CWTS
Meaningful metrics

Responsible Research Evaluation: lessons and perspectives

Czech Academy of Science

Prague, 18 November 2016

Paul Wouters

Universiteit Leiden

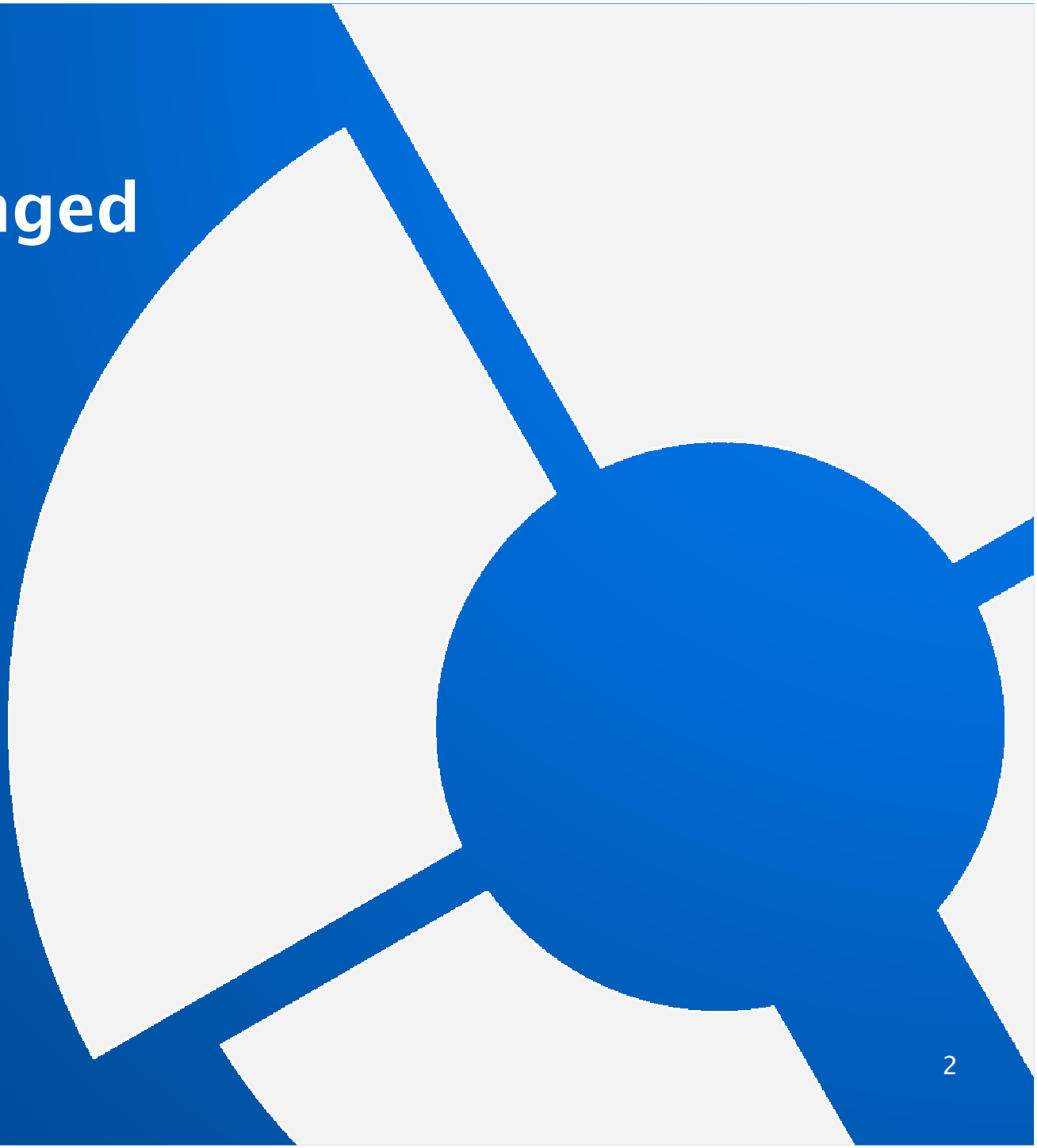
Centre for Science and Technology Studies (CWTS)

- Research center at Leiden University focusing on quantitative studies of science (bibliometrics and scientometrics)

- Bibliometric contract research
  - Monitoring & evaluation
  - Advanced analytics
  - Training & education

CWTS

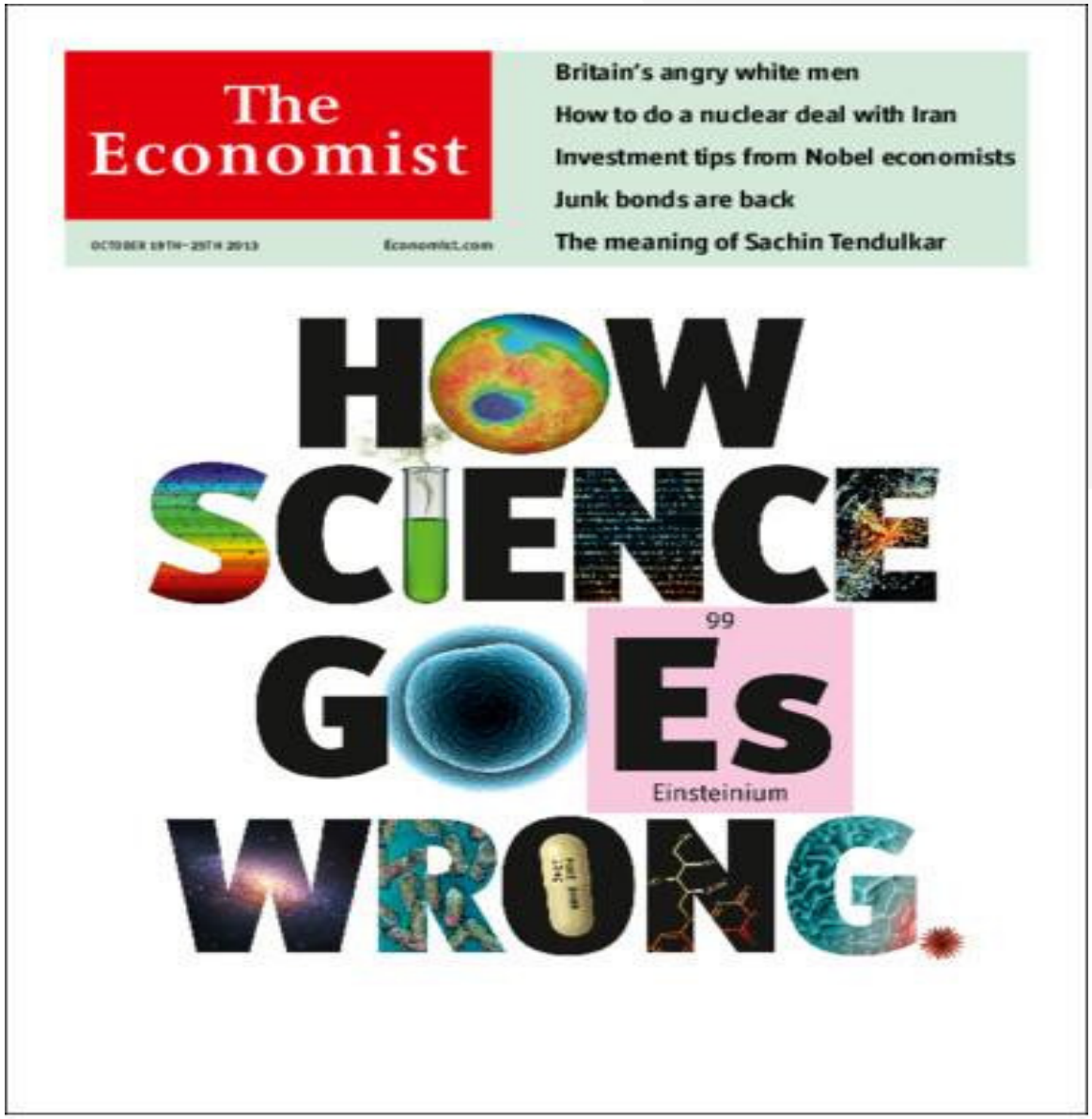
The Challenged University



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

CWTS



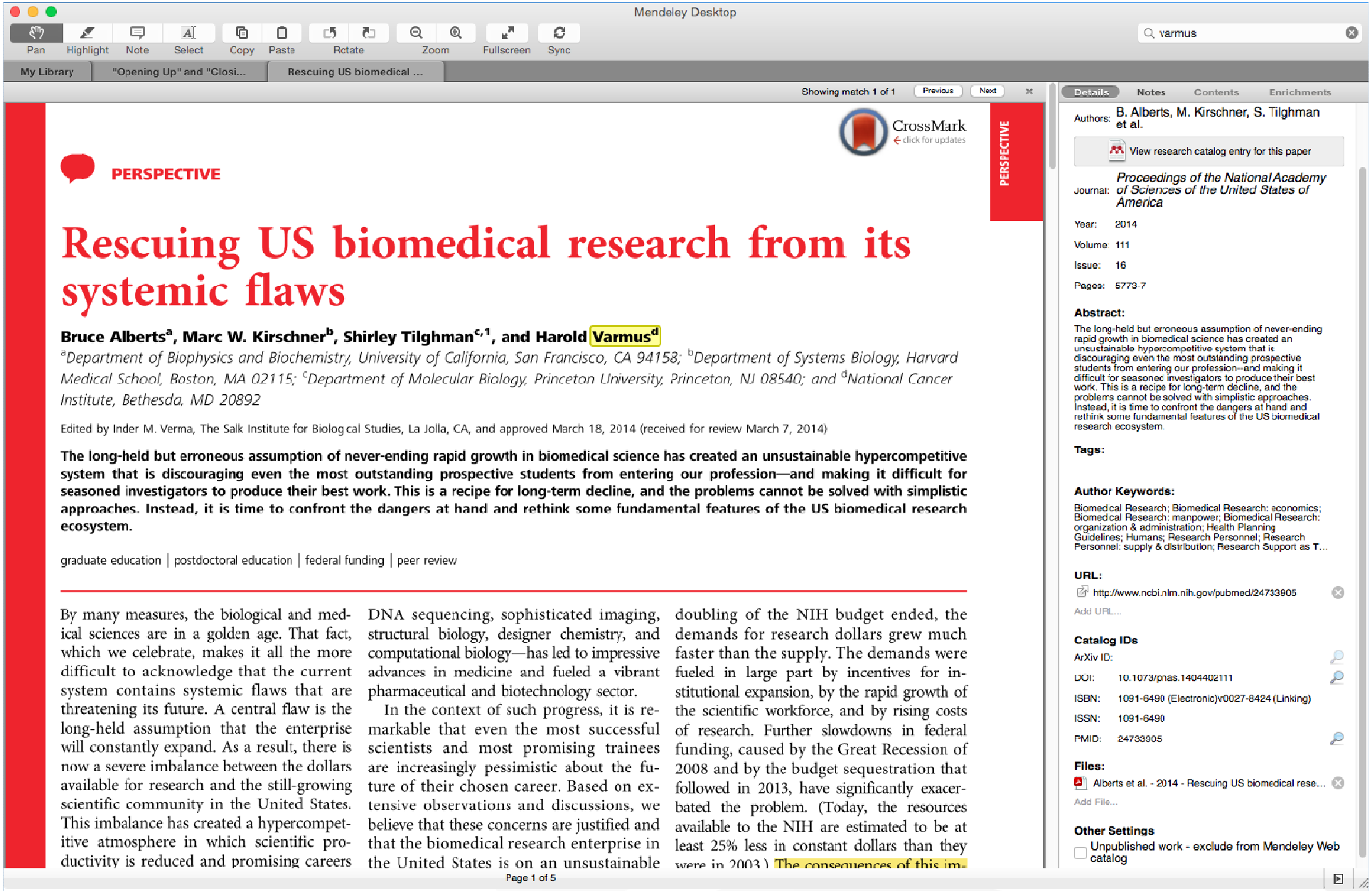
OCTOBER 19TH - 25TH 2013 Worldwide cover

CWTS

A SIMPLE idea underpins science: “trust, but verify”. Results should always be subject to challenge from experiment. That simple but powerful idea has generated a vast body of knowledge. Since its birth in the 17th century, modern science has changed the world beyond recognition, and overwhelmingly for the better. But success can breed complacency. Modern scientists are doing too much trusting and not enough verifying—to the detriment of the whole of science, and of humanity.

Too many of the findings that fill the academic ether are the result of shoddy experiments or poor analysis (see article (http://www.economist.com/news/briefing/21588057-scientists- think-science-self-correcting-alarming-degree-it-not-trouble)). A rule of thumb among biotechnology venture-capitalists is that half of published research cannot be replicated.

CWTS


Mendeley Desktop
Pan
Highlight
Note
Select
Copy
Paste
Rotate
Zoom
Fullscreen
Sync
varmus
My Library
"Opening Up" and "Closi...
Rescuing US biomedical ...
Showing match 1 of 1
Previous
Next
Details
Notes
Contents
Enrichments
CrossMark
click for updates
PERSPECTIVE
PERSPECTIVE
Rescuing US biomedical research from its systemic flaws
Bruce Alberts[a], Marc W. Kirschner[b], Shirley Tilghman[c,1], and Harold Varmus[d]
[a]Department of Biophysics and Biochemistry, University of California, San Francisco, CA 94158; [b]Department of Systems Biology, Harvard Medical School, Boston, MA 02115; [c]Department of Molecular Biology, Princeton University, Princeton, NJ 08540; and [d]National Cancer Institute, Bethesda, MD 20892
Edited by Inder M. Verma, The Salk Institute for Biological Studies, La Jolla, CA, and approved March 18, 2014 (received for review March 7, 2014)
The long-held but erroneous assumption of never-ending rapid growth in biomedical science has created an unsustainable hypercompetitive system that is discouraging even the most outstanding prospective students from entering our profession—and making it difficult for seasoned investigators to produce their best work. This is a recipe for long-term decline, and the problems cannot be solved with simplistic approaches. Instead, it is time to confront the dangers at hand and rethink some fundamental features of the US biomedical research ecosystem.
graduate education | postdoctoral education | federal funding | peer review
By many measures, the biological and medical sciences are in a golden age. That fact, which we celebrate, makes it all the more difficult to acknowledge that the current system contains systemic flaws that are threatening its future. A central flaw is the long-held assumption that the enterprise will constantly expand. As a result, there is now a severe imbalance between the dollars available for research and the still-growing scientific community in the United States. This imbalance has created a hypercompetitive atmosphere in which scientific productivity is reduced and promising careers
DNA sequencing, sophisticated imaging, structural biology, designer chemistry, and computational biology—has led to impressive advances in medicine and fueled a vibrant pharmaceutical and biotechnology sector.
In the context of such progress, it is remarkable that even the most successful scientists and most promising trainees are increasingly pessimistic about the future of their chosen career. Based on extensive observations and discussions, we believe that these concerns are justified and that the biomedical research enterprise in the United States is on an unsustainable
doubling of the NIH budget ended, the demands for research dollars grew much faster than the supply. The demands were fueled in large part by incentives for institutional expansion, by the rapid growth of the scientific workforce, and by rising costs of research. Further slowdowns in federal funding, caused by the Great Recession of 2008 and by the budget sequestration that followed in 2013, have significantly exacerbated the problem. (Today, the resources available to the NIH are estimated to be at least 25% less in constant dollars than they
Page 1 of 5
Authors: B. Alberts, M. Kirschner, S. Tilghman et al.
View research catalog entry for this paper
Journal: Proceedings of the National Academy of Sciences of the United States of America
Year: 2014
Volume: 111
Issue: 16
Pages: 5773-7
Abstract:
The long-held but erroneous assumption of never-ending rapid growth in biomedical science has created an unsustainable hypercompetitive system that is discouraging even the most outstanding prospective students from entering our profession--and making it difficult for seasonec investigators to produce their best work. This is a recipe for long-term decline, and the problems cannot be solved with simplistic approaches. Instead, it is time to confront the dangers at hand and rethink some fundamental features of the US biomedical research ecosystem.
Tags:
Author Keywords:
Biomedical Research; Biomedical Research: economics; Biomedical Research: manpower; Biomedical Research: organization & administration; Health Planning Guidelines; Humans; Research Personnel; Research Personnel: supply & distribution; Research Support as T...
URL:
http://www.ncbi.nlm.nih.gov/pubmed/24733905
Add URL...
Catalog IDs
ArXiv ID:
DOI: 10.1073/pnas.1404402111
ISBN: 1091-6490 (Electronic)\r0027-8424 (Linking)
ISSN: 1091-6490
PMID: 24733905
Files:
Alberts et al. - 2014 - Rescuing US biomedical rese...
Add File...
Other Settings
Unpublished work - exclude from Mendeley Web catalog

- A severe imbalance between the dollars available for research and the still-growing scientific community in the United States.
- The training pipe-line produces more scientists than relevant positions in academia, government, and the private sector are capable of absorbing
- Hyper-competition for the resources and positions that are required to conduct science suppresses the creativity, cooperation, risk-taking, and original thinking required to make fundamental discoveries.
- Overvaluing translational research is detracting from an equivalent appreciation of fundamental research of broad applicability
- As competition for jobs and promotions increases, the inflated value given to publishing in a small number of so-called “high impact” journals has put pressure on authors to rush into print, cut corners, exaggerate their findings, and overstate the significance of their work.
- Today, time for reflection is a disappearing luxury for the scientific community.
- The quality of evaluation has declined

CWTS

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Published research findings are sometimes refuted by subsequent evidence, with ensuing confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims [6–8]. However, this should not be surprising. It can be proven that most claimed research findings are false. Here I will examine the key factors that influence this problem and some corollaries thereof.

The Essay section contains opinion pieces on topics of broad interest to a general medical audience.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful. "Negative" is actually a misnomer, and the misinterpretation is widespread. However, here we will target relationships that investigators claim exist, rather than null findings.

As has been shown previously, the probability that a research finding is indeed true depends on the prior probability of it being true (before doing the study), the statistical power of the study, and the level of statistical significance [10,11]. Consider a 2 × 2 table in which research findings are compared against the gold standard of true relationships in a scientific field. In a research field both true and false hypotheses can be made about the presence of relationships. Let R be the ratio of the number of "true relationships" to "no relationships" among those tested in the field. R is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R + 1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2 × 2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2 × 2 table, one gets PPV = $(1 - \beta)R/(R - \beta R + \alpha)$. A research finding is thus

Citation: Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8): e124.

Abbreviation: PPV, positive predictive value

John P. A. Ioannidis is in the Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, and Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States of America. E-mail: jioannid@cc.uoi.gr

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020124

NATURE | NEWS

Over half of psychology studies fail reproducibility test

Largest replication study to date casts doubt on many published positive results.

Monya Baker

27 August 2015

Don't trust everything you read in the psychology literature. In fact, two thirds of it should probably be distrusted.

Brian Nosek's team set out to replicate scores of studies.

In the biggest project of its kind, Brian Nosek, a social psychologist and head of the Center for Open Science in Charlottesville, Virginia, and 269 co-authors repeated work reported in 98 original papers from three psychology journals, to see if they independently came up with the same results.

The studies they took on ranged from whether expressing insecurities perpetuates them to differences in how children and adults respond to fear stimuli, to effective ways to teach arithmetic.

According to the replicators' qualitative assessments, as previously reported by *Nature*, only 39 of

5 challenges

- Informatisation knowledge production
- Research funding system
- Publication system
- Career structures in science
- Research evaluation practices

CWTS

Strategic science

Research leaders face key questions

- How should we monitor our research?
- How can we profile ourselves to attract the right students and staff?
- How should we divide funds?
- What is our scientific and societal impact?
- What is actually our area of expertise?
- How is our research trans-disciplinary connected?

CWTS

Research leaders need 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
  - citation metrics and altmetrics

CWTS

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?

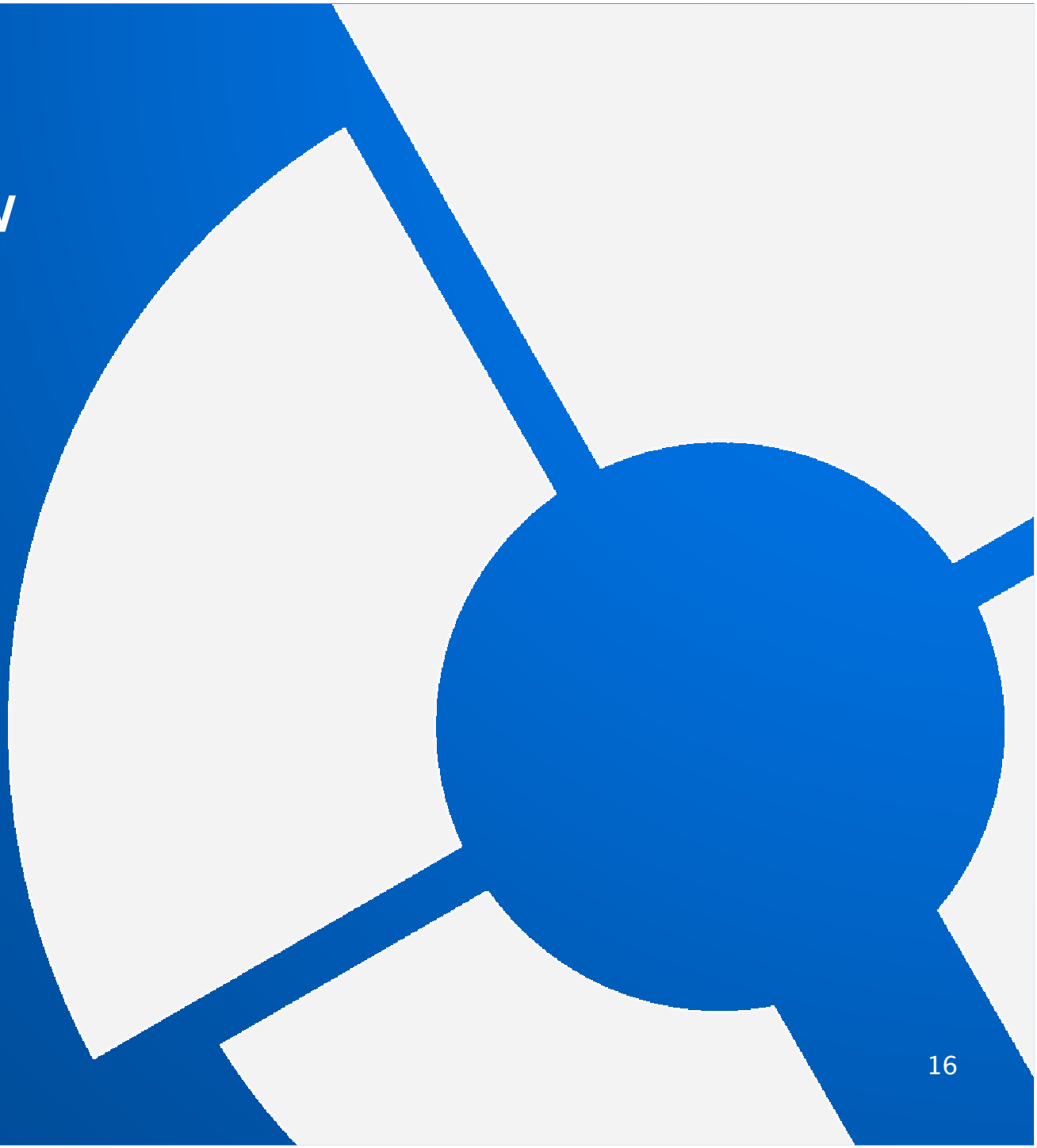
CWTS

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?

CWTS

Peer Review



Summary literature review I (on peer review)

- Peer review is an umbrella term: quite variable practices, procedures and criteria
  - Journal manuscript review
  - Funding proposal review
  - Career reviews
  - Postpublication reviews (like the REF)
- Generally, modestly positive correlations between peer review and bibliometric indicators but varies by type of review and choice of analytical dimensions
- Lack of common methodology in studies of peer review

CWTS

Summary literature review II

- Studies of relationship funding decisions and bibliometrics often suffer from circular reasoning
- Citation impact is not a measure of quality but a proxy measure of influence
- Quality is multi-dimensional, some aspects of which may be reflected in citation impact but not all
- Correlation strengths peer review and metrics vary considerably by field:
  - Weaker in humanities, technical and social sciences, and applied fields
  - Influenced by database coverage and citation culture

CWTS

Summary literature review III

- Peer review and bibliometric data not completely independent – intricate mutually shaping relationships:
  - Citation data based on citing decisions
  - Peer communities drawn from the citing and cited population
  - Citing decisions influenced by role of citation counts in assessments
  - Peer judgement influenced and shaped by “citation impressions”
- Strengths and weaknesses of peer review and bibliometrics may be complementary:
  - Bibliometrics may add value in post-publication assessments (like the REF) since peer review must be very selective

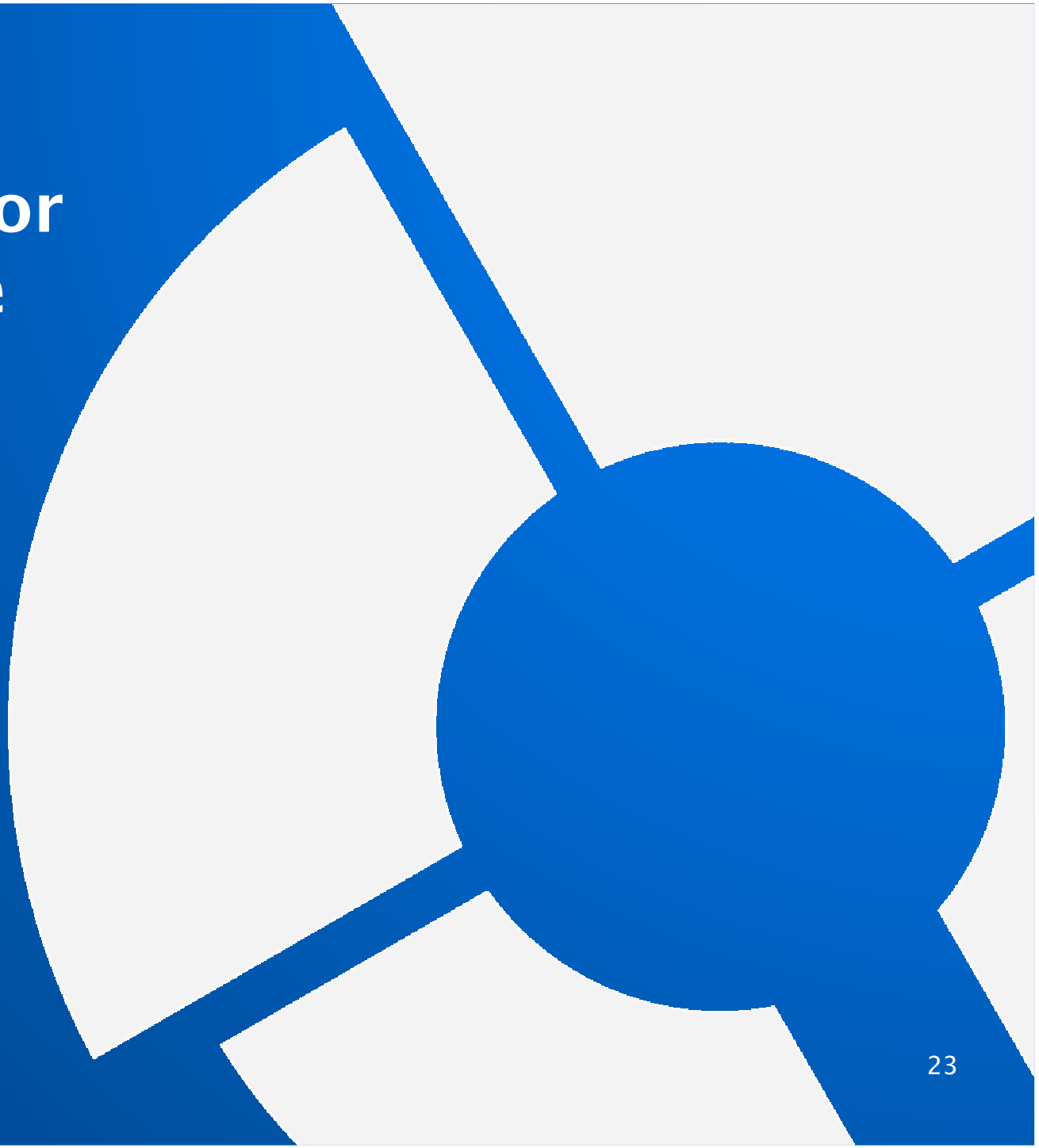
CWTS

Weaknesses of Peer Review	Strengths of Peer Review
• It is slow, inefficient and expensive, although most costs are hidden; • Human judgment is subjective – which may however also be seen as a strength;[173] • It is almost by definition not transparent; • It is inconsistent, sometimes characterised as a lack of inter-rater reliability;	• Its foundation in specialised knowledge of the subject, methodology and literature relevant for specific decisions; • Its social nature; • The subjectivity of this approach could be seen as a strength (as well as a weakness);

• It is a biased process (e.g. gender bias regarding career decisions, bias against negative studies in publication decisions, bias in favour of prestigious institutes, bias in favour of dominant paradigms); • Its bias is strengthened by the Matthew effect;[174] • The process can be abused (e.g. to block competitors, to plagiarise); • It is not very good at identifying errors in data or even in detecting fraudulent research;[175] • It cannot process the complete research output of a nation and will therefore result in distorted rankings (since rankings are sensitive to the selection of submissions to the assessments);	• It can help assess elements of research which are challenging to quantify e.g. novelty; • It can deliver more nuanced and detailed understandings of research in the context of research production.

- It cannot provide information about the productivity and efficiency of the research system;
- The selection of peer reviewers may create problems because of a variety of reasons (bias, lack of experts in emerging and interdisciplinary areas, lack of experts due to the speed of research areas, etc).

Principles for responsible metrics



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

CWTS

COMMENT

SUSTAINABILITY Data needed to drive UN development goals p.432

CONSERVATION Economics and environmental catastrophe p.434

ECOLOGY Questions raised over proposed Anthropocene dates p.436

HISTORY Music inspired Newton to add more colours to the rainbow p.436

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 metrics. 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 launched: 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).

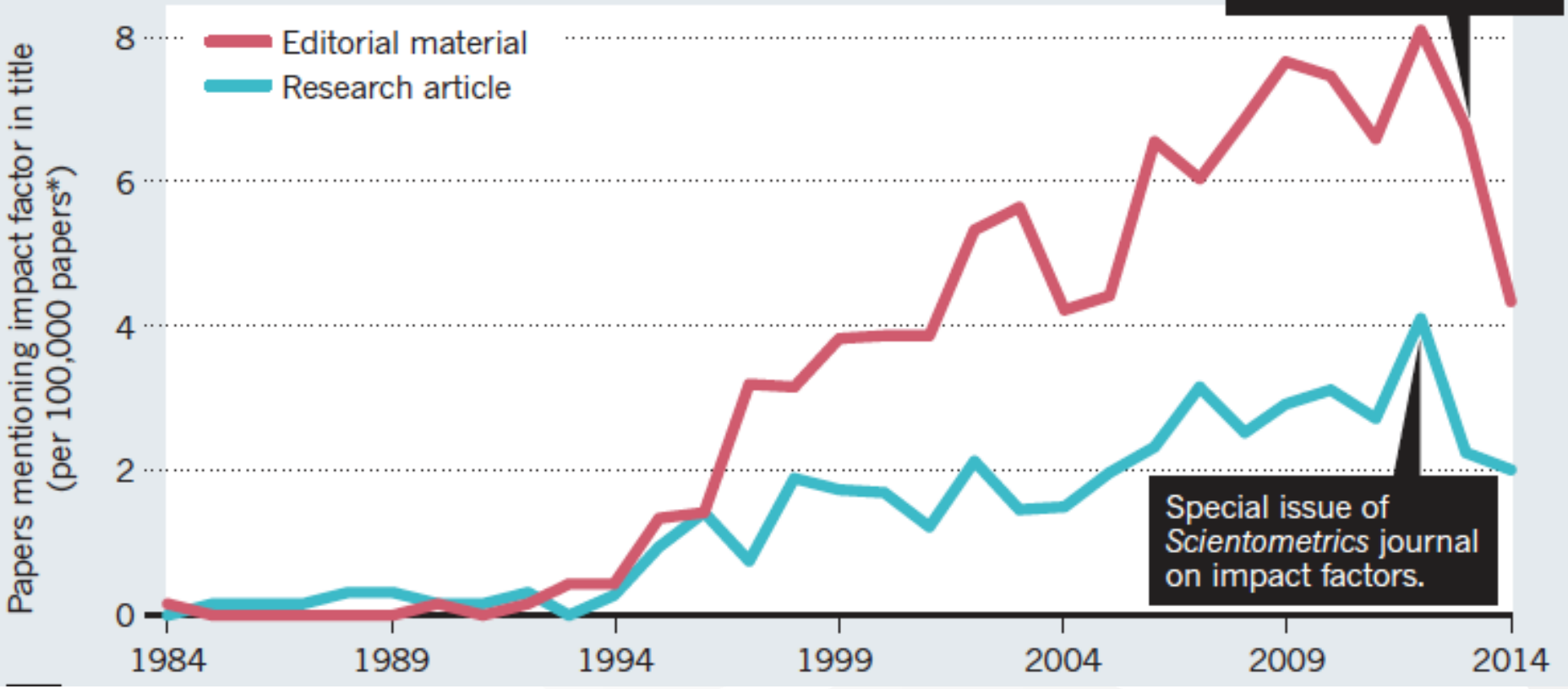
In 2005, Jorge Hirsch, a physicist at the University of California, San Diego, proposed the *h* index, popularizing citation counting for individual researchers. Interest in the journal impact factor grew steadily after 1995 (see ‘Impact factor obsession’).

Lately, metrics related to social usage ▶

IMPACT-FACTOR OBSESSION

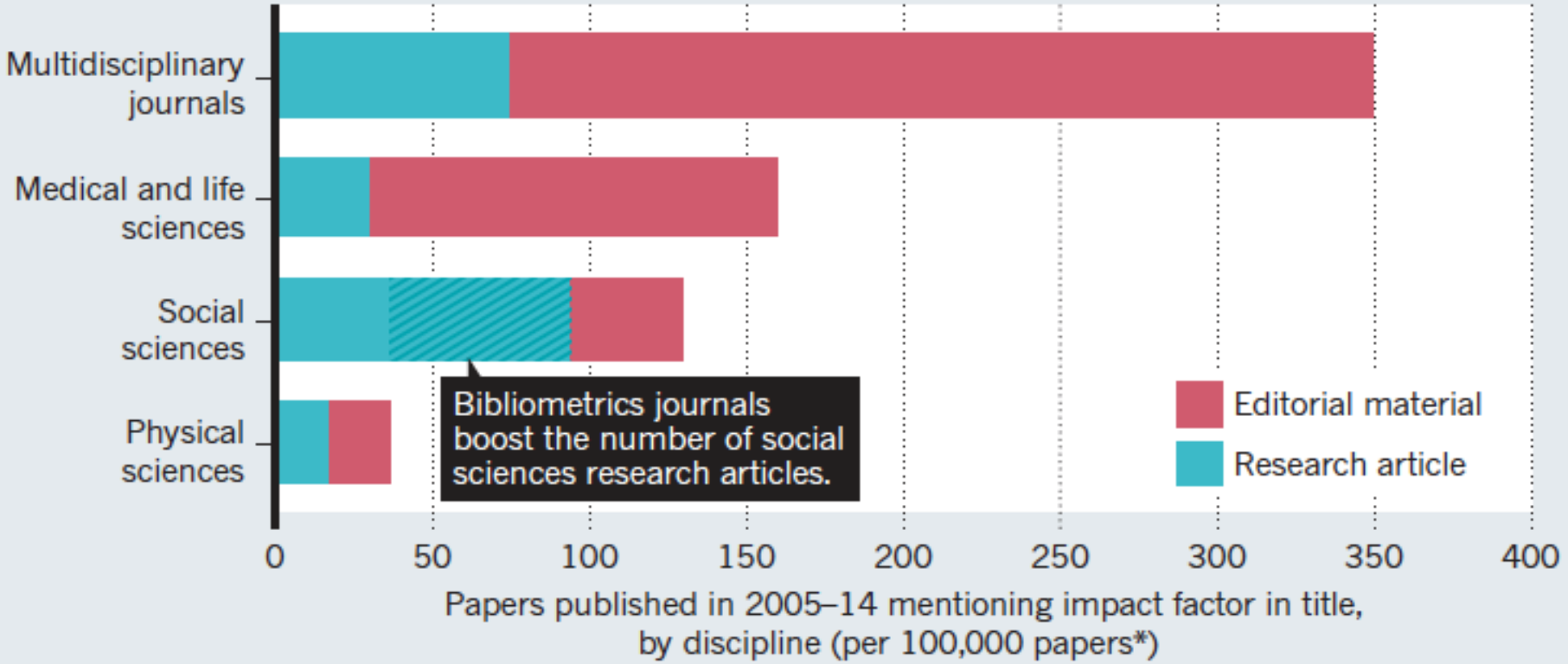
Soaring interest in one crude measure — the average citation counts of items published in a journal in the past two years — illustrates the crisis in research evaluation.

1 ARTICLES MENTIONING 'IMPACT FACTOR' IN TITLE



CWTS

2 WHO IS MOST OBSESSED?



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

CWTS

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

CWTS

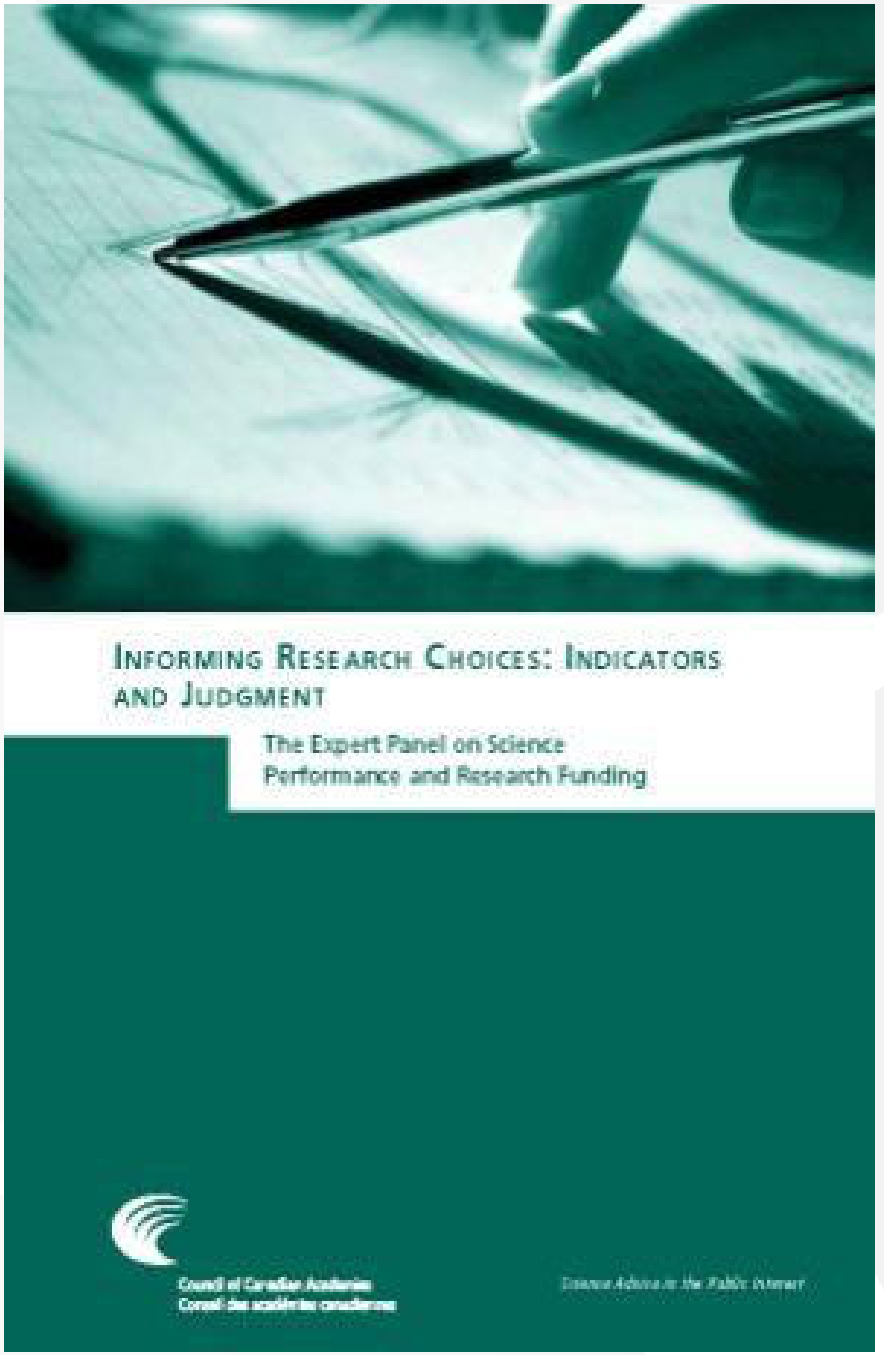
Diana Hicks (Georgia Tech), Paul Wouters (CWTS), Ismael Rafols (SPRU/Ingenio), Sarah de Rijcke and Ludo Waltman (CWTS) (2015) *Nature* 520: 429–31. doi:10.1038/520429a

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

CWTS

Peer review, despite its flaws and limitations, continues to command widespread support across disciplines. Metrics should support, not supplant expert judgement.

CWTS



Inappropriate indicators create perverse incentives. There is legitimate concern that some quantitative indicators can be gamed, or can lead to unintended consequences.

CWTS

newsblog

Nature brings you breaking news from the world of science

News & Comment › News Blog › Post

Previous post
Upsides and downsides of openness — the view from TEDGlobal

Next post
Mexico's new president aims high on science

NATURE NEWS BLOG

Record number of journals banned for boosting impact factor with self-citations

29 Jun 2012 | 19:53 BST | Posted by Richard Van Noorden | Category: Science communication

More research journals than ever are boosting their impact factors by self-citation.

Every year, Thomson Reuters, the firm that publishes the impact-factor rankings, takes action against the most extreme offenders by banning them from the latest lists. It lets them in again, suitably chastened, a couple of years later.

And this year, the apparent game playing has reached an all-time high. Thomson Reuters has excluded 51 journals from its 2011 list, published yesterday; 28 of the banned are new offenders, says Marie McVeigh, director of the firm's annual *Journal Citation Reports* (*JCR*), and the others remain blacklisted from last year. The full list is available here for subscribers to *JCR*.

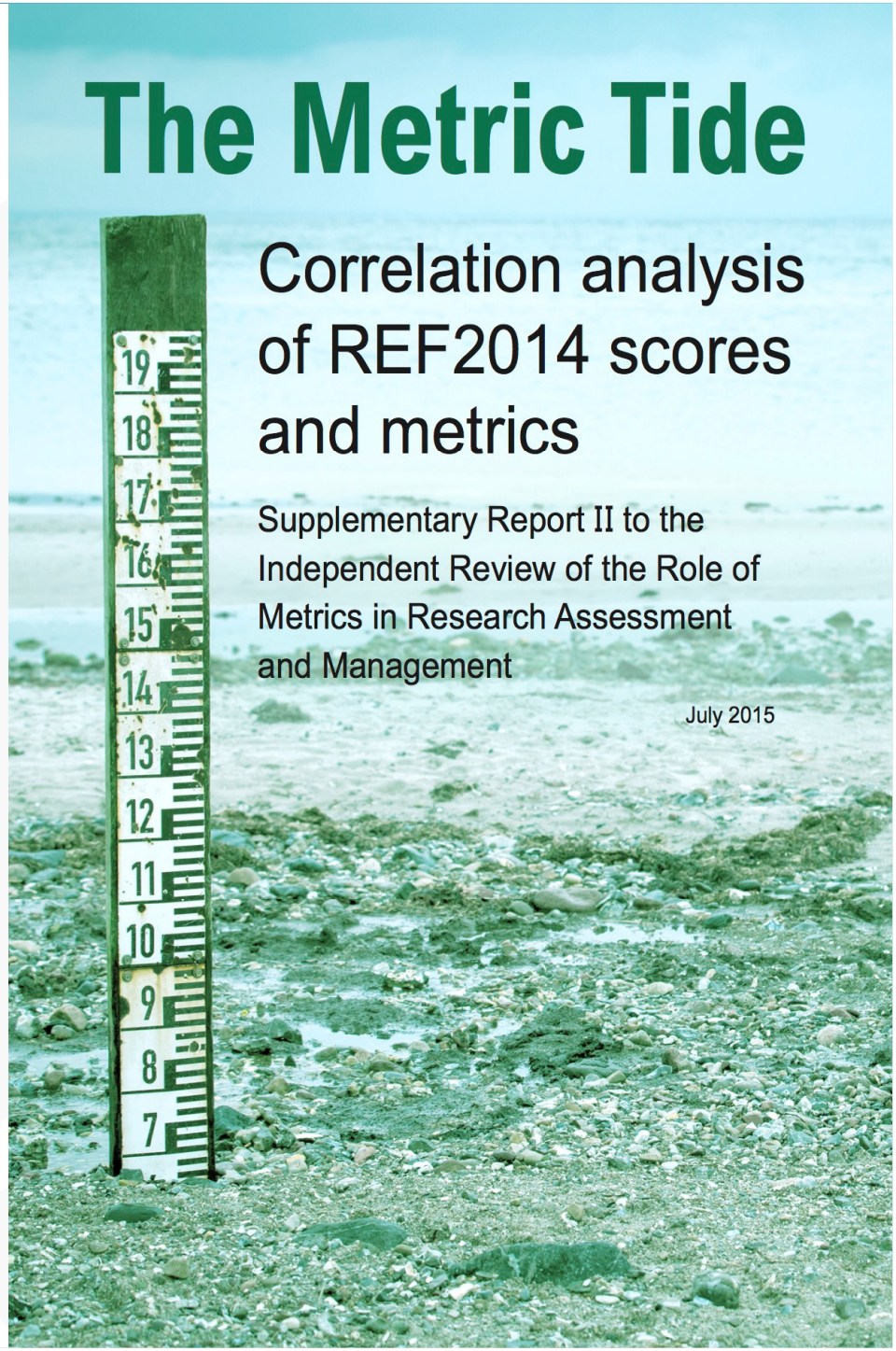
That's a substantial increase on previous years: 34 journals were excluded from the 2010 lists, compared to only 26 in 2009, 20 in 2008 and just 9 in 2007.

Almost all of those banned are excluded because of excessive self-citation, although three journals — *Cell Transplantation*, *Medical Science Monitor* and *The Scientific World Journal* — apparently worked together to cite each other and thus raise impact factors. That "cartel" was originally reported by Phil Davis on The Scholarly Kitchen, and he has today posted a follow-up article on that ban. McVeigh says that this incident, which she calls "an anomaly in citation stacking", is the only one of its kind that she has found.

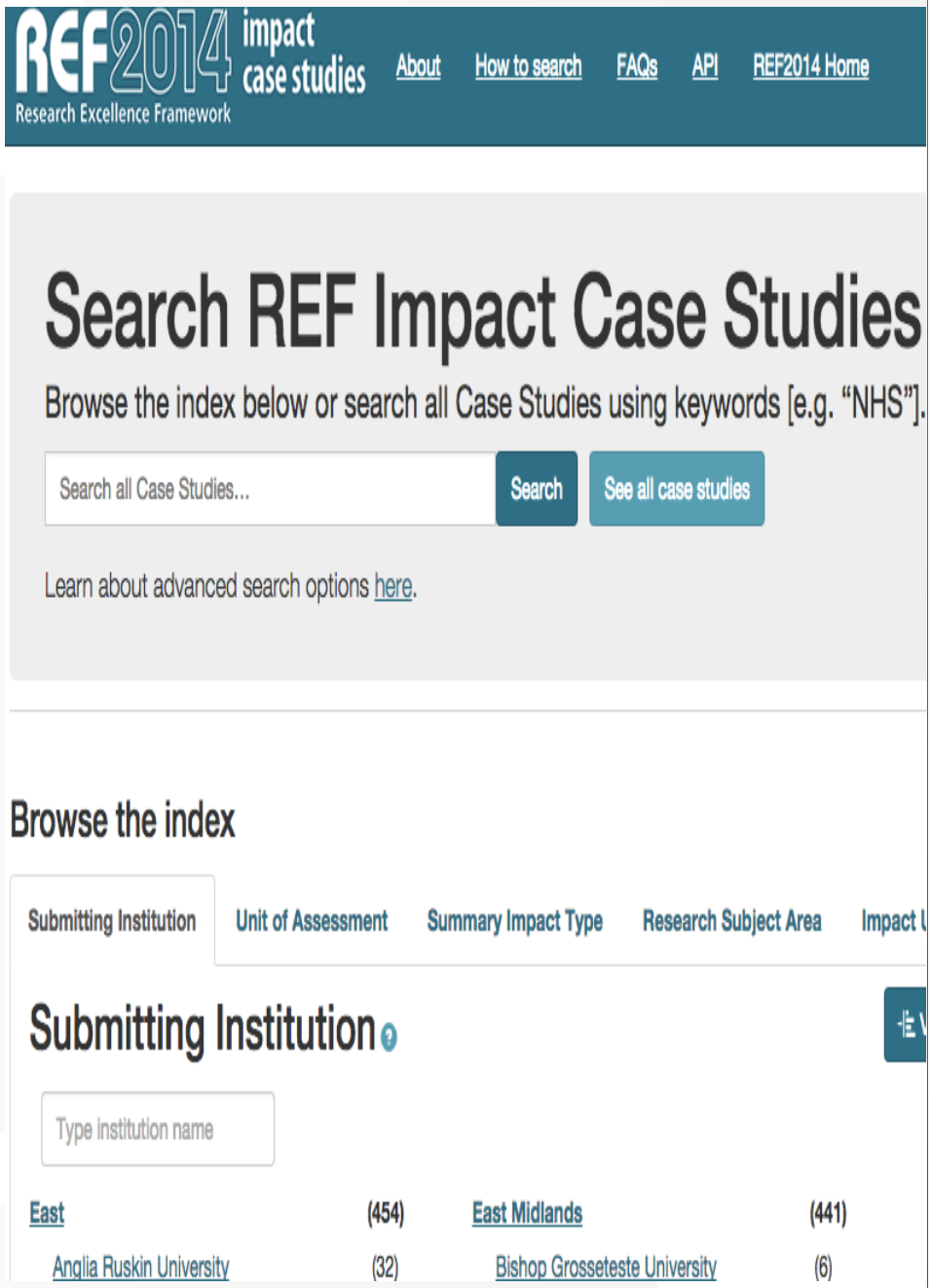
Indicators can only meet their potential if they are underpinned by an open and interoperable data infrastructure.

Our correlation analysis of the REF2014 results at output-by-author level has shown that individual metrics cannot provide a like-for-like replacement for REF peer review.

CWTS



Within the REF, it is not currently feasible to assess the quality of UOAs using quantitative indicators alone, or to replace narrative impact case studies, or the impact template.



CWTS

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.

CWTS

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

CWTS

Open Science

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)

CWTS

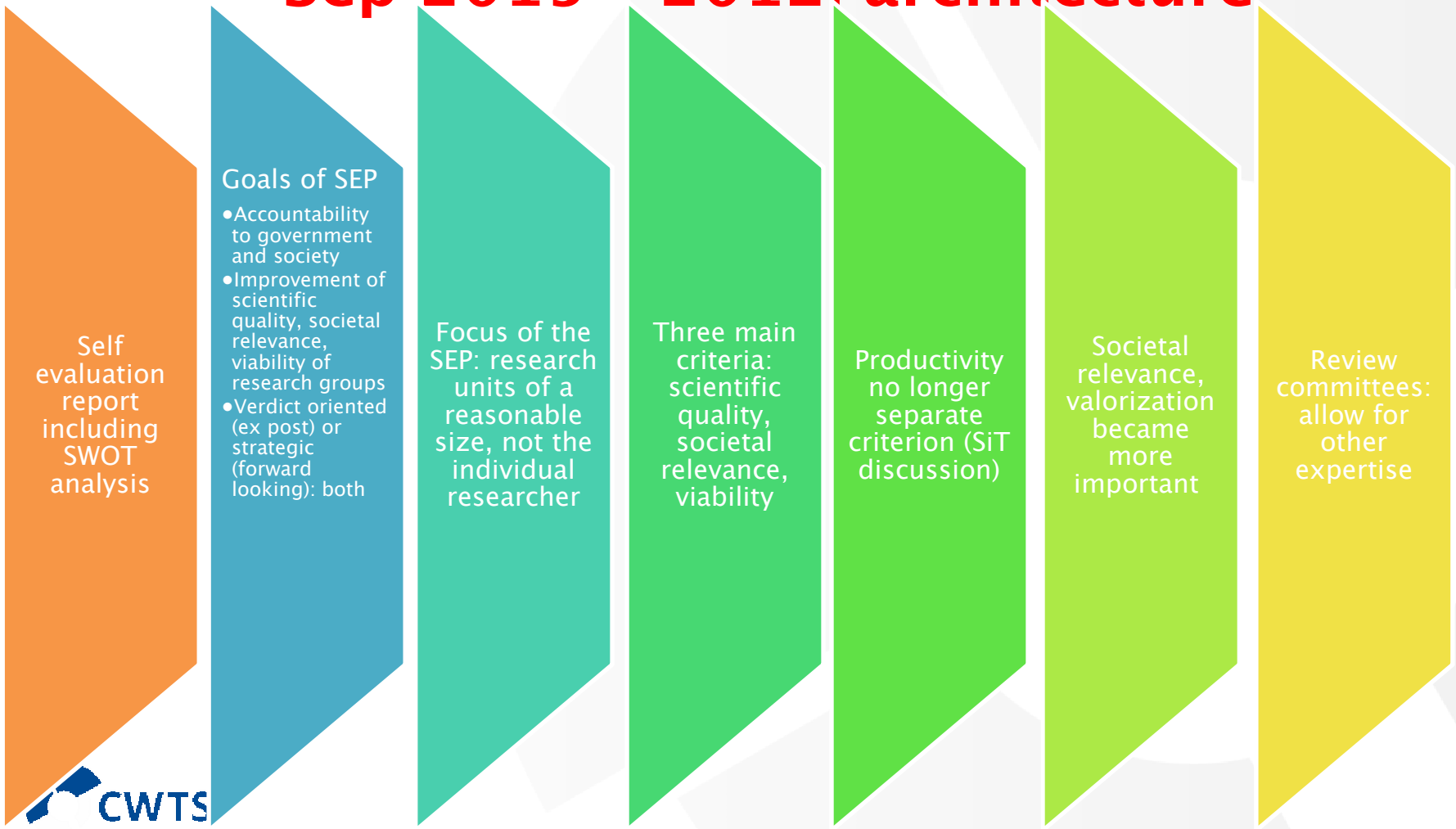
Context counts

- 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

CWTS

Standard Evaluation Protocol NL

Sep 2015 – 2012: architecture



CWTS

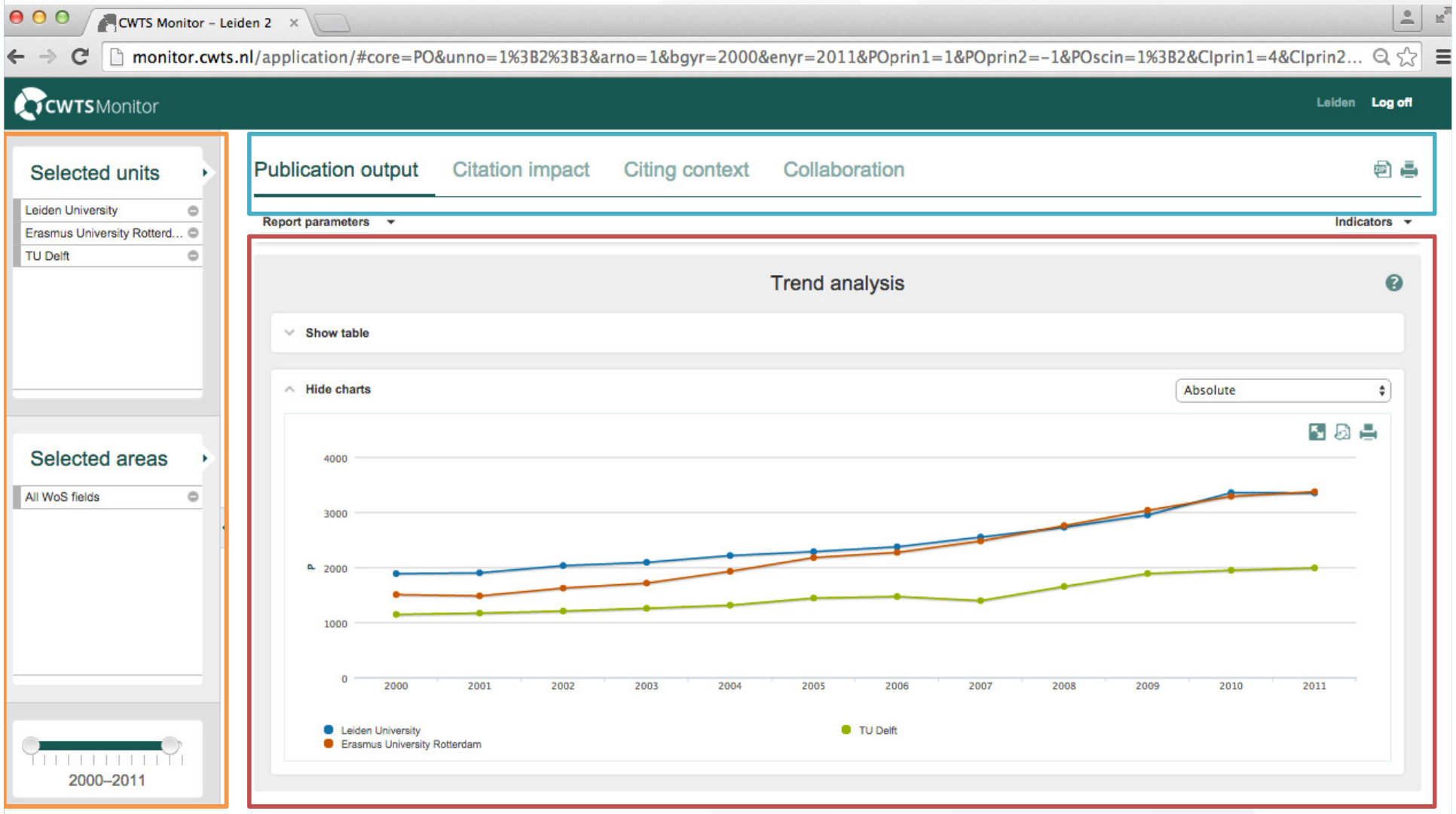
Context Sensitive Solutions

CWTS Monitor – Meaningful Metrics

- A new interactive way of bibliometric analyses
- Powerful web-based application:
  - User-friendly reporting interface
  - Robust cleaned WoS database run by CWTS
  - Fair and correct benchmarking by state-of-the-art indicators
  - Highly configurable to client's specific needs
- Professional bibliometric reporting in your hands
- Scientists affiliated to the CTWS Institute of Leiden University provide expert support

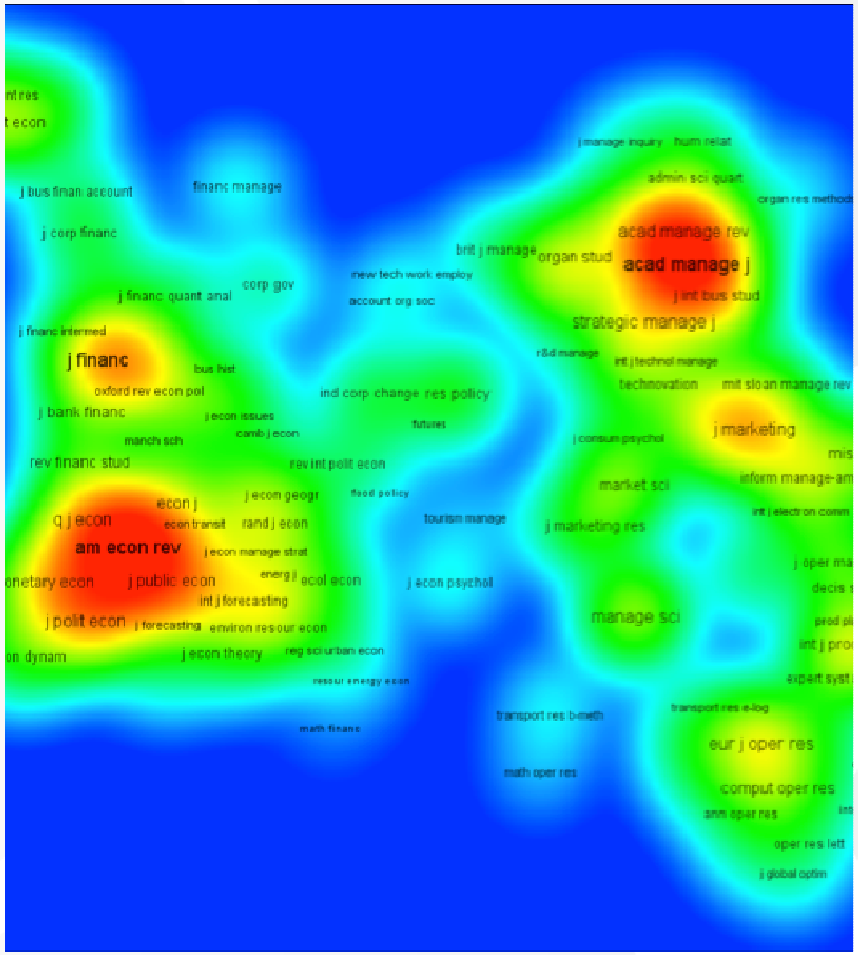
CWTS

CWTS Monitor: Select–Visualise–Conclude



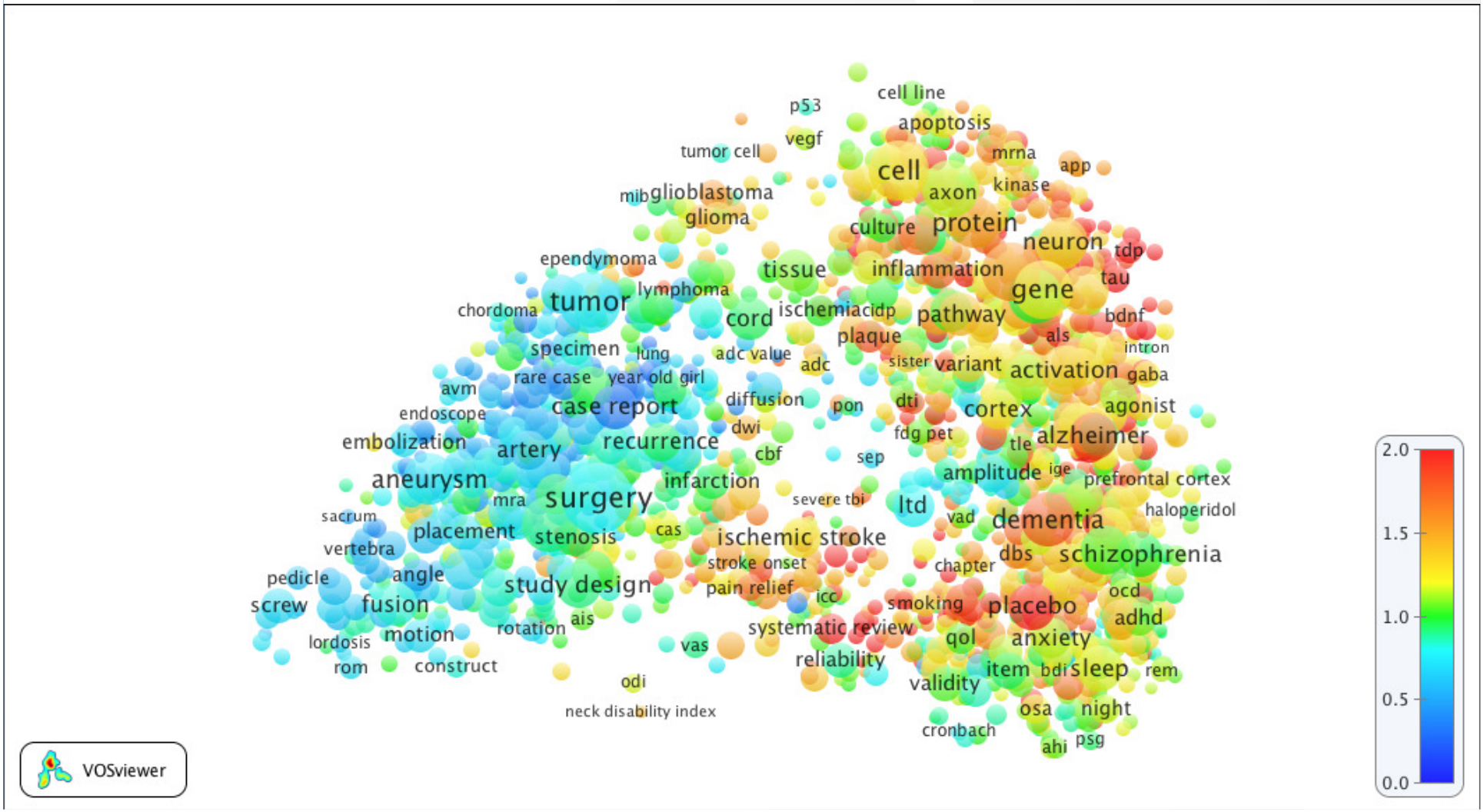
CWTS Advanced Analytics

- Tailor-made analysis based on network analysis, text mining and visualisation techniques
- Research strengths analysis
- Find blind spots/hot spots
- Identification of partners/potential new staff
- Enhanced collaborative network analysis

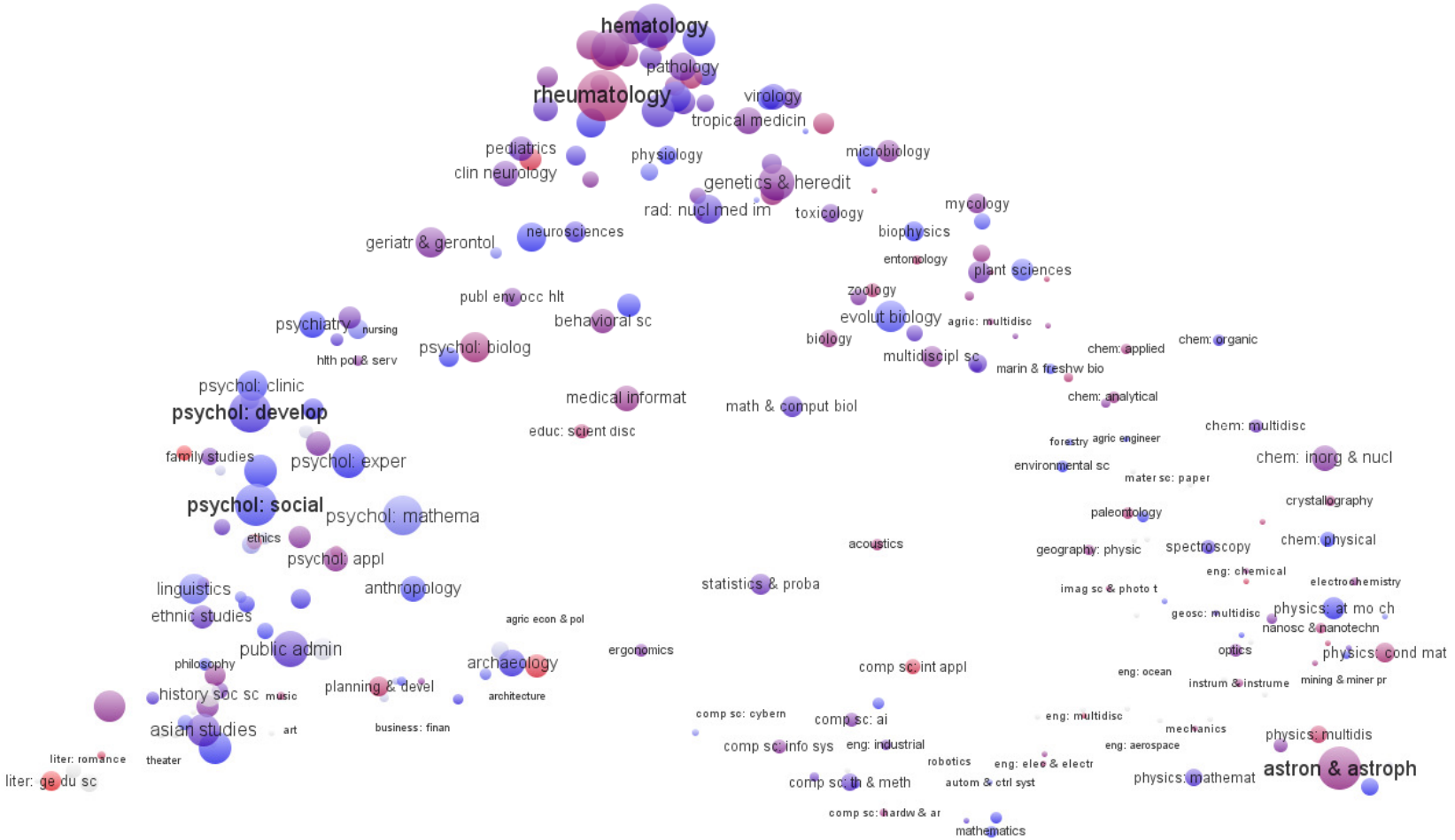


CWTS

Citation density map
Clinical neurology



Strengths and weaknesses
– University Profiles – Leiden



VOSviewer

ACUMEN portfolio

aim is to give researchers a voice in evaluation

- evidence based arguments
- shift to dialog orientation
- selection of indicators
- narrative component
- Good Evaluation Practices
- envisioned as web service

CWTS

ACUMEN Portfolio

Career Narrative

Links expertise, output, and influence together in an evidence-based argument; included content is negotiated with evaluator and tailored to the particular evaluation

Expertise

- scientific/scholarly
- technological
- communication
- organizational
- knowledge transfer
- educational

Output

- publications
- public media
- teaching
- web/social media
- data sets
- software/tools
- infrastructure
- grant proposals

Influence

- on science
- on society
- on economy
- on teaching

Evaluation Guidelines

- aimed at both researchers and evaluators
- development of evidence based arguments (what counts as evidence?)
- expanded list of research output
- establishing provenance
- taxonomy of indicators: bibliometric, webometric, altmetric
- guidance on use of indicators
- contextual considerations, such as: stage of career, discipline, and country of residence

CWTS

Narrative

The ACUMEN Portfolio contains a narrative that the academic can use to explain their academic value, backed by evidence from the rest of the portfolio, when possible.

- Highlight: achievements, ambitions and interests
- Link the three sub-portfolios together
- Present your self-perspective
- Situation dependent
- Not too long
  - Not more than 500 words

Examples: see handouts -application for full professor -application for horizon 2020-grant

CWTS

ACUMEN
academic careers understood through measurement and norms

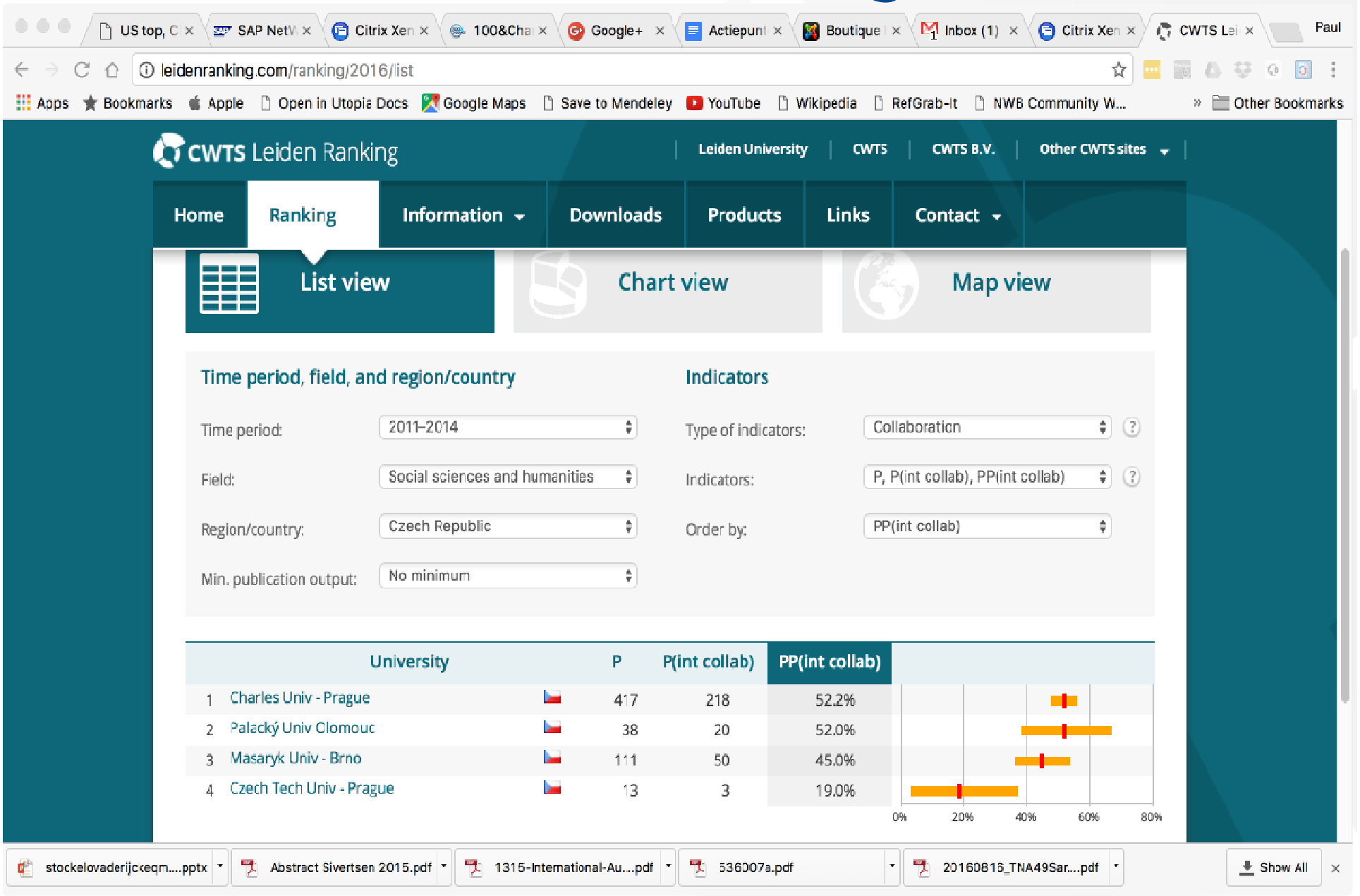
Portfolio – Summary

- The portfolio is modular. Consider only:
  - items relevant for the individual
  - Items relevant for the specific evaluation
- The aim of the portfolio is to provide a holistic view of someone's expertise, output and influence
- This version of the portfolio is built to supplement the traditional CV cause it highlights key achievements rather than giving an exhaustive list
- The use of a portfolio makes it easier for evaluators to compare people based upon their portfolios and to identify specific kinds of skills or expertise needed
- A slightly different portfolio could serve as a replacement or as a tool to create an extended CV

CWTS

ACUMEN
academic careers understood through measurement and norms

www.leidenranking.com



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 ▶

CWTS