



Responsible metrics for science

Stellenbosch, 2 November 2016

Paul Wouters



Universiteit
Leiden

SCISTIP

Evaluation Gap

- discrepancy between evaluation criteria and the social and economic functions of science
- evaluation methods (esp. qualitative) have not adapted to increased scale of research
- available quantitative measures are often not applicable at the individual level
- lack of recognition for new types of work that researchers need to perform

Principles for current metrics

An abstract graphic composed of several overlapping blue shapes. On the left, there is a large blue circle. To its right, a blue line extends from the top edge, curves downwards, and then connects to a blue circle. From the bottom of this circle, another blue line extends downwards and to the right, connecting to a larger blue circle. The background is white, and the overall style is clean and modern.

Across the research community, the description, production and consumption of 'metrics' remains contested and open to misunderstandings.



COMMENT

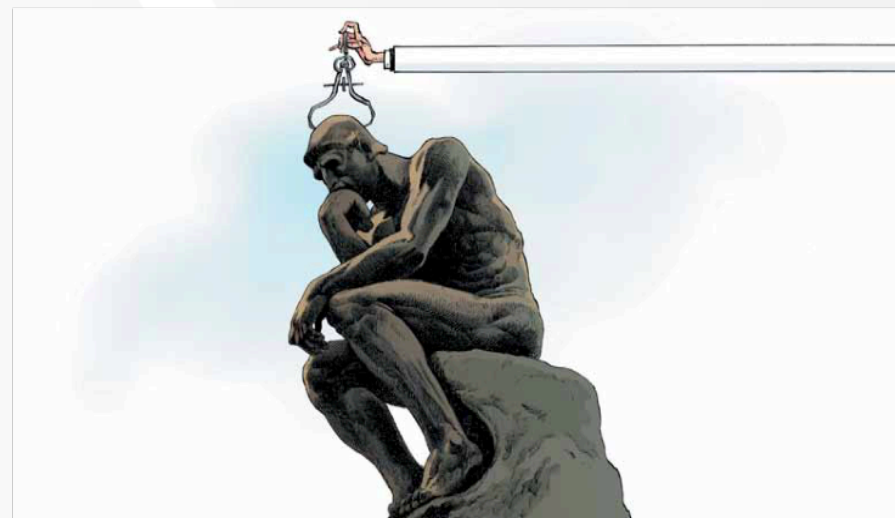
SUSTAINABILITY Data needed to drive UN development goals **p.432**



CONSERVATION Economics and environmental catastrophe **p.434**

GEOLOGY Questions raised over proposed Anthropocene dates **p.436**

HISTORY Music inspired Newton to add more colours to the rainbow **p.438**



The Leiden Manifesto for research metrics

Use these ten principles to guide research evaluation, urge **Diana Hicks, Paul Wouters** and colleagues.

Data are increasingly used to govern science. Research evaluations that were once bespoke and performed by peers are now routine and reliant on metrics¹. The problem is that evaluation is now led by the data rather than by judgement. Metrics have proliferated: usually well intentioned, not always well informed, often ill applied. We risk damaging the system with the very tools designed to improve it, as evaluation is increasingly implemented by organizations without knowledge of, or

advice on, good practice and interpretation. Before 2000, there was the Science Citation Index on CD-ROM from the Institute 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 (beta version released in 2004). Web-based tools to easily compare institutional research productivity and impact

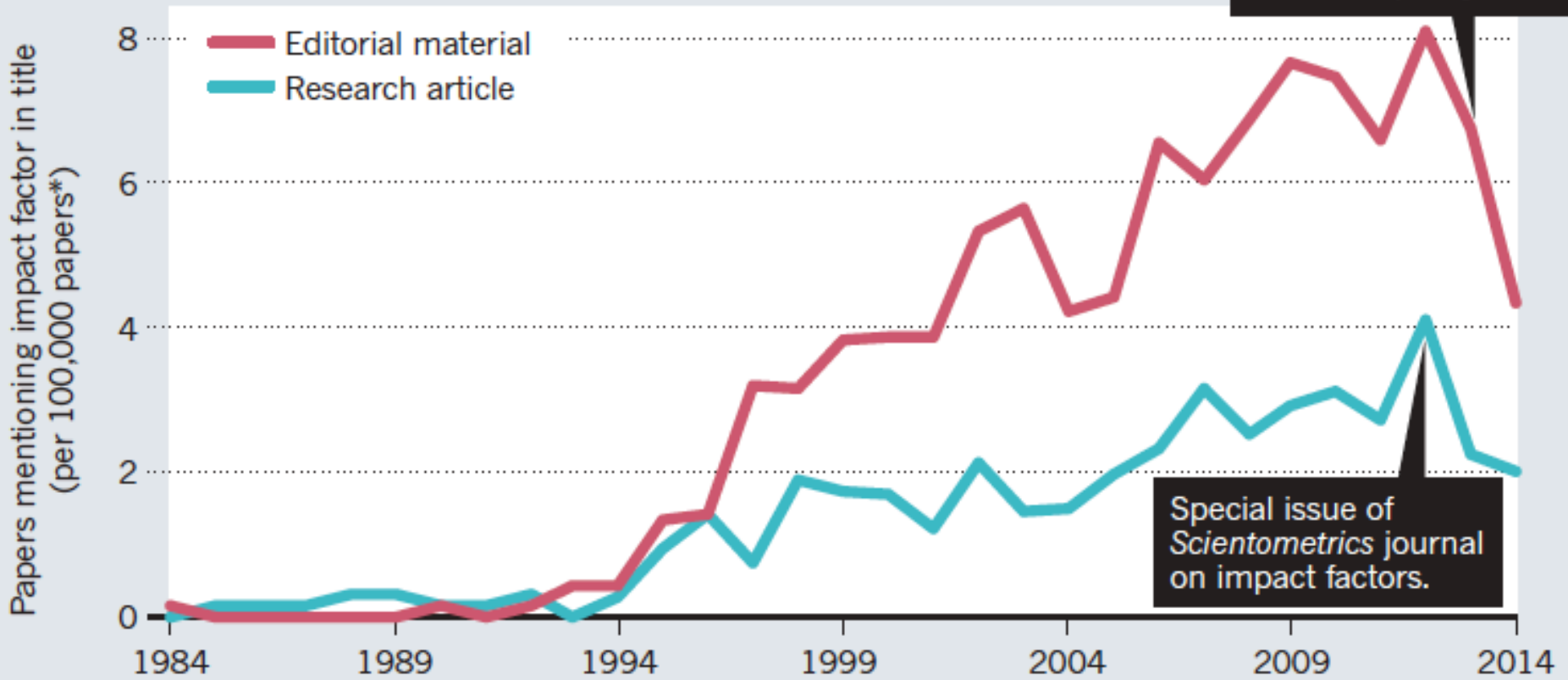
were introduced, such as InCites (using the Web of Science) and SciVal (using Scopus), as well as software to analyse individual citation profiles using Google Scholar (Publish or Perish, released in 2007). In 2005, Jorge Hirsch, a physicist at the University of California, San Diego, proposed the *h*-index, popularizing citation counting for individual researchers. Interest in the journal impact factor grew steadily after 1995 (see 'Impact factor obsession'). Lately, metrics related to social usage ▶

ILLUSTRATION BY DAVID WOODS

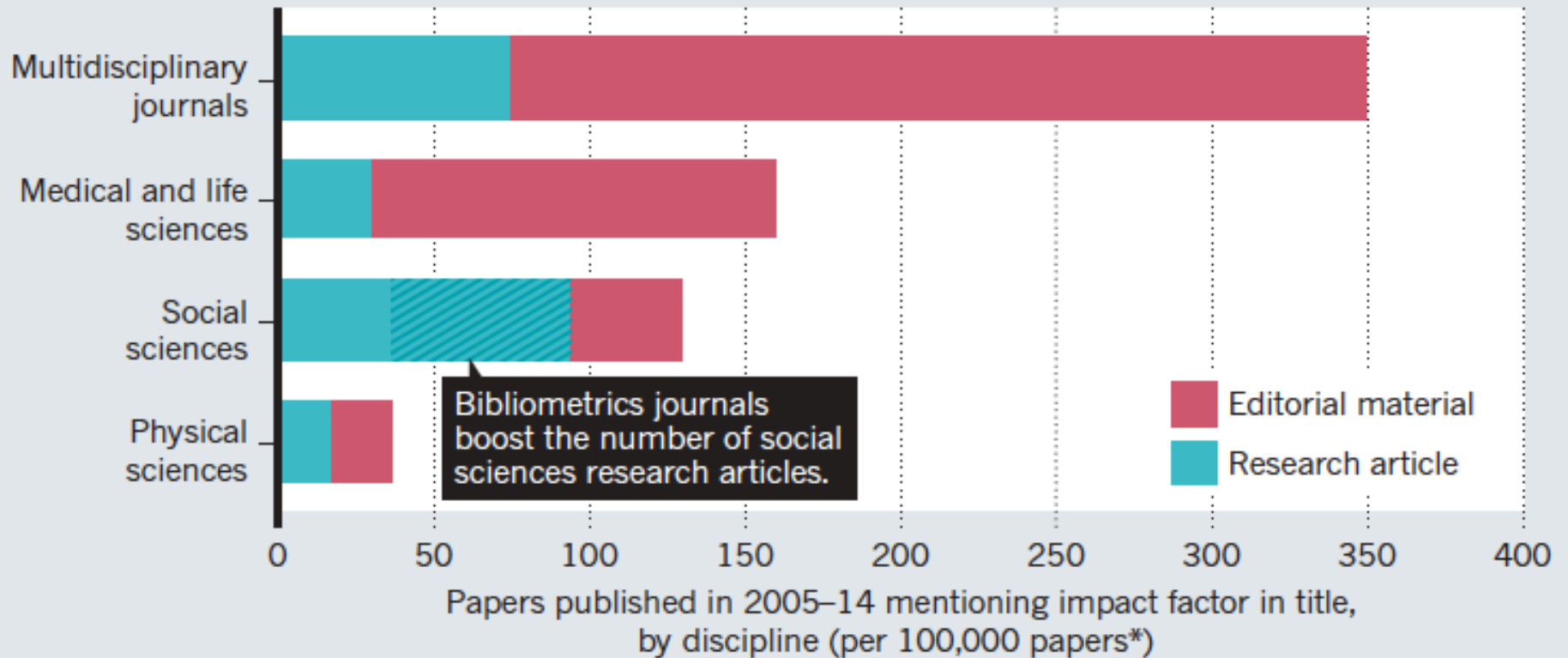
IMPACT-FACTOR OBSESSION

Soaring interest in one crude measure — the average citation counts of items published in a journal in the past two years — illustrates the crisis in research evaluation.

1 ARTICLES MENTIONING 'IMPACT FACTOR' IN TITLE



2 WHO IS MOST OBSESSED?



*Indexed in the Web of Science. †DORA, San Francisco Declaration on Research Assessment

The Leiden Manifesto

- Quantitative evaluation should support expert assessment.
- Measure performance in accordance with the research mission.
- Protect excellence in locally relevant research
- Keep data collection and analytical processes open, transparent and simple.
- Allow for data verification
- Account for variation by field in publication and citation practices
- Data should be interpreted taking into account the difficulty of credit assignment in the case of multi-authored publications.
- Base assessment of individual researchers on *qualitative* judgment.
- False precision should be avoided (eg. the JIF).
- Systemic effects of the assessment and the indicators should be taken into account and indicators should be updated regularly

The Metric Tide

<http://www.hefce.ac.uk/rsrch/metrics/>

Report of the Independent Review
of the Role of Metrics in
Assessment and Manag

The Metric Tide

Literature Review

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

July 2015

The Metric Tide

Correlation analysis of REF2014 scores and metrics

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

July 2015

Responsible metrics

Responsible metrics can be understood in terms of:

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



**But we need
more data**

Research leaders need more, not less, strategic intelligence

- Increasing demand for information about research:
 - hyper competition for funding
 - globalization
 - industry – academic partnerships
 - interdisciplinary research challenges
 - institutional demands on research & university management
- Increased supply of data about research:
 - web based research
 - deluge of data producing machines and sensors
 - increased social scale of research: international teams
 - large scale databases of publications, data, and applications

New trends in assessment

- Increased bibliometric services at university level available through databases
- Increased self-assessment via “gratis bibliometrics” on the web (h-index; publish or perish; etc.)
- Emergence of altmetrics
- Increased demand for bibliometrics at the level of the individual researcher
- Societal impact measurements required
- Career advice – where to publish?

Key challenges in research information system building

- Will the information infrastructure contain high quality data and indicators?
- Will it enable and support context- and mission-sensitive research assessments?
- Will it enable application of research information for primary research purposes (eg in VREs)?
- Will the public sector remain master in its own house or will it hand over control to the private sector?
- Will it be possible to truly open up the research agenda to all stakeholders – open science in a democratic society?

Open Science

An abstract graphic composed of blue geometric shapes. It features a large circle on the right side, with several thick blue lines radiating from its center towards the left and top-left. These lines intersect with a large, light blue circular arc that spans across the middle of the page. The background is a solid dark blue on the left and a light blue gradient on the right.

Measuring is changing

- What counts as excellence is shaped by how we measure and define “excellence”
- What counts as impact is shaped by how we measure and define “impact”
- *Qualities* and *interactions* are the foundation for “excellence” and “impact” so we should understand those more fundamental processes first
- We need different indicators at different levels in the scientific system to inform wise management that strikes the right balance between trust and control
- Context crucial for effective data standardization

Ambitions for Open Science

- More comprehensive measurement of traditional scientific publications (eg Mendeley)
- Recognizing and capturing the diversity of scientific output including new forms (eg software and blogs)
- Opening up the whole scientific publication system (open access) and more interactive communication
- Opening up the very core of knowledge creation and its role in higher education and innovation (participatory science)

Responsible Metrics for Open Science

- EU Expert Group on Altmetrics, chaired by James Wilsdon
- Aim: to develop a framework for responsible metrics and altmetrics for research management and evaluation, which can be incorporated into the successor framework to Horizon 2020
- Call for evidence:
https://ec.europa.eu/research/openscience/index.cfm?pg=altmetrics_eg
- A new type of metrics for a new type of science

Fewer numbers, better science

Scientific quality is hard to define, and numbers are easy to look at. But bibliometrics are warping science — encouraging quantity over quality. Leaders at two research institutions describe how they do things differently.

REDEFINE EXCELLENCE

Fix incentives to fix science

*Rinze Benedictus and
Frank Miedema*

An obsession with metrics pervades science. Our institution, the University Medical Center Utrecht in the Netherlands, is not exempt. On our website, we proudly declare that we

publish about 2,600 peer-reviewed scientific publications per year, with higher than average citation rates.

A few years ago, an evaluation committee spent hours discussing which of several faculty members to promote, only to settle on the two who had already been awarded particularly prestigious grants. Meanwhile, faculty members who spent time crafting policy advice had a hard time explaining how this added to their scientific output, even when it affected clinical decisions across the country.

Publications that directly influenced patient care were weighted no higher in evaluations than any other paper, and ▶

Context counts

- Responsible metrics is *not* supposed to be a universal standard
- Responsible metrics should be responsive and inclusive metrics
- Measuring means changing
- The context shapes what responsible metrics means:
 - the urgency of social problems (poverty, inequality, unemployment and corruption)
 - local research and educational missions
 - the local appropriation of “the global”
 - the values embedded in the policies and communities