

THE FUTURE OF RESEARCH EVALUATION:

A SYNTHESIS OF CURRENT DEBATES AND
DEVELOPMENTS

DISCUSSION PAPER

May 2023

AUTHORS AND ACKNOWLEDGEMENTS

This paper was authored by the members of the GYA-IAP-ISC Scoping Group, which worked intermittently between May 2021 and February 2023 (more detail in Appendix A):

- Sarah de Rijcke (Chair, Netherlands)
- Clemencia Cosentino (USA)
- Robin Crewe (South Africa)
- Carlo D'Ippoliti (Italy)
- Shaheen Motala-Timol (Mauritius)
- Noorsaadah Binti A Rahman (Malaysia)
- Laura Rovelli (Argentina)
- David Vaux (Australia)
- Yao Yupeng (China)

The Working Group thanks Tracey Elliott (ISC Senior Consultant) for her coordination and drafting work. Thanks also go to Alex Rushforth (Centre for Science and Technology Studies (CWTS), Leiden University, Netherlands) and Sarah Moore (ISC) for additional input and support.

The Working Group is also grateful to all those who were consulted in the preparation of this paper (Appendix B), who gave their time and shared their perspectives on research evaluation in their respective countries and regions, and to the reviewers nominated by the GYA, IAP and ISC:

- Karina Batthyány, Executive Director, Latin American Council of Social Sciences (CLACSO) (Uruguay)
- Richard Catlow, Research Professor, University College London (UK)
- Sibel Eker, Assistant Professor, Radboud University (Netherlands)
- Encieh Erfani, Scientific Researcher, International Centre for Theoretical Physics (Iran, Italy)
- Motoko Kotani, Executive Vice-President, Riken (Japan)
- Pradeep Kumar, Professor and Senior Researcher, University of Witwatersrand (South Africa)
- Boon Han Lim, Associate Professor, University of Tunku Abdul Rahman (UTAR) (Malaysia)
- Priscilla Kolibea Mante, Senior Lecturer, Kwame Nkrumah University of Science and Technology (KNUST) (Ghana)
- Alma Hernández-Mondragón, President, Mexican Association for the Advancement of Science (AMEXAC) (Mexico)

Khatijah Mohamad Yusoff, Senior Professor, University of Putra Malaysia (UPM) (Malaysia)

DOI: 10.24948/2023.06

DESIGN: Zhenya Tsoy

COVER PHOTO: Ambrose Chua via [Unsplash](#)

CONTENTS

Executive summary	4
Preface	6
1. Why research evaluation needs to be reformed	7
1.1 Maintaining research quality and protecting research integrity	7
1.2 Maximizing inclusion and diversity	9
1.3 Optimizing research as a global public good	11
1.4 Responding to a fast-changing world	12
2. The challenges for research evaluation reform	15
3. Significant efforts to reform research evaluation	17
3.1 Global manifestos, principles and practices	17
3.2 Regional perspectives and developments	19
3.2.1 Europe	20
3.2.2 Latin America and the Caribbean	24
3.2.3 North America	26
3.2.4 Africa	27
3.2.5 Asia-Pacific	30
4. Conclusions	34
5. Recommendations for action	35
APPENDIX A: Scoping Group members	38
APPENDIX B: Stakeholders consulted for this paper	39
APPENDIX C: GYA, IAP and ISC membership and missions	42
APPENDIX D: List of acronyms	44

EXECUTIVE SUMMARY

A dynamic and inclusive research system is profoundly important for both science and society to advance fundamental knowledge and understanding and to address increasingly urgent global challenges. But the research system is under pressure due to increasing expectations from multiple actors (including funders, governments and the publishing industry), tensions between dynamics of competition and cooperation, an evolving scholarly communication system, an aggressive – at times – publishing and data analytics industry and limited resources. The research enterprise must manage these demands and tensions while maintaining research quality, upholding research integrity, being inclusive and diverse and safeguarding both basic and applied research.

Over the past decade, these pressures on, and the need for responsiveness of, the science system have been accompanied by more critical reflections on systems of research evaluation and performance measurement. While appropriate, context-sensitive methodologies for assessing research quality and impact are important, debates have intensified about the wide-ranging, complex and ambiguous effects of current evaluation criteria and metrics on the quality and culture of research, the quality of evidence informing policymaking, priorities in research and research funding, individual career trajectories and researchers' well-being. In some parts of the world, there is a growing recognition that a narrow and simplistic set of evaluative metrics and indicators do not satisfactorily capture the quality, utility, integrity and diversity of research. Routinely used – often journal-based – metrics fail to capture important additional dimensions of high-quality research such as those found in mentorship, data-sharing, engaging with publics, nurturing the next generation of scholars and identifying and giving opportunities to underrepresented groups. In addition to being too narrow in scope, the issue of misapplication of metrics and indicators is also seen to distort incentives for achievement, disadvantage some disciplines (including vital interdisciplinary and transdisciplinary research) and fuel predatory and unethical publication practices.

Campaigns to curb the misapplication of metrics, broaden quality criteria and transform research culture more systematically through manifestos and statements, principles and reforms have set the stage for a global debate on the need to reform research assessment. These voices are now calling for a move from manifestos to action. This is happening against the background of transformational shifts in the ways in which research is being undertaken and communicated. The rise of open research frameworks and of social media, a shift to mission-oriented and transdisciplinary science, the growth in open peer review and the transformative potential of artificial intelligence (AI) and machine learning requires new thinking on how research and researchers are evaluated.

Against this backdrop, the [Global Young Academy \(GYA\)](#), the [InterAcademy Partnership \(IAP\)](#) and [the International Science Council \(ISC\)](#) joined forces to take stock of debates and developments in research evaluation worldwide, drawing on a scoping group of scientists and a series of regional consultations. New approaches are being developed and piloted by higher education institutions and research funders in some parts of the world, and several are included in this paper. In other parts of the world, these debates and actions are nascent or even absent. With research systems evolving at different rates, there is a risk of divergence and fragmentation. Such divergence may compromise the homogeneity needed to enable research collaboration and facilitate researcher mobility across different geographies, sectors and disciplines. However, one size cannot fit all and there is a need for context-sensitive efforts to reform evaluation, recognizing local challenges.

With a focus on public sector research and the evaluation of research and researchers, this discussion paper is global in its perspective, covering an agenda that is typically dominated by developments in Europe and North America: regional perspectives and examples of national

development and institutional reform are highlighted. The global and collective membership of the GYA, IAP and ISC represents a broad cross-section of the research ecosystem whose diverse mandates can facilitate genuine *systemic* change. This paper endeavours to serve as a stimulus for the GYA, IAP and ISC – as platforms for mutual learning, experimentation and innovation – to work with their members, other science institutions and key constituencies worldwide, to initiate and progress conversations, and mobilize more inclusive and joint action.

Recommendations for the GYA, IAP and ISC and their members (see section 5) are structured around their roles as advocates, exemplars, innovators, funders, publishers, evaluators and collaborators, with indicative timeframes for action. Most immediately, these actions include creating space for sharing lessons and outcomes from relevant initiatives to-date (to build a community of practice); in the medium term, co-convening multistakeholder fora with key constituencies to redesign and implement research evaluation in practicable, context-sensitive and inclusive ways; and, in the longer-term, instigating novel studies that contribute to futures thinking, sensitive to fast-moving developments in AI technologies, peer review methodologies and reform, and communications media.

PREFACE

The [Global Young Academy \(GYA\)](#), [InterAcademy Partnership \(IAP\)](#) and the [International Science Council \(ISC\)](#) came together in 2021 to take stock of challenges, debates and developments in research evaluation/assessment worldwide, across diverse research cultures and systems, and to explore ways in which they might participate in and influence the re-imagining of research evaluation/assessment for the 21st century, in an open and inclusive way.

An international scoping group (Appendix A) was convened to survey the field and advise the three organizations on how they could strengthen existing efforts to reform research evaluation. Central to this work was the premise that (1) a concerted, researcher-led initiative would give the global research community a stronger voice in shaping the future of research evaluation and (2) there are benefits to ‘evaluating with the evaluated’; thus, helping to chart a path to sustained, systemic change in evaluation cultures and practices.

Supplementing desk research, a series of regional consultations with experts identified by the scoping group and partners was conducted in late 2021. The discussion paper is the primary output of this work. It is intended to serve as a prospectus for exploratory conversations with multiple stakeholders, not least the global research community itself.

1. WHY RESEARCH EVALUATION NEEDS TO BE REFORMED

Research evaluation practices serve multiple objectives and are conducted by multiple stakeholders. They are used to assess research proposals for funding decisions, research papers for publication, researchers for recruitment or promotion and the performance of research institutions and universities. This paper focuses predominantly on the evaluation of researchers and research, and does not cover institutional evaluation or ranking, although all these areas of assessment are inextricably linked. Current practices rely heavily on quantitative and largely journal-based metrics, such as Journal Impact Factor (JIF), number of publications, number of citations, h-index and Article Influence Score (AIS). Other metrics include grant income targets, measures of input (such as research funding or size of research team), number of registered patents and, more recently, social media metrics (formerly ‘altmetrics’) such as social media shares or downloads. Together, these metrics profoundly influence institutional, research group and individual reputations, individual and collaborative research agendas, career trajectories and resource allocations.

Over the past two decades, global investment in research and development (R&D) has tripled – to around USD 2 trillion a year. The past years alone have brought the fastest growth in R&D expenditure since the mid-1980s, up by around 19% (UNESCO, 2021¹). This extra investment in research brings with it a culture of accountability that places pressure on research institutions and individuals, and can generate aberrations, or perverse incentives, in response. It has also led to greater aspirations: to maintain quality and reduce research waste, error and inefficiency; maximize inclusion and diversity; optimize research as a global public good; and promote more open and engaged scholarship. Without reform, research quality, integrity, diversity and utility are under threat.

1.1 Maintaining research quality and protecting research integrity

Quantitative metrics *can* form an important part of research evaluation, in the transition to a more open, accountable and public-facing research system (The Royal Society, 2012²). But they are also partly responsible for fuelling the ‘publish or perish’ research culture which exists worldwide, with deleterious effects on the quality of research outputs, the integrity and trustworthiness of research systems and the diversity of research communities (e.g. Haustein and Larivière, 2014³). This is because metrics are used as *proxies* for research quality by institutions, policymakers and research funders alike, but they are a measure of outputs and not of research quality or impact *per se*. As such, these actors do much to set the social and cultural context in which research occurs, and academia’s reward and promotion systems shape the choices of scientists at all stages of their career (Macleod et al., 2014⁴).

1 UNESCO. 2021. UNESCO Science Report: The Race Against Time for Smarter Development (Chapter 1). UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000377250>

2 The Royal Society. (2012). Science as an Open Enterprise. The Royal Society Science Policy Centre. <https://royalsociety.org/~media/policy/projects/sape/2012-06-20-saoe.pdf>

3 Haustein, S. and Larivière, V. 2014. The use of bibliometrics for assessing research: possibilities, limitations and adverse effects. I. Welp, J. Wollersheim, S. Ringelhan, M. Osterloh (eds.), *Incentives and Performance*, Cham, Springer, pp. 121–139.

4 Macleod, M., Michie, S., Roberts, I., Dirnagi, U., Chalmers, I., Ioannidis, J., Al-Shahi Salman, R., Chan, A. W. and Glasziou, P. 2014. Biomedical research: increasing value, reducing waste. *The Lancet*, Vol. 383, No. 9912, pp. 101–104.

The use of bibliometric indices... as proxy metrics for the performance of researchers is a convenient index of assessment but deeply flawed. Most place a relentless focus on individual achievement, thin out research support through a university's interest in high impact metrics, pressurize all to 'tick boxes' and conform, whilst they play an important role in distorting the journal publication market. There is urgent need for reform.

Opening the Record of Science (2021), the International Science Council

The other stakeholder community holding huge power and influence over research communication and knowledge production is the publishing sector. Journal-based metrics have become a powerful incentive to publish in commercial journals and can incentivize behaviour that can have serious side effects. Rather than judging the outcomes of research on its scientific merits, it is the perceived quality of the journal in which it is published that is routinely accepted as evidence of scientific quality, driving a highly commercial publishing market based on reputation rather than on science. Open access costs are largely incurred through author processing charges (APCs): these can be prohibitively high, particularly in some parts of the world, creating barriers to research publication for resource-poor researchers and potentially risking the fracture of the international science community. The risks of becoming more and more dependent on commercial providers and their terms of use in all stages of the research process creates a strong case for not-for-profit alternatives. Further, as bibliometric indicators have provided the dominant source of incentives in universities, they have diminished the value of educational and other forms of scientific work (such as teaching and policy advice). With research evaluation systems tending to favour those who secure large grants and publish in journals with high impact factors, there is evidence to suggest that researchers who have already succeeded are more likely to succeed again (the 'Matthew effect', [Bol et al., 2018](#)⁵).

When scholarly publishing becomes a means of assessment rather than communication, this disadvantages those who choose to communicate their research in other meaningful ways ([ISC's 2021 report](#)⁶). This includes common outputs (and arguably the main currency) of the Global Young Academy (GYA), the InterAcademy Partnership (IAP) and the International Science Council (ISC): reports, working papers, joint statements, opinion editorials, news items and webinars. Some disciplines are also disadvantaged: for example, researchers in engineering and computer science where (usually faster) communication through conferences and their proceedings are important; and those in humanities and social sciences who typically use monographs, books and professional magazines.

Others choose to publish in research-specific or local journals, or are unable to afford to publish their research (however high quality) in open access journals with high impact factors (and concomitant high APCs); the latter disadvantaging those in low-income countries, especially early career researchers (ECRs). These same researchers are under intense pressure for tenured academic posts and their behaviour is strongly conditioned by the quantitative criteria used by research funding agencies and institutional hiring and promotion boards. The temptation to think with indicators ([Muller and de Rijcke, 2017](#)⁷), and even 'game' the system, is a reality for all researchers everywhere in the world (e.g. [Ansedé, 2023](#)⁸).

5 Bol, T., de Vaan, M. and van de Rijdt, A. 2018. The Matthew effect in science funding. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 115, No. 19, pp. 4887–4890.

6 International Science Council. 2021. Opening the Record of Science: Making Scholarly Publishing Work for Science in the Digital Era. Paris, France, ISC. <https://doi.org/10.24948/2021.01>

7 Müller, R. and de Ricke, S. 2017. Thinking with indicators. Exploring the epistemic impacts of academic performance indicators in the life sciences. *Research Evaluation*, Vol. 26, No. 3, pp. 157–168.

8 Ansedé, M. 2023. One of the world's most cited scientists, Rafael Luque, suspended without pay for 13 years. El País. <https://english.elpais.com/science-tech/2023-04-02/one-of-the-worlds-most-cited-scientists-rafael-luque-suspended-without-pay-for-13-years.html>

Manifestations of this gaming include researchers (knowingly or inadvertently) using predatory journals and conferences to boost their publication count (IAP, 2022⁹; Elliott et al., 2022¹⁰), indulging in self-citation and falsifying peer reviews, plagiarism, impact factor inflation and ‘salami-slicing’ (partitioning a large study that could have been reported in a single research article into smaller published articles) (Collyer, 2019¹¹). Under pressure, researchers may be tempted to resort to predatory services with the sole purpose of getting their PhDs, being hired or promoted, or having their research projects financed (e.g. Abad-García, 2018¹²; Omobowale et al., 2014¹³). Metrics-driven academia and academic publishing systems drive insidious incentives: *where* a researcher publishes is more important than *what* they publish.

The impact on the quality and integrity of research is hugely concerning. The number of retracted scholarly articles has risen dramatically in recent years, due to research and publishing misconduct and poor or fraudulent datasets. Journals can take months to years to retract unreliable research, by which time it may already have been cited numerous times and be in the public domain (Ordway, 2021¹⁴).

1.2 Maximizing inclusion and diversity

The predominance of metrics-driven research evaluation is unequivocal and there are diverging trends globally when it comes to assessment reform, which risks leaving parts of the research community behind. In its analysis of the global landscape of research assessment (Curry et al., 2020¹⁵; submitted), it appears that many research and funding institutions in higher-income countries/regions are beginning to include a broader set of indicators, such as qualitative ‘impact’ measures, while bibliometrics remain predominant in institutions in the ‘Global South’¹⁶, across all disciplines. Without more inclusive action, there risks being a divergence of national assessment systems, potentially introducing yet further systemic bias and potential incompatibilities in research, evaluation, funding and publishing systems. This, in turn, may inhibit international research collaboration and the mobility of researchers. In creating barriers to north–south collaboration, it may also inhibit the concomitant strengthening of research ecosystems in the Global South – robust research evaluation strengthens research ecosystems and trust in them, reduces the likelihood of brain drain and helps establish a strong human capital for sustainable

9 IAP. 2022. Combatting Predatory Academic Journals and Conferences. Trieste, Italy, IAP. <https://www.interacademies.org/publication/predatory-practices-report-English>

10 Elliott, T., Fazeen, B., Asrat, A., Cetto, A-M., Eriksson, S., Looi, L. M. and Negra, D. 2022. Perceptions on the prevalence and impact of predatory academic journals and conferences: a global survey of researchers. *Learned Publishing*, Vol. 3, No. 4, pp. 516–528.

11 Collyer, T.A. 2019. ‘Salami slicing’ helps careers but harms science. *Nature Human Behaviour*, Vol. 3, pp. 1005–1006.

12 Abad-García, M. F. 2019. Plagiarism and predatory journals: a threat to scientific integrity. *Anales De Pediatría (English Edition)*, Vol. 90, No. 1, pp. 57.e1–57.e8.

13 Omobowale, A. O., Akanle, O., Adeniran, A. I. and Adegbayega, K. 2013. Peripheral scholarship and the context of foreign paid publishing in Nigeria. *Current Sociology*, Vol. 62, No. 5, pp. 666–684.

14 Ordway, D.-M. 2021. Academic journals, journalists perpetuate misinformation in handling research retractions. The Journalist’s Resource. <https://journalistsresource.org/home/retraction-research-fake-peer-review/>

15 Curry, S., de Rijcke, S., Hatch, A., Pillay, D., van der Weijden, I. and Wilsdon, J. 2020. *The Changing Role of Funders in Responsible Research Assessment: Progress, Obstacles and the Way Ahead*. London, UK, Research on Research Institute.

16 The Global North generally refers to industrialized or developed economies, as defined by the United Nations (2021), while the Global South, refers to economies newly industrialized or that are in the process of industrializing or in development, and that are frequently current or former subjects of colonialism.

development. Nevertheless, one-size-fits-all versions of what constitutes good performance generate forms of behaviour not necessarily conducive to excellence, fairness, transparency and inclusion. To measure the achievements of scholars who have thrived in supportive, well-resourced environments where opportunities abound, in the same way as those who have fought challenges and overcome hurdles in hostile and unsupportive environments is questionable at best (GYA, 2022¹⁷). Many scholars feel historical and geographical exclusion from the research community, fuelled in large part by the way they are assessed throughout their careers. In excluding some forms of research and failing to harness a diversity of ideas globally, there is a risk that current research assessment practices promote a mainstream/follower culture of dominant Western-conceived models.

Researchers in low-income countries and at early stages in their career need a voice so that they can help shape new evaluation models in context-sensitive ways that are fit-for-purpose and account for the challenges they face on a day-to-day basis. The GYA and growing number of National Young Academies give ECRs this voice, and the [GYA's Working Group on Scientific Excellence](#)¹⁸ offers its views on the reform of research evaluation (see text box).

Views from the early career researcher community

Early career researchers (ECRs) are especially concerned about the practices of research evaluation because their career prospects and pursuit of their research agenda crucially depend on how they are evaluated. This informs funding, hiring and promotion practices in ways that are not always perceived as fair and equitable.

While it is obvious that funding and human-resources decisions affect the composition of the labour force of researchers, it is not always recognized that, through its impact on funding, research evaluation shapes incentives for institutions and researchers to pursue a certain research trajectory, work in a certain field or join some networks over others. In this way, research evaluation shapes the development of science itself, and this is especially true in relation to its disproportionate impact on the prospects and expectations of ECRs.

Although science is a global enterprise, some scholars face higher barriers to enter and engage with the research community because of where they were born, their identity or socio-economic background. This is an issue of organization of the science industry and not of research evaluation per se, but many ECRs feel that evaluation criteria should not be blind to this reality of researchers' experience, and should not impose uniform and standardized criteria to different situations.

Research conducted by the Scientific Excellence Working Group of the GYA (forthcoming report) shows that research assessment might be driven more by a country's research policy than by cultural or scientific debates. Focusing on promotion criteria to full professorship (or equivalent) in academia, the report shows that national policies and institutions tend to have specific documents setting out their criteria for research assessment. Rather than encompassing a large and varied set of criteria that could be used to form a comprehensive view of a researcher, these documents tend to focus on a single dimension or priority. For example, some documents focus on the evaluation of a researcher's service activities (such as teaching and mentoring) or some on a researcher's accumulated output (for example, in terms of number of journal articles) – but rarely both.

17 InterAcademy Partnership. Session 12: Winning from Greater Inclusion: Relation Between Diversity and Academic Culture. IAP. <https://www.interacademies.org/page/session-12-winning-greater-inclusion-relation-between-diversity-and-academic-culture>

18 Global Young Academy. Scientific Excellence Working Group. Berlin, Germany, GYA. <https://globalyoungacademy.net/activities/scientific-excellence/>

There are two main implications of this finding. First, that research assessment is hierarchical and top-down. This creates a risk, insofar as both metrics and qualitative methods often ignore the diversity of researchers: both in their background and career paths, and – equally important – diversity in their methods and ideas. In contrast, ECRs represented in the GYA feel it would be important to recognize the diversity of activities that are necessary for the research enterprise, and to devise research evaluation schemes that foster diversity and pluralism rather than mandating conformity and homologation.

Second, differences between disciplines are less significant than differences according to the economic status of countries in which a researcher works. Lower-income countries seem to rely on quantitative metrics and reward ‘productivity’, while higher-income countries are increasingly more open to qualitative assessment of impact. Should this divergence develop further, it might constitute a further obstacle to the international mobility of scholars – which is especially important for ECRs.

In conclusion, the GYA report stresses that there is no silver bullet: research evaluation should be geared toward the objectives of evaluation, and ultimately the goals of an institution or a country’s research policy. Evaluation should allow for the diversity of researcher profiles and careers, and adopt a different focus depending on the purpose of the evaluation. Science being a global and self-critical conversation, an external evaluation may not always be necessary. Indeed, the use and real value of invidious rankings (of people, institutions, outlets or even whole countries) is often debated.

1.3 Optimizing research as a global public good

Today’s global challenges, many of which are articulated in the United Nations (UN) Sustainable Development Goals (SDGs), require transformative, crossdisciplinary and transdisciplinary research, which itself requires new modalities of research delivery and cooperation (ISC, 2021¹⁹). The urgency for inclusive, participatory, transformative, transdisciplinary research is not matched by how research is supported, assessed and funded – for research to deliver on its promise to society, it needs more open, inclusive, context-sensitive assessment systems (Gluckman, 2022²⁰). The embedded behaviour of academics, funders and publishers can make change difficult, so that investment can potentially be directed away from the very areas of greatest need.

The growth in interdisciplinary and transdisciplinary research and participatory or citizen science are important developments and vital in addressing global challenges. As research cuts across disciplinary and institutional boundaries and engages a wider set of stakeholders – including the user community to co-design urgent research questions for society – traditional academic research assessment criteria are insufficient and may even constrain transdisciplinary

19 ISC. 2021. Unleashing Science: Delivering Missions for Sustainability. Paris, France, ISC. doi: [10.24948/2021.04](https://doi.org/10.24948/2021.04)

20 ISC. 2022. An extract from Peter Gluckman’s speech to the Endless Frontier Symposium. Paris, France, ISC. <https://council.science/current/blog/an-extract-from-peter-gluckmans-speech-to-the-endless-frontier-symposium/>

research development and use (Belcher et al., 2021²¹). More appropriate principles and criteria are needed to guide transdisciplinary research practice and evaluation: an early example of a quality assessment framework is built around principles of relevance, credibility, legitimacy and utility (Belcher et al., 2016²²).

1.4 Responding to a fast-changing world

The ways in which research is commissioned, funded, conducted and communicated are evolving at pace and require the acceleration of research evaluation reform. They include the following:

(1) The transition to open science

The open science movement requires concomitant reform of research evaluation systems to improve openness and transparency. Many of the metrics and indicators used to measure research performance are themselves opaque and frequently calculated behind closed commercial doors. This lack of transparency compromises the autonomy of the research community – it restricts options for evaluating, testing, verifying and improving research indicators (Wilsdon et al., 2015²³). Responsible research assessment is becoming a core aspect of global moves towards open science, as witnessed, for example, in the United Nations Educational, Scientific and Cultural Organization (UNESCO) Recommendation on Open Science (UNESCO, 2021²⁴) – which includes the development of an Open Science Toolkit for its members to help them review and reform their research career assessments and evaluation criteria²⁵.

(2) Developments in peer review

The growth of open peer review – whether publishing peer review reports and/or public identification of reviewers – is an important development for research evaluation (Barroga, 2020²⁶; Woods et al., 2022²⁷). The growth of data infrastructure has enabled publishers to generate Digital Object Identifiers (DOIs) for peer review reports, link peer review reports to individual Open Researcher and Contributor IDs (ORCID) and publish papers as preprints. The number of preprints grew significantly during the global COVID pandemic and exposed the challenges posed in assessing research in rapid response mode. Nevertheless, open peer review practices – whether pre- or post-publication – may help to disrupt the control commercial publishers have over research communication and knowledge production processes, reducing the power of the scientific journal and associated metrics such as JIFs. Open records of peer review activities may also provide infrastructure to document – and in time generate greater value in – peer review

21 Belcher, B., Clau, R., Davel, R., Jones, S. and Pinto, D. 2021. A tool for transdisciplinary research planning and evaluation. Integration and Implementation Insights. <https://i2insights.org/2021/09/02/transdisciplinary-research-evaluation/>

22 Belcher, B. M., Rasmussen, K. E., Kemshaw, M. R. and Zornes, D. A. 2016. Defining and assessing research quality in a transdisciplinary context. *Research Evaluation*, Vol. 25, No. 1, pp. 1–17.

23 Wilsdon, J. et al. 2015. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. HEFCE.

24 UNESCO. UNESCO Recommendation on Open Science. Paris, France, UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000379949>

25 A UNESCO source revealed that this work is presently on hold because the debate is dominated by only a few and does not necessarily resonate with many: extensive dialogue must precede the development of recommendations.

26 Barroga, E. 2020. Innovative strategies for peer review. *Journal of Korean Medical Science*, Vol. 35, No. 20, pp. e138.

27 Woods, H. B., et al. 2022. Innovations in peer review in scholarly publishing: a meta-summary. SocArXiv, doi: [10.31235/osf.io/qaksd](https://doi.org/10.31235/osf.io/qaksd)

activities, which are a vital professional service often largely invisible and under-appreciated within academic assessments (Kaltenbrunner et al., 2022²⁸).

(3) The application of artificial intelligence and machine learning

Technological advances in artificial intelligence (AI) and machine learning are likely to have profound consequences for research evaluation, including peer review processes supporting it (e.g. Holm et al., 2022²⁹; Proctor et al., 2020³⁰). AI is already being used to streamline and strengthen peer review (Nature, 2015³¹; Nature, 2022³²), test the quality of peer review (Severin et al., 2022³³), test the quality of citations (Gadd, 2020³⁴), detect plagiarism (Foltýnek et al., 2020³⁵), catch researchers doctoring data (Quach, 2022³⁶) and find peer reviewers, who are increasingly in short supply because this work does not get the credit it deserves in researcher evaluation. ‘Conversational AI’, such as ChatGPT (Chat Generative Pre-Trained Transformer), has the capacity to design experiments, write and complete manuscripts, conduct peer review and support editorial decisions to accept or reject manuscripts (Nature, 2023³⁷). There is also potential for AI to improve the efficiency of peer review by using algorithms to ease the burden of peer reviewers as referees of research output (Nature, 2022³⁸). The use of AI is already being piloted in China to find referees (Nature, 2019³⁹).

All of these AI applications can reduce this burden and allow experienced experts to focus their judgement on research quality and more complex assessments (Thelwall, 2022⁴⁰). But they also risk propagating biases because they are predictive technologies that reinforce existing data which may be biased (for example by gender, nationality, ethnicity or age): indeed, the use of AI itself could benefit from a deeper understanding of what constitutes ‘quality’ research (Chomsky et al., 2023⁴¹; ISI, 2022⁴²).

28 Kaltenbrunner, W., Pinfield, S., Waltman, L., Woods, H. B. and Brumberg, J. 2022. Innovating peer review, reconfiguring scholarly communication: An analytical overview of ongoing peer review innovation activities. SocArXiv, doi: [10.31235/osf.io/8hdxu](https://doi.org/10.31235/osf.io/8hdxu)

29 Holm, J., Waltman, L., Newman-Griffis, D. and Wilsdon, J. 2022. *Good Practice in the Use of Machine Learning & AI by Research Funding Organizations: Insights from a Workshop Series*. London, UK, Research on Research Institute. <https://doi.org/10.6084/m9.figshare.21710015.v1>

30 Procter, R., Glover, B. and Jones, E. 2020. *Research 4.0 Research in the Age of Automation*. London, UK, DEMOS.

31 Baker, M. 2015. Smart software spots statistical errors in psychology papers. *Nature*, <https://doi.org/10.1038/nature.2015.18657>

32 Van Noorden, R. 2022. The researchers using AI to analyse peer review. *Nature* 609, 455.

33 Severin, An., Strinzel, M., Egger, M., Barros, T., Sokolov, A., Mouatt, J. and Muller, S. 2022. *Arxiv*,

34 Gadd, E. 2022. AI-based citation evaluation tools: good, bad or ugly? The Bibliomagician. <https://thebibliomagician.wordpress.com/2020/07/23/ai-based-citation-evaluation-tools-good-bad-or-ugly/>

35 Foltýnek, T., Meuschke, N. and Gipp, B. 2020. Academic plagiarism detection: a systematic literature review. *ACM Computing Surveys*, Vol. 52, No. 6, pp. 1–42.

36 Quach, K. 2022. Publishers use AI to catch bad scientists doctoring data. *The Register*. https://www.theregister.com/2022/09/12/academic_publishers_are_using_ai/

37 Van Dis, E., Bollen, J., Zuidema, van Rooji, R and Bockting, C. 2023. ChatGPT: five priorities for research. *Nature*, Vol. 614, pp. 224–226.

38 Chawla, D. 2022. Should AI have a role in assessing research quality? *Nature*, <https://doi.org/10.1038/d41586-022-03294-3>

39 Cyranoski, D. 2019. Artificial intelligence is selecting grant reviewers in China. *Nature*, Vol. 569, pp. 316–317.

40 Mike, T. 2022. Can the quality of published academic journal articles be assessed with machine learning? *Quantitative Science Studies*, Vol. 3, No. 1, pp. 208–226.

41 Chomsky, N., Roberts, I. and Watumull, J. 2023. The false promise of ChatGPT. *The New York Times*. <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>

42 Clarivate. 2022. Research Assessment: Origins, Evolution, Outcomes. Clarivate. <https://clarivate.com>

Vitality, however, all forms of AI and machine learning are open to abuse (Blauth et al., 2022⁴³; Bengio, 2019⁴⁴). Academic and research communities will need to build preparedness and resilience to this, working with government, industry and civil society leadership governing this space.

(4) The rise of social media

Conventional quantitative measures of research impact fail to account for the rise in social media engagement and socially networked researchers/academics (Jordan, 2022⁴⁵). Many academics use social media platforms to engage communities, policymakers and publics throughout the lifetime of their research project; to positively engage with, test and inform their research, and bring a diversity of ideas and inputs, rather than simply publishing the final output as a *fait accompli* for a recipient audience. This engagement is not picked up by conventional forms of research assessment yet can lead to wider influencing and outreach opportunities. Social media metrics ('altmetrics') are being developed as a contribution to responsible metrics (Wouters et al., 2019⁴⁶) and include Twitter or Facebook mentions and the number of followers on ResearchGate, for example. On the one hand, these altmetrics can help open up, create space and broaden evaluation (Rafols and Stirling, 2021⁴⁷) but on the other hand – like other indicators – can also be used irresponsibly and/or be seen to impose another layer of metrics in evaluation systems.

[com/lp/research-assessment-origins-evolutions-outcomes/](https://doi.org/10.1080/17439884.2022.2065298)

43 Blauth, T. F., Gstrein, O. J. and Zwitter, A. 2022. Artificial intelligence crime: an overview of malicious use and abuse of AI. *IEEE Access*, Vol. 10, pp. 77110–77122.

44 Castelvocchi, D. 2019. AI pioneer: 'The dangers of abuse are very real'. *Nature*, doi: <https://doi.org/10.1038/d41586-019-00505-2>

45 Jordan, K. 2022. Academics' perceptions of research impact and engagement through interactions on social media platforms. *Learning, Media and Technology*, doi: 10.1080/17439884.2022.2065298

46 Wouters, P., Zahedi, Z. and Costas, R. 2019. Social media metrics for new research evaluation. Glänzel, W., Moed, H.F., Schmoch U., Thelwall, M. (eds.), *Springer Handbook of Science and Technology Indicators*. SpringerLink.

47 Rafols, I. and Stirling, A. 2020. Designing indicators for opening up evaluation. Insights from research assessment. ResearchGate, doi: [10.31235/osf.io/h2fxp](https://doi.org/10.31235/osf.io/h2fxp)

2. THE CHALLENGES FOR RESEARCH EVALUATION REFORM

Challenges to the reform of research evaluation are manifold. A few of the most significant are illustrated here.

Any reform that includes more qualitative measures must – at the same time – **safeguard the quality of basic and applied research**. There is anecdotal evidence that some scientists may themselves oppose reform, perhaps especially advanced career researchers who have thrived in the current system, because they fear it risks fuelling mediocre research, or that more qualitative forms of evaluation may favour applied over basic research. Reform of research evaluation criteria tends to be framed around moves towards mission-oriented, societally impactful research which appeal to public and political support in a way that less tangible basic or blue-skies research may not. Some argue that a more nuanced interpretation of research ‘value’ is required to underpin innovation, as the future requires continued investment in fundamental, curiosity-driven research and a wider appreciation of the crucial role it plays in the capacity to respond to global challenges (GYA, 2022⁴⁸).

The **lack of consistency in the meaning and use of research terminology**, more generally, is a barrier to change. The conceptual framework for research evaluation has not changed substantively over time, nor has the language supporting it: the research system is still stuck in old dichotomies such as ‘basic’ and ‘applied’ science, and terms like ‘impact’, ‘quality’ (unhelpfully equated with productivity) and ‘excellence’ are not clearly defined in a way that avoids geographic, disciplinary, career stage and gender bias (Jong et al., 2021⁴⁹): this may be particularly acute in decision-making panels lacking diversity (Hatch and Curry, 2020⁵⁰).

Like metrics-driven evaluation, more qualitative forms of evaluation are also imperfect. Making the argument that peer review processes and expert judgement are at least as important as bibliometrics is not straightforward. They can be biased due to the lack of clarity and transparency in the peer review process. Peer review committees, for example, have been criticized as mechanisms which preserve established forms of power and privilege by enabling ‘old boys’ networks’ and homophily (evaluators seeking out those like themselves) to persist, while also being vulnerable to groupthink dynamics. Quantitative metrics, however imperfect, are seen in some parts of the world as a defence against nepotism and bias. Similar arguments can be applied to the peer review of research papers, with the use of more qualitative assessment potentially opening the door to other forms of discriminatory behaviour.

The lack of professional recognition of, and training for, peer review in any form creates disincentives to serve as a peer reviewer, thereby reducing capacity. Further, as demand exceeds supply, it can create incentives to cut corners and reduce rigour. Increasing peer review transparency (whether fully open, anonymized or hybrid) and training, fostering and rewarding good peer review practice are all required; as is further research on models for its evolution as research outputs diversify (IAP, 2022⁵¹) and AI technologies advance.

Debates on research evaluation reform are complex and not binary. Qualitative and quantitative information have often been combined in peer review contexts: statements like the

48 Rich, A., Xuereb, A., Wrobel, B., Kerr, J., Tietjen, K., Mendisu, B., Farjalla, V., Xu, J., Dominik, M., Wuite, G., Hod, O. and Baul, J. 2022. Back to Basics. Halle, Germany, Global Young Academy.

49 Jong, L., Franssen, T. and Pinfield, S. 2021. *Excellence in the Research Ecosystem: A Literature Review*. London, UK, Research on Research Institute.

50 Hatch, A. and Curry, S. 2020. Research Culture: Changing how we evaluate research is difficult, but not impossible. *eLife*, Vol. 9, p. e58654.

51 IAP. 2022. *Combatting Predatory Academic Journals and Conferences*. Trieste, Italy, IAP.

Leiden Manifesto for Research Metrics (Hicks et al., 2015⁵²) call for ‘informed peer review’ in which expert judgement is supported – but not led by – appropriately selected and interpreted quantitative indicators and by qualitative information. The debate on research evaluation is not a binary ‘qualitative versus quantitative’ choice of evaluation tools, but how to ensure the best combination of multiple forms of information.

Finally, any reform must also be **convenient and practicable**. The research system is already showing signs of collapsing under itself, as the volume of publications rises exponentially and the burden of review falls unevenly across the research enterprise (e.g. Publons, 2018⁵³; Kovanis et al., 2016⁵⁴; Nature, 2023⁵⁵). Journal-based metrics and the h-index, together with qualitative notions of publisher prestige and institutional reputation, can provide convenient shortcuts for busy evaluators and present obstacles to change that have become deeply entrenched in academic evaluation (Hatch and Curry, 2020⁵⁶). Quantitative metrics are hailed in some countries as providing relatively clear and unambiguous routes for appointment and promotion. In the ‘Global South’, average impact factors are routinely used to shortlist applicants, and any alternative must be equally implementable and able to draw on the additional resources inevitably required to broaden the scope of evaluation. The convenience of using simple quantitative metrics in research evaluation is likely to be a major obstacle for change, and the introduction of new evaluation systems may even create *more* global inequity due to lack of capacity or competence in some countries.

52 Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. and Rafols, I. 2015. Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, Vol. 520, pp. 429–431.

53 Publons. 2018. *Global State of Peer Review*. London, UK, Clarivate. <https://doi.org/10.14322>

54 Kovanis, M., Porcher, R., Revaud, P. and Trinquart, L. 2016. The global burden of journal peer review in the biomedical literature: strong imbalance in the collective enterprise. *PLoS ONE*, Vol. 11, No. 11, p. e0166387.

55 Forrester, B. 2023. Fed up and burnt out: ‘quiet quitting’ hits academia. *Nature*, Vol. 615, pp. 751–753.

56 Hatch, A. and Curry, S. 2020. Research culture: changing how we evaluate research is difficult, but not impossible. *eLife*, Vol. 9, p. e58654.

3. SIGNIFICANT EFFORTS TO REFORM RESEARCH EVALUATION

Over the past decade, there have been a series of high-profile manifestos and principles on research evaluation to address these challenges, including the Leiden Manifesto (developed by a group of international experts), the Hong Kong Principles (Moher et al., 2020⁵⁷) (developed at the 6th World Conference on Research Integrity in 2019) and *The Metric Tide*⁵⁸ and *Harnessing the Metric Tide*⁵⁹ reports (developed in the context of a review of the UK's research and evaluation framework, REF). There are at least 15 distinct efforts urging key stakeholders – whether policymakers, funders or heads of higher education institutions (HEIs) – to minimize the potential harm of current assessment systems. All of these initiatives have reached a wide audience and are progressive in their focus on responsible metrics as a prerequisite for improving research culture and bringing equality, diversity, inclusion and belonging into the research community. But there is a growing concern from some architects of these initiatives that, while helpful, they detract from tangible practical action: the act of being a signatory is only effective if followed up with practical implementation (Nature, 2022⁶⁰).

There is increasing support for 'responsible research assessment or evaluation' and 'responsible metrics' (DORA, 2012⁶¹; Hicks et al. 2015⁶²; Wilsdon et al., 2015⁵⁸) that move away from purely quantitative metrics to a wider variety of measures to enable researchers to describe the economic, social, cultural, environmental and policy impact of their research; to account for issues the research community values: 'data for good' or 'value-led indicators' that address wider attributes (Curry et al., 2022⁶³). In recent years, innovative and progressive approaches to responsible research assessment have been developed and piloted by some HEIs and research funders in regions and countries around the world. Some are highlighted here.

3.1 Global manifestos, principles and practices

Of the global initiatives mentioned above, the 2013 San Francisco '*Declaration on Research Assessment*'⁶⁴ (DORA) is perhaps the most active global initiative. It has catalogued problems caused by using journal-based indicators to evaluate the performance of individual researchers and provides 18 recommendations to improve such evaluation. DORA categorically discourages the use of journal-based metrics to assess a researcher's contribution or when looking to hire, promote or fund. As of mid-April 2023, the declaration has been signed by 23,059 signatories (institutions and individuals) in 160 countries, committing to reform. With a focus on navigating

57 Moher, D., Bouter, L., Kleinert, S., Glasziou, P., Har Sham, M., Barbour, V., Coriat, A. M., Foeger, N. and Dirnagi, U. 2020. The Hong Kong Principles for assessing researchers: fostering research integrity. *PLoS Biology*, Vol. 18, No. 7, p. e3000737.

58 Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R. and Kerridge, S. 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

59 Curry, S., Gadd, E. and Wilsdon, J. 2022. *Harnessing the Metric Tide: Indicators, Infrastructures & Priorities for UK Responsible Research Assessment*. London, UK, Research on Research Institute.

60 Nature Editorial. 2022. Support Europe's bold vision for responsible research assessment. *Nature*, Vol. 607, p. 636.

61 Declaration on Research Assessment (DORA). <https://sfdora.org/about-dora/>

62 Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. and Rafols, I. 2015. Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, Vol. 520, pp. 429–431.

63 Curry, S., Gadd, E. and Wilsdon, J. 2022. *Harnessing the Metric Tide: Indicators, Infrastructures & Priorities for UK Responsible Research Assessment*. London, UK, Research on Research Institute. https://rori.figshare.com/articles/report/Harnessing_the_Metric_Tide/21701624

64 DORA. *San Francisco Declaration on Research Assessment*. <https://sfdora.org/read/>

the intrinsic challenges and innate biases of qualitative assessment, DORA is developing [Tools to Advance Research Assessment \(TARA\)](#)⁶⁵ to help put the declaration into practice: these tools include a dashboard to index and classify innovative policies and practices in career assessment and a toolkit of resources to help de-bias committee composition and to recognize different, qualitative forms of research impact.

Additionally, DORA is funding ten projects – in Argentina, Australia, Brazil, Colombia (2), India, Japan, Netherlands, Uganda and Venezuela – to test different ways of promoting reform in research evaluation in their local contexts, as well as [compiling examples of good practice](#): for example, awareness raising, developing new policy or practice, training and practical guidance for job applicants ([DORA](#)⁶⁶). Demand for grants of this kind has been high – over 55 applicants from 29 countries – indicating a growing recognition of the need for reform.

Professional research management associations like the International Network of Research Management Societies (INORMS) have also been actively developing resources to guide organizational change, including the [SCOPE Framework Research Evaluation Group | INORMS - The INORMS SCOPE framework for research evaluation](#)⁶⁷ which starts by defining what is valued, who is being evaluated and why (a useful explanatory poster [here](#)⁶⁸).

The international development sector has offered new perspectives on research evaluation, a prime example being [Research Quality Plus | IDRC - International Development Research Centre](#)⁶⁹, which measures what matters to people at the receiving end of research. The Research Quality Plus (RQ+) tool recognizes that scientific merit is necessary but not sufficient, acknowledging the crucial role of the user community in determining whether research is relevant and legitimate. It also recognizes that research uptake and influence begin *during* the research process. Research applications are often assessed by highly interdisciplinary panels, also containing development experts from outside academia (e.g. a government department or nongovernment organization (NGO)), practitioners and in-country representatives: this reinforces the importance of the user community/non-subject experts needing to understand the research and how it can be applied in practice. Research in complex, low-income or fragile settings can be accompanied by an ethics toolkit or framework, designed to inform and support ethical choices in the research lifecycle, from inception to dissemination and impact, e.g. [Reid et al., 2019](#)⁷⁰. ‘Theory of Change’ approaches are widely used in international development research by donors, NGOs and multilateral agencies, where applicants must articulate pathways to impact, supported by monitoring, evaluation and learning frameworks, e.g. [Valters, 2014](#)⁷¹. The academic research community can potentially learn from the development community.

Recognizing the role of funders in shaping the strategies of HEIs, the [Global Research Council’s \(GRC\) Responsible Research Assessment](#) (RRA) initiative⁷² has been incentivizing major research

65 DORA. *Tools to Advance Research Assessment*. DORA. <https://sfdora.org/project-tara/>

66 DORA. *DORA Community Engagement Grants: Supporting Academic Assessment Reform* <https://sfdora.org/dora-community-engagement-grants-supporting-academic-assessment-reform/>

67 Inorms. *SCOPE Framework for Research Evaluation*. <https://inorms.net/scope-framework-work-for-research-evaluation/>

68 Inorms. *The SCOPE Framework*. <https://inorms.net/scope-framework-for-research-evaluation/>

69 Torfin, S. 2018. *Research Quality Plus*. International Development Research Centre. <https://www.idrc.ca/en/rqplus>

70 Reid, C., Calia, C., Guerra, C. and Grant, L. 2019. *Ethical Action in Global Research: A Toolkit*. Edinburgh, Scotland, University of Edinburgh. <https://www.ethical-global-research.ed.ac.uk/>

71 Valters, C. 2014. *Theories of Change in International Development: Communication, Learning, or Accountability?* [The Asia Foundation](#). <https://www.alnap.org/system/files/content/resource/files/main/jsrp17-valters.pdf>

72 Fraser, C., Nienaltowski, M. H., Goff, K. P., Firth, C., Sharman, B., Bright, M. and Dias, S. M. 2021. *Responsible Research Assessment*. Global Research Council. <https://globalresearchcouncil.org/>

funders all over the world to work towards RRA ambitions in their own regional and national contexts and to develop effective evaluation frameworks to assess impact (explanatory video [here](#)⁷³). Commissioning a working paper on RRA (Curry et al., 2020⁷⁴), the GRC called for its members to embed RRA principles and take concrete action to fulfil them, and to learn from each other through collaboration and sharing of good practice. An [international working group](#)⁷⁵ is providing guidance and support to GRC members, helping them transition from movement to action.

In large part through the efforts of the GYA, ECRs are also beginning to mobilize themselves around this agenda. Its [Working Group on Scientific Excellence](#)⁷⁶ is working to identify research environments conducive to ‘unleashing curiosity and creativity in science and to foster the development of human potential through diversity and inclusion’. Their work calls for the ECR community to challenge definitions of ‘excellence’ used by their organizations, to get involved in initiatives to reform research evaluation and to join the Young Academies movement. It also calls on funding and hiring bodies to involve ECRs in research evaluation debates and acknowledge a broader diversity of contributions to, and careers in, research.

Although some universities and other HEIs are signatories to DORA and/or joining the European movement (described below), they do not appear to be organizing themselves collectively around research evaluation in the way that other key constituencies are.

3.2 Regional perspectives and developments

Problems created by evaluation systems that are almost exclusively quantitative are largely seen and diagnosed from a ‘Global North’ perspective, with the ‘Global South’ at risk of playing catch-up. At the risk of over-generalization, there are major systemic issues in the ‘Global North’ around the lack of diversity, equity and inclusion that are exacerbated by evaluation systems. In the ‘Global South’, there is a lack of local and regional definition of what constitutes ‘quality’ and ‘impact’, widely varying evaluation systems (even across departments at the same university), and relatively little in the way of challenge to the *status quo*. The world over, problems derive from the overemphasis on quantitative indicators, the link between evaluation and resource allocation, the highly competitive funding system and pressure to publish, and the disregard for other, less quantifiable dimensions of research and academic life.

Peer-reviewed literature on comparative studies of research evaluation reform is sparse. A rare exception is a comparison of research evaluation interventions in six different geographies (Australia, Canada, Germany, Hong Kong, New Zealand and UK), which observes that the indexed performance of all six *appears* to improve after multiple types of intervention (at least using conventional bibliometric indicators) (ISI, 2022⁷⁷). DORA provides (largely institutional)

[news/responsible-research-assessment/](#)

73 The Global Research Council. GRC Responsible Research Assessment. [YouTube](#). <https://www.youtube.com/watch?v=CnsqDYHGdDo>

74 Curry, S., de Rijcke, S., Hatch, A., Dorsamy, P., van der Weijden, I. and Wilsdon, J. 2020. *The Changing Role of Funders in Responsible Research Assessment*. London, UK, Research on Research Institute. <https://doi.org/10.6084/m9.figshare.13227914.v1>

75 Global Research Council. *Responsible Research Assessment Working Group*. [GRC](#). <https://globalresearchcouncil.org/about/responsible-research-assessment-working-group/>

76 Global Young Academy. *Scientific Excellence*. GYA. <https://globalyoungacademy.net/activities/scientific-excellence/>

77 Adams, J., Beardsley, R., Bornmann, L., Grant, J., Szomszor, M. and Williams, K. 2022. *Research Assessment: Origins, Evolution, Outcomes*. Institute for Scientific Information. <https://clarivate.com/ISI-Research-Assessment-Report-v5b-Spreads.pdf>

case studies on its webpage (DORA⁷⁸) and in a report (DORA, 2021⁷⁹) designed to inspire others to act, but these are predominantly European examples.

Here, the authors provide regional overviews and national examples of experimentation and reform for further insight – these are not intended to be comprehensive or exhaustive.

3.2.1 Europe

The [EU Coalition on Reforming Research Assessment](#)⁸⁰, or CoARA, approved in July 2022, is the largest initiative on research evaluation reform in the world. Four years in the making and developed by 350 organizations in 40 (largely European) countries, the European University Association and Science Europe (a network of the continent's science funders and academies), in concert with the European Commission, have developed an agreement or set of principles (a 'reform journey'), for more inclusive and responsible research assessment (CoARA, 2022⁸¹). The agreement focuses on three levels of assessment: institutions, individual researchers and the research itself. While governed by European partners, the coalition has ambitions to become global and both DORA and the GYA are already signatories. Signatories undertake to commit resources to improve research evaluation, develop new criteria and tools for evaluation, and raise awareness and provide training on research evaluation (e.g. to peer reviewers). This development has been described as 'the most hopeful sign yet of real change' (Nature, 2022⁸²).

The EU is also funding some exciting new initiatives designed to support research evaluation reform: notably, Open and Universal Science (OPUS⁸³) – to develop a 'comprehensive suite' of indicators across multiple research processes and outputs, and thereby incentivize European researchers to practice open science – and the open science assessment dataspace GraspOS⁸⁴ – to build an open dataspace to support policy reform for research assessment.

The European Research Council (ERC), which supports frontier research across all fields (with a budget of €16 billion for 2021–2027) has signed CoARA and has amended its evaluation forms and processes to build in more narrative descriptions, including accounting for less conventional career paths and 'exceptional contributions' to the research community. Proposals will be judged more on their merit than on the past achievements of the applicant and will continue to be evaluated by peer review panels composed of leading scholars using the sole criterion of scientific excellence (ERC, 2022⁸⁵).

Some European academies are also engaged. The Board of ALLEA, the European Federation of Academies of Sciences and Humanities, representing nine of the 50-plus national academies in

78 DORA. Resource Library. <https://sfdora.org/resource-library>

79 Saenen, B., Hatch, A., Curry, S., Proudman, V. and Lakoduk, A. 2021. Reimagining Academic Career Assessment: Stories of Innovation and Change. DORA. https://eua.eu/downloads/publications/eua-dora-sparc_case%20study%20report.pdf

80 Coalition for Advancing Research Assessment (CoARA). <https://coara.eu/>

81 CoARA. 2022. Agreement on Reforming Research Assessment. https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf

82 Nature Editorial. 2022. Support Europe's bold vision for responsible research assessment. *Nature*, Vol. 607, p. 636.

83 Open and Universal Science. OPUS Home - Open and Universal Science (OPUS) Project. <https://opusproject.eu/>

84 Vergoulis, T. 2023. GraspOS Moving Forward to a More Responsible Research Assessment. OpenAIRE. <https://www.openaire.eu/graspos-moving-forward-to-a-more-responsible-research-assessment>

85 European Research Council. 2022. ERC Scientific Council Decides Changes to the Evaluation Forms and Processes for the 2024 Calls. ERC. <https://erc.europa.eu/news-events/news/erc-scientific-council-decides-changes-evaluation-forms-and-processes-2024-calls>

40 European countries, has endorsed the CoARA movement. ALLEA has undertaken to establish a dedicated task force to collect, exchange and promote good practice for admitting new Academy Fellows and to contribute to ‘meaningful cultural exchange’ of research assessment, based on principles of quality, integrity, diversity and openness. In its [October 2022 statement](#)⁸⁶, ALLEA calls on member academies to do the following:

- 1. Recognise the diversity of contributions to, and careers in, research in accordance with the needs and nature of the research; in the case of Academy fellows, selection procedures should (1) take into consideration gender balance and the unique challenges of early career researchers, (2) support diversity of cultures and disciplines, (3) value a variety of competency areas and talents, and (4) promote interdisciplinarity and multilingualism.*
- 2. Base research assessment primarily on qualitative evaluation for which peer review is central, supported by responsible use of quantitative indicators; assessment of excellence and impact regarding candidate fellows’ work should be based on qualitative peer review that meets the fundamental principles of rigor and transparency and takes into consideration the specific nature of the scientific discipline.*
- 3. Abandon inappropriate use of journal- and publication-based metrics in research assessment; in particular, this means moving away from using metrics like the Journal Impact Factor (JIF), Article Influence Score (AIS) and h-index as dominant proxies for quality and impact.*

[Allea Statement On Reforming Research Assessment Within The European Academies](#)

In their [joint response](#)⁸⁷ to the EU Agreement and CoARA Coalition, the ECR community in the GYA have also welcomed this commitment and offer ways of implementing its principles. These include practices that are inclusive and reflect the diversity of national specificities and disciplines’ characteristics, with researchers of all career stages receiving training, incentives and rewards, with mandatory training on open science for researchers, staff and committee members being vital.

Research-intensive universities in Europe have also got behind the reform of research evaluation as a pathway for ‘multidimensional’ research careers ([Overlaet, B., 2022](#)⁸⁸). They have developed a common framework to inspire and support universities to recognize a diversity of contributions in research, education and service to society.

At the national level, several countries are now piloting different assessment models: for example, national funding agencies in [Belgium, the Netherlands, Switzerland](#) and [UK](#) are all using ‘narrative CVs’. Narrative CVs look more holistically at academic achievement: contribution to the generation of knowledge, to the development of individuals, to the wider research community and to broader society ([Royal Society](#)⁸⁹). While there is widening support for these types of CVs, there is also some concern that they force academics to be good at everything, and thus risk compromising deep expertise in the

86 All European Academies. 2022. ALLEA Statement on Reforming Research Assessment within the European Academies. ALLEA. <https://allea.org/wp-content/uploads/2022/10/ALLEA-Statement-R-RA-in-the-Academies.pdf>

87 Eurodoc, MCAA, YAE, ICoRSA and GYA. 2022. Joint Statement on the EU Council conclusions on Research Assessment and the Implementation of Open Science. Zenodo, doi: 10.5282/zenodo.7066807.

88 Overlaet, B. 2022. A Pathway towards Multidimensional Academic Careers - A LERU Framework for the Assessment of Researchers. LERU, Leuven, Belgium. https://www.leru.org/files/Publications/LERU_PositionPaper_Framework-for-the-Assessment-of-Researchers.pdf

89 Royal Society. Résumé for Researchers. <https://royalsociety.org/topics-policy/projects/research-culture/tools-for-support/resume-for-researchers/>

pursuit of all-rounder status ([Grove, J., 2021⁹⁰](#)).

Four examples of national research systems that are coordinating *nationwide* reforms in career-oriented academic assessments are included in the following text boxes.

National example: The UK

The UK Research Evaluation Framework (REF) measures research impact through two dimensions: ‘significance’ (the tangible difference a project makes) and ‘reach’ (the quantifiable extent to which it does so) ([UKRI](#)). Impact here is defined 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’ but beyond this it is very open-ended, discipline-variable and arguably ambiguous, failing to account adequately for public engagement, for example.

The UK’s REF is being evaluated in 2022–2023 under the [Future Research Assessment Programme](#) to explore possible new approaches to the assessment of UK higher education research performance, and includes understanding international research assessment practice. The next iteration of the REF will possibly account for a more diverse set of outputs and perhaps even reduce the importance attached to them. The current model attaches 60% importance to outputs, 25% to research impact and 15% to research culture/environment. If these were more evenly weighted, then the REF would look very different, with more importance attached to research culture, research integrity and team working ([Grove, 2020](#)).

National example: Finland

In 2020, Finland’s Federation of Learned Societies coordinated a taskforce of research funders, universities and unions that published the statement [Good Practice in Research Assessment](#). This sets out guidance for following a responsible process for assessment of individual academics, including five general principles of evaluation: transparency, integrity, fairness, competence and diversity. *Good Practice in Research Assessment* calls on research integrity, education and mentorship, and scientific service (e.g. peer review) to be better acknowledged in assessing individuals’ academic contributions. The statement sees assessments as not simply about producing summative judgements: it also encourages evaluators to share feedback with individuals being evaluated to facilitate feedback and learning.

Research-performing organizations and research funding organizations have all committed to implementing *Good Practice in Research Assessment* and generating their own local variations on the guidance, and a national researcher portfolio CV model is being developed. *Good Practice in Research Assessment* commits to regular reviews and refinements.

90 Grove, J. 2021. Do narrative CVs tell the right story? Times Higher Education (THE). <https://www.timeshighereducation.com/depth/do-narrative-cvs-tell-right-story>

National example: The Netherlands

In the Netherlands, the national *Recognition and Rewards* programme commenced in 2019, with publication of the position statement [Room for Everyone's Talent](#). This nationwide collaboration between the Royal Netherlands Academy of Arts and Sciences (KNAW – an IAP and ISC member), research funders, universities and medical centres states that systems-wide modernization of research assessment cultures needs to occur. In doing so, it lays out five ambitions for change in assessment procedures: greater career path diversity, recognizing individual *and* team performance, prioritizing quality of work over quantitative indicators, open science and academic leadership.

Since 2019, Dutch universities have moved to enact their own local translations of the national vision statement. Simultaneously, funding agencies have initiated more 'narrative CV' formats and ceased requesting bibliometric information, citing San Francisco DORA as an inspiration. The Dutch Research Council has very recently moved to an '[evidence-based' CV](#) in which some quantitative information may be used. The KNAW also developed its own [three-year plan](#) to implement the *Recognition and Rewards* agenda internally. A full-time program manager and team have been appointed to facilitate the *Recognition and Rewards* reform programme, and a 'Recognition and Rewards Festival' is held annually among major reform stakeholders to support community-wide learning.

Finally, funded by a DORA Community Engagement grant, the [Young Scientist in Transition initiative for PhD students](#), based in Utrecht, has developed a new evaluation guide for PhDs, in an effort to change research culture.

National example: Norway

In 2021, Norway, Universities Norway and the Norwegian Research Council published [NOR-CAM – A toolbox for recognition and rewards in academic assessments](#). NOR-CAM provides a matrix framework for improving transparency and broadening the evaluation of research and researchers away from narrow bibliometric-informed indicators. NOR-CAM stands for Norwegian Career Assessment Matrix, and was adapted from a 2017 [report](#) by the European Commission which presented the Open Science Career Assessment Matrix. Like its European predecessor, NOR-CAM also posits ways for better integrating open science practices into assessments. The matrix aims to guide evaluators and candidates for academic positions, research funding applications and national evaluators evaluating Norwegian research and education. It is also intended to act as a general guide for individual career development.

The matrix includes six main areas of competence: research outputs, research process, pedagogical competencies, impact and innovation, leadership and other competencies. The matrix then provides suggestions to enable career planning and assessment recognition around each of the criteria – examples of results and skills, means of documentation and prompts for reflection about each criterion. Candidates are not expected to perform equally on all criteria.

NOR-CAM was created by a working group of research performing and funding organization stakeholders, coordinated by Universities Norway, meaning in principle it has buy-in from members of all Norwegian universities. Workshops involving Norwegian universities have been held subsequently to co-develop ways of integrating NOR-CAM into appointment and promotion assessment procedures, and an 'automatic' CV system is being developed to retrieve data from multiple national and international sources in development to reduce administrative burden. Coordinators from the abovementioned three national-level reform schemes have met regularly to exchange their experiences and shared learning.

3.2.2 Latin America and the Caribbean

Latin America and the Caribbean (LAC) contrast in many ways to other parts of the world. Here, science is considered a global public good and its research and academic publishing systems and infrastructure are publicly owned (funded) and non-commercial: but these regional strengths and traditions are not yet reflected in evaluation systems. Key stakeholders who can effect change are the national research councils, ministries of science and main research universities – the role of HEIs is vital, given that more than 60% of researchers are located in universities (RiCyT, 2020⁹¹). There is the potential to align evaluation systems more closely with the SDGs and with open science and citizen science movements, which have a flourishing tradition in the region.

Currently, there is high fragmentation of research assessment systems nationally, locally and institutionally, putting research in competition with other functions, such as teaching, extension and coproduction. Research evaluation and researcher award systems in LAC generally favour a notion of excellence anchored in methodologies of the ‘Global North’, based exclusively on the impact factor of journals and university rankings (CLACSO, 2020⁹²). Recognition of different forms of knowledge production and communication, and the multiplicity of academic careers (e.g. teaching, training and mentoring, citizen science and public communication of science) are largely absent in research evaluation practices. This is especially problematic for researchers in social sciences and humanities, where monographs and local languages are used extensively (CLACSO, 2021⁹³). Regional journals and indicators are devalued or not recognized in such evaluation processes. All of this is exacerbated by weak information systems and weak interoperability of (especially community-owned) infrastructures, underfunded because scarce funds are directed to APC payments for open access journals.

Nevertheless, some universities in the region are beginning to implement evaluation practices that deploy a combination of qualitative and quantitative methodologies, especially in the assessment of researchers and missions-oriented research (Gras, 2022⁹⁴). The transition to more comprehensive research evaluation schemes will require the co-design of more qualitative criteria; the responsible use of quantitative data and strengthening peer review processes; incremental changes that harmonize and coordinate policies and methodologies toward shared principles on responsible research assessment and open science; new methodologies and data for better assessing inter/ transdisciplinary science, environmental and local issues; shared, interoperable, sustainable, federated infrastructures that support bibliodiversity and multilingualism; and participatory, bottom-up designs that broaden participation of citizens and social movements and the inclusion of underrepresented research groups.

91 RiCYT. Researchers by employment sector (FTE) 2011-2020. app.ricyt.org/ui/v3/comparative.html?indicator=INVESTIJCSEPER&start_year=2011&end_year=2020

92 CLACSO. 2020. Evaluating Scientific Research Assessment. Towards a Transformation of Scientific Research Assessment in Latin America and the Caribbean Series from The Latin American Forum for Research Assessment (FOLEC). CLACSO, Buenos Aires, Argentina. <https://www.clacso.org/wp-content/uploads/2020/05/FOLEC-DIAGNOSTICO-INGLES.pdf>

93 CLACSO. 2021. Towards a Transformation of Evaluation Systems in Latin America and the Caribbean, Tools to Promote New Evaluation Policies. Series from The Latin American Forum for Research Assessment (FOLEC). CLACSO, Buenos Aires, Argentina. <https://www.clacso.org/wp-content/uploads/2022/02/Documento-HERRAMIENTA-2-ENG.pdf>

94 Gras, N. 2022. Forms Of Research Assessment Oriented At Development Problems. Practices And Perspectives From National Science And Technology Organizations And Higher Education Institutions In Latin America And The Caribbean. FOLEC. CLACSO, Buenos Aires, Argentina.

To address these challenges, the region has adopted a set of principles and guidelines for research assessment. The [CLACSO-FOLEC Declaration of Principles for Research Assessment](#)⁹⁵, approved in June 2022, sets out to guarantee and protect quality and socially relevant science, and embrace the principles of DORA and open science, the diversity of research outputs and research careers, the value of regional journals and indexing services, and of interdisciplinarity, local language and indigenous knowledge. To-date, it has over 220 adherents and there are already positive trends in responsible research assessment and examples of reform. Some national examples are provided in the following text boxes.

National example: Colombia

Funded by a DORA Community Engagement Award, the Colombian Associations of Universities, University Publishers, Research Managers and a network of science and technology management, among others, have been working together on the opportunities and challenges of responsible metrics in Colombia. Through a series of workshops and consultations, including with international organizations as benchmarks, they have developed a rubric to help Colombian institutions design their own REFs. This rubric endeavours to account for challenges identified at the local level, which – for HEIs – include the lack of knowledge on research evaluation alternatives, the nature of the national research evaluation ecosystem and resistance to change. A [dedicated website](#) has been developed, together with infographics to assist researchers, and dissemination and learning continues to be shared across the country.

More information: [The Colombian responsible metrics project: towards a Colombian institutional, methodological instrument for research assessment | DORA \(sfdora.org\)](#)

National example: Argentina

An interesting attempt of reform in the [National Council for Scientific and Technical Research](#) (Consejo Nacional de Investigaciones Científicas y Técnicas – CONICET) has been the creation of a special resolution for the social and human sciences that puts journals indexed in the mainstream circuit on the same level with journals indexed in regional bases such as [SciELO](#), [Redalyc](#) or [Latindex-Catálogo](#). The regulation is currently under review, to clarify some ambiguities in its implementation and expand its criteria. In turn, in 2022, CONICET's Directors Board adhered to the San Francisco DORA, publicly acknowledging its commitment to improve research by strengthening the evaluation and continuous improvement of its processes.

The [National Agency for the Promotion of Research, Technological Development and Innovation](#) (Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación – AGENCIA I+D+i), under the [Ministry of Science, Technology and Innovation](#), is the main research funder in the country due to the diversity and scope of its highly competitive calls. At present, AGENCIA is implementing a [programme](#) to strengthen research assessment processes in their main financial funds. Current improvements include the remuneration of peer reviewers to stimulate their commitment with these processes, incentive to open access as project outcomes should be destined to the public domain through publications or documents of open circulation (in accordance with the obligations of [‘Open Access Institutional Digital Repositories’ National Law 26.899](#)) and the incorporation of equity and inclusivity dimensions through gender, underrepresented generational groups and/or institutional strengthening equalizer mechanisms in research assessment processes. Nevertheless, in various disciplinary

95 [CLACSO is the Council for Social Sciences](#) in the region and a leading champion for socially relevant and responsible science. The [Latin American Forum on Research Assessment](#) (FOLEC) is a regional space for debate and sharing of good practice, and is developing [regional guidelines for research assessment](#) to support these principles. Both provide strong regional leadership.

committees, curricular background of leading researchers responsible for the proposals is still assessed by their peers with the use of citation impact indicators.

Finally, funded by a DORA Community Engagement Grant, the Faculty of Psychology at the Universidad Nacional de la Plata hosted a [virtual event](#) in September 2022 on assessment in psychology and social sciences that attracted over 640 (largely undergraduates) from 12 countries, demonstrating the interest of young people on the continent. The event has helped shape the faculty's four-year management plan and will inform a book on the reform of academic evaluation in the Spanish-speaking context.

National example: Brazil

Research evaluation is hotly debated in Brazil among research institutions and researchers, if not state and federal governments. However, despite the highest number of institutional signatories for DORA in the world, examples of research evaluation reform are surprisingly few. Following a survey of in-country DORA signatories, institutional consultations and a public event, funded by a DORA Community Engagement Grant, a [guide](#) has been prepared for university leaders to explore responsible evaluation practices.

The guide focuses on three main actions: (1) raising awareness of responsible evaluation in all its forms; (2) training and capacity building of evaluators and those being evaluated; and (3) implementation and appraisal. The next steps are to build a network of practitioners – or ten university intelligence offices – to effect change in evaluation practices and pilot context-sensitive models, and ultimately develop a roadmap for Brazilian institutions who wish to bring about change.

Projeto Métricas (2022). Institutional challenges and perspectives for responsible evaluation in Brazilian Higher Education: Projeto Métricas DORA partnership summary of findings. University of São Paulo, Brazil. <https://metricas.usp.br/institutional-challenges-and-perspectives-for-responsible-evaluation-in-brazilian-higher-education/>

3.2.3 North America

There is an ongoing shift away from purely quantitative indicators in North America, accelerated by the open science agenda. Open science and open review are helping to make evaluation practices more transparent, providing an opportunity for self-reflection and surfacing problems, e.g. self-citation and cronyism on hiring, promotion and peer review panels, as well as innate gender and other biases. Debates are ongoing about the need to develop smarter, intelligent indicators and mixed methods of evaluation, with potential for a hybrid, convergent model of evaluation that serves basic science (advancing knowledge) and applied science (societal impact).

There is also a recognition that universities need academic space and freedom to wean themselves off the tools they currently use for evaluation, without any 'first mover disadvantage', and that the user community should be part of the evaluation process to help measure the usability of knowledge, its uptake and impact. But there is also a contingent resistance to change (a 'wilful blindness') from the top and bottom of the research ecosystem – from those benefiting from the status quo and those who have recently entered it. Very few US universities have signed DORA and a new DORA project is endeavouring to understand why this is the case ([TARA](#)). Nevertheless, in both Canada and the USA there are some interesting examples of national and institutional initiatives designed to bring about systemic change (see the following text boxes).

National example: USA

In the US, the National Science Foundation is a leading voice for change through its [Advancing Research Impact in Society](#) programme and accompanying [broader impacts toolkit](#) for researchers and evaluators. Equity, diversity and inclusion, including engaging indigenous and traditionally marginalized communities, are key drivers. An IAP and ISC member, the US National Academies of Sciences is also looking to stimulate broad reform, providing a platform for information exchange and learning on reforming the traditional researcher CV ([NAS Strategic Council, 2022](#)). Borne out of the US academies' work, the [Higher Education Leadership Initiative for Open Scholarship](#) is a cohort of more than 60 colleges and universities committed to collective action to advance open scholarship, including rethinking research evaluation to reward openness and transparency.

The National Institute for Health, for example, designed a new [biosketch \(SciENcv\)](#) for personnel in grant applications to minimize systemic bias and reporting burden, and at the same time be more impacts-driven.

National example: Canada

In Canada, there are multiple conversations about the reform of research evaluation, driven by DORA; all three federal research councils are signatories. The Natural Sciences and Engineering Council has [redefined criteria for research quality](#)⁶², dispensing with bibliometrics, citations and the h-index, in line with DORA principles: quality metrics now include good research data and data access management, equity, diversity and inclusion, and training responsibilities. The other two research councils are likely to follow suit.

Canadian researchers tend to focus on 'knowledge mobilization', an intentional effort to advance the societal impact of research, through co-production with user communities ([ISI, 2022](#)⁶³). [Research Impact Canada](#)⁶⁴ is a network of over 20 universities that aims to build institutional capacity through impact literacy, or the ability to 'identify appropriate impact goals and indicators, critically appraise and optimize impact pathways, and reflect on skills needed to tailor approaches across contexts' in order to maximize the impact of research for the public good.

It is worth noting that very few Canadian universities have signed DORA. The main motivator to any change is likely to be embracing indigenous scholarship: this has become a moral imperative in Canada.

3.2.4 Africa

Research incentive and reward systems in Africa tend to reflect 'international', primarily Western, norms and conventions. African institutions endeavour to follow these when developing their approach to 'quality' and 'excellence' in research but they are not always appropriate for local knowledge and needs. Research 'quality', 'excellence' and 'impact' are not well-defined on the continent, and some researchers are not used to a culture of 'research impact'.

Evaluation systems in Africa tend not to account for research for societal benefit, teaching, capacity-building, research administration and management. Publication models are not context-sensitive, with APCs creating barriers to African research output. Reform of research evaluation systems could help correct asymmetries in the contribution African research can make to societal challenges, as well as improve access to resources to help the African research community do this. Breaking down barriers to cross-sectoral and cross-disciplinary cooperation is imperative to

enable a diversity of views and knowledge systems to thrive and help interpret what constitutes research quality for Africa. Mechanisms that integrate local, indigenous and ‘conventional’ world views about the assessment of research quality and excellence need to be considered in any reform.

Strong partnerships are being built around RRA on the continent. Funded by an international consortium of development agencies, the [Science Granting Councils Initiative \(SGCI\)](#)⁹⁶, engaging 17 African countries, conducted a study on research excellence in Africa, looking at science funding agencies and researcher evaluation from a Global South perspective ([Tijssen and Kraemer-Mbula, 2017](#)^{97,98}). It explored the issue of research excellence in sub-Saharan Africa and the need for an approach which expands the notion of excellence beyond publications ([Tijssen and Kraemer-Mbula, 2018](#)⁹⁹); producing a guidelines document, that is currently being updated, on good practices in implementing research competitions ([SGCI](#)¹⁰⁰). At the World Science Forum in 2022, under the auspices of the SGCI and the GRC, South Africa’s National Research Foundation (NRF) and Department of Science and Innovation convened international and local partners to discuss the role of funding agencies in advancing RRA, and to share experiences, advance good practice and evaluate progress in capacity-building and collaboration ([NRF, 2022](#)¹⁰¹).

The [African Evidence Network](#), a pan-African, cross-sectoral network of over 3,000 practitioners has conducted some work on assessment of transdisciplinary research ([African Evidence Network](#)¹⁰²) but the extent to which this has been embedded in national and regional assessment systems is not yet clear. The [Africa Research and Impacts Network](#)¹⁰³ has been working on a scorecard comprising a collection of indicators to evaluate the quality of science, technology and innovation (STI) assessment in Africa, which it hopes to develop into a web-based decision-making tool to guide STI investment decisions.

At the national level, incremental changes have begun – some examples are given in the following text boxes. Other countries where research funding agencies are taking the lead include Tanzania (COSTECH), Mozambique (FNI) and Burkina Faso (FONRID). The GRC’s RRA initiative is proving to be an important platform for change on the continent, as is learning from the international development sector, most notably, IDRC’s [Research Quality Plus \(RQ+\) Assessment Framework](#)¹⁰⁴, with the distinction that it has already been applied, studied and improved. The Africa-based [International Evaluation Academy](#)¹⁰⁵ may also provide an interesting opportunity.

96 SGCI. Science Granting Councils Initiative (SGCI) in Sub-Saharan Africa. <https://sgciafrica.org/>

97 SGCI. Tijssen, R. and Kraemer-Mbula, E. 2017. Policy brief: Perspectives on research excellence in the Global South – assessment, monitoring and evaluation in developing country contexts. SGCI. https://sgciafrica.org/wp-content/uploads/2022/03/Policy-Brief-Perspectives-on-research-excellence-in-the-Global-South_-_Assessment-monitoring-and-evaluation-in-developing-country-contexts.pdf

98 Tijssen, R. and Kraemer-Mbula, E. 2018. Research excellence in Africa: Policies, perceptions, and performance. SGCI. <https://sgciafrica.org/research-excellence-in-africa-policies-perceptions-and-performance/>

99 Tijssen, R. and Kraemer-Mbula, E. 2018. Research excellence in Africa: Policies, perceptions, and performance. *Science and Public Policy*, Vol. 45 No. 3, pp. 392–403. <https://doi.org/10.1093/scipol/scx074>

100 SGCI. Good Practice Guideline on the Quality of Research Competitions. <https://sgciafrica.org/eng-good-practice-guideline-on-the-quality-of-research-competitions/>

101 NRF. [The NRF Hosts Strategic Meetings to Advance Research Partnerships in Africa - National Research Foundation](#)

102 Belcher, B. M., Rasmussen, K. E., Kemshaw, M.R. and Zornes, D. A. 2016. Defining and assessing research quality in a transdisciplinary context, *Research Evaluation*, Vol. 25, pp: 1–17, <https://doi.org/10.1093/reseval/rvv025>

103 ARIN. 2020. [Science Technology and Innovation \(STI\) Metrics – Africa Research & Impact Network \(arin-africa.org\)](#)

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

105 [International Academy for Monitoring and Evaluation](#)

National example: Côte d'Ivoire

At the heart of Côte d'Ivoire's [Programme Appui Stratégique à la Recherche Scientifique \(PASRES\)](#) (Strategic Support Program for Scientific Research) is the belief that excellence in research must transcend the number of research publications and include the 'research uptake' dimension. Adapting to national context, the research evaluation process is based on criteria related to scientific and social relevance, the involvement of partners, student training, knowledge mobilization and feasibility. Evaluation panels involve scientific experts (to judge the quality of the research performed), the private sector (to judge economic enrichment) and other institutions (to measure the cultural and social potential of the research).

PASRES has established two local journals (one for social sciences and linguistics and the other for environment and biodiversity) and meets the entire cost of publishing in these. Finally, PASRES funds capacity-building activities and thematic conferences to enable researchers to present their research to the private sector and to civil society.

More information:

Ouattara, A. and Sangaré, Y. 2020. Supporting research in Côte d'Ivoire: processes for selecting and evaluating projects. E. Kraemer-Mbula, R. Tijssen, M. L. Wallace, R. McLean (Eds.), *African Minds*, pp. 138–146

[PASRES || Programme d'Appui Stratégique Recherche Scientifique \(csrs.ch\)](#)

National example: South Africa

Research evaluation in South Africa (SA) is predominantly focused around bibliometrics. Since 1986, when the Department of Higher Education (DHET) introduced a policy of paying subsidies to universities for research publications published in journals of accredited indexes, university research publication output grew concomitantly with the Rand value awarded per publication. In an effort to secure research funding and advance their careers, SA researchers published as many articles as quickly as they could, creating perverse and unintended consequences.

The Academy of Science of South Africa (ASSAf) commissioned a report on scholarly publishing in the country (2005–2014) and found indications of questionable editorial practices and predatory publishing (ASSAf, 2019). Using a nuanced system of categorization, an estimated figure of 3.4% of total articles over the past ten years were judged to be predatory, with figures rising more steeply from 2011. Journals judged to be predatory were included on the DHET 'acceptable for funding' list and academics in all SA universities were found to be involved (Mouton and Valentine, 2017).

The ASSAf report made recommendations at systemic, institutional and individual levels and ensuing countermeasures by the DHET, NRF and some universities appears to have curbed predatory practices in SA with the incidence of predatory publishing by SA academics (in DHET accredited journals) peaking in 2014–2015 and subsequently declining. There were also concerns among researchers that DHET policies in SA discouraged collaboration and failed to recognize the contribution of individuals within large research teams, requiring a revision of performance appraisal/research evaluation schemes. The use of the publication unit system is now recognized as a poor proxy for the assessment of research quality and productivity and for the selection and promotion of academics.

More information:

Academy of Science of South Africa (ASSAf). 2019. Twelve years: Second ASSAf Report on Research Publishing in and from South Africa. Pretoria, ASSAf.

Mouton, J. and Valentine, A. 2017. The extent of South African authored articles in predatory journals. South African Journal of Science, Vol. 113, No. 7/8, pp. 1–9.

Mouton, J. et al. 2019. The Quality of South Africa's Research Publications. Stellenbosch.

[2019_assaf_collaborative_research_report.pdf](#)

National example: Nigeria

Universities in Nigeria evaluate researchers in three main areas: teaching, research productivity and community services. Of these, research productivity is more heavily weighted, with an emphasis on published peer-reviewed research articles and taking into consideration the number and roles of authors (first authorship and/or corresponding authorship) in these publications. In an effort to become more globally competitive, most universities assign more importance to journals indexed by International Scientific Indexing or SCOPUS, to place more emphasis on quality and international collaboration; and use the percentage of articles in these journals as promotion criteria.

An unfortunate consequence of this is that many researchers, especially those in the humanities, lack adequate funding and/or capacity to publish in these journals. Instead, they publish more review rather than research articles, or they feel compelled to include influential, senior colleagues as co-authors, by virtue of their financial rather than intellectual contribution. Plagiarism rises, as does predatory publishing. However, the overall global ranking of Nigerian universities has increased, thereby satisfying the government and funding agencies, and being seen as a success. Nigeria is not alone in this regard.

The Nigerian Academy of Science has re-established its own peer-reviewed journal as a flagship journal in which academics can publish (currently for free) and be rated highly by their institutions.

3.2.5 Asia-Pacific

Highly competitive, quantitative metrics-driven assessment systems dominate the region, with Anglophone countries typically shaping assessment frameworks and other countries following suit. In Australia, for example, there is a competitive funding system based on bibliometrics and university rankings: 'even the SDGs are being turned into performance indicators'. Similar challenges exist in Malaysia and Thailand, and other ASEAN countries are likely to follow. An important exception is China where the government is playing a significant role in creating major systemic change and which could have profound implications globally (see text box).

Encouragingly, there is a growing awareness and concern amongst the research community in the region about the limits of current research evaluation systems and their threat to research integrity. Although ECRs, including National Young Academies and the ASEAN network of Young Scientists, together with grass roots movements, are increasingly engaging on this issue, they are struggling to be heard. Government and funding communities, including university leadership, are largely absent from the debate: they attach importance to quantitative metrics but do not appreciate the implications for research. Indeed, consultees report that more quantitative criteria are being added, to the extent that institutions and researchers are beginning to game the system, fuelling research misconduct.

But there are significant opportunities for change, as exemplified in the following text boxes.

National example: China

Now the most research-productive country in the world ([Tollefson, 2018](#); [Statista, 2019](#)), and the second in terms of research investment ([OECD, 2020](#)), what happens in China has the potential to effect real systemic change. A new state-level policy aims to restore ‘the scientific spirit, innovation quality and service contribution’ of research and to ‘promote the return of universities to their original academic aims’ ([MOST, 2020](#)). Web of Science indicators will no longer be a predominant factor to evaluation or funding decisions, nor will the number of publications and JIFs. Publications in high-quality Chinese journals will be encouraged and their development supported. ‘Representative publications’ – 5–10 choice papers rather than exhaustive lists – are being sought in evaluation panels, together with criteria that assess the contribution research has made to resolving important scientific questions, providing new scientific knowledge or introducing innovations to, and genuine advancements of, a particular field.

In developing a research quality and excellence assessment system more attuned to its own needs, China’s largest funding agency for basic research, the National Natural Science Foundation of China (NSFC), has carried out systematic reforms since 2018 to reflect shifts in science: changing global science landscapes, the importance of transdisciplinarity, the combination of applied and basic research and the interplay between research and innovation ([Manfred Horvat, 2018](#)), moving away from bibliometrics to a system that strengthens the local relevance of research in China ([Zhang and Sivertsen, 2020](#)). It has improved its peer review system for proposal evaluation to better fit curiosity-driven disruptive research, burning problems at the frontiers of research, excellent science applied to economic and social demands, and transdisciplinary research dealing with grand challenges. In 2021, 85% of proposals were submitted and reviewed using these categories. Recently, in November 2022, a two-year pilot reform plan for science and technology talent evaluation was announced, engaging eight ministries, twelve research institutes, nine universities and six local governments. Its objective will be to explore evaluation indicators and methods for science and technology talents engaged in different parts of the innovation system.

Subregional example: Australia and New Zealand

Both Australia and New Zealand are currently at important junctures. In Australia, ongoing concurrent reviews of the Australian Research Council, the Excellence in Research in Australia and Gold Open Access negotiations cumulatively present a window of opportunity ([Ross, 2022](#)).

Following public consultation on the future of science funding, New Zealand is developing a new [systemic programme](#) for the future of its national research and innovation system. Both Australia and New Zealand have contributed to the development of a metrics system for their indigenous research groups ([CARE Principles](#)).

National example: India

The Department of Science and Technology’s Centre for Policy Research (DST-CPR) has conducted recent studies on research assessment and its reform in India, leading workshops with key stakeholders (national funding agencies, research institutions and academies), interviews and surveys. It has found that, while universities and many institutions of national policy significance (like agriculture) focus almost exclusively on quantitative metrics, some funding agencies and institutions like Indian Institutes of Technology have been adopting more qualitative measures too. This more qualitative approach at top-tier institutions is already serving to divert more funding towards research on national priorities, although it is too early to say if it is having any quantifiable effect on research quality and impact.

The primary benchmark for assessment is peer review based on expert committee opinion, but only after initial screening of applications based wholly on quantitative metrics. Fundamental challenges also exist with these committees: the lack of diversity and understanding of open science practices, little consideration of societal impacts of research, and poor capacity and bias. These problems, and methodologies for assessment more generally, are poorly understood and there is a dearth of guidelines and literature on the subject.

Nevertheless, there is a growing awareness of the need to reform research assessment. Funded by a DORA community engagement grant, the Indian National Young Academy of Sciences partnered with the Indian Institute of Science (IISc) and DST-CPR to explore ways in which research assessment can be improved – their deliberations have been shared with key stakeholders with a view to stimulating a national conversation on the need to reform and ultimately change India’s research culture so that its research is more innovative and/or societally relevant. DST-CPR anticipates developing a framework for research excellence that could be integrated into its National Institutional Ranking Framework.

More information:

Battacharjee, S. 2022. [Does the Way India Evaluates Its Research Doing Its Job? – The Wire Science](#)

[DORA IdeasForAction.pdf \(dstcpriisc.org\)](#).

Suchiradipta, B. and Koley, M. 2022. Research Assessment in India: What Should Stay, What Could be Better? DST-CPR, IISc.

National example: Japan

Research evaluation protocols are highly devolved in Japan: while there are ‘National Guidelines for Evaluating R&D’, issued by the Cabinet Office’s Council for Science, Technology and Innovation, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and other ministries have also developed their own guidelines. On top of this, universities and research institutes have their own research evaluation systems in place for research and researchers, which – like in many parts of the world – have become linked to institutional performance and budgetary allocation.

There have been growing concerns about overreliance on quantitative evaluation. In response, the Science Council of Japan has prepared a [recommendation on the future of research evaluation in Japan](#) (2022) calling for less emphasis on quantitative and more on qualitative measures, more recognition of research diversity and responsibility in research evaluation and the monitoring of international trends in the reform of research evaluation practices. Ultimately, research interests and promotion should be at the heart of research evaluation, and every effort made to prevent fatigue, demotivation and excessive pressure on researchers.

A survey by MEXT on evaluation indicators found that the JIF is one of many indicators and, as such, has not had a strong impact in Japanese research, although this is discipline-dependent: for example, JIF use is higher in the medical sciences – and less-traditional research activities, such as open data, are less likely to be evaluated.

More information:

[RECOMMENDATION - Toward Research Evaluation for the Advancement of Science: Challenges and Prospects for Desirable Research Evaluation \(scj.go.jp\)](#)

In conclusion, there is a growing momentum for the reform of research evaluation in some regions, countries and institutions. Examples illustrated here include *nationwide* reforms, building consortia or coalitions of like-minded institutions seeking change, the targeting/steering of specific sectors and interventions to tackle perverse incentives and behaviours.

This is not yet a coherent and inclusive global conversation, nor are practices and insights necessarily shared openly. Some GYA, IAP and ISC members are already proactive in this space and opportunities could usefully be found to help them share their learning and good practice with each other and with the wider membership. The launch of the Global Observatory of Responsible Research Assessment (AGORRA) by the Research on Research Institute (RoRI) later in 2023 will provide a further platform for sharing learning, for the comparative analysis of national and international reform systems and to accelerate the two-way exchange and testing of good ideas across these systems.

4. CONCLUSIONS

This paper has set out the major drivers, opportunities and challenges for research evaluation reform and collated illustrative examples of change happening at global, regional, national and institutional levels. The purpose of this is to mobilize the GYA, IAP and ISC and their respective members, as important constituencies of the global research ecosystem.

Building on the past decade of scientific literature and advocacy work, there are five main conclusions.

1. The imperative to rethink the way in which research individuals, institutions and outputs are evaluated is clear and urgent. Maintaining research integrity and quality, maximizing diverse, inclusive and non-discriminatory science, and optimizing science for the global public good are major drivers, set in the context of a fast-changing world.
2. The way in which research is commissioned, funded, delivered and communicated is evolving at pace. Moves towards mission-oriented and transdisciplinary science, open science frameworks, evolving models of peer review, the use of AI and machine learning and the rapid rise of social media are changing traditional ways of doing and communicating research, requiring new thinking on research evaluation systems and the metrics and peer review processes underpinning it. More, and urgent, research is needed to future-proof these systems.
3. There is an imperative for more balanced research evaluation systems with both quantitative and qualitative indicators that value multiple forms of research output, processes and activities. However, stating that qualitative peer review processes are at least as important as bibliometrics is not straightforward and is further complicated by different parts of the world being at different stages in developing their assessment systems: in some, debates on research evaluation reform are quite advanced, in others they are nascent or absent.
4. A concerted and genuinely global and inclusive initiative is required to mobilize key stakeholder communities to develop and implement coherent ways of assessing and funding research; learning from each other and from other sectors (notably the research funders and development agencies). Collective, inclusive action towards transformative change will need to recognize interconnectedness rather than internationalization or universalization, i.e. be context-sensitive, cognizant of different challenges faced by different parts of the world and the rich heterogeneity of research ecosystem, *while at the same time* ensuring sufficient homogeneity to enable compatible research and funding systems and researcher mobility, to minimize divergence and fragmentation. A partial, exclusive conversation risks further biasing and disadvantaging those who have historically been excluded.
5. Change is required at all levels – global, regional, national and institutional – because metrics cascade through the whole research ecosystem and all these levels are interconnected. *All* stakeholders need to play their part as partners not adversaries – including funders, universities, university and research institute associations, intergovernmental organizations (IGOs), governments and government networks, academies, science policy makers, research and innovation managers and individual researchers. The GYA, IAP and ISC membership, collectively, covers a large part of this rich landscape (Figure 1, Appendix C).

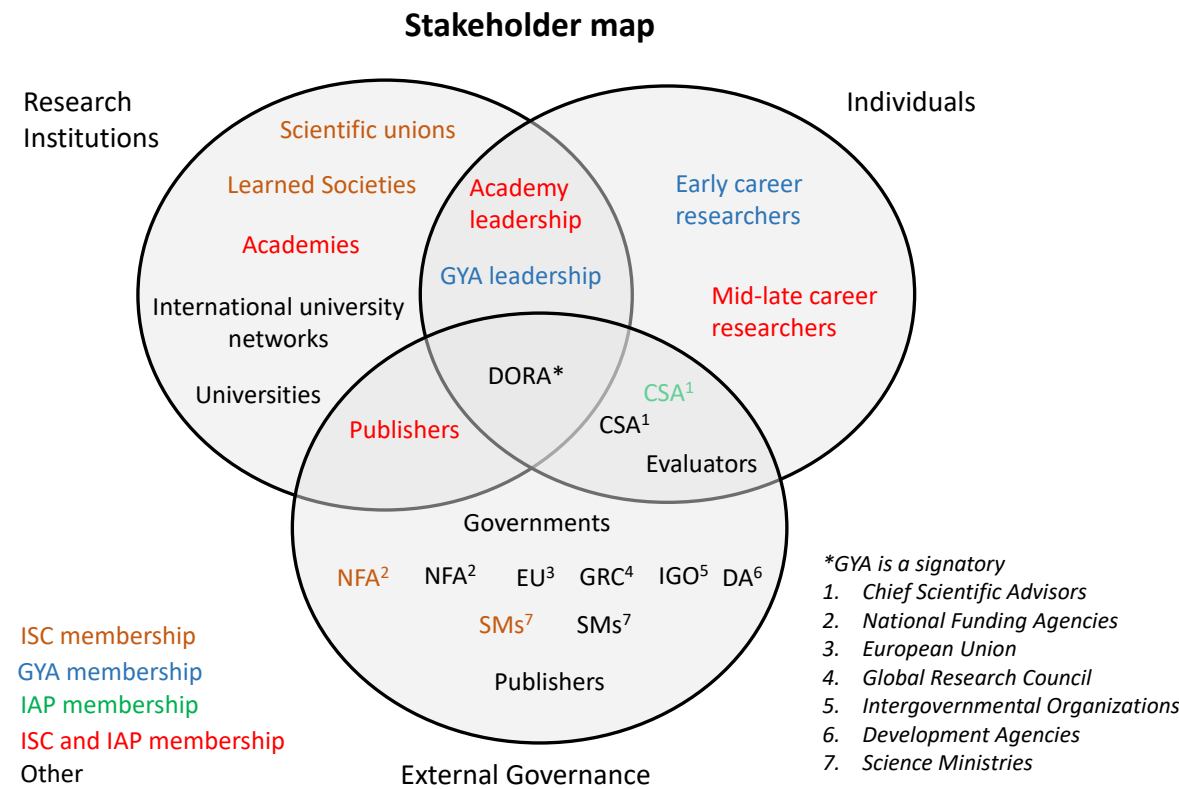


Figure 1: Stakeholder map relative to GYA, IAP and ISC membership

5. RECOMMENDATIONS FOR ACTION

The convening power of organizations like GYA, IAP and the ISC can help bring together a diversity of views and experiences across much of the research ecosystem: experimenting with, learning from, and building on, existing and new initiatives. Critically, they can **connect with the key stakeholders in instigating change – governments, research funders and universities, and vital global movements like DORA** – to help mobilize an architecture of actors. Collectively, they can serve as:

- **advocates** – raising awareness of research assessment debates, developments and reforms in recognition that their members serve as (i) mentors and supervisors of junior colleagues, (ii) leaders of HEIs, (iii) board members of funding and publishing governance bodies and (iv) advisers to policymakers;
- **innovators** – exploring different approaches to valuing basic and applied research in inclusive and innovative ways;
- **exemplars** – changing their own institutional culture – refreshing their membership, awards, publishing and conferencing practices, and leading by example;
- **evaluators** – capitalizing on the role of members at both institutional and individual levels whose business is to evaluate researchers, research and institutions, and those who have publishing, editorial and peer review roles;
- **funders** – drawing on the funding agencies represented in ISC, in particular, and members who manage and disperse large national and international grants;
- **collaborators** – supporting already established campaigns for reform, e.g. DORA, the EU's CoARA and UNESCO's open science commitment.

The authors of this paper encourage the GYA, IAP and the ISC, and organizations like them, to engage in the following ways:

ACTION 1: Share learning and good practice

This paper highlights examples of interventions and innovations from around the world. Space for sharing experiences and building a strong and inclusive 'coalition of the willing' is vital.

1.1: Provide a platform for members who are already proactive in this space to share their learning and build strategic connections, especially at the national level. Use these examples to help populate [DORA's dashboard](#)¹⁰⁶ of learning and good practice.

1.2: Survey and map member-led developments in research evaluation reform to identify institutional, national and regional approaches, and to find and share good practice. Convene those who have already led/engaged in major national and international initiatives to build advocacy and learning across the membership.

ACTION 2: Lead by example

GYA, IAP and ISC membership covers many parts of the research ecosystem, and each can play an important role in shaping what success as a scientist looks like.

2.1: Transition to more progressive research evaluation methodologies across the broad membership. Lead by example and help change the culture of research evaluation through their own membership philosophies and practices, while drawing on learning from

106 DORA. TARA Dashboard. <https://sfdora.org/tara-landing-page/>

DORA and the GRC. Academies, as traditionally elite organizations, have a particular role to play here – they should be encouraged to broaden their own criteria for election and selection to reflect broader and more plural understanding of research quality and impact, in order to reflect this pluralism (and with it more inclusion and diversity) in their membership.

2.2: Stimulate regional cooperation and leadership. Encourage the regional networks of GYA members and National Young Academies, IAP's regional academy networks and the ISC's Regional Focal Points to consider emulating the ALLEA Board's [initiative](#), tailored to their own contexts.

ACTION 3: Build strategic partnerships with key constituencies.

The three primary actors responsible for driving research evaluation reform are governments, research funders and universities. The GYA, IAP and ISC can each help bring the research community into their efforts to reform and bridge the disconnects which presently exist.

3.1: Engage with GRC leadership to explore ways of working together – essentially to stimulate members and their respective GRC national representatives to explore how their research communities can get involved.

3.2: Engage with global and regional networks of universities, such as the International Association of Universities (IAU), to develop new training tools for the research community; use HEI leadership within the collective membership of the GYA, IAP and ISC as advocates.

3.3: Connect member institutions in [DORA grant countries](#) (Argentina, Australia, Brazil, Colombia, India, Japan, Netherlands, Uganda and Venezuela) with DORA grant-leads to share ideas and potentially scale-up these local initiatives.

3.4: Build relations with leading international development agencies who are already deploying innovative and impactful strategies for research evaluation in low- and middle-income countries and least developed countries.

3.5: Work with UNESCO to help shape national research evaluation commitments under its [Recommendation on Open Science](#).

ACTION 4: Provide intellectual leadership on the future of research evaluation.

Focusing on specific and urgent challenges for research evaluation reform is imperative. The GYA, IAP and ISC, and international networks like them, can draw on their respective convening powers, the intellectual weight and influence of their members and connections with key constituencies.

4.1: Co-convene with key constituencies a series of multistakeholder discussion fora or 'Transformation Labs' to rethink and implement research evaluation reform – engage leaders of HEIs and their global (e.g. IAU and IARU) and regional networks (e.g. LERU and AAU¹⁰⁷), research funders (including GRC national representatives), international development agencies and leading publishers, among others. Raise new or deploy existing resources to fund this work (see Appendix D for some preliminary ideas).

107 IARU, International Association of Research-Intensive Universities; LERU, League of European Research Universities; AAU, African Association of Universities

4.2: Develop a novel study on an important aspect of the future development of research evaluation such as (1) the **impact of technological advances on research evaluation and peer review (including both use and abuse)**, and how these could evolve in future and (2) **reforming the peer review system** more broadly (in terms of its transparency, openness, capacity, recognition and training). Both issues are integral to the reliability of knowledge and trustworthiness of science.

At the heart of all these efforts should be three fundamental things:

- Expanding evaluation criteria for scientific research and researchers beyond traditional academic metrics to include multiple forms of research output and function, including quantitative criteria that can measure the social impact of research.
- Encouraging leaders of HEIs and research funders to adopt and foster these new evaluation criteria as measures of research quality and value.
- Working with these leaders on new forms of awareness-raising and training for future generations of researchers to equip them with the necessary skills to communicate and engage effectively with policymakers, publics and other key constituencies; and to foster diversity and inclusion in the research enterprise.

The authors of this paper conclude that networks like the GYA, IAP and ISC, together with and supporting other key constituencies, can help build a coherent, participatory, global initiative to mobilize research communities, universities and other HEIs around this agenda, and to consider how to operationalize new ways of assessing and funding research to make it more efficient, fair, inclusive and impactful.

APPENDIX A: SCOPING GROUP MEMBERS

Name	Position	Region
Sarah de Rijcke	Chair of Scoping Group and Director, Centre for S&T Studies (CWTS), Leiden University	Netherlands
Clemencia Cosentino	Director of Food and Agriculture Organization (FAO) Office of Evaluation (OED), and formerly Chief Evaluation Officer and Evaluation and Assessment Capability Section Head, US National Science Foundation	USA
Robin Crewe	Senior Research Fellow, Centre for the Advancement of Scholarship, University of Pretoria	South Africa
Carlo D'Ippoliti	Associate Professor of Political Economy, Department of Statistical Sciences, Sapienza University of Rome	Italy
Shaheen Motala-Timol	Acting Head, Regulatory Affairs and Accreditation Division, Higher Education Commission	Mauritius
Noorsaadah Binti A. Rahman	Professor, Deputy Vice-Chancellor (Research and Innovation, University of Malaya	Malaysia
Laura Rovelli	Coordinator of the Latin American Forum for Research Assessment (FOLEC-CLACSO) and member of the DORA Advisory Board, researcher at CONICET	Argentina
David Vaux	Laboratory Head; Deputy Director, Science Integrity and Ethics, WEHI Institute of Medical Research, University of Melbourne	Australia
YAO Yupeng	Director of Policy Department, National Natural Science Foundation of China (NSFC)	China

Additional support from [Dr Alex Rushforth](#), Assistant Professor at CWTS, Leiden University

Secretariat support:

- Tracey Elliott, International Science Council (ISC), Senior Consultant
- Anna-Maria Grammatte, Global Young Academy (GYA), Senior Project Officer
- Rania Kosti, InterAcademy Partnership (IAP), Executive Director (US Office)
- Sarah Moore, ISC, Director of Operations
- Beate Wagner, GYA, Managing Director

APPENDIX B: STAKEHOLDERS CONSULTED FOR THIS PAPER

Region	Experts	Affiliation/position
Global	<p>Geoffrey Boulton Masresha Fetene Peter Gluckman Peggy Hamburg Priscilla Mante Prosper Ngabonziza</p> <p>Haley Hazlett</p> <p>Hannah Flewitt</p>	<p>ISC Governing Board Member (UK) IAP Co-Chair (Ethiopia) ISC President (New Zealand) IAP Co-Chair (US) GYA Co-Chair (Ghana) GYA Co-Chair (US) Acting Program Director, Declaration on Research Assessment (DORA)</p> <p>IARU, International Projects Coordinator, IARU Secretariat</p>
Asia-Pacific	<p>Michael Barber</p> <p>Ginny Barbour</p> <p>Chai Lay Ching</p> <p>Fiona Fidler</p> <p>Moumita Koley</p> <p>John Mattick</p> <p>Orakanoke Phanraksa</p>	<p>Member of the Steering Committee for the International Science Council's project the Future of Scientific Publishing.</p> <p>Co-Lead, Office for Scholarly Communications, Queensland University of Technology; Director of Open Access Australasia; Past President COPE</p> <p>Associate Professor Biological Sciences, University of Malaya, Chairperson Young Scientists Network – Academy of Sciences Malaysia</p> <p>Professor School of BioSciences and School of Historical and Philosophical Studies, University of Melbourne; leader of the Interdisciplinary Meta-Research Group (IMeRG); Association for Interdisciplinary Meta-Research and Open Science. https://amos.community/)</p> <p>STI Policy Researcher, Department of Science and Technology's Centre for Policy Research (DST-CPR), Indian Institute of Science, India</p> <p>Professor RNA Biology, University of New South Wales, Australia</p> <p>Senior Intellectual Property Consultant at Technology Licensing Office, Technology Management Center, National Science and Technology Development Agency (NSTDA), Thailand, past Co-Chair of the Global Young Academy, Interim Co-Chair of the ASEAN Young Scientists Network</p>

China	GUO, Zhengtang	Professor, Director of Department of Earth Science, National Natural Science Foundation; Academician of Chinese Academy of Sciences; Former Vice-President, University of Chinese Academy of Sciences
	LI, Jinghai	President of the National Natural Science Foundation of China (NSFC)
	LI, Xiaoxuan	Professor, Institute of Policy and Management, Chinese Academy of Sciences; Director, Management Innovation and Evaluation Research Center, Chinese Academy of Sciences
	REN, Shengli	Professor; Executive Editor, Science China Press
Africa	Julie Makani	Wellcome Trust Research Fellow and Associate Professor, Muhimbili University of Health and Allied Sciences, Tanzania
	Zenda Ofir	Independent evaluator, South Africa
	Emmanuel Unuabonah	Fellow of the Nigerian Academy of Science and former President of Nigerian Young Academy
Europe	Priya Bondre-Beil	Internationale Zusammenarbeit, Europäische Wissenschaftspolitik, Deutsche Forschungsgemeinschaft (DFG)
	Stephen Curry	Chair of DORA; Professor of Structural Biology, Faculty of Natural Sciences, Department of Life Sciences, Imperial College London, UK; Assistant Provost for Equality, Diversity and Inclusion, Imperial College London
	Elisabeth (Lizzie) Gadd	Head of Research Operations, University of Glasgow, UK
	Karen Stroobants	Policy Adviser, Royal Society of Chemistry, UK
	James Wilsdon	Director, Research on Research Institute (RoRI) UK; Professor of Research Policy, UCL-STEaPP, Department of Science, Technology, Engineering and Public Policy (STEaPP), University College London
North America	Amy Brand	Director, The MIT Press, USA
	Thomas Ferguson	Research Director, Institute for New Economic Thinking, USA
	David Phipps	Assistant Vice-President, Research Strategy & Impact, Office of the Vice-President Research & Innovation at York University and director of Research Impact Canada (RIC)

Latin America	Graciela Díaz de Delgado	IUCr Executive Committee, Universidad de Los Andes
	Judith Sutz	Researcher, University of the Republic, Uruguay; Coordinator of the Academic Unit of the Scientific Research Sector Commission (CSIC); member of the National Academy of Sciences of Uruguay
	Hebe Vessuri	Centro de Investigaciones en Geografía Ambiental (CIGA-UNAM , Mexico)

APPENDIX C: GYA, IAP AND ISC MEMBERSHIP AND MISSIONS

PARTNER	MEMBERSHIP	MISSION STATEMENTS
GYA and the National Young Academies	<p>200 young scientists from all disciplines, typically 3–10 years after their PhD, aged 30–40 years and in the early stages of their independent academic careers. Selected for their scientific excellence and commitment to engage with society, they serve five-year terms before joining an alumnus of 327 alumni (as of 2021). Members and alumni together hail from 94 countries. The GYA is also a conduit to (presently) 54 National Young Academies plus additional young researcher networks.</p> <p>The GYA is a full member of the IAP and affiliate member of the ISC, and has numerous partnerships with other research, funding, publishing and development agencies. It also sits on the Advisory Board of the UN Major Group for Children and Youth.</p>	<p>To promote science for all, science for the future and giving a voice to young scientists around the world.</p> <p>To lead international, interdisciplinary and intergenerational dialogue for evidence-informed and inclusive global, regional and national decision-making.</p> <p>Four thematic areas: Science and Society; GYA & the SDGs; Research Environment; and Science Education & Outreach.</p>
IAP	<p>150 national, regional and global member academies who represent leading scientists, engineers and health professionals in over 100 countries. Academicians are nominated by their peers, based on their demonstrable excellence in science. IAP has four regional networks – in Africa, Asia, the Americas and Europe.</p> <p>Academies can apply for IAP membership if they meet the eligibility criteria.</p> <p>The IAP is an observer organization to the Intergovernmental Panel on Climate Change, the Belmont Forum and the ISC, and works closely with the WHO.</p>	<p>To play a vital role in ensuring that science serves society inclusively and equitably and underpins global sustainable development.</p> <p>To work collaboratively to address issues of global, regional and national importance: to advance sound policies, improve public health, promote excellence in science education and achieve other critical development goals.</p>

ISC	<p>An institutionally diverse network of over 200 international scientific unions and associations, as well as national and regional scientific organizations including academies, research institutes and foundations. Its members cover both natural and social sciences. ISC has an office in New York, Colombia (to cover LAC), and (soon) Australia (to cover Asia-Pacific) and has plans to set up an office in Africa.</p> <p>Institutions can apply to become ISC members if they meet the eligibility criteria, adhere to ISC principles and values, and pay a membership fee. The GYA is a full (non-paying) member and the IAP an observer of ISC.</p> <p>The ISC is the co-convenor for the S&T Community Major Group, representing the global science community in the UN system, and as such has privileged access to UN meetings. It has an MOU with the WHO and works closely with UNEP and UNDP.</p>	<p>To act as the global voice for science in the UN system.</p> <p>Three principal areas of work:</p> <ul style="list-style-type: none"> (i) wide-ranging science-for-policy issues: to stimulate, support and communicate policy-relevant science; (ii) policy-for-science issues: to enable science to contribute more effectively to the global public good; and (iii) scientific freedom and responsibility to defend the free and responsible practice of science. <p>ISC leads a broad range of co-sponsored international research programmes, networks and committees, well-being and disaster risk.</p>
-----	---	---

APPENDIX D: LIST OF ACRONYMS

ACRONYM	MEANING
AAU	African Association of Universities
AGENCIA I+D+i	Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación (Argentina)
AGORRA	A Global Observatory of Responsible Research Assessment
AI	Artificial intelligence
AIS	Article Influence Score
ALLEA	All European Academies (the European Federation of Academies of Sciences and Humanities)
APC	Author processing charge
ASSAf	Academy of Science of South Africa
ChatGPT	Chat Generative Pre-Trained Transformer
CLACSO-FOLEC	The Latin American Council of Social Sciences – Latin American Forum on Research Assessment
CoARA	Coalition on Reforming Research Assessment
CONICET	Consejo Nacional de Investigaciones Científicas y Técnicas (Argentina)
DOI	Digital Object Identifier
DORA	Declaration on Research Assessment
DST-CPR	Department of Science and Technology - Centre for Policy Research (India)
ECR	Early career researcher
ERC	European Research Council
GraspOS	Open science assessment dataspace
GRC	Global Research Council
GYA	Global Young Academy
HEI	Higher education institution
IAP	InterAcademy Partnership
IARU	International Alliance of Research Universities

IAU	International Association of Universities
IDRC	International Development Research Centre (Canada)
IISc	Indian Institute of Science
INORMS	International Network of Research Management Societies
ISC	International Science Council
JIF	Journal Impact Factor
KNAW	Royal Netherlands Academy of Arts and Sciences
LAC	Latin America and the Caribbean
LERU	League of Research-Intensive Universities
MEXT	Ministry of Education, Culture, Sports, Science and Technology (Japan)
NSFC	National Natural Science Foundation of China
NGO	Nongovernment Organization
NOR-CAM	Norwegian Career Assessment Matrix
NRF	National Research Foundation (South Africa)
OPUS	Open and Universal Science
ORCID	Open Researcher and Contributor ID
PASRES	Programme Appui Stratégique à la Recherche Scientifique (Côte d'Ivoire)
REF	Research Evaluation Framework
R&D	Research and development
RoRI	Research on Research Institute
RRA	Responsible Research Assessment
SA	South Africa
SGCI	Science Granting Councils Initiative
SCOPE	Start–Context–Options–Probe–Evaluate
SDGs	Sustainable Development Goals
STI	Science, technology and innovation

TARA	Tools to Advance Research Assessment
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization

