RoRI RESEARCH ON RESEARCH INSTITUTE

Metrics, merit & maximising impact:
where next for responsible research evaluation?

James Wilsdon, RoRI & University College London

Science for Society, Masaryk University, 31 May 2023

@RoRInstitute @jameswilsdon

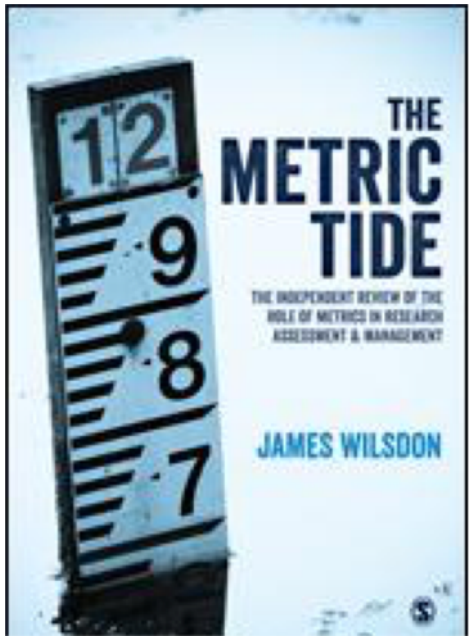
HARNESSING THE METRIC TIDE:
indicators, infrastructures and priorities for responsible research assessment in the UK

Stephen Curry, Elizabeth Gadd and James Wilsdon

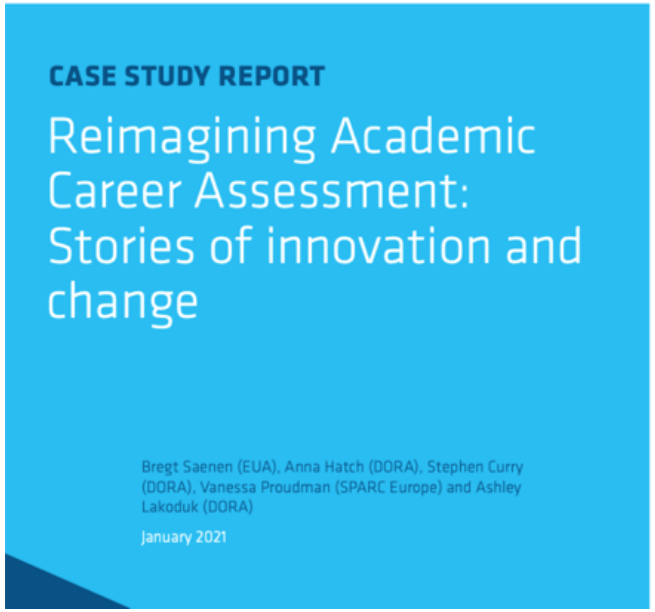
Report of *The Metric Tide Revisited* panel
December 2022

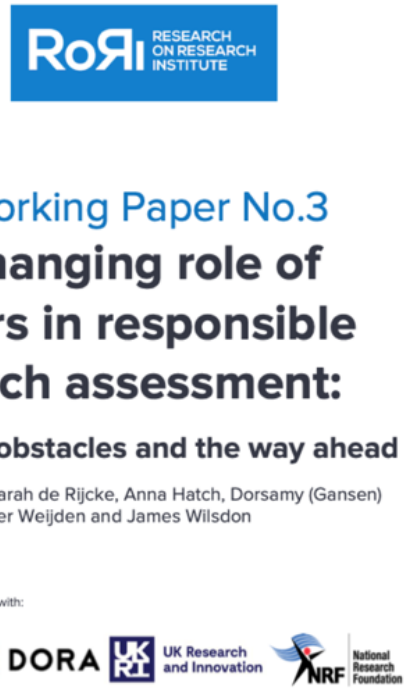
What I'll cover:

- The rise of responsible research assessment & evaluation
- Experiments in RRA: some interim results
- Some insights from the UK's REF (Research Excellence Framework)
- 8 priorities for the next decade of RRA in Czechia & beyond



From responsible metrics….





...to responsible research assessment

Defining RRA

Responsible research assessment (RRA) is an umbrella term for approaches to assessment which incentivise, reflect and reward the plural characteristics of high-quality research, in support of diverse and inclusive research cultures.

RRA draws on broader frameworks for responsible research and innovation (RRI) and applies these to the development and application of evaluation, assessment and review processes.

While RRI is commonly used as a broad framework for the governance of research and innovation, and notions of 'responsible metrics' can be applied at a micro level to indicators themselves, the idea of RRA encourages funders, research institutions, publishers and others to focus attention on the methodologies, systems and cultures of research assessment.

What's the problem?

Concern has intensified over several long-standing problems linked to research evaluation:

- the **misapplication of narrow criteria and indicators of research quality or impact**, in ways that distort incentives, create unsustainable pressures on researchers, & exacerbate problems with research integrity & reproducibility.

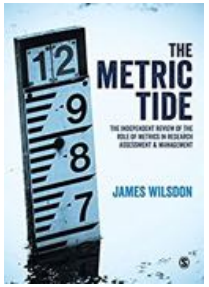
- this narrowing of criteria and indicators has **reduced the diversity of research missions and purposes**, leading institutions and researchers to adopt similar strategic priorities, or to focus on lower-risk, incremental work.

- **systemic biases against those who do not meet—or choose not to prioritise—narrow criteria and indicators** of quality or impact, have reduced the diversity, vitality and representative legitimacy of the research community.

- a **diversion of policy & managerial attention to things that can be measured**, at the expense of less tangible or quantifiable qualities, impacts, assets and values – a trend exacerbated by flawed university league tables.

RRA movers and shapers

DORA



HuMetricsHSS
HUMANE METRICS INITIATIVE

inorms

SCIENCE GRANTING COUNCILS INITIATIVE SGCI IOSRS INITIATIVE DES ORGANISMES SUBVENTIONNAIRES DE LA RECHERCHE SCIENTIFIQUE

GLOBAL YOUNG ACADEMY

Helsinki Initiative on Multilingualism
Hi!

FOLEC Latin American Forum for Research Assessment
CLACSO

SCIENCE EUROPE
Shaping the future of research

eua EUROPEAN UNIVERSITY ASSOCIATION

wellcome

CoARA

About Agreement Coalition News Resources Contact Sign

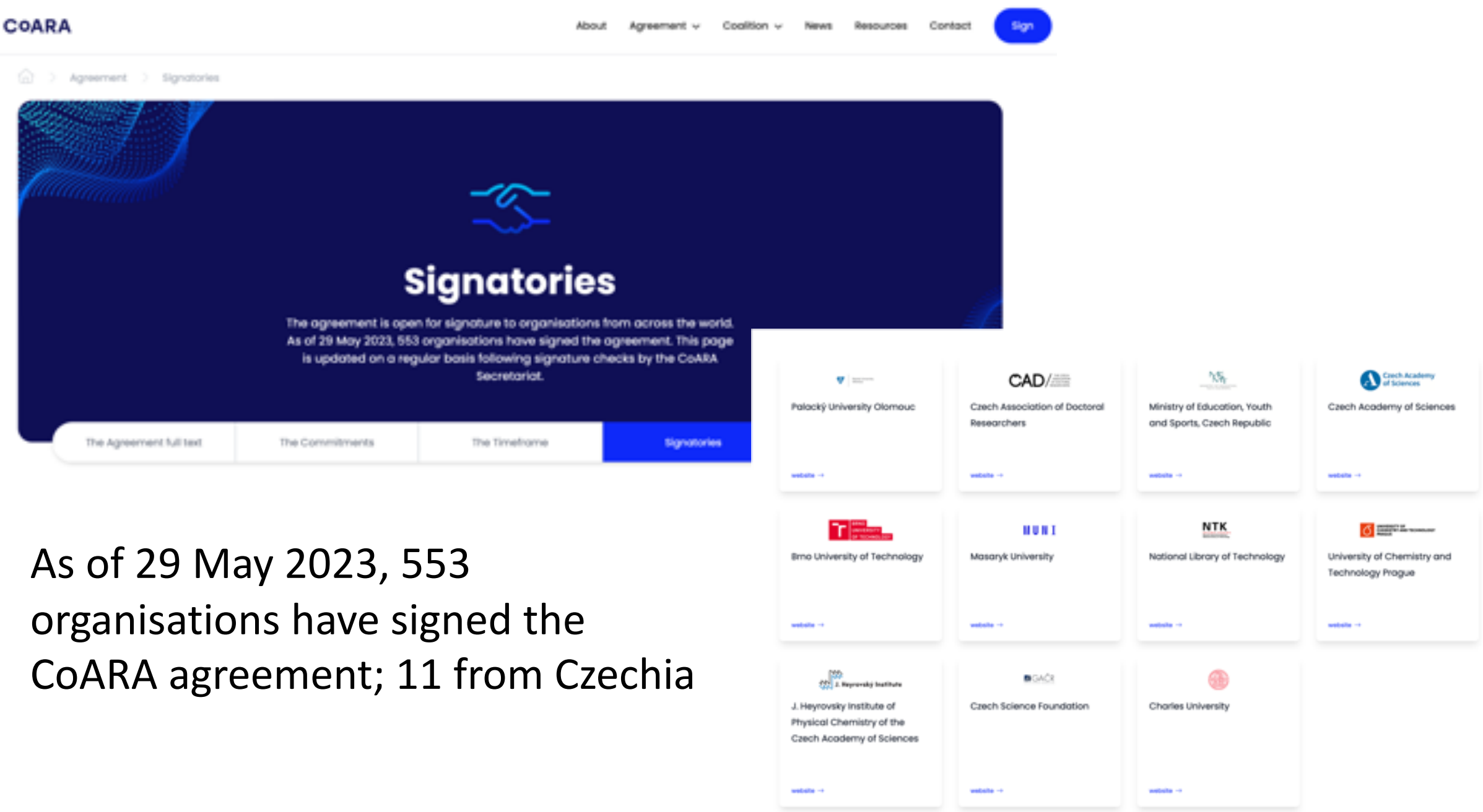
Coalition for Advancing Research Assessment

Our vision is that the assessment of research, researchers and research organisations recognises the diverse outputs, practices and activities that maximise the quality and impact of research. This requires basing assessment primarily on qualitative judgement, for which peer review is central, supported by responsible use of quantitative indicators.

The Agreement

Based on 10 commitments, establishes a common direction for research assessment reform, while respecting organisations' autonomy. The Agreement on Reforming Research Assessment sets a shared direction for changes in assessment practices for research, researchers and research performing organisations, with the overarching goal to maximise the quality and impact of research.

RoRI RESEARCH ON RESEARCH INSTITUTE



As of 29 May 2023, 553 organisations have signed the CoARA agreement; 11 from Czechia

OTHER

The Future of Research Evaluation: A Synthesis of Current Debates and Developments

May 2023

Policy for Science

LEADING AUTHORING INSTITUTION

The Interacademy Partnership (IAP)

Download PDF

Sample Web

SHARE ON

A question widely debated by stakeholders around the world is whether current research evaluation systems are effective in identifying high-quality research and in supporting the advancement of science. Over recent years, concerns have risen about the limitations and potential biases of traditional evaluation metrics which often fail to capture the full range of research impact and quality. Consequently there has been an increased demand by

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.

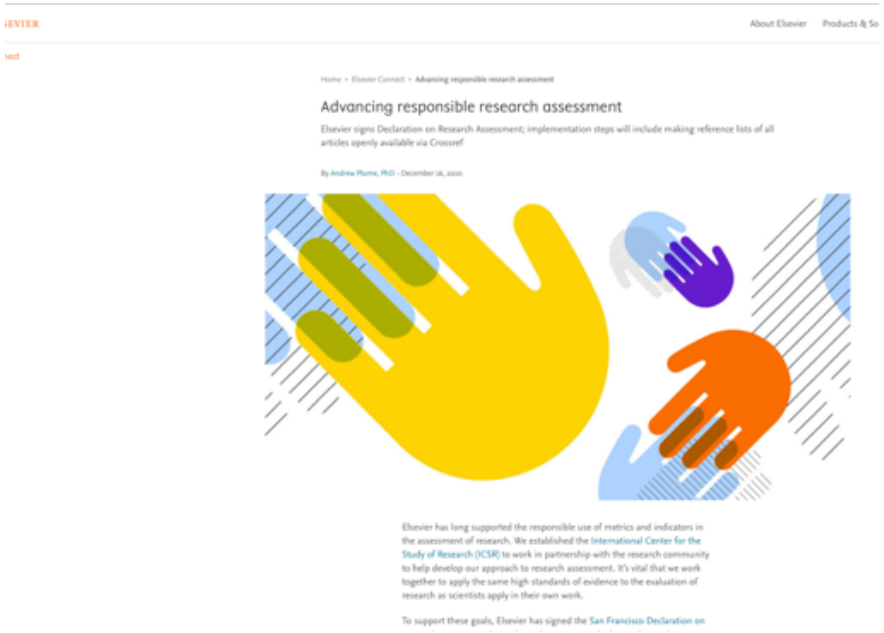
4. A concerted and genuinely global and inclusive initiative is required to mobilize key stakeholder communities to develop and implement coherent ways of assessing and funding research; learning from each other and from other sectors (notably the research funders and development agencies). Collective, 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.

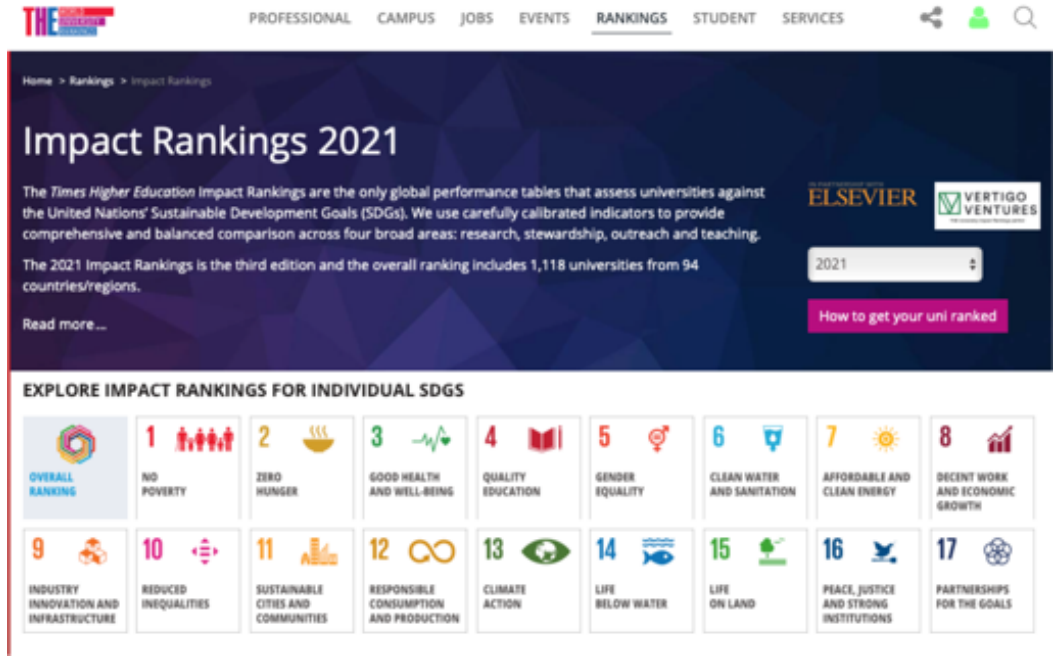
5. Change is required at all levels – global, regional, national and institutional – because metrics cascade through the whole research ecosystem and all these levels are interconnected. *All* stakeholders need to play their part as partners not adversaries – including funders, universities, university 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).


CHEERFUL
WHISTLING
PERMITTED

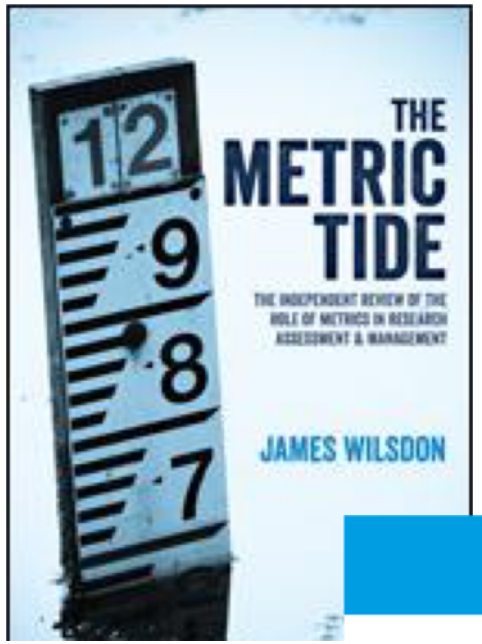
Experiments in RRA: some interim results

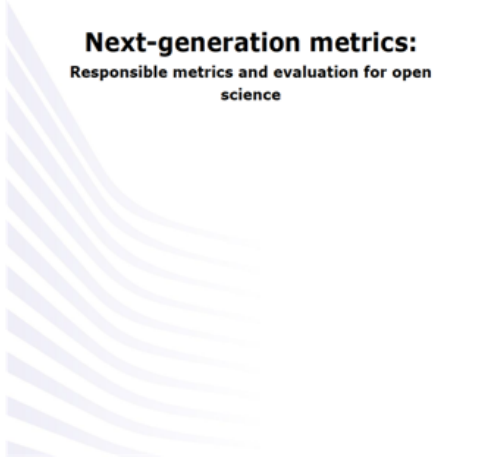
- Cosmetic appropriation
- Calibrating the machine
- Advocacy coalitions
- Institutional culture change
- System change..?





Cosmetic appropriation?





Calibrating the machine

RECOMMENDATIONS from Next-Generation Metrics (2017)

#1: Ahead of the launch of its ninth research framework programme (FP9), the EC should provide clear guidelines for the responsible use of metrics in support of open science.

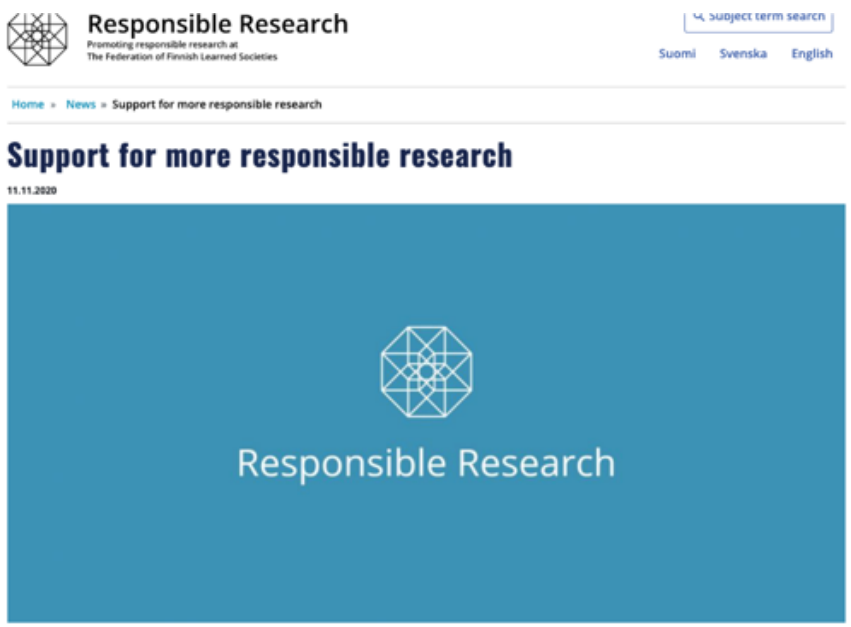
#2: The EC should encourage the development of new indicators, and assess the suitability of existing ones, to measure and support the development of open science.

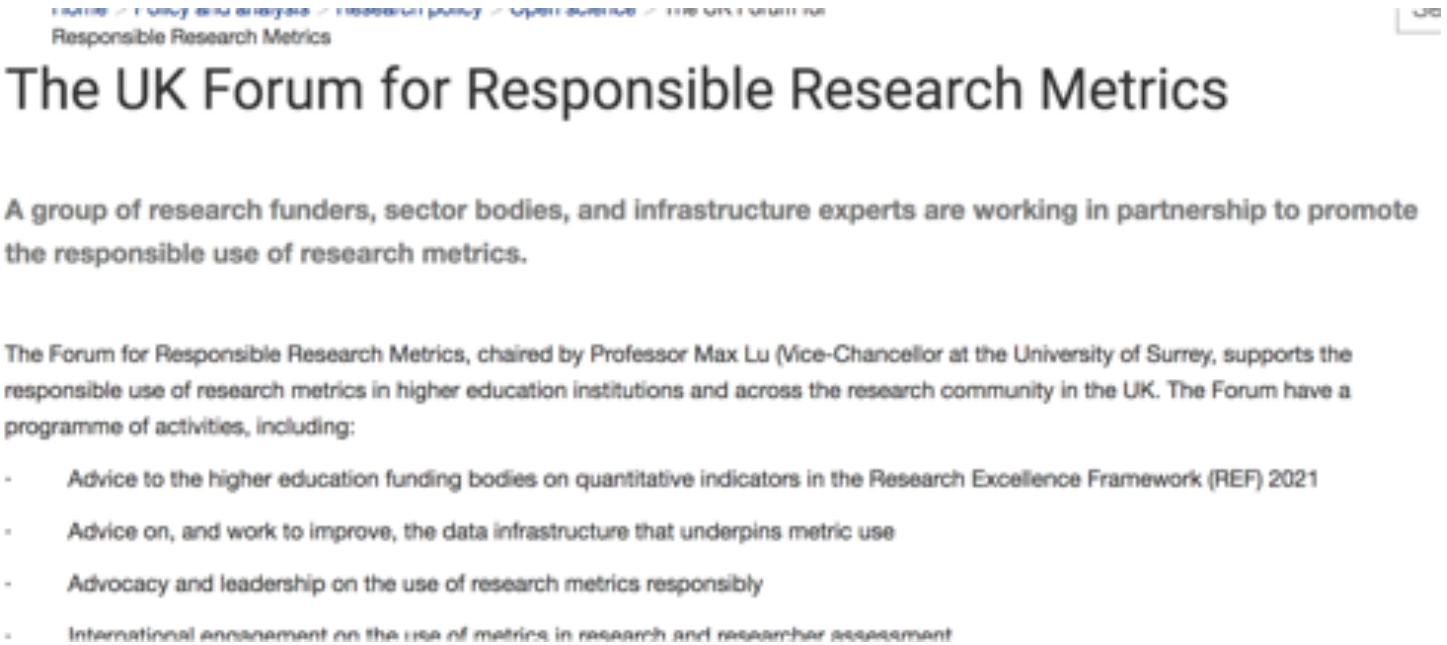
#3: Before introducing new metrics into evaluation criteria, the EC needs to assess the likely benefits and consequences as part of a programme of 'meta-research'.

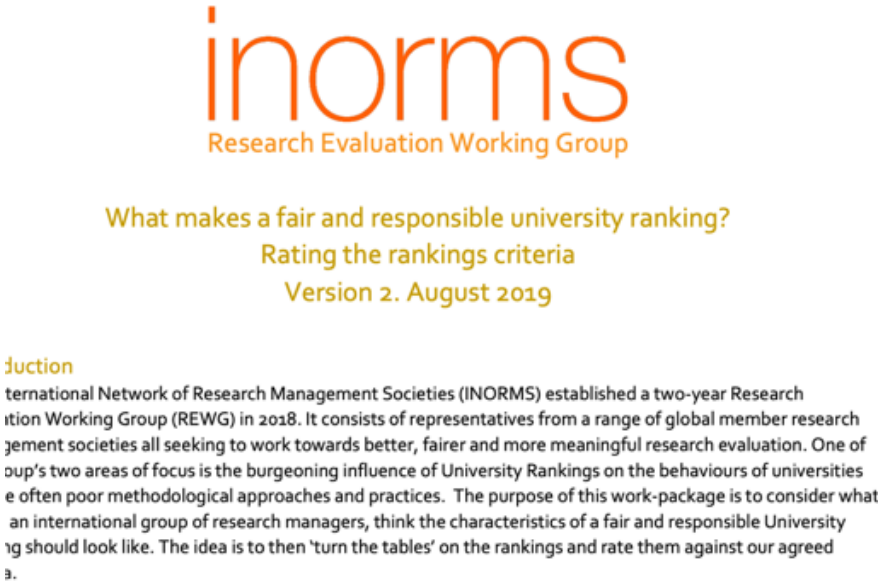
#4: The adoption and implementation of open science principles and practices should be recognised and rewarded through the European research system

#5: The EC should highlight how the inappropriate use of indicators (whether conventional or altmetrics or next generation metrics) can impede progress towards open science.

##10: The EC should identify mechanisms for promoting best practices, frameworks and standards for responsible use of metrics in support of open science

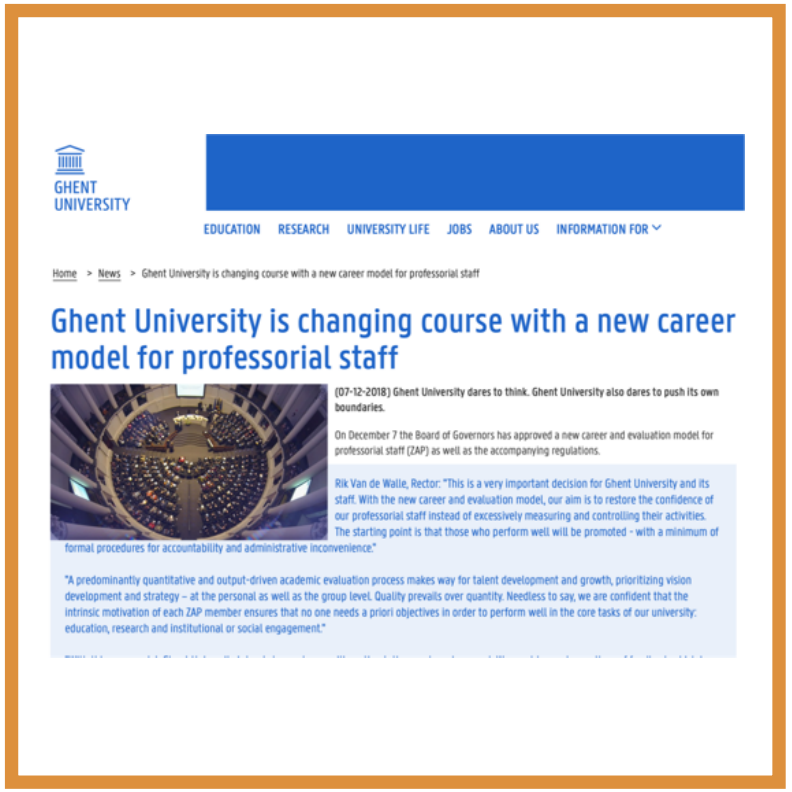






Advocacy coalitions

Institutional culture change


GHENT UNIVERSITY

EDUCATION RESEARCH UNIVERSITY LIFE JOBS ABOUT US INFORMATION FOR

Home > News > Ghent University is changing course with a new career model for professorial staff

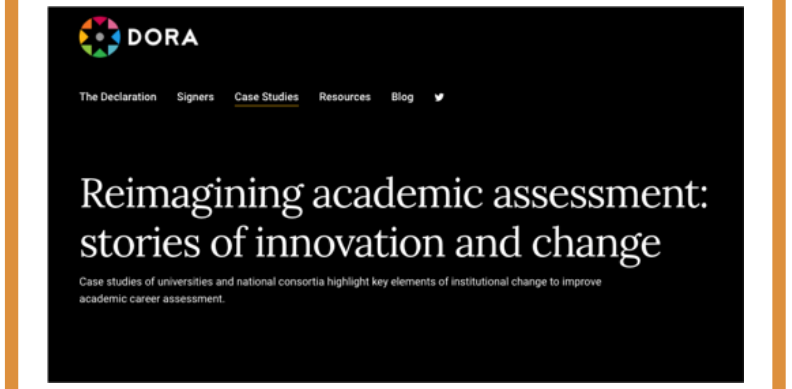
Ghent University is changing course with a new career model for professorial staff

(07-12-2018) Ghent University dares to think. Ghent University also dares to push its own boundaries.

On December 7 the Board of Governors has approved a new career and evaluation model for professorial staff (ZAP) as well as the accompanying regulations.

Rik Van de Walle, Rector: "This is a very important decision for Ghent University and its staff. With the new career and evaluation model, our aim is to restore the confidence of our professorial staff instead of excessively measuring and controlling their activities. The starting point is that those who perform well will be promoted - with a minimum of formal procedures for accountability and administrative inconvenience."

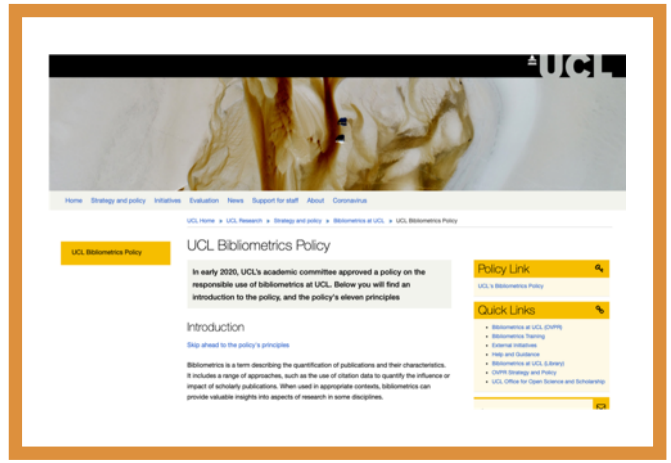
"A predominantly quantitative and output-driven academic evaluation process makes way for talent development and growth, prioritizing vision development and strategy – at the personal as well as the group level. Quality prevails over quantity. Needless to say, we are confident that the intrinsic motivation of each ZAP member ensures that no one needs a priori objectives in order to perform well in the core tasks of our university: education, research and institutional or social engagement."


DORA

The Declaration Signers Case Studies Resources Blog

Reimagining academic assessment: stories of innovation and change

Case studies of universities and national consortia highlight key elements of institutional change to improve academic career assessment.


UCL Bibliometrics Policy

In early 2020, UCL's academic committee approved a policy on the responsible use of bibliometrics at UCL. Below you will find an introduction to the policy, and the policy's eleven principles

Introduction

Skip ahead to the policy's principles

Bibliometrics is a term describing the quantification of publications and their characteristics. It includes a range of approaches, such as the use of citation data to quantify the influence or impact of scholarly publications. When used in appropriate contexts, bibliometrics can provide valuable insights into aspects of research in some disciplines.

Policy Link

Quick Links


44. Research England encourages providers to support the principles of open research in their research environment. Most Research England funding is deployed by universities at their discretion and is not intended to lead to specified outputs. In such cases, outputs cannot be attributed directly to Research England funding and no acknowledgement of Research England funding is expected or necessary. Such outputs are therefore out of scope of the UKRI Open Access policy. Where funding is given for particular purposes, and where that funding leads directly to particular research outputs, those outputs will be subject to the UKRI Open Access policy and providers will be required to include acknowledgement of Research England's funding.

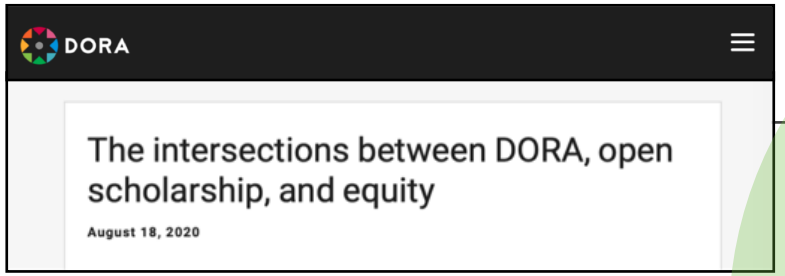
Responsible research assessment

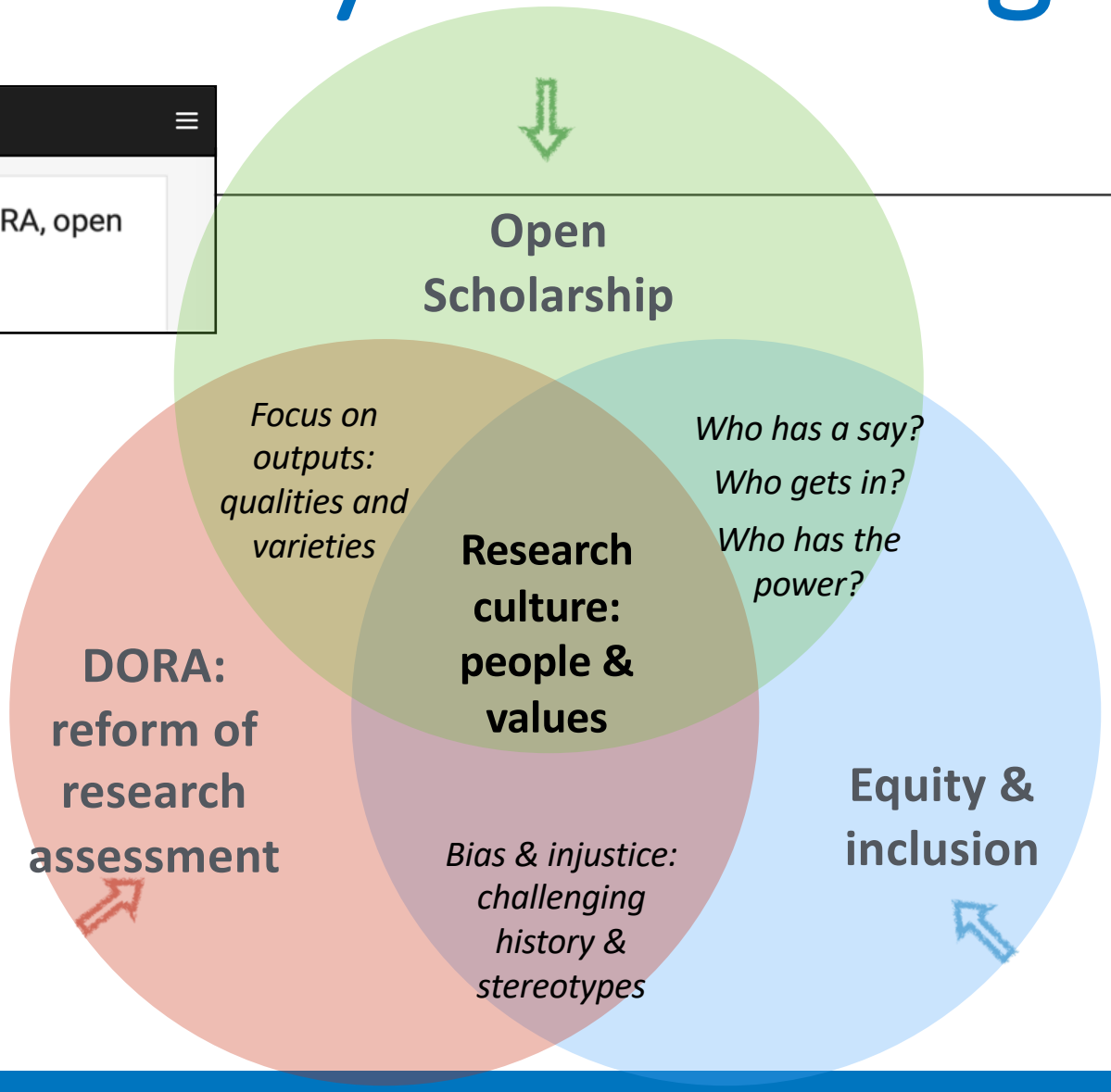
45. Our expectation is the providers we fund will comply with the principles of the San Francisco Declaration on Research Assessment (DORA)[8], Leiden Manifesto[9] or equivalent. Research England commits to assessing the intrinsic merit of research and will not consider the publication channel, its impact factor (or other journal metrics), or the publisher when assessing quality.

Equality, diversity and inclusion

46. We expect higher education providers to ensure that equality, diversity and inclusion is considered and supported in the use of our funding, taking into account UK Research and Innovation policies and principles[10] for equality, diversity and inclusion. Providers' approaches to supporting equality, diversity and inclusion are expected to exceed all relevant legal obligations, including but not limited to those of the Equality Act 2010.

Culture and system change





Insights from the UK's Research Excellence Framework (REF)

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

Future Research Assessment Programme

This information is hosted by Jisc on behalf of the four UK higher education funding bodies.

About the programme

The Future Research Assessment Programme aims to explore possible approaches to the assessment of UK higher education research performance. It has been initiated at the request of the UK and devolved government ministers and funding bodies. This significant piece of work will be led by the four UK higher education funding bodies:

- Research England
- Scottish Funding Council
- Higher Education Funding Council for Wales

What the FRAP happens next? Four priorities for reforming the REF

The next exercise should clarify its purpose and language, relax its disciplinary focus and refine research culture, says James Wilsdon

May 26, 2022

James Wilsdon

Twitter: @jameswilsdon

Source: Getty (edited)

There is, it seems, no rest between Research Excellence Frameworks.

Barely 72 hours after the release of the REF 2021 results, the first email landed. Sent on behalf of an anonymous university working group "set up to look specifically at data capture for the next REF cycle", it linked me to an Excel spreadsheet. This contained 27 columns, each with a detailed question about research collaborations, talks and

Harnessing the Metric Tide

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

STEPHEN CURRY, ELIZABETH GADD AND JAMES WILSDON

DECEMBER 2022

The Brief

- Revisit the conclusions of The Metric Tide (2015), and assess progress on its recommendations
- Consider the potential of infrastructure, methodological and metric developments since 2015
- Look afresh at the possible roles of metrics in any future REF
- Advise on how best to support uptake of responsible metrics and responsible research assessment

Harnessing the Metric Tide: 10 headline recommendations

1: Put principles into practice.

2: Evaluate with the evaluated.

3: Redefine responsible metrics.

4: Revitalise the UK Forum.

5: Avoid all-metric approaches to REF.

6: Reform the REF over two cycles.

7: Simplify the purposes of REF.

8: Enhance environment statements.

9: Use data for good.

10: Rethink university rankings.

1: Put principles into practice

- UK stakeholders should participate in the growing global movement to **implement** responsible research assessment (RRA).
- We strongly encourage participation in the recently formed Coalition for Advancing Research Assessment **(CoARA)**.


THE HONG KONG PRINCIPLES
FOR ASSESSING RESEARCHERS

FOSTERING RESEARCH INTEGRITY

What are the HKP?
The Hong Kong Principles (HKP) were developed as part of the 6th World Conference on Research Integrity. They were developed to reinforce the need to ensure that researchers are rewarded for specific behaviors that promote trustworthy research. The HKP have been developed with the idea that implementation of them could assist in how researchers are assessed for career advancement with a view to strengthen research integrity.

PRINCIPLE | IMPLEMENTATION EXAMPLES

1 Assess responsible research practices.
2 Value complete reporting.
3 Reward the practice of open science.
4 Acknowledge a broad range of research activities.
5 Recognize essential other tasks such as peer reviewing and mentoring.


nature

Explore content | About the journal | Publish with us | Subscribe

nature › editorials › article

EDITORIAL | 27 July 2022

Support Europe's bold vision for responsible research assessment

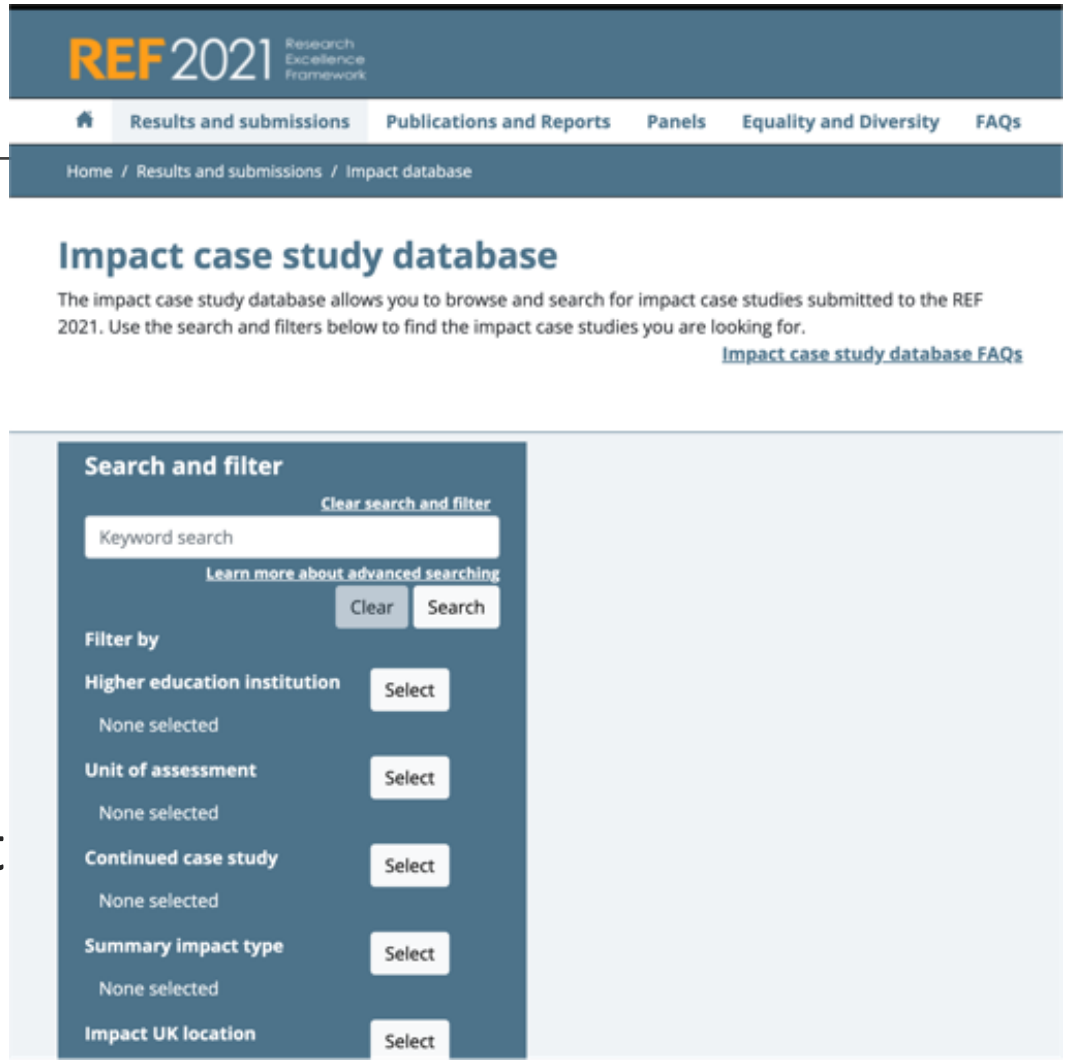
There have been many initiatives to combat the distorting effect of research assessment exercises. The latest looks like it might work

2: Evaluate with the evaluated

Stakeholders involved in research evaluation should enable and incentivise the **co-design and co-interpretation** of research assessments with research-active and research-enabling staff.

5: Avoid all-metric approaches

- It is **unlikely that an all-metric approach will deliver** what stakeholders need from REF.
- …particularly with regards to the assessment of **research impacts.**
- Metrics might be more viable, in combination with qualitative modes of assessment, at **higher levels** of assessment in future cycles of the REF.



7: Simplify the purposes of REF

- UK research funding bodies should **agree on a simplified statement of REF purposes.**
- We propose **renaming the REF**—for example as the 'Research Qualities Framework (RQF)'—in order to **replace the contested and ill-defined term 'excellence'**.

8: Enhance focus on research environment

- There should be **greater weight** overall on research environments.
- Statements should reflect additional dimensions of **research culture**, and to draw responsibly on data, indicators and other evidence.
- Replace ‘environment statements’ with '**people and culture statements'** to capture important aspects of research activity that can be assessed in a size-independent way.

REF 2021 Research Excellence Framework

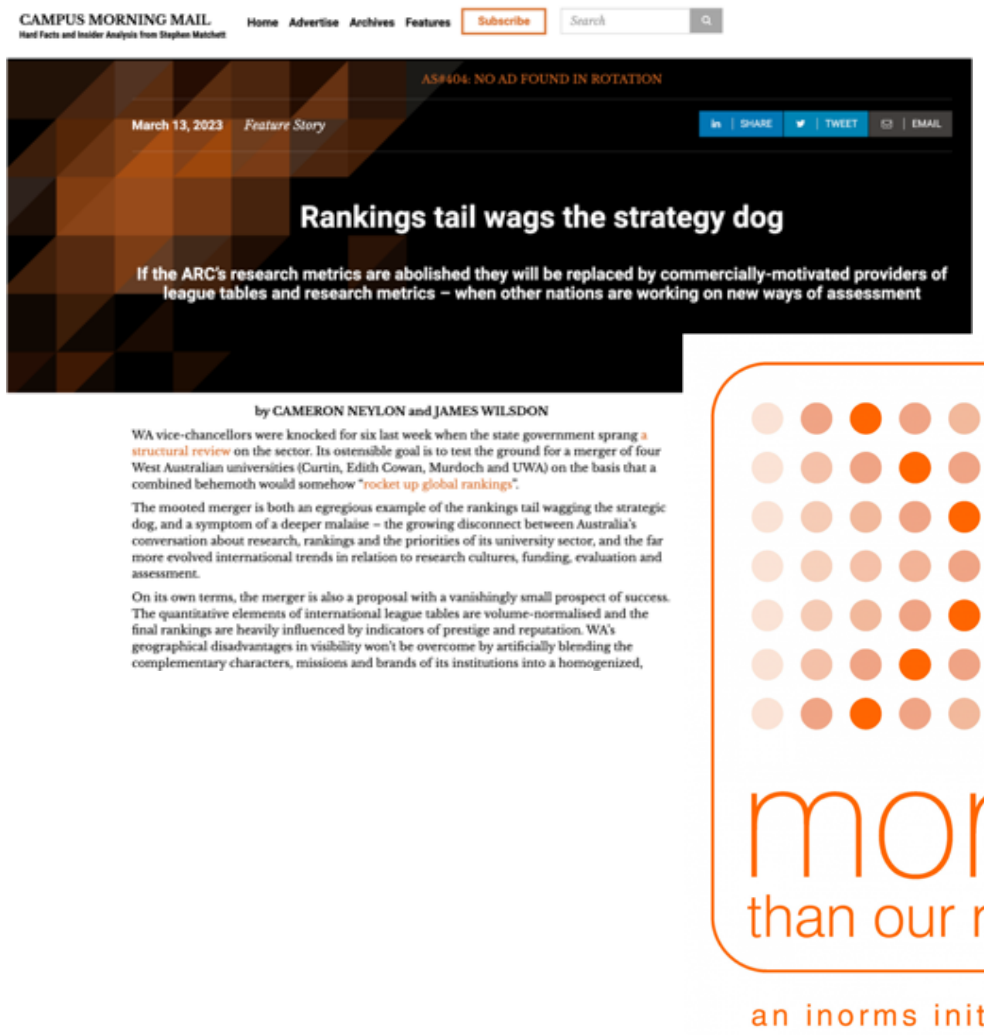
Overall profile	Outputs	Impact	Environment
	60% of overall	25% of overall	15% of overall
	Learn about outputs	Learn about impact	Learn about environment

9: Use data for good

DATA FOR GOOD

- Gender pay gaps for research staff;
- % of research staff on short term contracts;
- Measures of research staff wellbeing and contentment in surveys of workplace culture
- Volume of teamwork; collaborations; co-produced research (with users);
- Open research indicators;
- Policy impacts e.g. via citations in policy literatures;
- Peer review work;
- Citizenship contributions (from workload models);
- Measures of support for EDI;
- Effective measures for dealing with bullying and harassment.

10: Rethink university rankings



CAMPUS MORNING MAIL
Hard Facts and Insider Analysis from Stephen Matchett

Home Advertise Archives Features Subscribe Search

AS#404: NO AD FOUND IN ROTATION

March 13, 2023 *Feature Story*

SHARE TWEET EMAIL

Rankings tail wags the strategy dog

If the ARC's research metrics are abolished they will be replaced by commercially-motivated providers of league tables and research metrics – when other nations are working on new ways of assessment

by CAMERON NEYLON and JAMES WILSDON

WA vice-chancellors were knocked for six last week when the state government sprang a structural review on the sector. Its ostensible goal is to test the ground for a merger of four West Australian universities (Curtin, Edith Cowan, Murdoch and UWA) on the basis that a combined behemoth would somehow "rocket up global rankings".

The mooted merger is both an egregious example of the rankings tail wagging the strategic dog, and a symptom of a deeper malaise – the growing disconnect between Australia's conversation about research, rankings and the priorities of its university sector, and the far more evolved international trends in relation to research cultures, funding, evaluation and assessment.

On its own terms, the merger is also a proposal with a vanishingly small prospect of success. The quantitative elements of international league tables are volume-normalised and the final rankings are heavily influenced by indicators of prestige and reputation. WA's geographical disadvantages in visibility won't be overcome by artificially blending the complementary characters, missions and brands of its institutions into a homogenized,

more than our rank

an inorms initiative

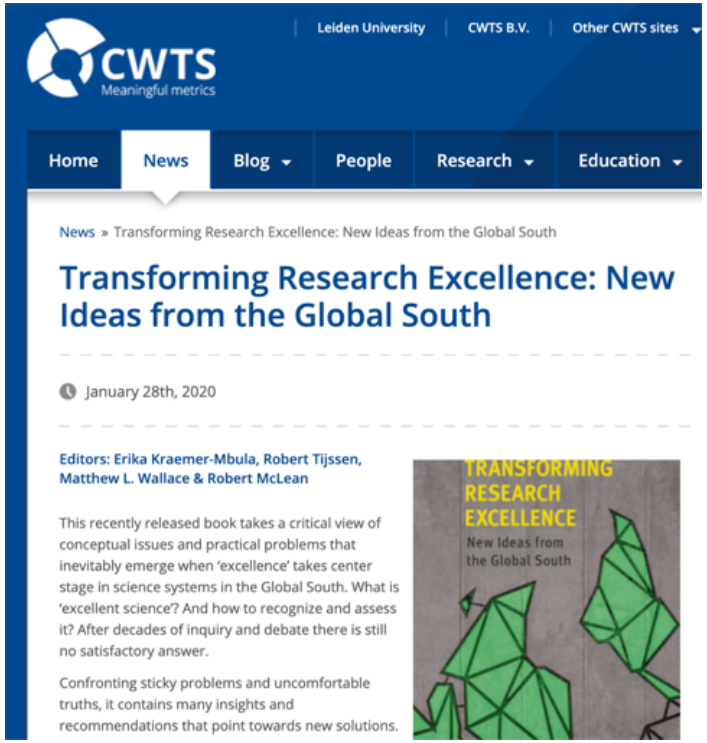
How to protect the 'national interest' at a systemic level and not surrender measures of value, quality or impact to unaccountable data and league table providers?

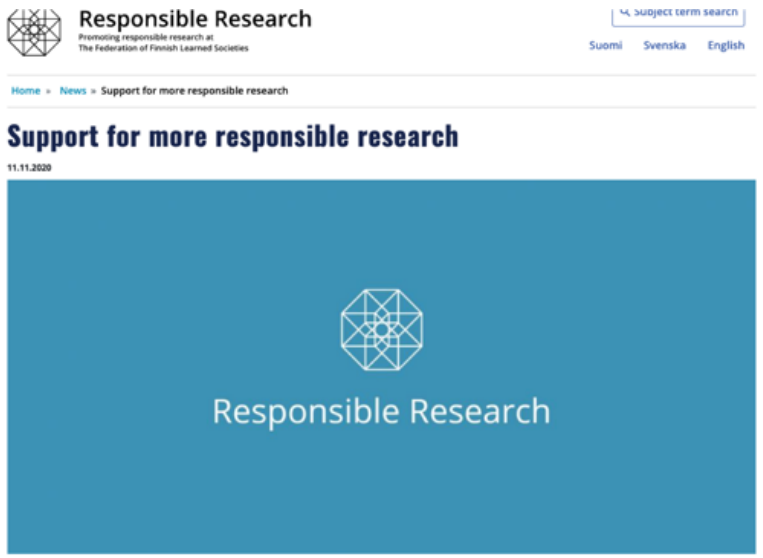
HEIs should be encouraged to take a more responsible approach to their engagement with and promotion of university **league tables.** This may include becoming a signatory to the **INORMS More Than Our Rank** initiative.

Priorities for the next decade of RRA in Czechia & beyond

Coalition for Advancing Research Assessment

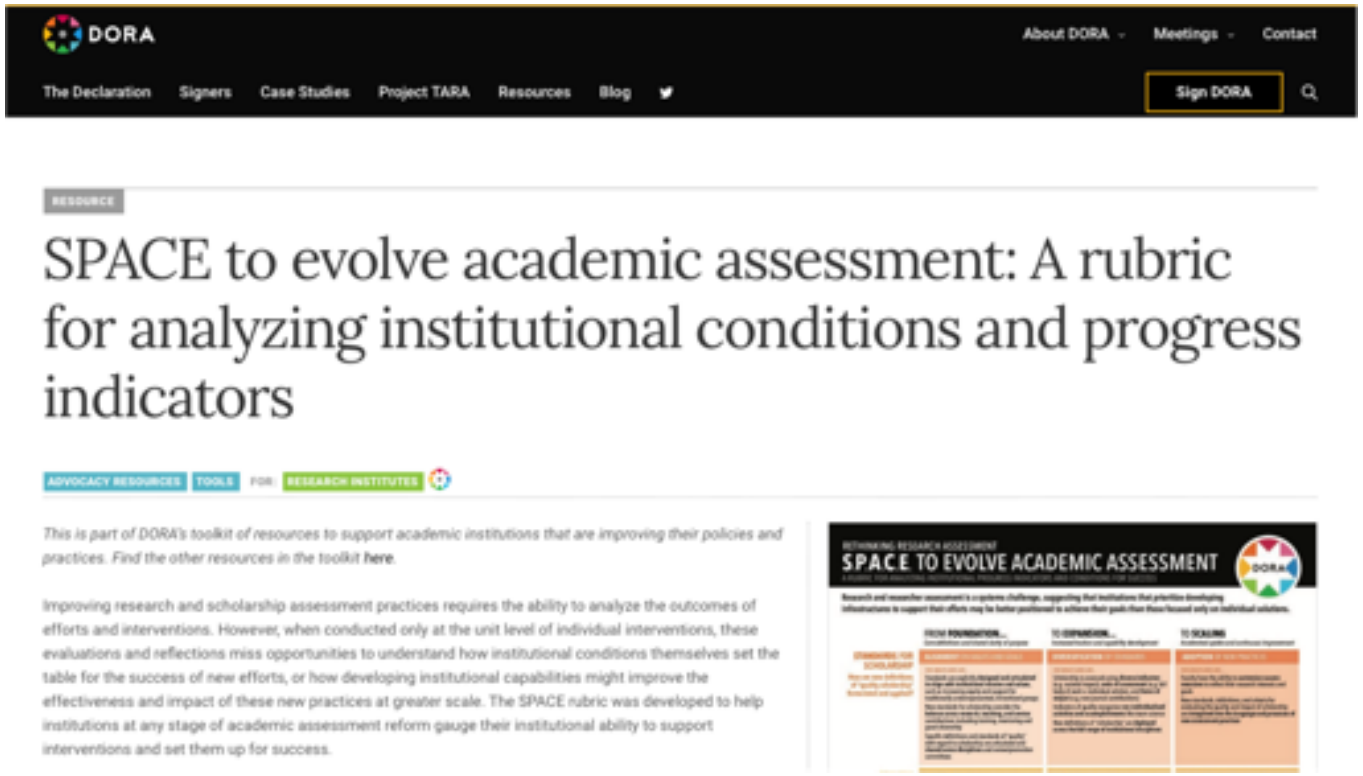
Our vision is that the assessment of research, researchers and research organisations recognises the diverse outputs, practices and activities that maximise the quality and impact of research. This requires basing assessment primarily on qualitative judgement, for which peer review is central, supported by responsible use of quantitative indicators.





Priority 1: Continue to build national and international coalitions for responsible research assessment

Priority 2: Implement & translate principles to institutional policy & practices



The international journal of science / 22 July 2021

nature

Responsible assessment faces the acid test

The University of Liverpool is planning lay-offs using controversial measures. How should the movement for responsible research respond?

A leading UK university has become mired in a public dispute over how it is assessing researchers' performance. The evolving situation at the University of Liverpool is being watched closely by concerned academics around the world – and is raising questions about whether more needs to be done to ensure that universities assess their researchers equitably. At the end of last month, the leaders of some of the world's foremost responsible-research initiatives – the Hong Kong Principles, the INORMS Research Evaluation Group, the Leiden Manifesto and the Metric Tide – wrote a strongly worded letter arguing that the University of Liverpool's proposals remain

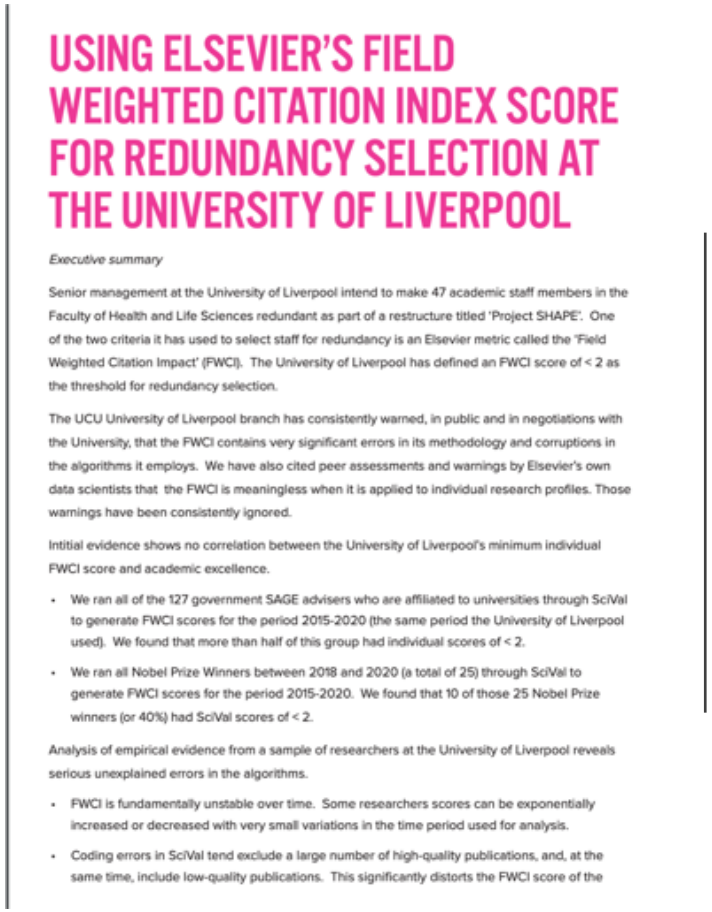
"Does the research community need a body with the

redundancy. In response to the threat of redundancies, researchers took industrial action during May, June and July.

One influential initiative is choosing to negotiate privately with the university. This is the organization behind the San Francisco Declaration on Research Assessment (DORA), an international voluntary agreement through which research organizations vow to conduct research assessment responsibly.

DORA's signatories pledge not to use metrics such as the Journal Impact Factor to evaluate researchers, and to be transparent in the criteria used to make decisions on matters such as hiring and promotion. Liverpool is one of some 2,200 organizations that have signed the declaration. DORA is in talks with the university, but choosing not to reveal further details. A statement on DORA's website says that it expects signatories to abide by their pledges, while also reiterating that it is not a regulatory body.

DORA's approach – to resolve disputes constructively but without publicity – has had some effect. Liverpool initially included the field-weighted citation metric on its criteria for redundancies, but dropped that after consultation with DORA. However, there are conflicting views of whether this puts Liverpool in the clear. The university told *Nature* its amended criteria are "in keeping with the principles of DORA". In response, a DORA spokesperson said there are "ongoing concerns". Such mixed messages show



USING ELSEVIER'S FIELD WEIGHTED CITATION INDEX SCORE FOR REDUNDANCY SELECTION AT THE UNIVERSITY OF LIVERPOOL

Executive summary

Senior management at the University of Liverpool intend to make 47 academic staff members in the Faculty of Health and Life Sciences redundant as part of a restructure titled 'Project SHAPE'. One of the two criteria it has used to select staff for redundancy is an Elsevier metric called the 'Field Weighted Citation Impact' (FWCI). The University of Liverpool has defined an FWCI score of < 2 as the threshold for redundancy selection.

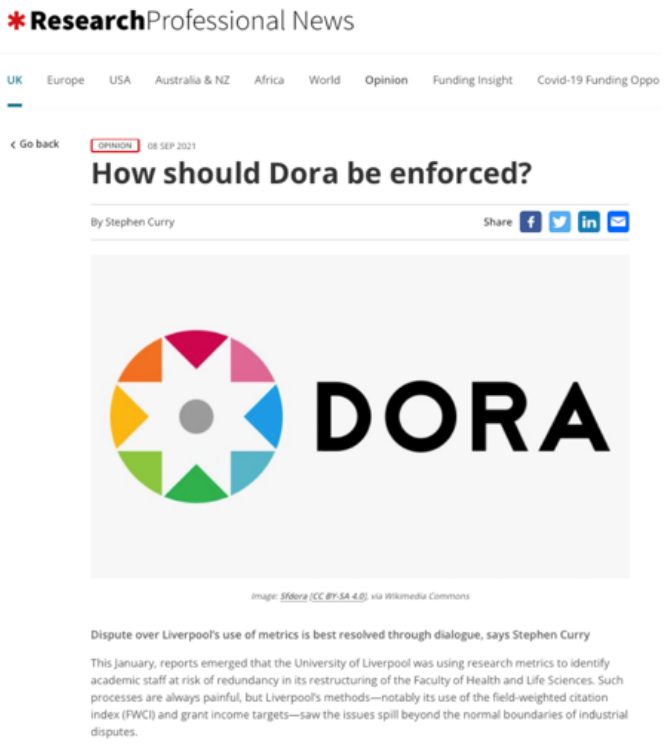
The UCU University of Liverpool branch has consistently warned, in public and in negotiations with the University, that the FWCI contains very significant errors in its methodology and corruptions in the algorithms it employs. We have also cited peer assessments and warnings by Elsevier's own data scientists that the FWCI is meaningless when it is applied to individual research profiles. Those warnings have been consistently ignored.

Intitial evidence shows no correlation between the University of Liverpool's minimum individual FWCI score and academic excellence.

- We ran all of the 127 government SAGE advisers who are affiliated to universities through SciVal to generate FWCI scores for the period 2015-2020 (the same period the University of Liverpool used). We found that more than half of this group had individual scores of < 2.
- We ran all Nobel Prize Winners between 2018 and 2020 (a total of 25) through SciVal to generate FWCI scores for the period 2015-2020. We found that 10 of those 25 Nobel Prize winners (or 40%) had SciVal scores of < 2.

Analysis of empirical evidence from a sample of researchers at the University of Liverpool reveals serious unexplained errors in the algorithms.

- FWCI is fundamentally unstable over time. Some researchers scores can be exponentially increased or decreased with very small variations in the time period used for analysis.
- Coding errors in SciVal tend exclude a large number of high-quality publications, and, at the same time, include low-quality publications. This significantly distorts the FWCI score of the



ResearchProfessional News

UK Europe USA Australia & NZ Africa World Opinion Funding Insight Covid-19 Funding Oppo

‹ Go back

OPINION 08 SEP 2021

How should Dora be enforced?

By Stephen Curry

Share

DORA

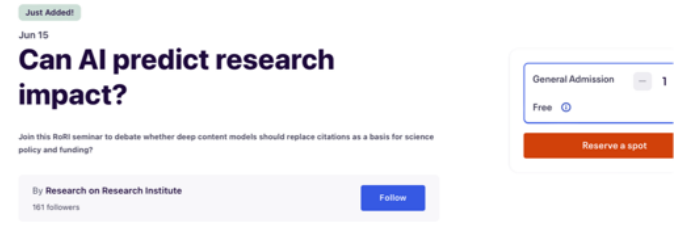
Image: Sfdora (CC BY-SA 4.0), via Wikimedia Commons

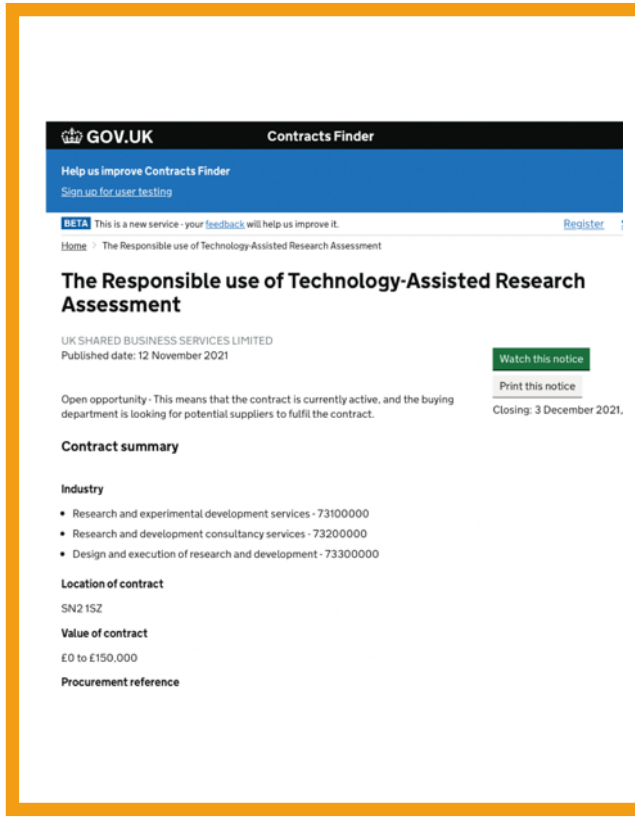
Dispute over Liverpool's use of metrics is best resolved through dialogue, says Stephen Curry

This January, reports emerged that the University of Liverpool was using research metrics to identify academic staff at risk of redundancy in its restructuring of the Faculty of Health and Life Sciences. Such processes are always painful, but Liverpool's methods—notably its use of the field-weighted citation index (FWCI) and grant income targets—saw the issues spill beyond the normal boundaries of industrial disputes.

Priority 3: Develop more sophisticated frameworks for compliance, accountability & enforcement

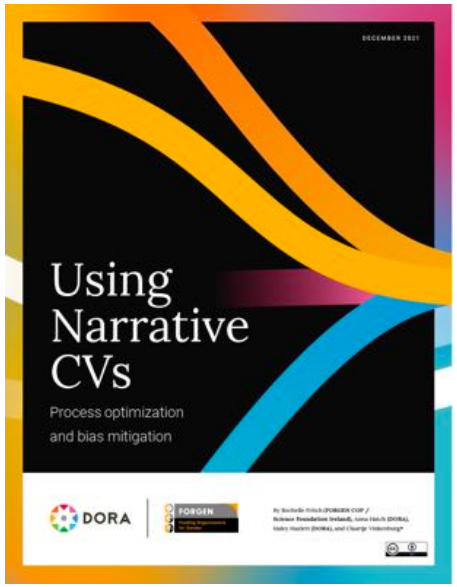
Priority 4: RRA needs to anticipate and keep pace with new tools and technologies of assessment and evaluation

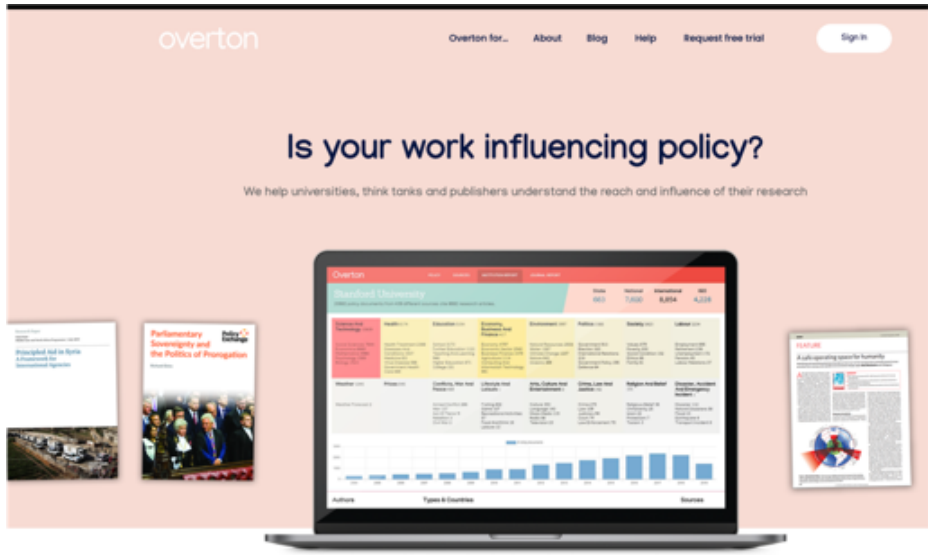




Priority 5: Strengthen weak proxies

More work is needed to overcome the significant methodological limitations of metrics for impact, and to embrace more creative mixed methods approaches.

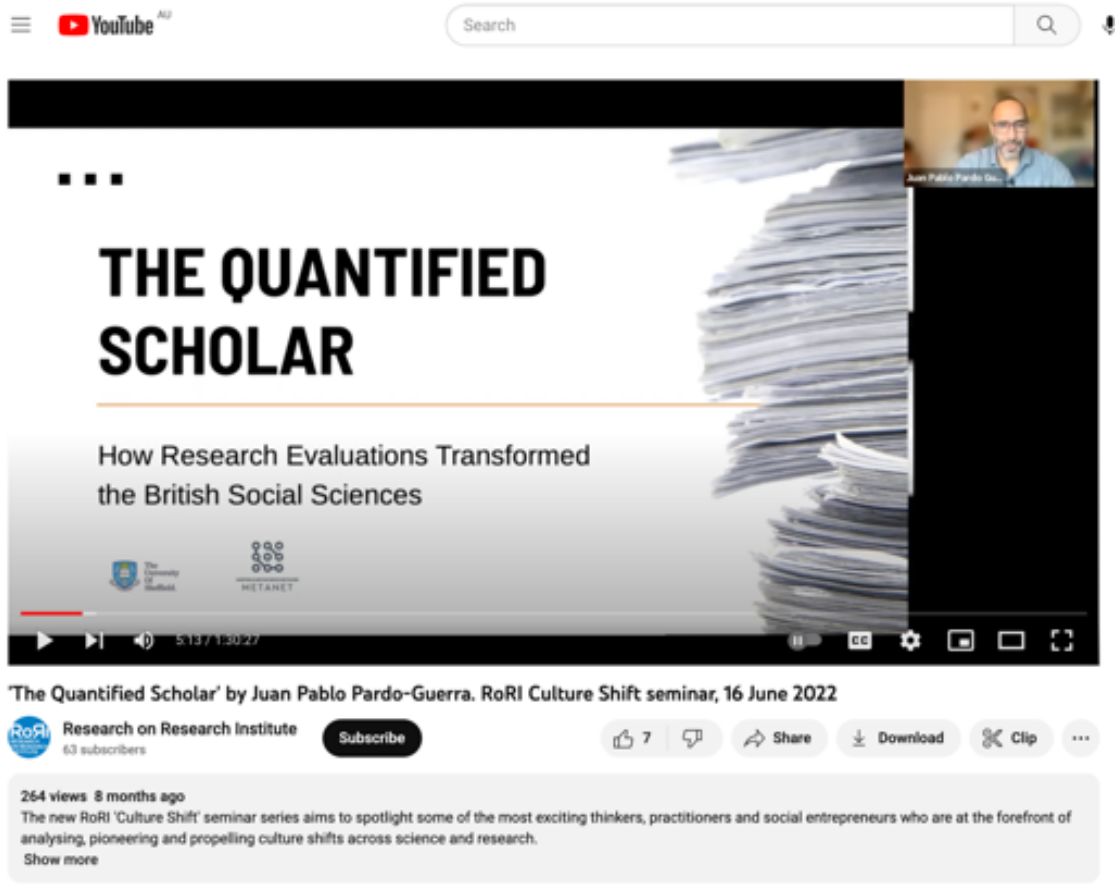




Track your policy impact

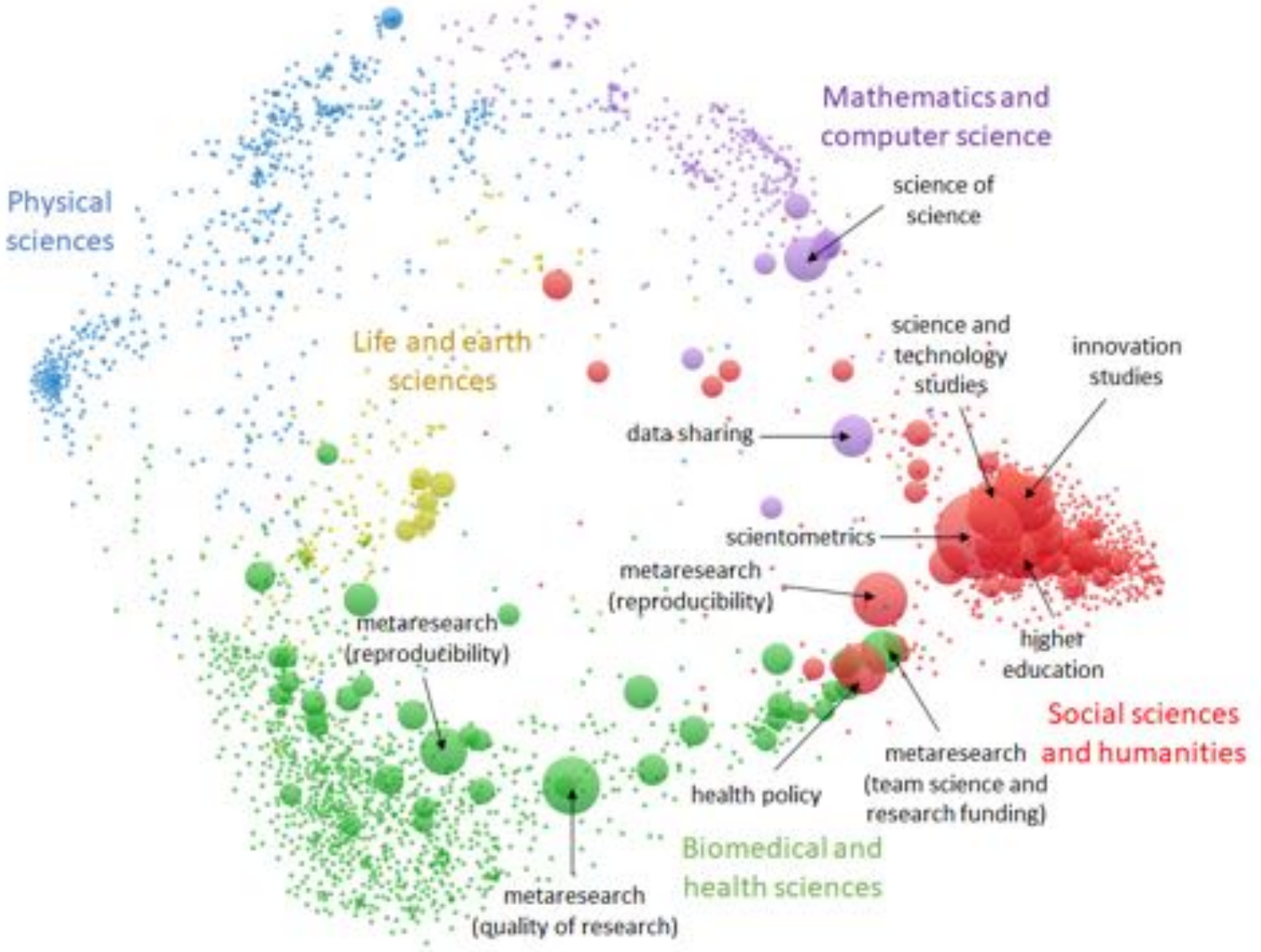
Priority 6: Move RRA ‘upstream’

We need to better understand and manage unintended shaping effects of assessment regimes on more fundamental priorities and hierarchies of knowledge production

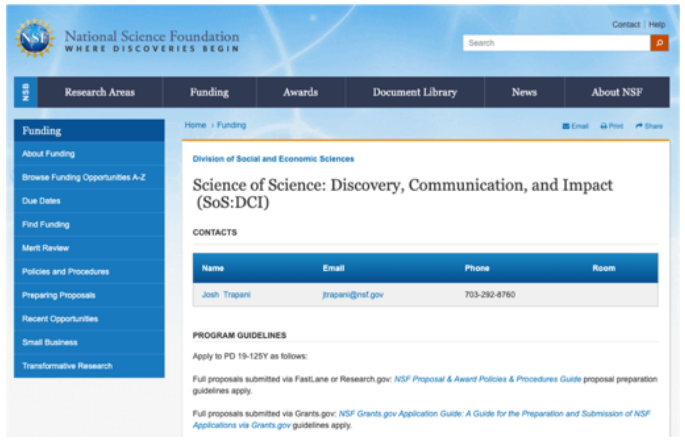


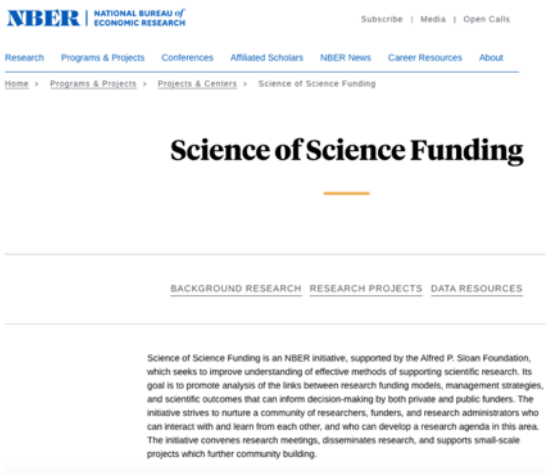
Priority 7: Harness collective intelligence

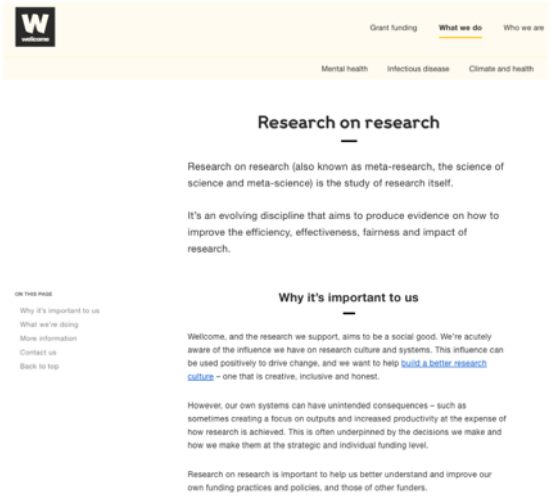
We need to get smarter at tapping into distributed data and insight into research system design, measurement and management



Priority 8: Invest & build capacity for metascience







RoRI RESEARCH ON RESEARCH INSTITUTE

www.researchonresearch.org

@RoRInstitute