



# Responsible research assessment (RRA): the state of play

JINSHA webinar & workshop, 5 February 2021

James Wilsdon, RoRI & University of Sheffield  
j.wilsdon@sheffield.ac.uk;  @jameswilsdon  
<http://www.researchonresearch.org/>

責任ある研究評価：  
その現状

本文中の日本語メモは、発表者の了解のもとに付記した仮訳ですのでご了承ください。



# What I'll aim to cover:

## 今日のお話

- From responsible metrics to responsible research assessment
  - Momentum, movers and shapers
  - Experiments in responsible assessment: some interim results
  - Global Research Council: autumn 2020 survey
  - Where next? Three priorities
- 「責任ある測定基準」から、「責任ある研究評価」へ
  - 機運、立役者
  - 責任ある評価の実験と中間的な結果
  - グローバル・リサーチ・カウンシル 2020年秋の調査結果
  - 次は何か？3つの優先事項

## A Celebrates Five Years!

18



**Join us as we discuss hiring decisions at research institutions**  
Live Monday, May 14 – 10:00 to 10:30 EDT #sfDOR

**Sandra Schmid, PhD**  
Cecil H. Green Distinguished Professor in Cellular and Molecular Biology, Chair, Cell Biology Department, UT Southwestern Medical Center

**Anna Manich, PhD**  
DORA Community

declaration was published in 2013, it has collected signature  
izations and 12,000 individuals. DORA has increased aware  
f the Journal Impact Factor and inspired change in the scient  
ions have started referencing the declaration in **research ass**  
at guide hiring, promotion, and funding decisions.



**COMMENT**

123456789101112  
131415161718192021222324252627282930313233343536373839404142434445464748495051525354555657585960616263646566676869707172737475767778798081828384858687888990919293949596979899100



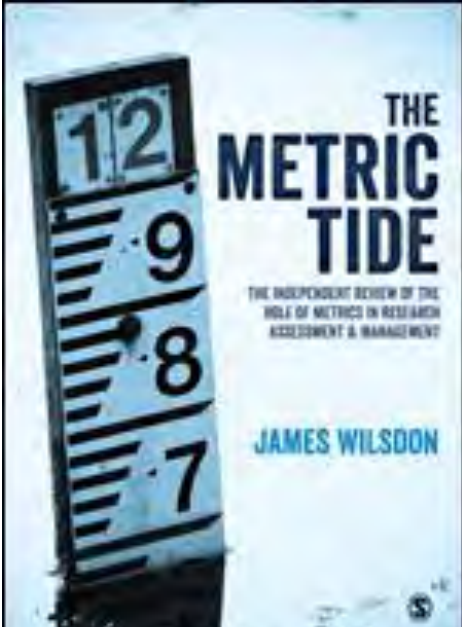
**The Leiden Manifesto for research metrics**  
Use these ten principles to guide research evaluation, urge Diana Hicks, Paul Wouters and colleagues.

**D**ata are increasingly used to govern science. Research evaluation that were once heuristic and performed now by peer and non-peer reviewers and metrics? The problem is that evaluation is now led by the data rather than by judgement. Metrics have proliferated, usually self-administered, and always well-intentioned, often ill-applied. We still denigrate the system with the very tools designed to improve it, as evaluation is increasingly implemented by organizations without knowledge of, or

admission, generative and interpretation. Before 2008, there was the Science Citation Index and 20,000 items for journals for Scientific Information (ISI), used by experts for specialist analyses. In 2002, Thomson Reuters launched an integrated web platform, making the Web of Science database widely accessible. Competing citation indices were created: Elsevier's Scopus (released in 2004) and Google Scholar (first version released in 2004). Web-based tools (usually compare institutional research productivity and impact

www.knowledgenet.org.uk/leiden/using-the-web-of-science-and-scopus/using-scopus-as-well-as-software-to-analyse-individual-citation-profiles-using-google-scholar (published in Paris, released in 2007)

In 2015, Anna Manich, a principal at the University of California, San Diego, proposed that to include pre-qualifying criteria consisting for individual researchers, later used in the Journal Impact Factor (first available after 1999 (see 'Impact factor observation'), Latin, metrics related to social usage. >



**THE METRIC TIDE**  
THE INDEPENDENT REVIEW OF THE WOLF OF METRICS IN RESEARCH ASSESSMENT & MANAGEMENT  
JAMES WILSDON



European Commission  
RESEARCH & INNOVATION  
Open Science

European Commission > Research & Innovation > Open Science > Expert Group on Altmetrics

Home Open Access European Open Science Cloud Open Science Policy Platform Expert

### Expert Group on Altmetrics

**NEW: Final Report of the Expert Group on Altmetrics is available**

Publication date: 20 March 2017

The Expert Group on Altmetrics outlines in this report how to advance a next-generatic metrics in the context of Open Science and delivers an advice corresponding to the following policy lines of the Open Science Agenda: Fostering Open Science, Removing barriers to Open Science, Developing research infrastructures and Embed Open Science in society.

The report will be presented and discussed at the Open Science Policy Platform on 20 March 2017

The report can be downloaded here  796 KB

# From responsible metrics....

## 「責任ある測定基準」から...

## CASE STUDY REPORT

# Reimagining Academic Career Assessment: Stories of innovation and change

Bregt Saeren (EUA), Anna Hatch (DORA), Stephen Curry (DORA), Vanessa Proudman (SPARC Europe) and Ashley Lakoduk (DORA)

January 2021

## RoRI Working Paper No.3 The changing role of funders in responsible research assessment: progress, obstacles and the way ahead

Stephen Curry, Sarah de Rijcke, Anna Hatch, Dorsamy (Gansen) Pillay, Inge van der Weijden and James Wilsdon

November 2020

Produced in partnership with:



## The European University Association and Science Europe Join Efforts to Improve Scholarly Research Assessment Methodologies

14 May 2019

Evaluating research and assessing researchers is fundamental to the research enterprise and core to the activities of research funders and research performing organisations, as well as universities. The European University Association (EUA) and Science Europe are committed to building a strong dialogue between their members, who share the responsibility of developing and implementing more accurate, open, transparent and responsible approaches, that better reflect the evolution of research activity in the digital era.

Today, the outcomes of scholarly research are often measured through methods based on quantitative, albeit approximate, indicators such as the journal impact factor. There is a need to move away from reductionist ways of assessing research, as well as to establish systems that better assess research potential. Universities, research funders and research performing organisations are well-placed to explore new and improved research assessment approaches, while also being indispensable in turning these innovations into systemic reforms.

EUA and Science Europe are committed to working together on building a strong dialogue between their members, with a view to:

- support necessary changes for a better balance between qualitative and quantitative research assessment approaches, aiming at evaluating the merits of scholarly research. Furthermore, novel criteria and methods need to be developed towards a fairer and more transparent assessment of research, researchers and research teams, conducive to selecting excellent proposals and researchers.
- recognise the diversity of research outputs and other relevant academic activities and their value in a manner that is appropriate to each research field and that challenges the overreliance on journal-based metrics.
- consider a broad range of criteria to reward and incentivise research quality as the fundamental principle of scholarly research, and ascertain assessment processes and methods that accurately reflect the vast dimensions of research quality and credit all scientific contributions appropriately.

EUA and Science Europe will launch activities to further engage their members in improving and strengthening their research assessment practices. Building on these actions, both associations commit to maintaining a continuous dialogue and explore opportunities for joint actions, with a view to promoting strong synergies between the rewards and incentives structures of research funders and research performing organisations, as well as universities.

# ...to responsible research assessment

...「責任ある研究評価」へ

# 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 scaffold for the governance of research, 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 fundamental aspects—methodologies, systems and cultures—of research assessment.

## 「責任ある研究評価」の定義

「責任ある研究評価」とは、多様で包摂的な研究文化のもとで、複数の異なる特性を有する質の高い研究を促し、把握し、報奨するような評価のアプローチを指す包括的用語である。

RRA は、責任ある研究とイノベーション (RRI) のためのより広範なフレームワークに基づいている。また、評価、レビュープロセスの開発と適用にも適用される。

RRI は研究のガバナンスのための広い足場として一般的に使用され、また「責任ある測定基準」の概念はマイクロレベルで指標自体にも適用できるのに対し、RRA の考え方は、資金提供者、研究機関、出版社などが研究評価の基本的な側面 (方法論、システム、文化) に注目するよう促すものである。

# A moment of opportunity?

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

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

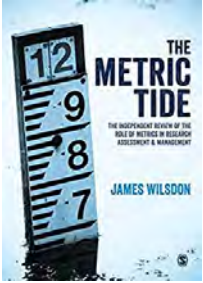
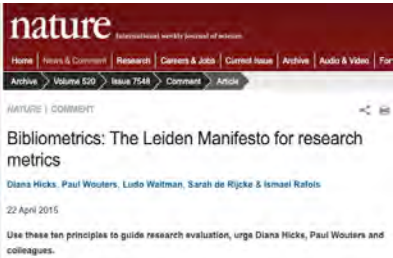
## チャンスの瞬間？

研究評価に関連したいくつかの長年の問題に対する懸念が強まっている。

- 研究の質やインパクトに関する狭い基準や指標が、インセンティブを歪め、研究者に持続不可能なプレッシャーを与え、研究の完全性や再現性の問題を悪化させるような形で誤用されていることである。
- このような基準や指標の狭めは、**研究のミッションや目的の多様性を低下させ**、研究機関や研究者が同じような戦略的優先事項を採用したり、リスクの低いincrementalな研究に集中したりするようになった。
- 質や影響力に関する狭い基準や指標を満たさない、あるいはそれに優先的に従わない選択した研究者に対する体系的な偏見が、研究コミュニティの多様性、活力、代表的な正統性を低下させている。
- 政策や経営上の注意を測定可能なものに向けてすることで、あまり明確でない、定量化できない質、インパクト、資産、価値が犠牲になり、大学のリーグテーブル(ランキング)によりこの傾向は更に悪化している。

# Fifteen movers and shapers

15の立役者



A yellow rectangular sign with a black border of diagonal stripes. The text is written in bold, black, sans-serif capital letters, centered on the sign.

**CHEERFUL  
WHISTLING  
PERMITTED**



# Experiments in responsible assessment: interim results

## 責任ある評価の実験： 中間結果



RoRI Working Paper No.3

**The changing role of funders in responsible research assessment:**

**progress, obstacles and the way ahead**

Stephen Curry, Sarah de Rijcke, Anna Hatch, Dorsamy (Gansen) Pillay, Inge van der Weijden and James Wilsdon

November 2020

Produced in partnership with:



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

- 表面的な適用
- システムの較正
- 缶切り：議論のきっかけ
- アドボカシー連合
- 組織文化の変革
- システムの変化...？

THE WORLD UNIVERSITY RANKINGS

PROFESSIONAL JOBS EVENTS RANKINGS STUDENT ABOUT US

Home > Rankings > University Impact Rankings

## University Impact Rankings 2019

The *Times Higher Education* University Impact Rankings are the only global performance tables that assess universities against the United Nations' Sustainable Development Goals. We use carefully calibrated indicators to provide comprehensive and balanced comparisons across three broad areas: research, outreach, and stewardship.

This first edition includes more than 450 universities from 76 countries.

Read more ...

How to get your university ranked

### EXPLORE UNIVERSITY IMPACT RANKINGS FOR INDIVIDUAL SDGS

Show me universities in **any country / region** offering **any subject**

Or, find specific universities **by name**

ELSEVIER

Home > Elsevier Connect > New metrics will make...

## New metrics will make journal assessment more complete and transparent

CiteScore metrics reveal the citation impact of more than 22,200 academic journals on Scopus

By Andrew Plume, PhD and Lisa Colledge, DPhil December 8, 2016

Elsevier Connect

330 disciplines

Largest Subject Area

1,549 titles

Analytical Chemistry

Microbiology

Physical Sciences

Decision Sciences

Embryology

Histology

Automotive Engineering

Geology

Social Sciences

Bioengineering

Conservation

Life Sciences

Immunology

General Medicine

Medicine

Anatomy

f t in e Comment

Subscribe to Elsevier Connect

# Cosmetic appropriation? 表面的な適用?

# Calibrating the machine

## システムの較正

### RECOMMENDATIONS

- #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

**Next-generation metrics:**  
Responsible metrics and evaluation for open  
science

「次世代の指標：  
オープン・サイエンスのための  
責任ある評価指標と評価」



Tin openers

缶切り：議論のきっかけ

Universities UK

Contact us Jobs

Home About us Facts and stats Policy and analysis Our work in parliament News and blog Events Internat

Home Policy and analysis Research policy Open science The UK Forum for Responsible Research Metrics

## The UK Forum for Responsible Research Metrics

A group of research funders, sector bodies, and infrastructure experts are working in partnership to promote the responsible use of research metrics.

The Forum for Responsible Research Metrics, chaired by Professor Max Lu (Vice-Chancellor at the University of Surrey), supports the responsible use of research metrics in higher education institutions and across the research community in the UK. The Forum have a programme of activities, including:

- Advice to the higher education funding bodies on quantitative indicators in the Research Excellence Framework (REF) 2021
- Advice on, and work to improve, the data infrastructure that underpins metric use
- Advocacy and leadership on the use of research metrics responsibly
- International engagement on the use of metrics in research and researcher assessment

The group was established in 2016, on the recommendation of the independent review on the role of metrics in research assessment and management. The review panel, chaired by Professor James Wilsdon, published their final report [The Metric Tide](#), which identified 20 specific recommendations for further work and action by stakeholders across the UK research system.

Advice, reports, and meeting papers will be made available on this webpage in due course. Full membership can be found below:



### What makes a fair and responsible university ranking? Rating the rankings criteria Version 2: August 2019

#### Introduction

The International Network of Research Management Societies (INORMS) established a two-year Research Evaluation Working Group (REWG) in 2018. It consists of representatives from a range of global member research management societies all seeking to work towards better, fairer and more meaningful research evaluation. One of the group's two areas of focus is the burgeoning influence of University Rankings on the behaviours of universities despite often poor methodological approaches and practices. The purpose of this work-package is to consider what we, as an international group of research managers, think the characteristics of a fair and responsible University Ranking should look like. The idea is to then 'turn the tables' on the rankings and rate them against our agreed criteria.

## Support for more responsible research

11.11.2020



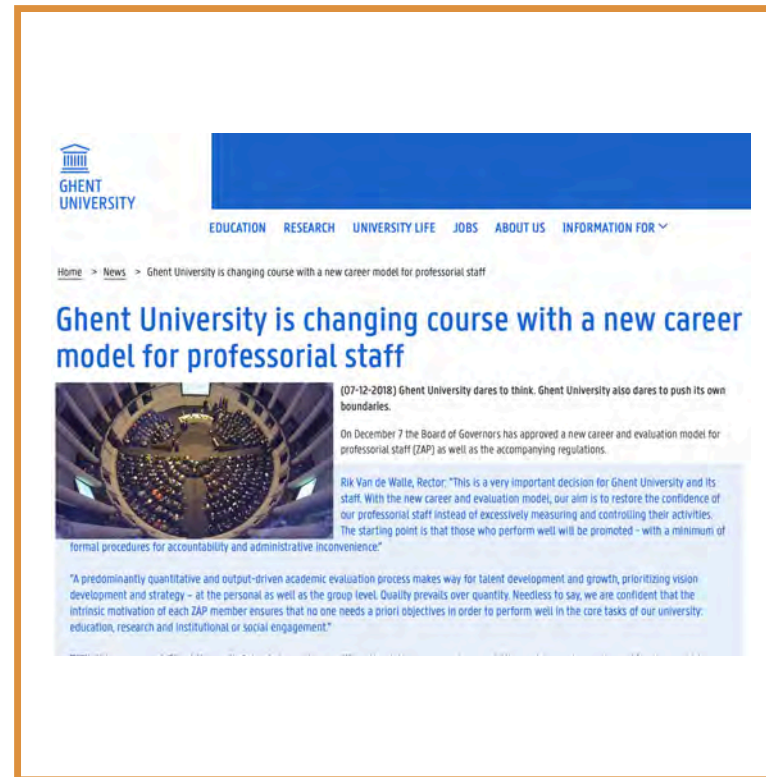
The Responsible Research website is a one-stop source for information and tools to support responsible research in Finland.

# Advocacy coalitions

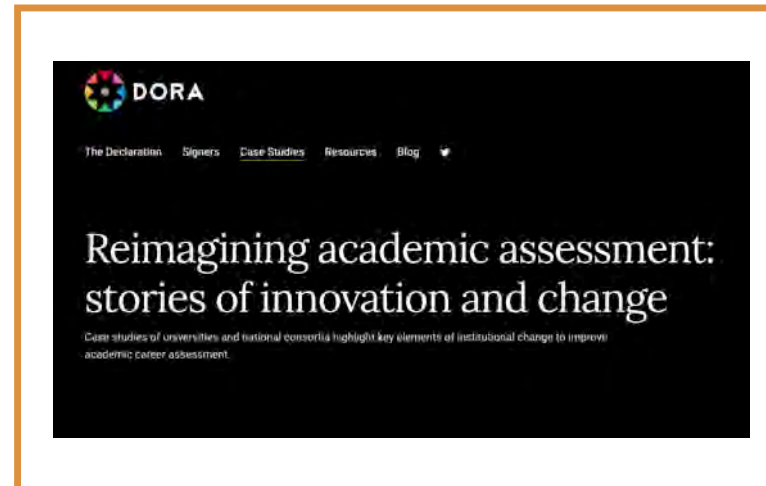
# アドボカシー連合

# Institutional culture change

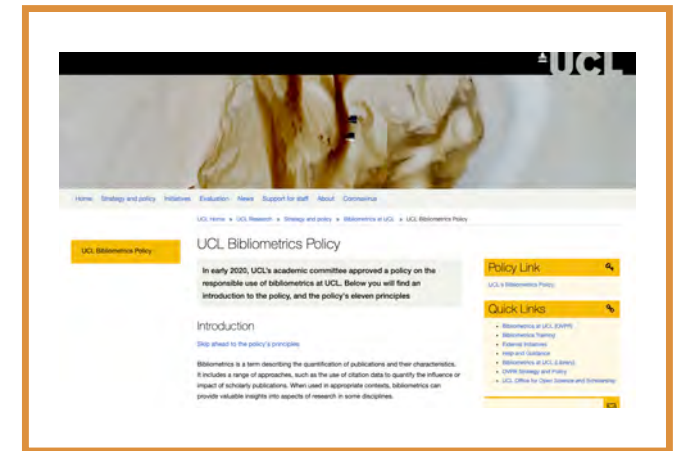
## 組織文化の変革



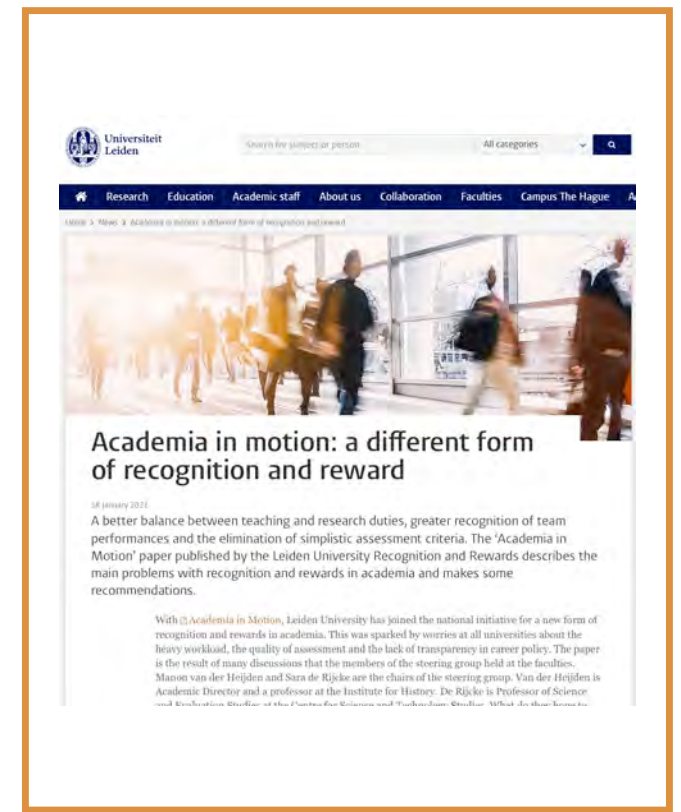
The screenshot shows a news article on the Ghent University website. The header includes the Ghent University logo and navigation links for Education, Research, University Life, Jobs, About Us, and Information For. The article title is "Ghent University is changing course with a new career model for professorial staff". The text discusses a decision by the Board of Governors on December 7 to approve a new career and evaluation model for professorial staff (ZAP). It quotes Rik Van de Walle, Rector, stating that the goal is to restore confidence in the professional staff and to promote those who perform well. A quote from the ZAP members is also included: "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."



The screenshot shows the DORA (Declaration on Research Assessment) website. The header includes the DORA logo and navigation links for The Declaration, Signers, Case Studies, Resources, and Blog. The main heading is "Reimagining academic assessment: stories of innovation and change". Below the heading, it states: "Case studies of universities and national consortia highlight key elements of institutional change to improve academic career assessment."

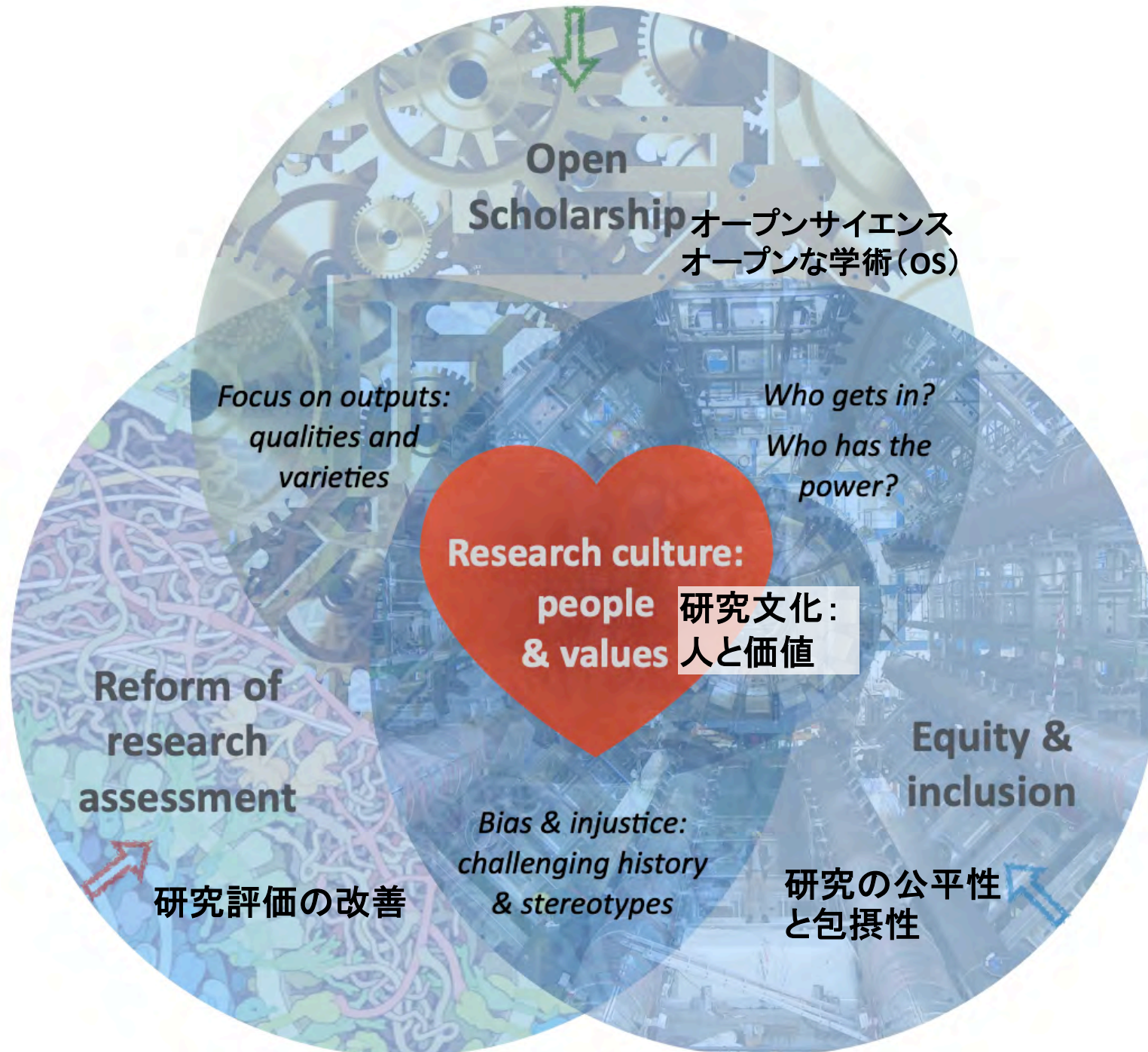


The screenshot shows the UCL Bibliometrics Policy page. The header includes the UCL logo and navigation links for Home, Strategy and policy, Initiatives, Evaluation, News, Support for staff, About, and Contact us. The article title is "UCL Bibliometrics Policy". The text states: "In early 2020, UCL's 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". The page includes a "Policy Link" section and a "Quick Links" section with links to Bibliometrics at UCL, Bibliometrics Training, External Evidence, Research Evidence, Bibliometrics at UCL & ERAC, UCL Strategy and Policy, and UCL Office for Open Science and Scholarship.



The screenshot shows a news article on the Leiden University website. The header includes the Universiteit Leiden logo and navigation links for Research, Education, Academic staff, About us, Collaboration, Faculties, and Campus The Hague. The article title is "Academia in motion: a different form of recognition and reward". The text states: "A better balance between teaching and research duties, greater recognition of team performances and the elimination of simplistic assessment criteria. The 'Academia in Motion' paper published by the Leiden University Recognition and Rewards describes the main problems with recognition and rewards in academia and makes some recommendations." The article is dated 18 January 2021. A quote from the paper is included: "With Academia in Motion, Leiden University has joined the national initiative for a new form of recognition and rewards in academia. This was sparked by worries at all universities about the heavy workload, the quality of assessment and the lack of transparency in career policy. The paper is the result of many discussions that the members of the steering group held at the faculties. Maartje van der Heijden and Sara de Rijcke are the chairs of this steering group. Van der Heijden is Academic Director and a professor at the Institute for History. De Rijcke is Professor of Science and Evaluation Director of the Centre for Evidence and Evidence-based Studies. What do these lessons..."

# System change?

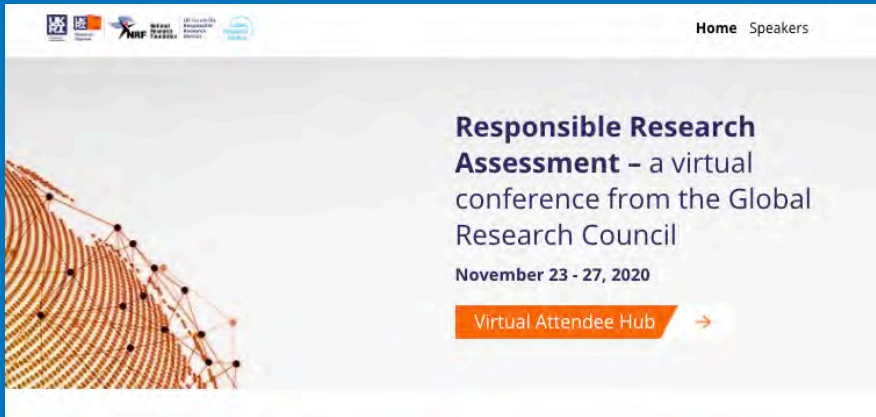


NB. This slide is used with thanks to Stephen Curry, and is adapted from a paper on the intersections between DORA, open scholarship and equity <https://sfdora.org/2020/08/18/the-intersections-between-dora-open-scholarship-and-equity/>

## システムの変化？

# Global Research Council Survey methodology

## グローバル・リサーチ・ カウンシルの調査： 方法



Online survey: 23 questions  
Open from September-October 2020  
Completed by 55 organisations / 46%  
response rate

オンライン調査: 質問数23  
2020年9月～10月実施  
回答数55機関 / 回答率 46%

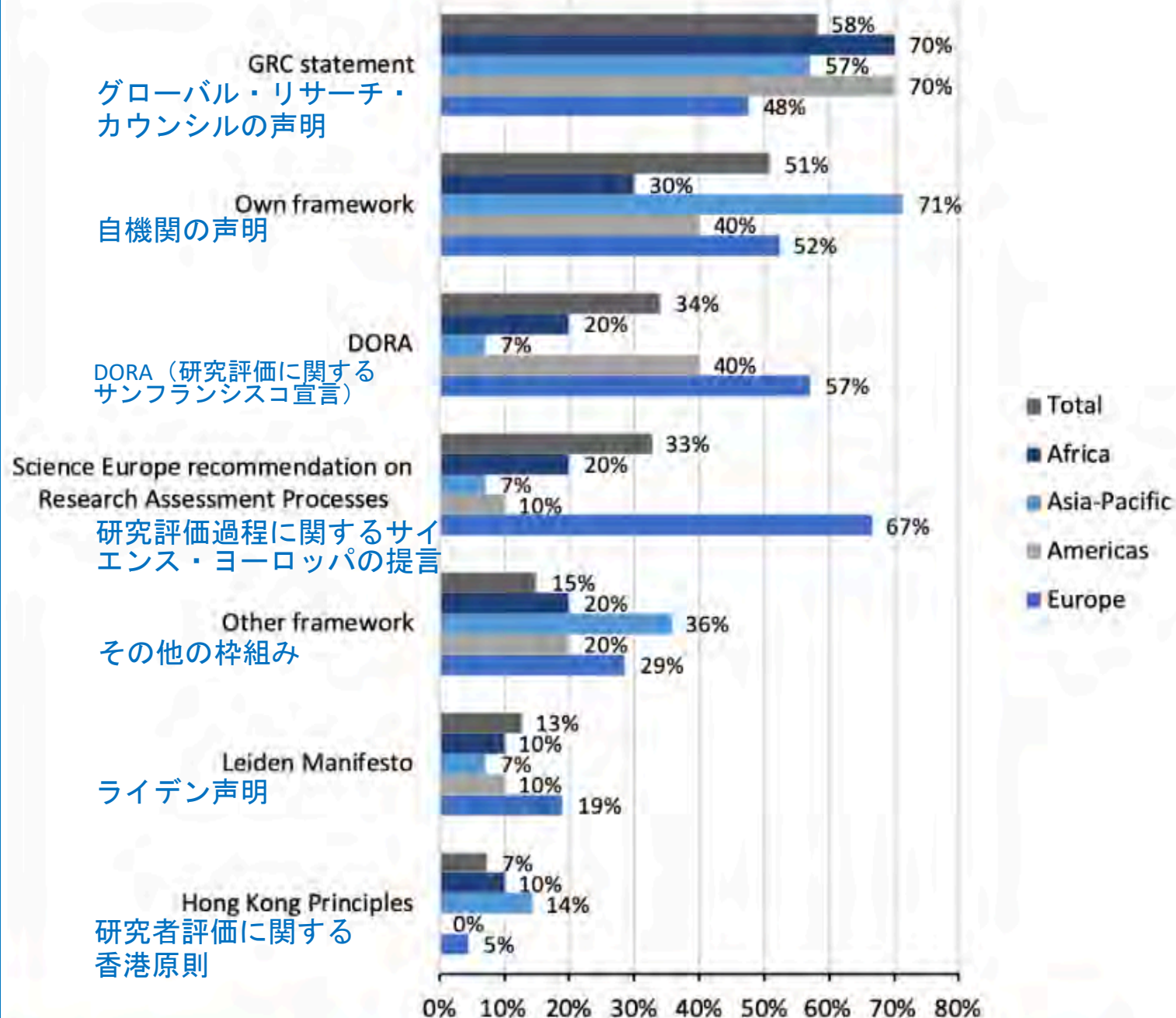
|   | N  | %    |
|---|----|------|
| Africa and Middle-East<br>アフリカ・中東<br>(Sub-Saharan Africa, North Africa & Middle East) | 10 | 18.2 |
| Asia-Pacific<br>アジア・太平洋   | 14 | 25.5 |
| Americas<br>南北アメリカ  | 10 | 18.2 |
| Europe<br>ヨーロッパ   | 21 | 38.2 |
| Total<br>合計   | 55 | 100  |

Table 1: Respondents by geographical region



# Endorsements of existing RRA Frameworks

## 既存の責任ある研究評価の枠組みの支持



# Research Assessment Indicators

# 研究評価指標

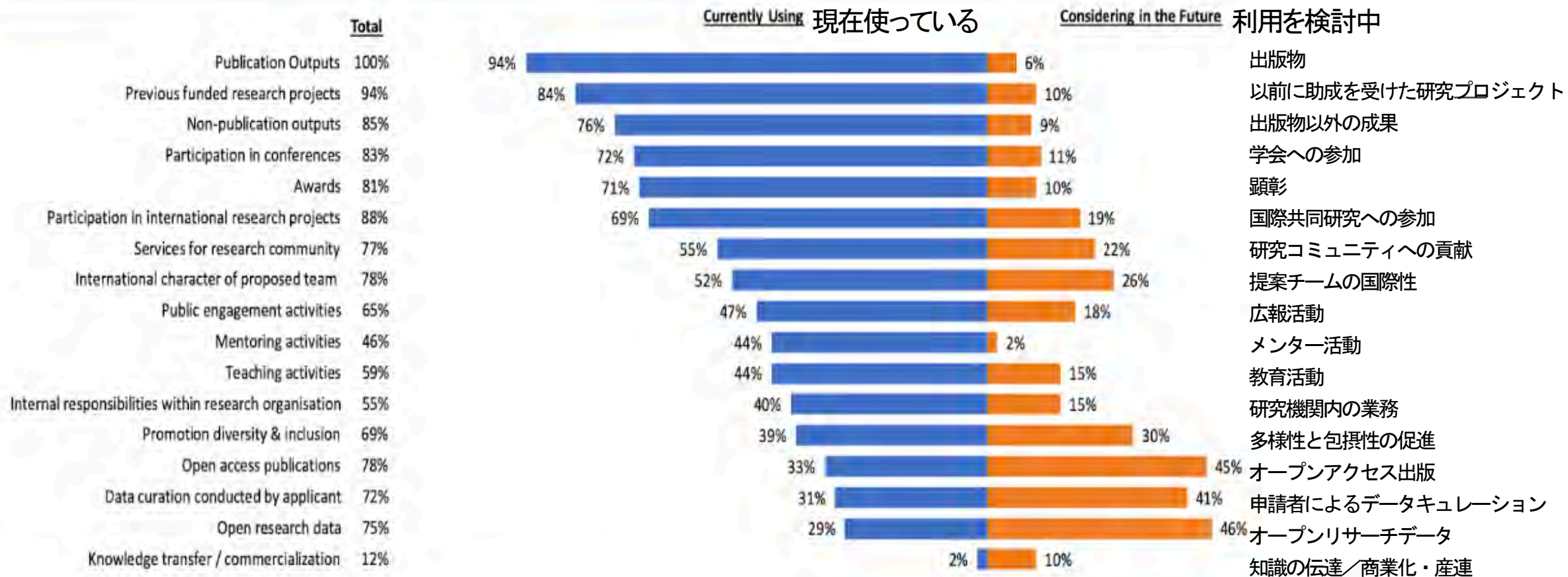
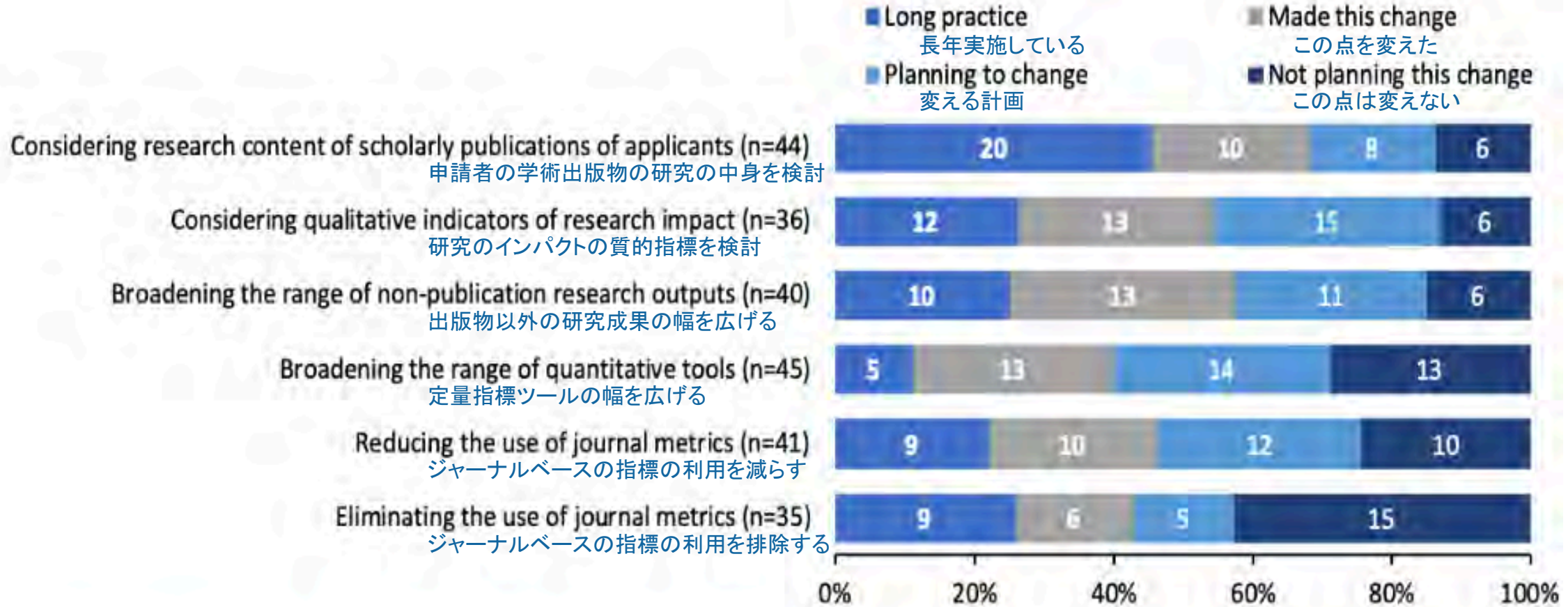
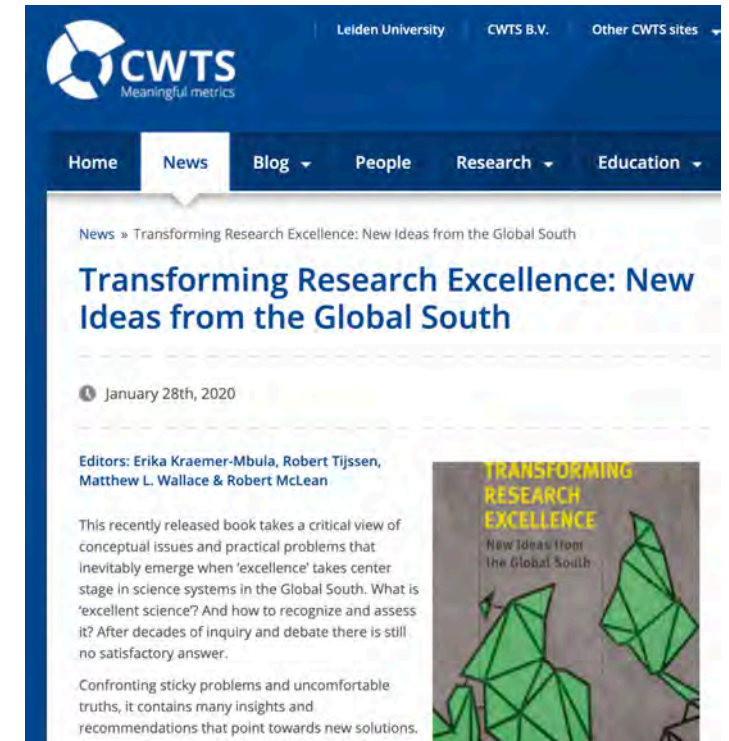


Figure 3: Research assessment indicators (to be) used by GRC participating organisations who responded to the survey (n=50, missing n=5)

# Changes in the way research proposals are assessed

## 研究計画調書の評価方法の変化





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

優先課題1： 責任ある研究評価のための国際的な連合を継続して築く

# Priority 2: Experiment, evaluate and amplify good practices

# 優先課題2: 実験し、 評価し、 グッド・プラクティスを 増幅する

**PLOS BIOLOGY** FIFTEENTH ANNIVERSARY

Browse Publish About

OPEN ACCESS

PERSPECTIVE

## Assessing scientists for hiring, promotion, and tenure

David Moher, Florian Naudet, Ioana A. Cristea, Frank Miedema, John P. A. Ioannidis, Steven N. Goodman

Version 2 Published: March 29, 2018 • <https://doi.org/10.1371/journal.pbio.2004089>

| Article                                | Authors | Metrics | Comments | Related Content |
|--|---------|---------|----------|-----------------|
| <a href="#">Introduction</a>           |         |         |          |                 |
| <a href="#">Methods</a>                |         |         |          |                 |
| <a href="#">Results</a>                |         |         |          |                 |
| <a href="#">Supporting information</a> |         |         |          |                 |
| <a href="#">Acknowledgments</a>        |         |         |          |                 |
| <a href="#">References</a>             |         |         |          |                 |
| <a href="#">Reader Comments (2)</a>    |         |         |          |                 |
| <a href="#">Media Coverage (3)</a>     |         |         |          |                 |
| <a href="#">Figures</a>                |         |         |          |                 |

**Abstract**

Introduction

Assessment of researchers is necessary for decisions of hiring, promotion, and tenure. A burgeoning number of scientific leaders believe the current system of faculty incentives and rewards is misaligned with the needs of society and disconnected from the evidence about the causes of the reproducibility crisis and suboptimal quality of the scientific publication record. To address this issue, particularly for the clinical and life sciences, we convened a 22-member expert panel workshop in Washington, DC, in January 2017. Twenty-two academic leaders, funders, and scientists participated in the meeting. As background for the meeting, we completed a selective literature review of 22 key documents critiquing the current incentive system. From each document, we extracted how the authors perceived the problems of assessing science and scientists, the unintended consequences of maintaining the status quo for assessing scientists, and details of their proposed solutions. The resulting table was used as a seed for participant discussion. This resulted in six principles for assessing scientists and

**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.

**ANNUAL REVIEWS**

JOURNALS & E JOURNAL INFO PRICING & SUBSCRIPTIONS

Home > Annual Review Statistics with Applications > Volume 7 (2019) > Highlights

## Calibrating the Scientific Ecosystem Through Meta-Research

Annual Review of Statistics and Its Application

Vol. 7 | Volume published late March 2019

Raymond K. Goldstein, David J. Foray, and Steven N. Goodman

Tom E. Harlow, Stylianos Sargiotto, Perrine Janneau, Valentin Danchev, Sophia Crivello, Steven N. Goodman, and John P. A. Ioannidis

**Abstract**

While some scientists study insects, molecules, brains, or clouds, other scientists study science itself. Meta-research, or research on research, is a fast-growing discipline that investigates efficiency, quality, and bias in the scientific ecosystem. Some meta-researchers have become especially relevant and widespread concerns about the credibility of the scientific literature. Meta-research may help calibrate the scientific ecosystem toward higher standards by providing empirical evidence that allows the scientific generation and refinement of robust solutions. We introduce a foundational framework that involves (1) identifying problems, (2) investigating problems, (3) developing solutions, and (4) evaluating solutions. In each of these areas, we review key meta-research methods and discuss some strategies of error and ongoing work. The scientific ecosystem is perpetually evolving; the discipline of meta-research presents an

**c&en** TOPICS MAGAZINE COLLECTIONS VIDEOS JOBS

## Research on research gains steam

New metascience institute aims for larger studies

by **Dalmeet Singh Chawla**, special to C&EN

OCTOBER 1, 2019

**I**n 2005, John Ioannidis, a professor of medicine at Stanford University, opened a can of worms. In a paper published in *PLOS Medicine*, he argued that most published scholarly literature is false (DOI: [10.1371/journal.pmed.0020124](https://doi.org/10.1371/journal.pmed.0020124)).

To date, Ioannidis's "landmark study" has attracted thousands of citations and helped solidify a whole field in its own right, says Jelle Wicherts, who studies research methodology at Tilburg University.

The use of scientific methodology to study science itself is called metascience. The discipline has become mainstream in recent years, tackling some of the thorniest problems science faces, including a lack of reproducibility of academic literature, biases in peer review, and the fair allocation of research funding. "Metascience is now a distinct species," although it has ancestors in medical science, psychology, and other disciplines, Wicherts says.

Ioannidis, who launched the **Meta-Research Innovation Center at Stanford (METRICS)** in 2014, however, is hesitant to frame metascience as a separate field. "In a way, every researcher is a metascientist, since the issues involved are at the core of how to do science and apply the scientific method and maximize the yield of reproducible and useful information," he says.

Credit: Courtesy of Jelle Wicherts  
Jelle Wicherts, founding director of the Research on Research Institute



## GRANTS AI is selecting reviewers in China

The tool is already saving time for the country's major grant funding agency.

BY DAVID CYRANOSKI

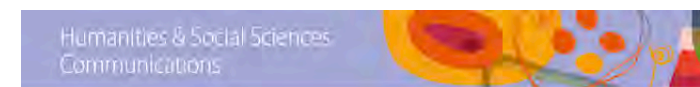
China's largest funder of basic science is piloting an artificial intelligence (AI) tool that selects researchers to review grant applications, in an attempt to make the process more efficient, faster and fairer. Some researchers say the approach by the National

Natural Science Foundation of China (NSFC) is world-leading, but others are sceptical about whether AI can improve the process. Choosing researchers to peer review project proposals or publications is time-consuming and prone to bias. Several academic publishers are experimenting with AI tools to select reviewers and carry out other tasks. And a few

funding agencies, including some in North America and Europe, have trialled simple AI systems, some of which match keywords in grant applications to those in publications of other scientists to identify potential reviewers. The NSFC is building a more sophisticated system that will crawl online scientific literature databases and scientists' personal

318 | NATURE | VOL 569 | 16 MAY 2019

© 2019 Springer Nature Limited. All rights reserved.



ARTICLE

<https://doi.org/10.1038/s41599-020-00703-8>

OPEN

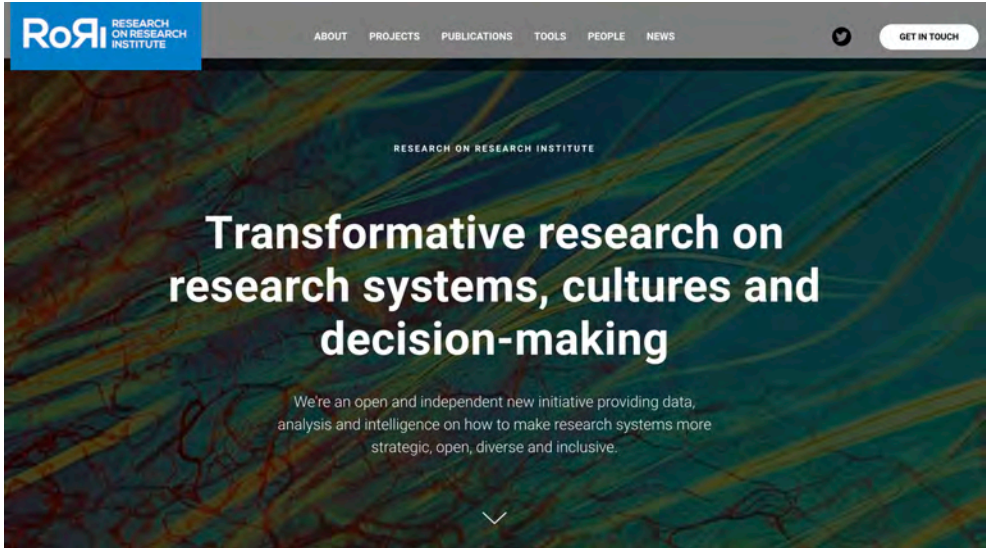
## AI-assisted peer review

Alessandro Checco<sup>1</sup>, Lorenzo Bracciale<sup>2,3</sup>, Pierpaolo Loreti<sup>2</sup>, Stephen Pinfield<sup>1,3</sup> & Giuseppe Bianchi<sup>2</sup>

The scientific literature peer review workflow is under strain because of the constant growth of submission volume. One response to this is to make initial screening of submissions less time intensive. Reducing screening and review time would save millions of working hours and potentially boost academic productivity. Many platforms have already started to use automated screening tools, to prevent plagiarism and failure to respect format requirements. Some tools even attempt to flag the quality of a study or summarise its content, to reduce reviewers' load. The recent advances in artificial intelligence (AI) create the potential for (semi) automated peer review systems, where potentially low-quality or controversial studies could be flagged, and reviewer-document matching could be performed in an automated manner. However, there are ethical concerns, which arise from such approaches, particularly associated with bias and the extent to which AI systems may replicate bias. Our main goal in this study is to discuss the potential, pitfalls, and uncertainties of the use of AI to approximate or assist human decisions in the quality assurance and peer-review process associated with research outputs. We design an AI tool and train it with 3300 papers from three conferences, together with their reviews evaluations. We then test the ability of the AI in predicting the review score of a new, unobserved manuscript, only using its textual content. We show that such techniques can reveal correlations between the decision process and other quality proxy measures, uncovering potential biases of the review process. Finally, we discuss the opportunities, but also the potential unintended consequences of these techniques in terms of algorithmic bias and ethical concerns.

Priority 3: RRA needs to anticipate and keep pace with new tools and technologies of measurement and evaluation

優先課題3: RRAは、測定・評価の新しいツールや技術を予測し、それに対応していく必要がある。



<http://researchonresearch.org>  
[j.wilsdon@sheffield.ac.uk](mailto:j.wilsdon@sheffield.ac.uk)  
[@RoRIInstitute](https://twitter.com/RoRIInstitute)