CWTS
Meaningful metrics

Towards best practices for authorship and research evaluation

Effects of performance metrics & the Leiden Manifesto

Sarah de Rijcke

Munin Conference on Scholarly Publishing

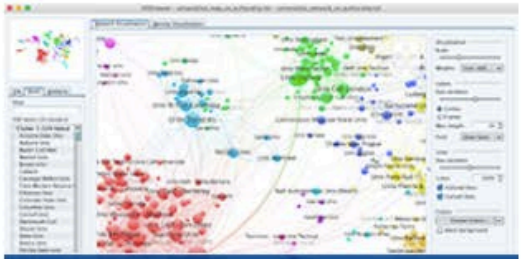
Tromsø, 23 November 2017

Universiteit Leiden

Research

Research at CWTS is organized into three research groups, three chairs (partly integrated in the research groups), and a working group. In addition, CWTS pays special attention to three research themes that are of major importance for research management and research policy.

Research groups



Quantitative Science Studies

Science and Evaluation Studies

Science, Technology, and Innov...

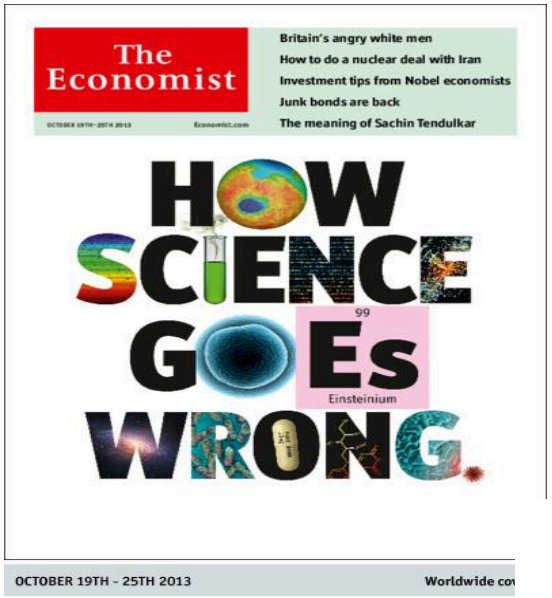
SCIENCE AND EVALUATION STUDIES RESEARCH GROUP

CONSTITUTIVE EFFECTS OF EVALUATION

Evaluation and metrics influence quite routine activities at various stages of the research process. This research line charts these

Some initial observations

- Research has become a strategic enterprise in which permanent communication is crucial
- The relative professional autonomy of science and scholarship has weakened considerably
- Both “quality/excellence” and “impact” have become crucial for success at all levels of the scientific system
- Quality and impact are mutually dependent in complex ways
- Peer & expert review and indicator based assessment have become intimately intertwined and mutually shape each other


The Economist
Britain's angry white men
How to do a nuclear deal with Iran
Investment tips from Nobel economists
Junk bonds are back
The meaning of Sachin Tendulkar
HOW SCIENCE GOES WRONG
99
Einsteinium
OCTOBER 19TH - 25TH 2013


ESA
European Sociological Association


Call to Action
Sign the petition addressed to European Commission for more support of Social Sciences and Humanities in Europe
Click here to sign the petition
1 2 3 4 5 6


theguardian
News Sport Comment Culture Business Money Life & style Travel
Jobs
News Science People in science
Nobel winner declares boycott of top science journals
Randy Schekman says his lab will no longer send papers to Nature, Cell and Science as they distort scientific process
Ian Sample, science correspondent


COMMENT
The Leiden Manifesto for research metrics


San Francisco
DORA
Declaration on Research Assessment


The Metric Tide
Report of the Independent Review of the Role of Metrics in Research Assessment and Management
July 2015

Four main problems

1. The funding system
2. The career structure
3. The publication system
4. The evaluation system

The Evaluation Gap

- discrepancy between evaluation criteria and the social and economic functions of science

- lack of recognition for new types of work that researchers perform

CWTS

nature human behaviour

PERSPECTIVE

PUBLISHED: 10 JANUARY 2017 | VOLUME: 1 | ARTICLE NUMBER: 0021

OPEN

A manifesto for reproducible science

Marcus R. Munafò[1,2*], Brian A. Nosek[3,4], Dorothy V. M. Bishop[5], Katherine S. Button[6], Christopher D. Chambers[7], Nathalie Percie du Sert[8], Uri Simonsohn[9], Eric-Jan Wagenmakers[10], Jennifer J. Ware[11] and John P. A. Ioannidis[12,13,14]

Improving the reliability and efficiency of scientific research will
and accelerate discovery. Here we argue for the
ods, reporting and disseminati
empirical studi

A hallmark of scientific creativity is the ability to see novel and unexpected patterns in data. John Snow's identification of links

‘Old-world’ metrics sustain perverse publishing incentives

Two widely (mis)used bibliometric indicators: Journal Impact Factor and Hirsch Index

- Definition of JIF:
  - The mean citation score of a journal, determined by dividing all citations in year T by all citable documents in years T-1 and T-2

- Definition of h-index:
  - The ‘impact’ of a researcher, determined by the number of received citations of an oeuvre, sorted by descending order, where the number of received citations equals the rank position

Some conceptual problems with JIF

- Inflates impact of **all** researchers publishing in same journal
- Promotes journal publishing
- Stimulates one-indicator thinking
- Ignores other scholarly virtues

Some conceptual problems with H-index

- Biased against young researchers
- Biased against selective researchers
- Invites strategic behavior
- Ignores other elements of scholarly activity
- Promotes one-indicator thinking


1
3
4
4

- Appropriation and expropriation

- Peer & expert review and indicator based assessment have become intimately intertwined and mutually shape each other

- Research assessment is not a measurement problem, because assessments are performative →

Indicators acquire (additional) meaning through contexts of use (Dahler-Larsen 2013)

Messed-up practices

Thinking with Indicators in life sciences

- Müller & De Rijcke (2017). Thinking with Indicators. Exploring the epistemic impacts of academic performance indicators in the life sciences. *Research Evaluation*, 26(3), 157-168.
- Rushforth & De Rijcke (2015). Accounting for Impact? The Journal Impact Factor and the Making of Biomedical Research in the Netherlands. *Minerva*, 53(2), 117-139.

The Journal Impact Factor

"Nobody's going to give you a grant if you have four papers in an impact factor 1 journal, but you may get a grant based on a paper that you published in an impact factor 12 journal or higher, right?

And so at that time, we said, ''We have to change the requirement for getting the PhD,'' and now, we set that bar at 15 impact points. So if you get a paper in an impact factor 15 journal, basically, you're done. And we've really noticed a change in that stimulating people for quality, and go for that one nice paper."

Rushforth and De Rijcke, *Minerva*, 2015

Assessing work-in-progress manuscripts
Grading for novelty and quality

PI goes to computer. "*Any alternatives? Any journals?*"

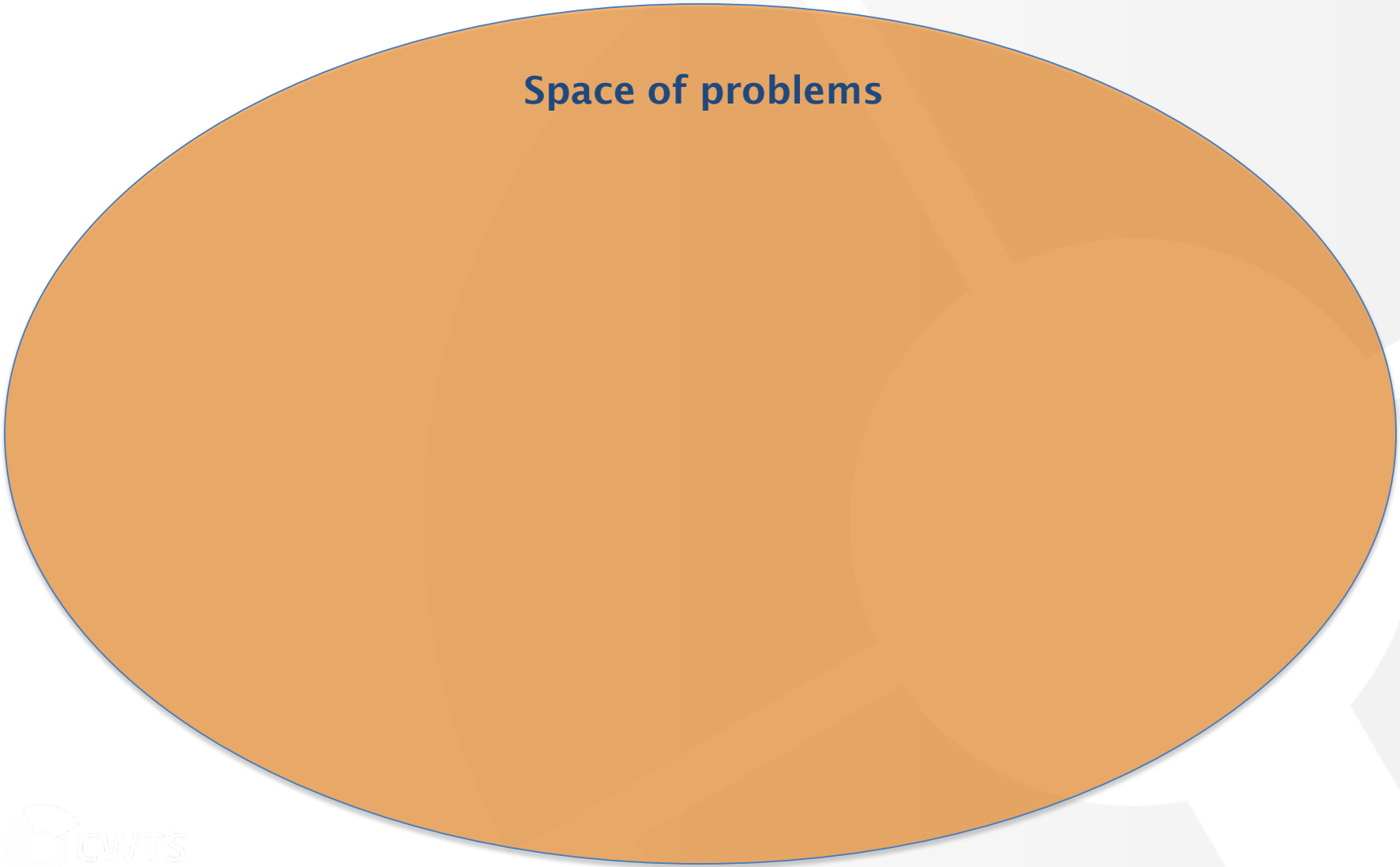
PhD: *Hmm maybe [Journal C]. They are similar in impact right?*

Post-doc: *Yeah seven-ish. It's difficult because some papers are descriptive and some have mechanism. So for this paper it could actually go one step higher than Journal C because you're going a bit beyond description. They also have priority reports in [Journal B].*

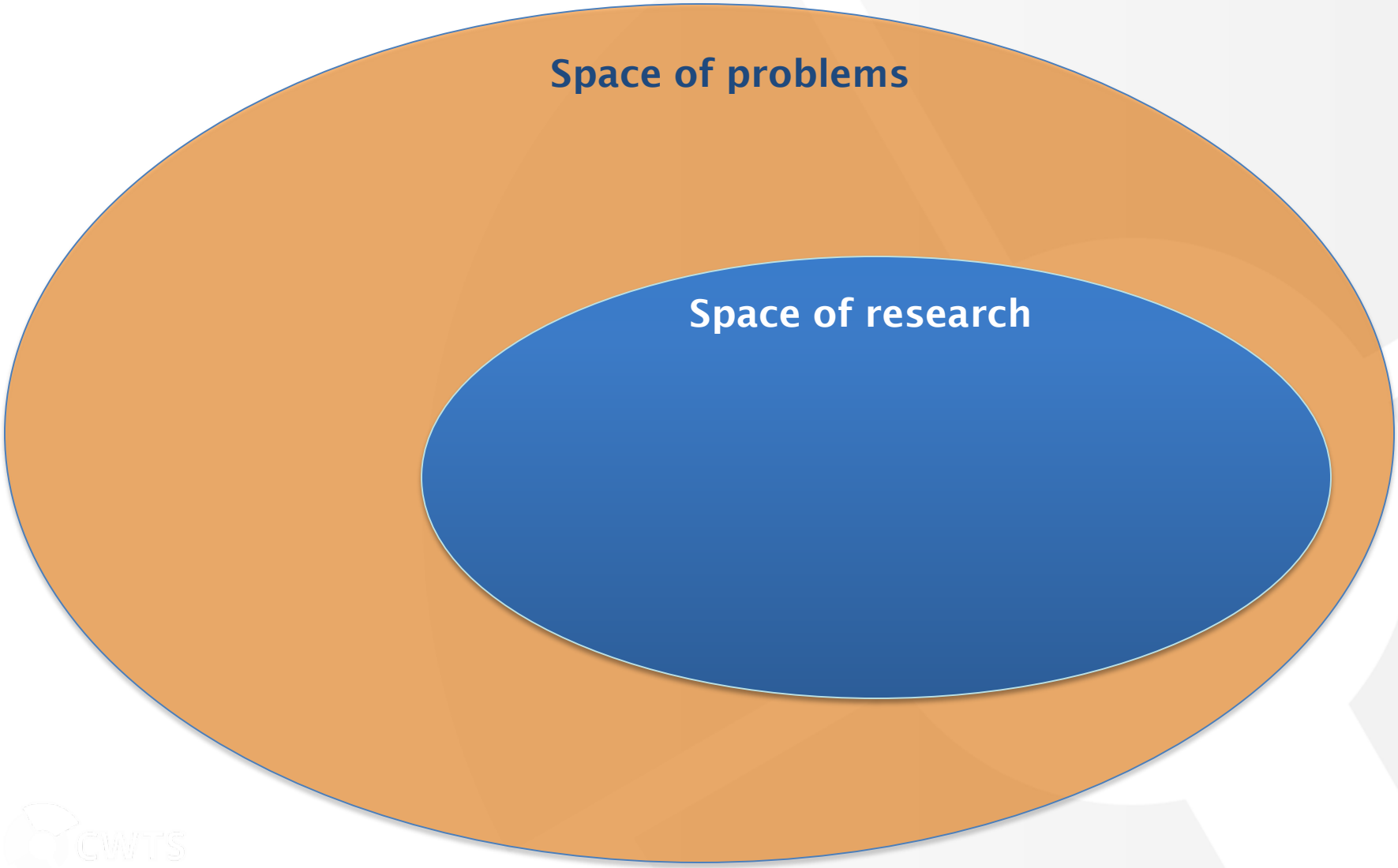
PI: *[Journal D] also has very fast publishing periods from date of submission- if they like it of course.*

(Fieldnote 22 July 2014)

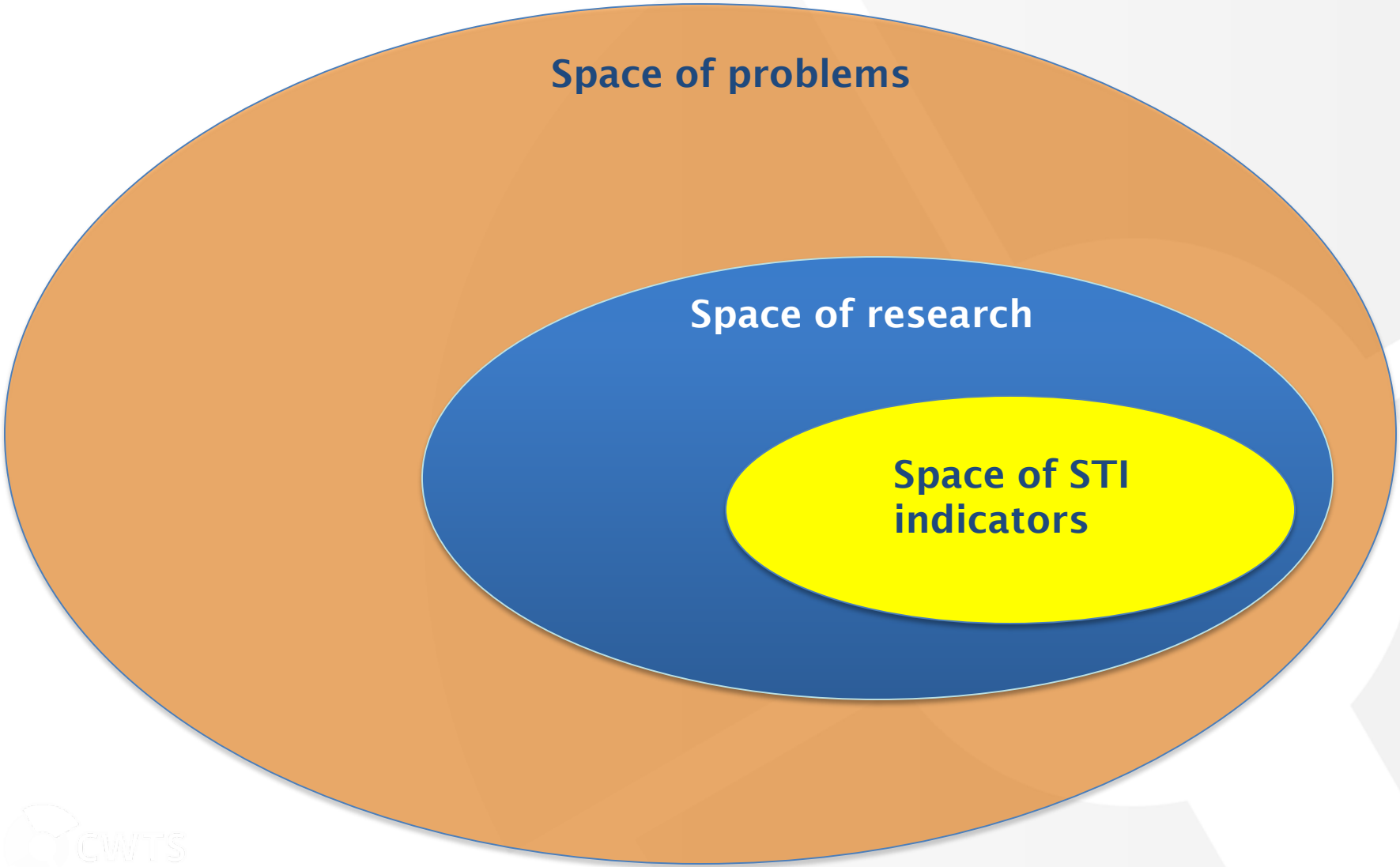
Problems, research and indicators



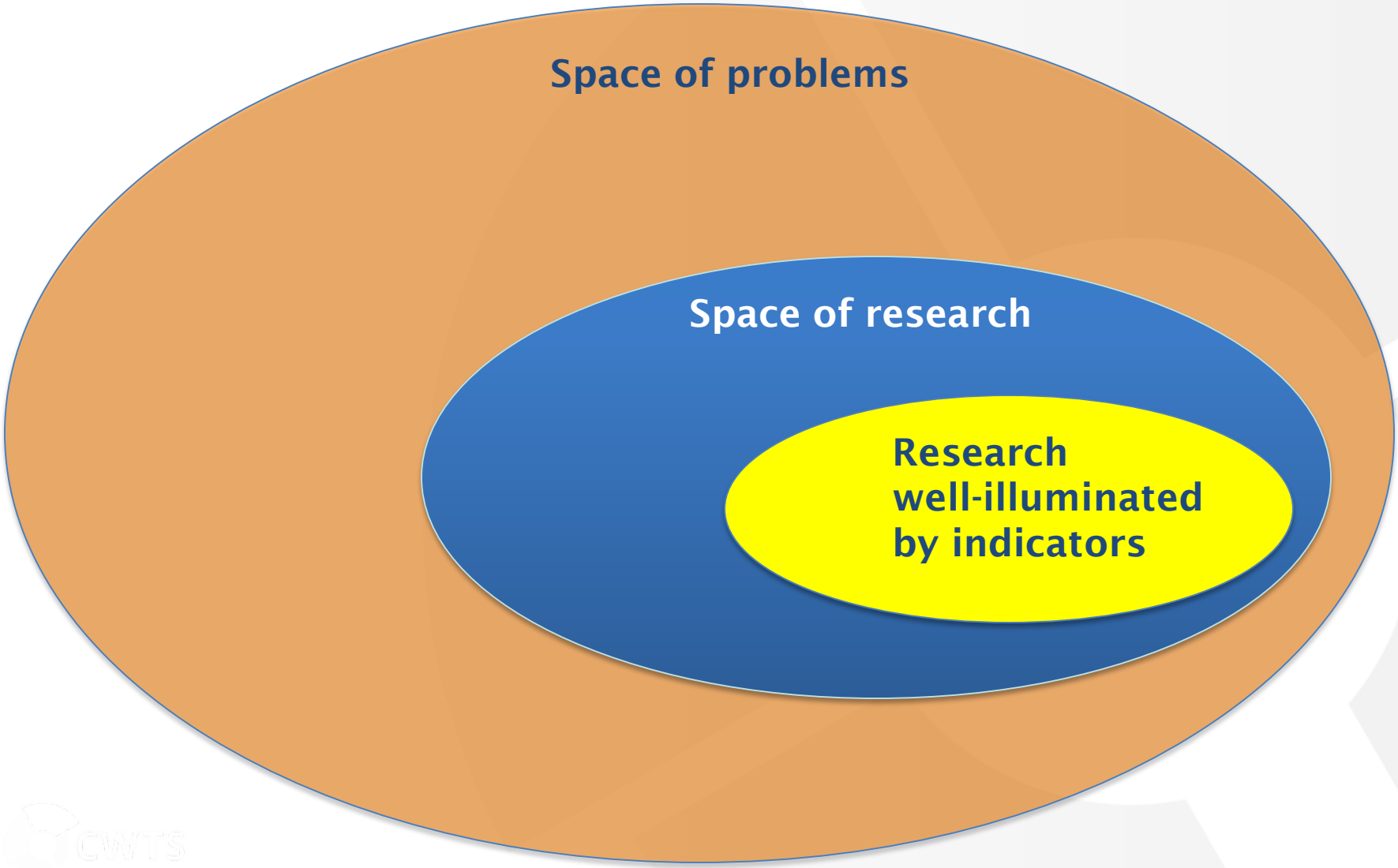
Problems, research and indicators



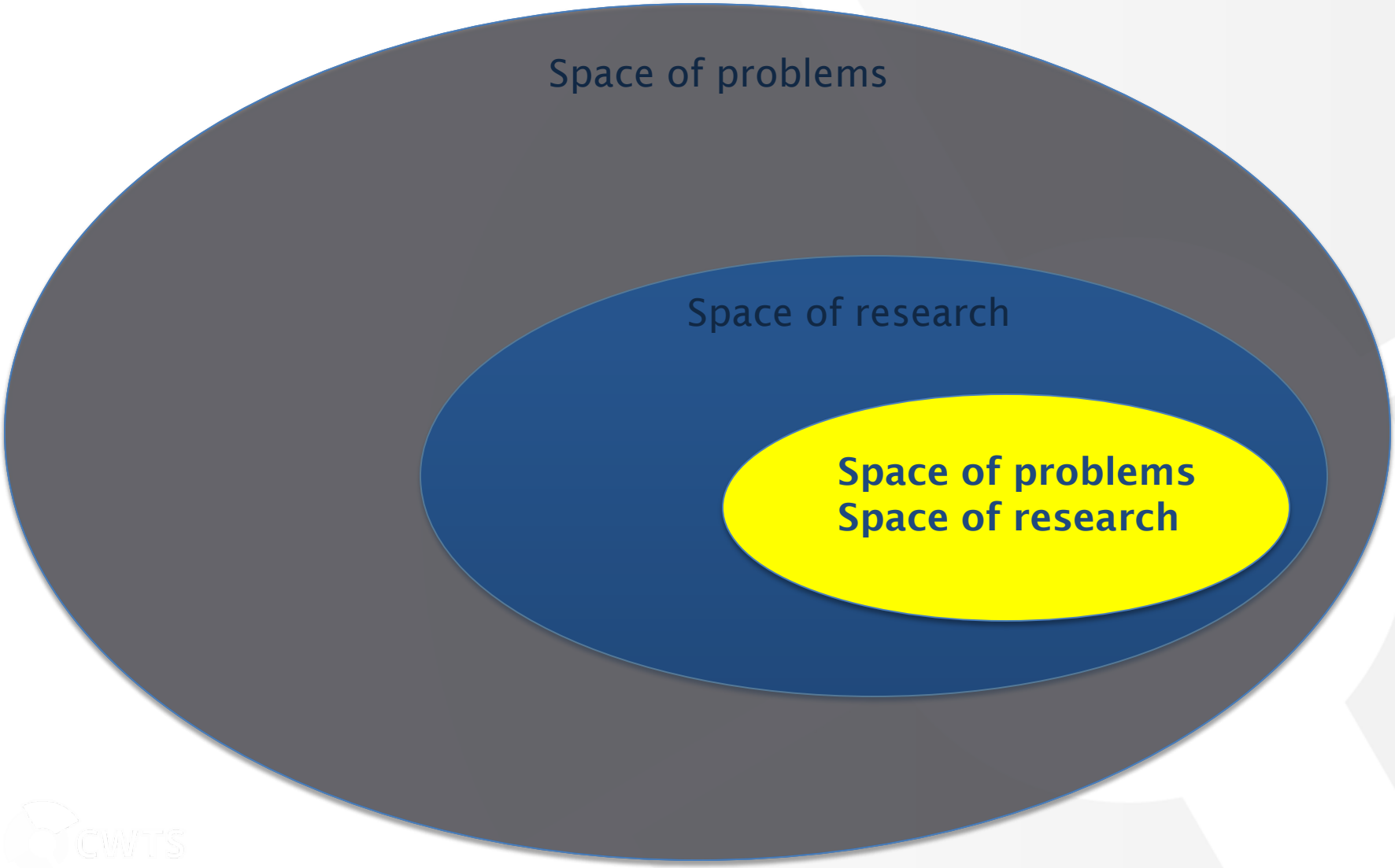
Problems, research and indicators



Streetlight effect indicators



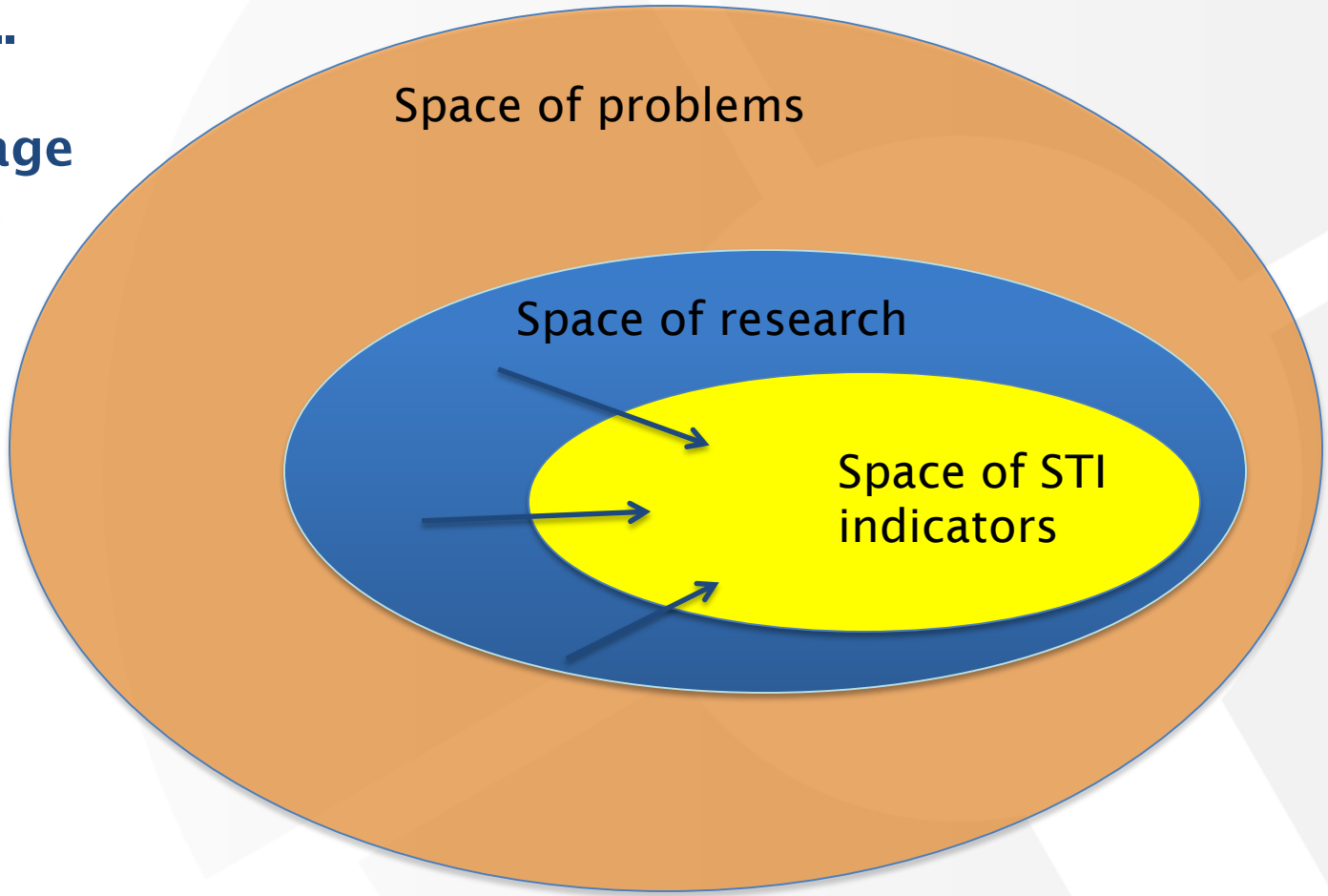
Streetlight effect in indicators: mistaking light with "problems"

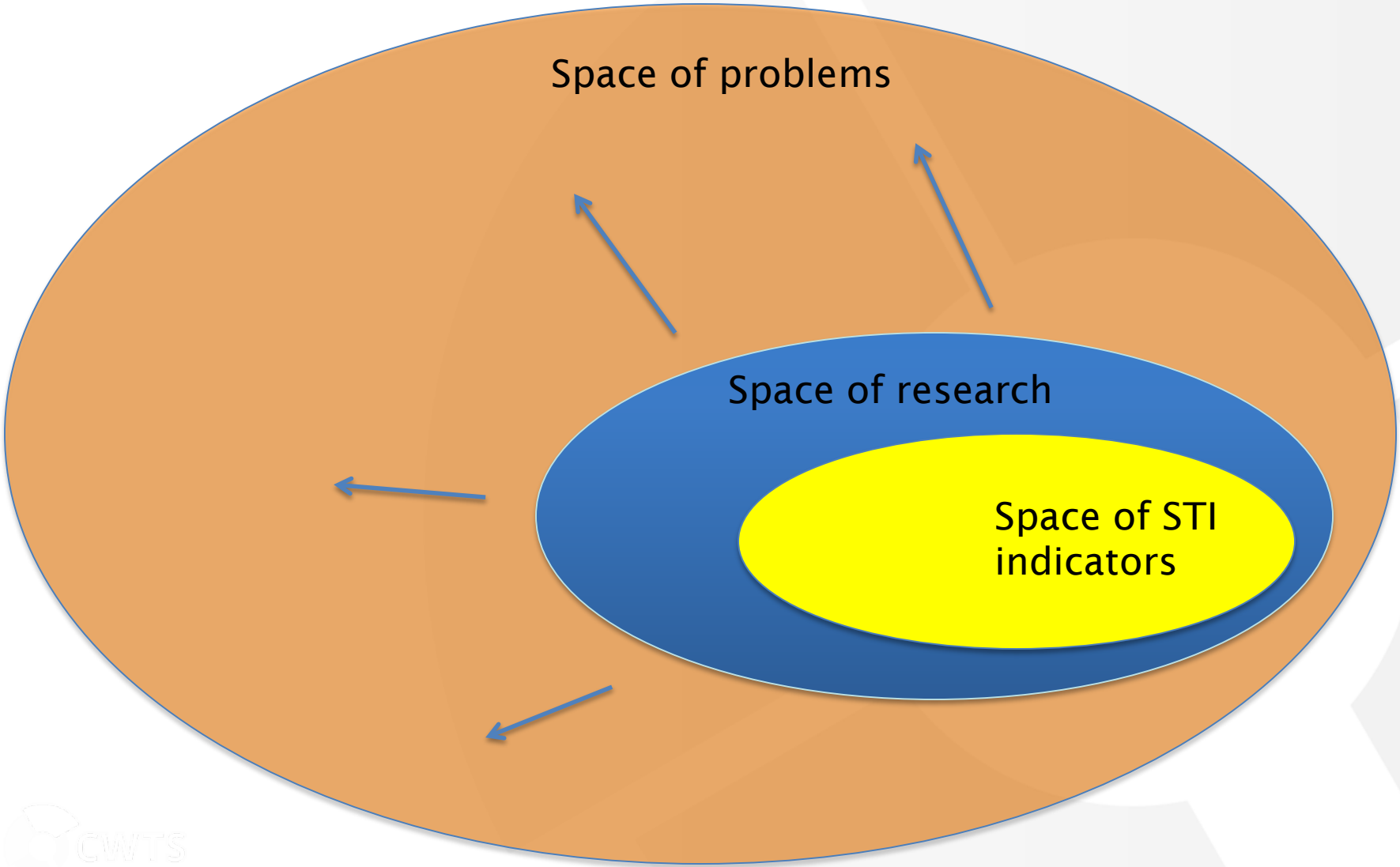


Questions dealt by research under streetlight will be better rewarded.

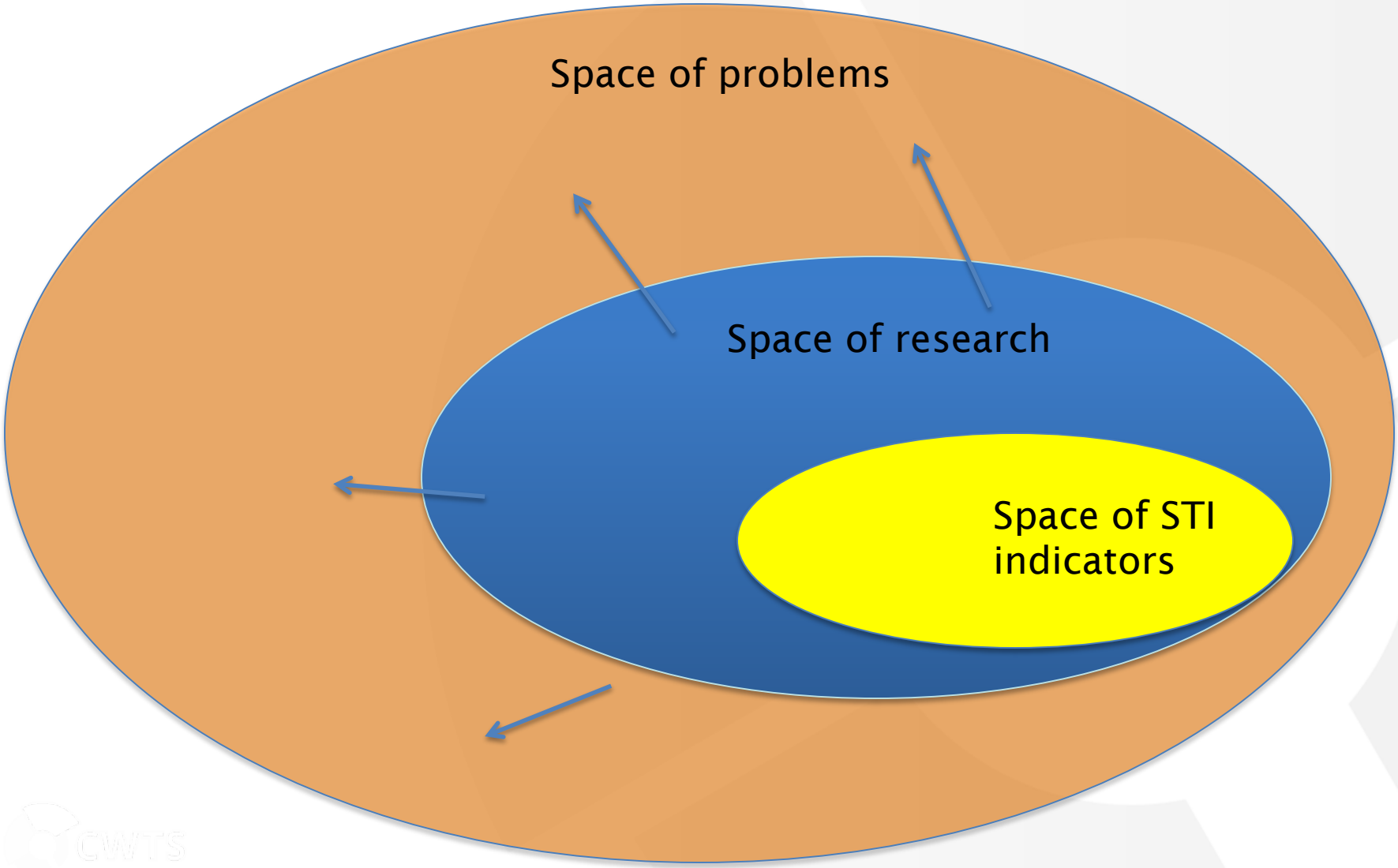
Reduced diversity of research efforts…

…reduced coverage of societal needs



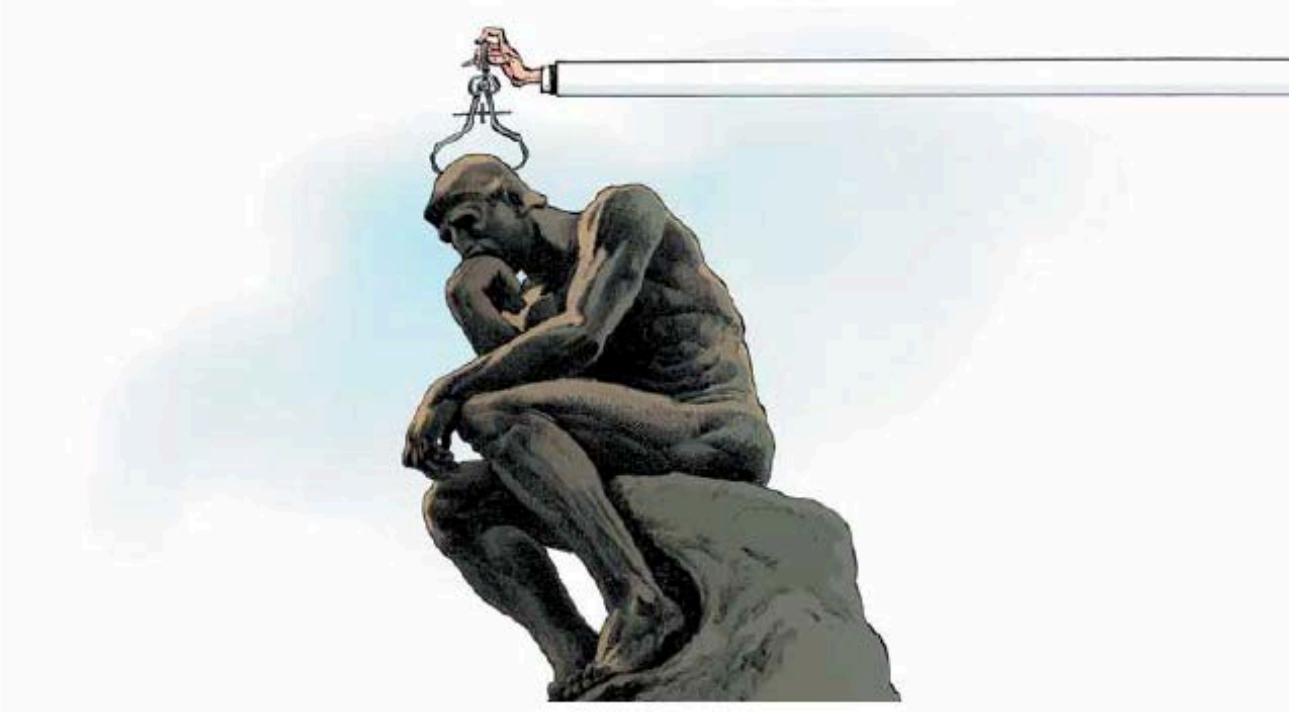

Space of problems
Space of research
Space of STI indicators

This is the move we should facilitate:



Responsible Metrics

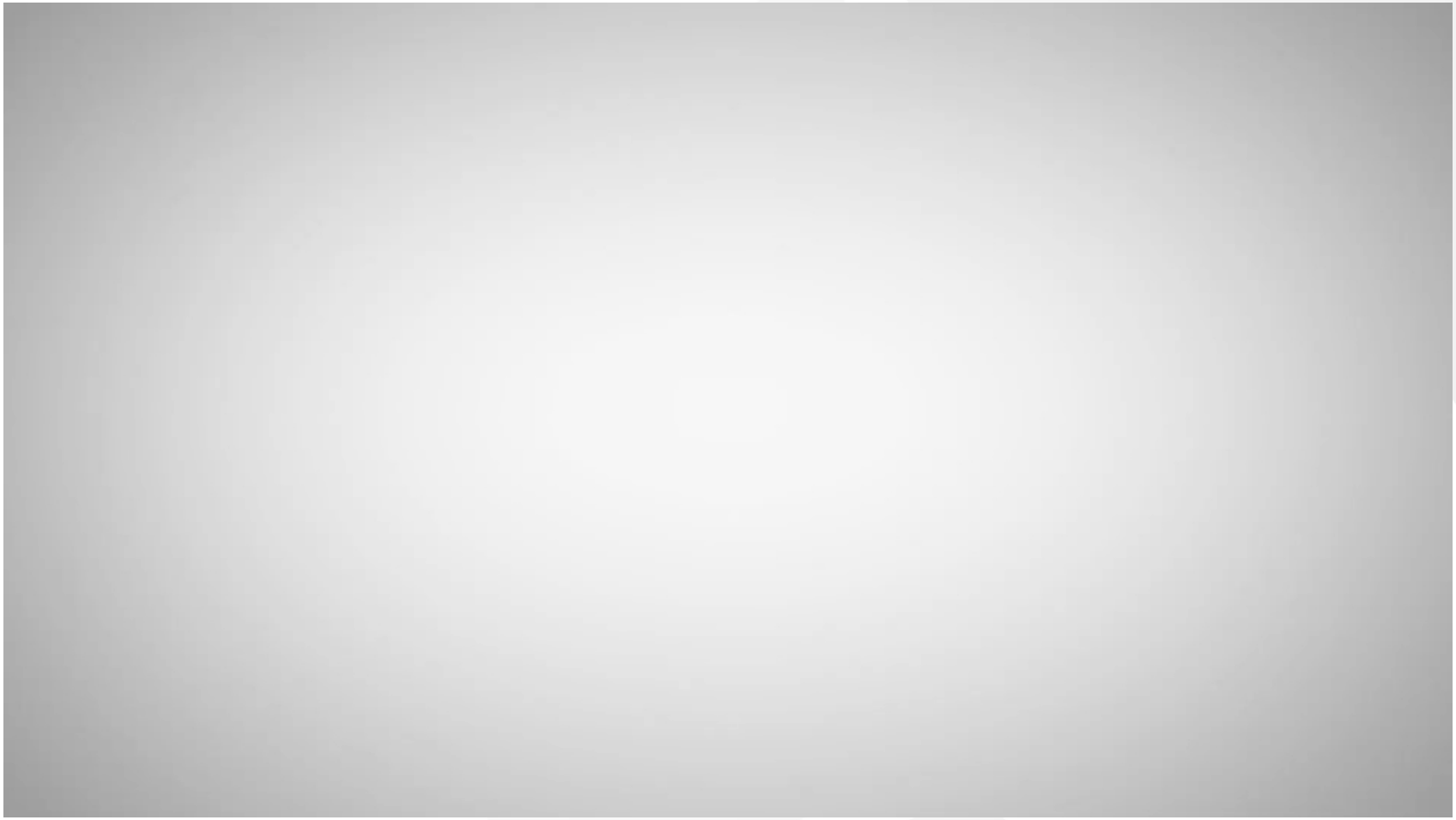
nature



The Leiden Manifesto for research metrics

A collaboration between Diana Hicks (Georgia Tech), Paul Wouters (CWTS), Ismael Rafols (SPRU/Ingenio), Sarah de Rijcke and Ludo Waltman (CWTS)

CWTS



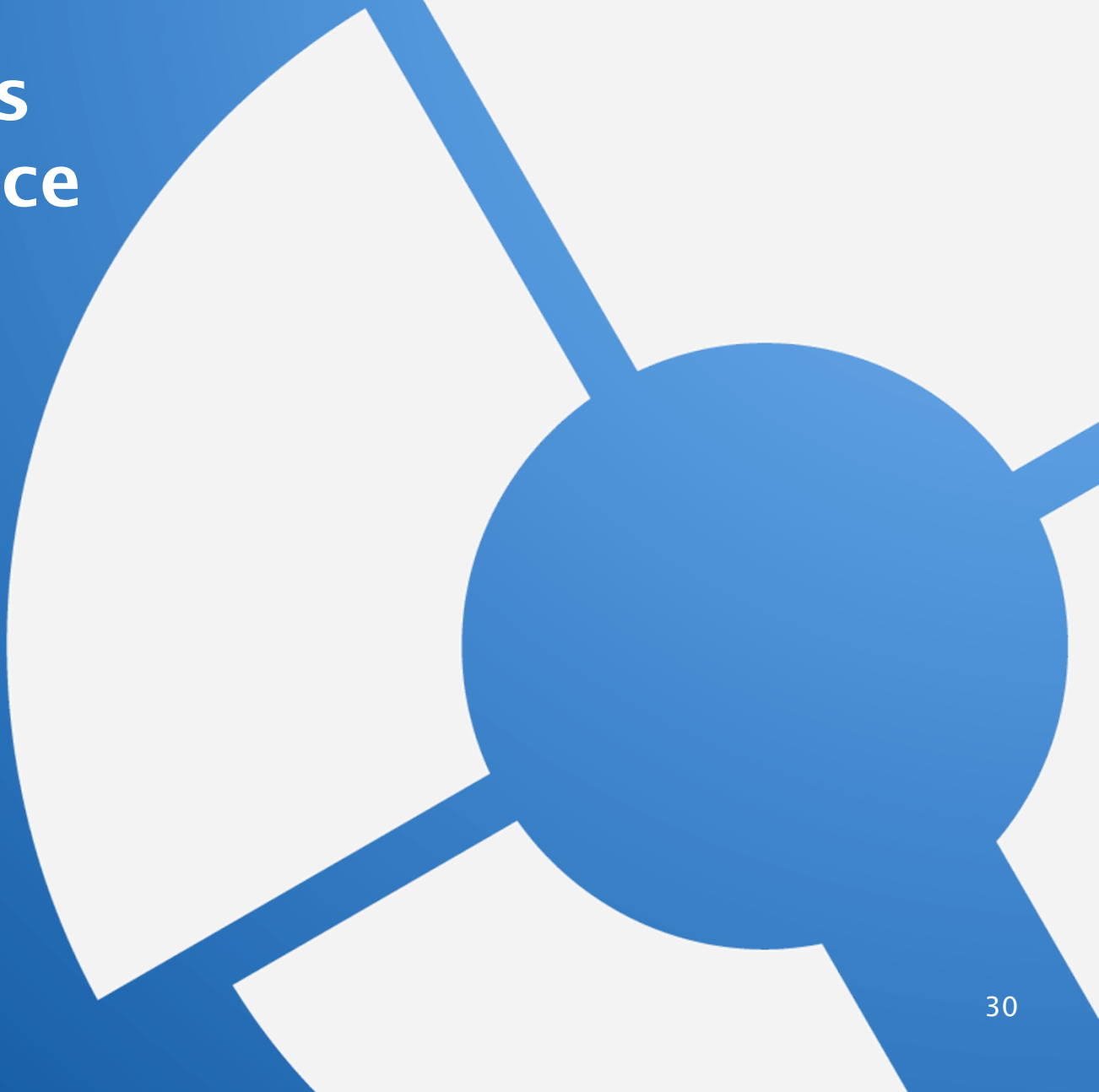
The Leiden Manifesto

- 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

Experiments with evidence in context



Solutions?

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 ▶

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

Examples next generation metrics

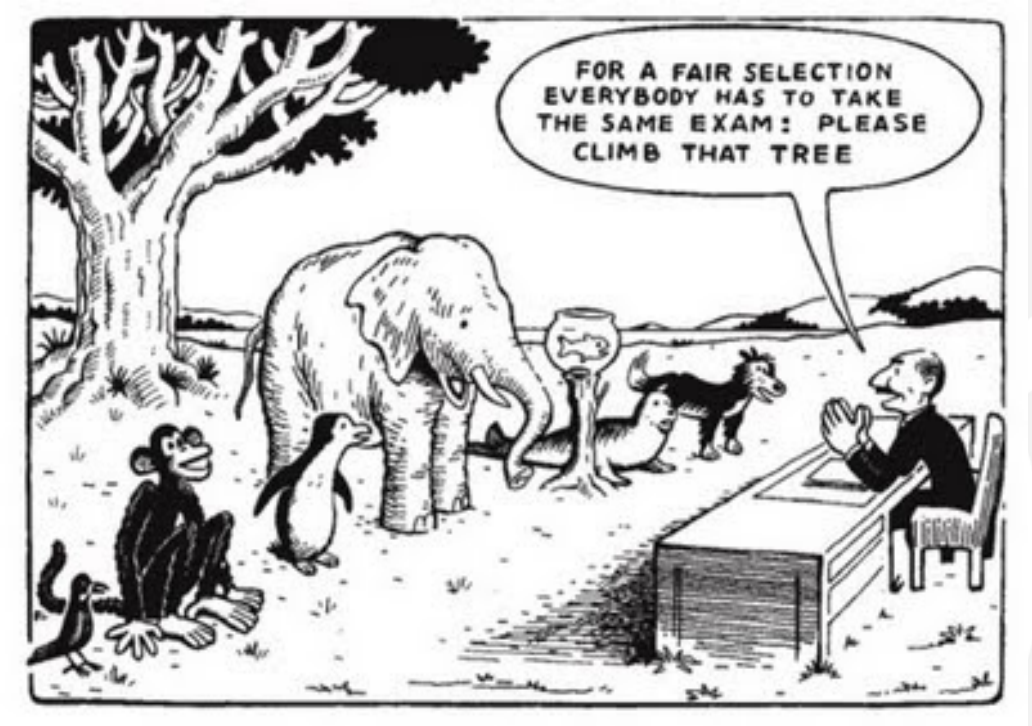
- PCRST indices (Ioannidis 2014)
- Productivity index:
  - preprints as evidence of productivity in grant applications
- “S-index” (Olfson 2017) – proportion of papers accompanied by shareable material, data, protocols
- Did applicant’s previous studies follow quality standards (fi for reporting; EQUATOR)?
- ‘Open data index’

Context counts

Responsible metrics are *not supposed to be* a universal standard

Responsible metrics should be responsive and inclusive metrics

The context shapes what responsible metrics means



Thank you for your attention