to-expand-horizons-with-community-driven-approach/

Neylon, C. (2016). Submission to the European Commission Expert Group on Altmetrics. *Science in the Open - The online home of Cameron Neylon* (2 August). http://cameronneylon.net/blog/submission-to-the-european-commission-expert-group-on-altmetrics/

Nieuwe Erkennen en waarderen schaadt Nederlandse wetenschap. (2021). *ScienceGuide* (19 july). https://www.scienceguide.nl/2021/07/nieuwe-erkennen-en-waarderen-schaadt-nederlandse-wetenschap/

NIHR. *Generic supporting information for applicants*. https://www.nihr.ac.uk/documents/generic-supporting-information-for-applicants/28196#orcid

Nuffield Council of Bioethics. (2014). *The culture of scientific research in the UK*. https://nuffieldbioethics.org/publications/the-culture-of-scientific-research

OpenAlex. https://openalex.org/

OpenCitations. *COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations*. https://opencitations.net/index/coci

ORCID. https://orcid.org/

ORCID. *Funders' ORCID Policies*. https://info.orcid.org/funders-orcid-policies/

OSEC. (2022). *Paris Call on Research Assessment*. https://osec2022.eu/paris-call/
Overton. https://www.overton.io/
PLOS. *PLOS and DORA*. https://plos.org/publish/dora/
Plum Analytics. *About PlumX Metrics*. https://plumanalytics.com/learn/about-metrics/
Poot, R. & Mulder, W. (2021). Banning journal impact factors is bad for Dutch science. *Times Higher Education* (3 August). https://www.timeshighereducation.com/opinion/banning-journal-impact-factors-bad-dutch-science
Powell, K. (2018). These labs are remarkably diverse — here's why they're winning at science. 558,, 19-22, doi:10.1038/d41586-018-05316-5
PREReview. https://prereview.org/
PubPeer. https://pubpeer.com/
REF 2021 Decisions on staff and outputs. (2017 (Updated 2018)). Bristol, England. https://www.ref.ac.uk/media/1034/ref-2017_04-decisions-updated-11042018.pdf
Refresh of the Journal Citation Reports data. (2018). *Clarivate Blog* (17 October). https://clarivate.com/blog/refresh-of-the-journal-citation-reports-data/
Research Activity Identifier. https://www.raid.org.au
Research Quality Plus. (2018). *IDRC* (14 June). https://www.idrc.ca/en/rqplus
Rochmyaningsih, D. (2018). Showcase scientists from the global south. *Nature* 553, 251, doi:10.1038/d41586-018-00662-w
ROR. https://ror.org/
Ross, M. B. *et al.* (2022). Women are credited less in science than men. *Nature* 608, 135-145, doi:10.1038/s41586-022-04966-w
Royal Society. *Résumé for Researchers*. https://royalsociety.org/topics-policy/projects/research-culture/tools-for-support/resume-for-researchers/
Saenen, B. & Borell-Damián, L. (2019). *EUA Briefing - Reflections on University Research Assessment: Key concepts, issues and actors*. Brussels, Belgium. https://eua.eu/resources/publications/825:reflections-on-university-research-assessment-key-concepts,-issues-and-actors.html
San Francisco Declaration on Research Assessment. https://sfdora.org/read/
Schimanski, L. & Alperin, J. (2018). The evaluation of scholarship in academic promotion and tenure processes: Past, present, and future [version 1; peer review: 2 approved]. *F1000Research* 7, doi:10.12688/f1000research.16493.1
Science Europe. (2020). *Position Statement and Recommendations on Research Assessment Processes*. https://doi.org/10.5281/zenodo.4916155
Science Europe & European University Association. (2019). *Joint Statement on Research Assessment*. doi:10.5281/zenodo.4925722. https://doi.org/10.5281/zenodo.4925722
ScienceOpen.com. https://www.scienceopen.com/
Scimago Journal & Country Rank. https://www.scimagojr.com
scite. https://www.scite.ai
Shu, F., Liu, S. & Larivière, V. (2022). China's Research Evaluation Reform: What are the Consequences for Global Science? *Minerva* 60, 329-347, doi:10.1007/s11024-022-09468-7
Singh Chawla, D. (2022). Swiss funder unveils new CV format to make grant evaluation fairer. *Nature* 606, 1033-1034, doi:10.1038/d41586-022-01599-x
SNIP. https://journalinsights.elsevier.com/journals/0969-806X/snip
Sørensen, M. P., Bloch, C. & Young, M. (2016). Excellence in the knowledge-based economy: from scientific to research excellence. *European Journal of Higher Education* 6, 217 - 236
Strinzel, M., Kaltenbrunner, W., van der Weijden, I., von Arx, M. & Hill, M. (2022). SciCV, the Swiss National Science Foundation's new CV format. *bioRxiv*, doi:10.1101/2022.03.16.484596
Sugimoto, C. R. & Larivière, V. (2018). *Measuring research What everyone needs to know*. (Oxford University Press). isbn:978-0-19-064012-5 http://ebookcentral.proquest.com/lib/umontreal-ebooks/detail.action?docID=5160923
Sutton, E. (2020). The increasing significance of impact within the Research Excellence Framework (REF). *Radiography* 26, S17-S19, doi:https://doi.org/10.1016/j.radi.2020.02.004
Tay, A., Martín-Martín, A. & Hug, S. E. (2021). Goodbye, Microsoft Academic – Hello, open research infrastructure? *LSE Impact Blog* (27 May).

https://blogs.lse.ac.uk/impactofsocialsciences/2021/05/27/goodbye-microsoft-academic-hello-open-research-infrastructure/
Teixeira da Silva, J. A. & Memon, A. R. (2017). CiteScore: A cite for sore eyes, or a valuable, transparent metric? *Scientometrics* 111, 553-556, doi:10.1007/s11192-017-2250-0
The Council of Canadian Academies. (2012). *Informing Research Choices: Indicators and Judgment*. Report No. 978-1-926558-42-4, Ottawa, Canada.
UNESCO. (2021). *UNESCO Recommendation on Open Science*. doi:SC-PCB-SPP/2021/OS/UROS, Paris. https://en.unesco.org/science-sustainable-future/open-science/recommendation
Universities Norway. (2021). *NOR-CAM - A toolbox for recognition and rewards in academic careers*. Oslo. https://www.uhr.no/en/_f/p3/i86e9ec84-3b3d-48ce-8167-bbae0f507ce8/nor-cam-a-tool-box-for-assessment-and-rewards.pdf
van Eck, N. J. & Waltman, L. (2022). Crossref as a source of open bibliographic metadata. doi:10.31222/osf.io/smxe5
Vertesy, D. & Tarantola, S. (2012). *Composite Indicators of Research Excellence*. Report No. JRC72592, Publications Office of the European Union, Luxembourg (Luxembourg). https://publications.jrc.ec.europa.eu/repository/handle/JRC72592
Viglione, G. (2020). Are women publishing less during the pandemic? Here's what the data say. *Nature* 581, 365-366, doi:10.1038/d41586-020-01294-9
VSNU, NFU, KNAW, NWO & ZonMw. (2019). *Room for everyone's talent*. The Hague, The Netherlands. https://vsnu.nl/recognitionandrewards/wp-content/uploads/2019/11/Position-paper-Room-for-everyone's-talent.pdf
Watson, C. (2022). Rise of the preprint: how rapid data sharing during COVID-19 has changed science forever. *Nature Medicine* 28, 2-5, doi:10.1038/s41591-021-01654-6
We moeten af van telzucht in de wetenschap. (2021). *ScienceGuide* (21 july). https://www.scienceguide.nl/2021/07/we-moeten-af-van-telzucht-in-de-wetenschap/
Wellcome. *Open Researcher and Contributor ID (ORCID)*. https://wellcome.org/grant-funding/open-researcher-and-contributor-id-orcid
Wellcome. (2021). *Guidance for research organisations on how to implement responsible and fair approaches for research assessment*. https://wellcome.org/grant-funding/guidance/research-organisations-how-implement-responsible-and-fair-approaches-research
Why we're removing the RG Score (and what's next). (2022). *ResearchGate Updates* (29 March). https://www.researchgate.net/researchgate-updates/removing-the-rg-score
Williams, W. M. & Ceci, S. J. (2015). National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences* 112, 5360-5365, doi:10.1073/pnas.1418878112
Wilsdon, J., Bar-Ilan, J., Frodeman, R., Peters, I. & Wouters, P. (2017). *Next-generation metrics: Responsible metrics and evaluation for open science*. doi:10.2777/337729, European Commission. https://data.europa.eu/doi/10.2777/337729
Wilsdon, J. *et al.* (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. doi:10.13140/RG.2.1.4929.1363, HEFCE. https://re.ukri.org/documents/hefce-documents/metric-tide-2015-pdf/
Zhang, L. & Sivertsen, G. (2020). The New Research Assessment Reform in China and Its Implementation. *Scholarly Assessment Reports* 2, 3, doi:10.29024/sar.15

APPENDIX C: Roundtable series, July 2022

Following the launch of the Metric Tide Revisited as a rapid review, we organised a series of three roundtables in July 2022. to seek advice and input from experts and stakeholder groups. Our approach to these rountables was shaped by a commitment to openness and inclusivity. Anyone was welcome to register to attend and contribute thoughts via the chat function on the day. Participants were also invited to send thoughts by email before or after the roundtables on this address: responsiblemetrics@gmail.com
The roundtables focused on the following groups and topics:

1: Research managers, librarians & planners (Monday 4 July)

Speakers included:

- Katy McKen, University of Bath
- Zosia Beckles, University of Bristol
- Julie Bayley, University of Lincoln
- Jackie Njoroje, Salford University & HESPA
- Jenni Stergiou, University of Northumbria & ARMA

The first roundtable brought together research managers, library and information professionals and university planners to explore:

- The extent to which the recommendations and conclusions of the Metric Tide report as they pertained to research-enabling professionals have been implemented or superseded since 2015.
- What limitations, challenges and concerns over uses of metrics in research management and assessment persist, or have intensified?
- In what contexts might indicators play a helpful role in the management or assessment of UK research environments, outputs, and/or impacts?
- What infrastructures, literacies, and incentives might be required to support the further roll-out of responsible use of metrics in UK research assessment and management?

2: Institutions, sector bodies & learned societies (Tuesday 12 July)

Speakers included:

- Prof. Colin Bain, PVC Research, Durham University
- Ms Sarah Cowan, Head of Policy (HE), British Academy
- Prof. Silke Machold, Dean of Research, Univ. Wolverhampton
- Ms Rachel Bruce, Head of Open Research, UKRI
- Dr Helen Ewles, Head of R&I Policy, RAEng
- Prof Rachael Gooberman-Hill, Chair, UK Committee on Research Integrity

The second roundtable brought together university leaders, researchers, and representatives from sector bodies and learned societies. Our goal was to understand better:

- The extent to which the recommendations and conclusions of the Metric Tide report as they pertained to university leaders, researchers and other bodies have been implemented or superseded since 2015.

- Examples of changing practice in research assessment within universities subject to the REF
- What limitations, challenges and concerns over uses of metrics in research management and assessment persist, or have intensified?
- The extent to which external pressures constrain the implementation of responsible research assessment practices
- How do developments, opportunities, uncertainties and concerns relate to broader agendas for equality, diversity and inclusion (EDI) and improving research cultures?

3: Metrics experts, data and infrastructure providers (Tuesday 19 July)

Speakers included:

- Dr Andrew Plume, Vice President of Research Evaluation at Elsevier and President of the International Center for the Study of Research (ICSR)
- Jonathan Adams, Director at the Institute for Scientific Information, Clarivate Analytics.
- Dr Juergen Wastl, Director Academic Relations and Consultancy, Digital Science
- Dr Vincent Traag & Professor Ludo Waltman, CWTS, Leiden University
- Mike Thelwall, Professor of Data Science and head of the Statistical Cybermetrics and Research Evaluation Group, University of Wolverhampton
- Cameron Neylon, Professor of Research Communications at the Centre for Culture and Technology at Curtin University
- Alis Oancea, Professor of Philosophy of Education and Research Policy, University of Oxford
- Dr Michael Ochsner, senior researcher at FORS (Swiss Centre of Expertise in the Social Sciences) and president of the European Network for Research Evaluation in the Social Sciences and Humanities, ENRESSH
- Sir Peter Gluckman, President of the International Science Council and Chair of the FRAP International Advisory Group

This third roundtable explored:

- What has changed in the technical, data and infrastructural landscape?
- What opportunities, limitations and concerns over uses of metrics in research management and assessment persist, or have reduced or intensified?
- How do developments, opportunities, uncertainties and concerns relate to broader agendas of responsible research assessment (RRA) and improving research cultures– and to options for the future design of the REF?