

**DRAFT**

**inorms**  
Research Evaluation Working Group

Better  
decision  
making  
through  
responsible  
research  
evaluation

Research Administration  
Briefing Material

# Why is responsible evaluation important?

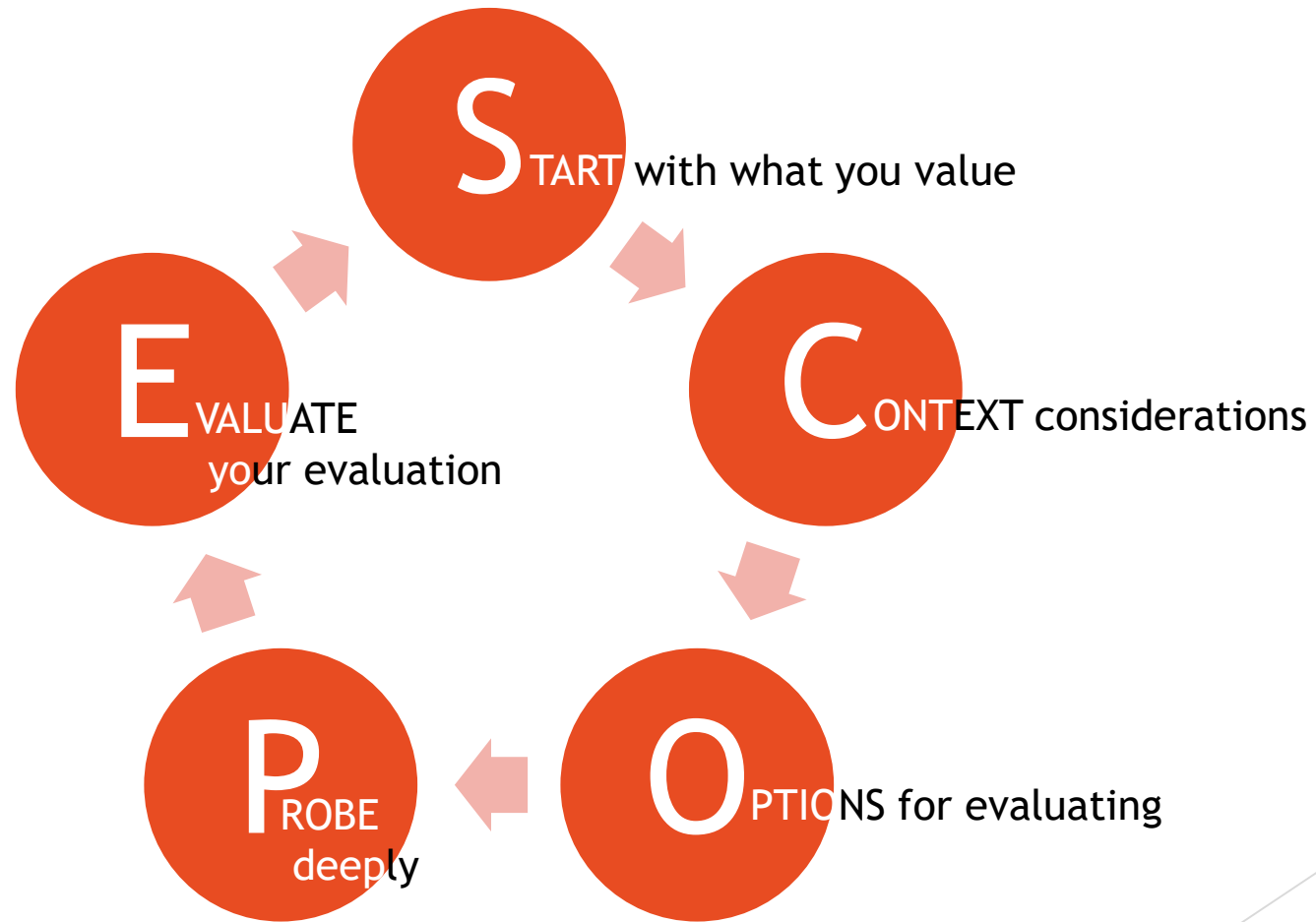
- ▶ External drivers, such as rankings, hand out values and missions to universities and use non-transparent methods to evaluate them
- ▶ Evaluations impact researchers and research organizations, scientific community has become more aware of the pitfalls of irresponsible evaluation (e.g. DORA, Leiden Manifesto, Plan S, Wellcome Trust)
- ▶ Responsible research evaluation leads to sensible decision making
- ▶ Organizations face a reputational risk around poor use of metrics

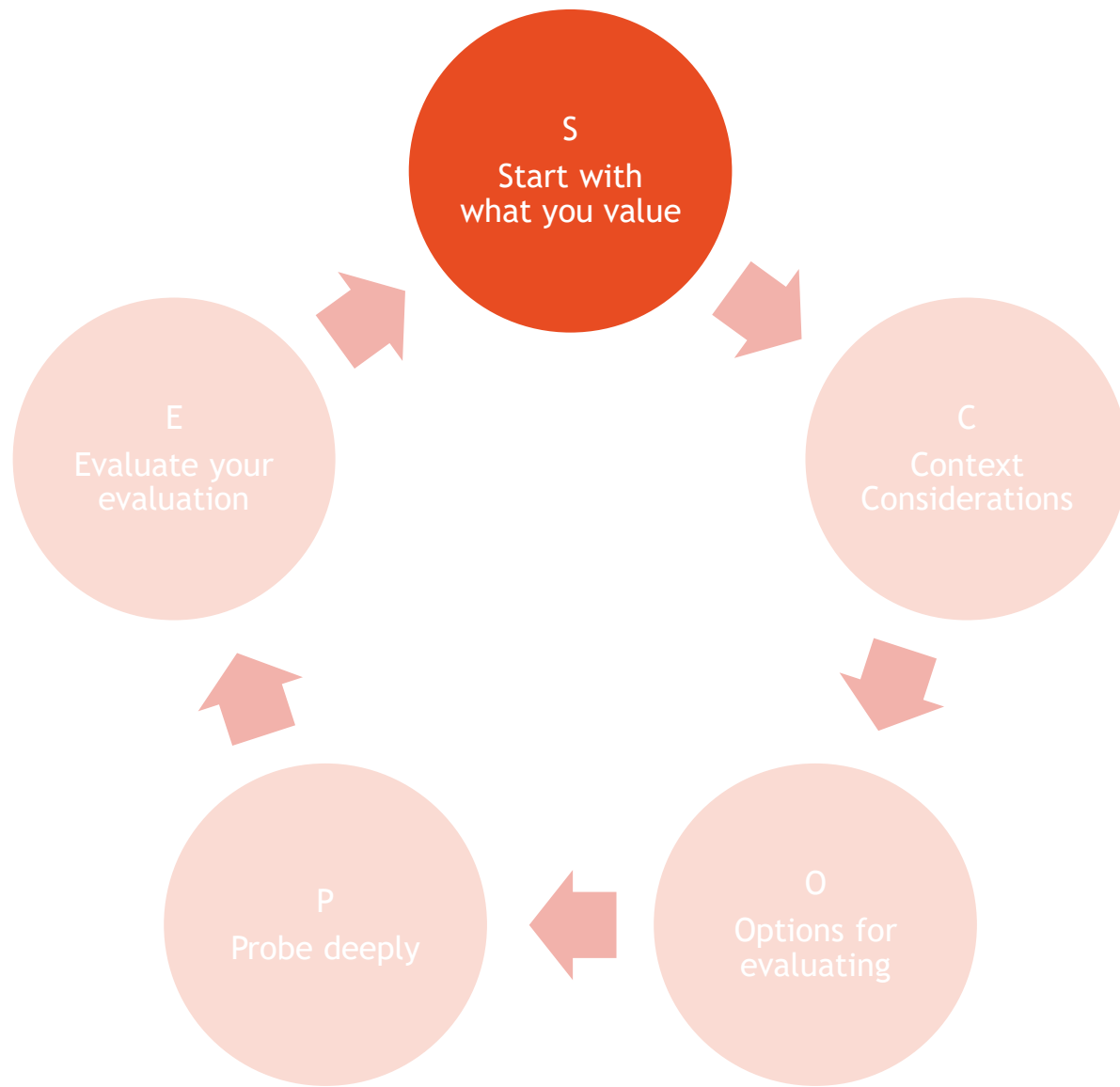
## Before starting any evaluation, expertise is needed

It's surprising how academic rigour is so often abandoned when it comes to management practice.

- ▶ Remember that producing and interpreting metrics requires expertise.
- ▶ Understand and accept the limitations of both the indicators and the data used as basis for analyses.
- ▶ Make sure that metrics only support expert judgement, never replace it.
- ▶ Understand that using metrics has consequences: you get what you measure.

# SCOPE: 5 stages for doing evaluation responsibly





© Mike Baldwin / Corneer



“What are they worth? Well, it’s difficult to put a price on such extraordinary works of art. How much you got?”

Do we value what is important  
for us or for others? Or both?  
Or none?

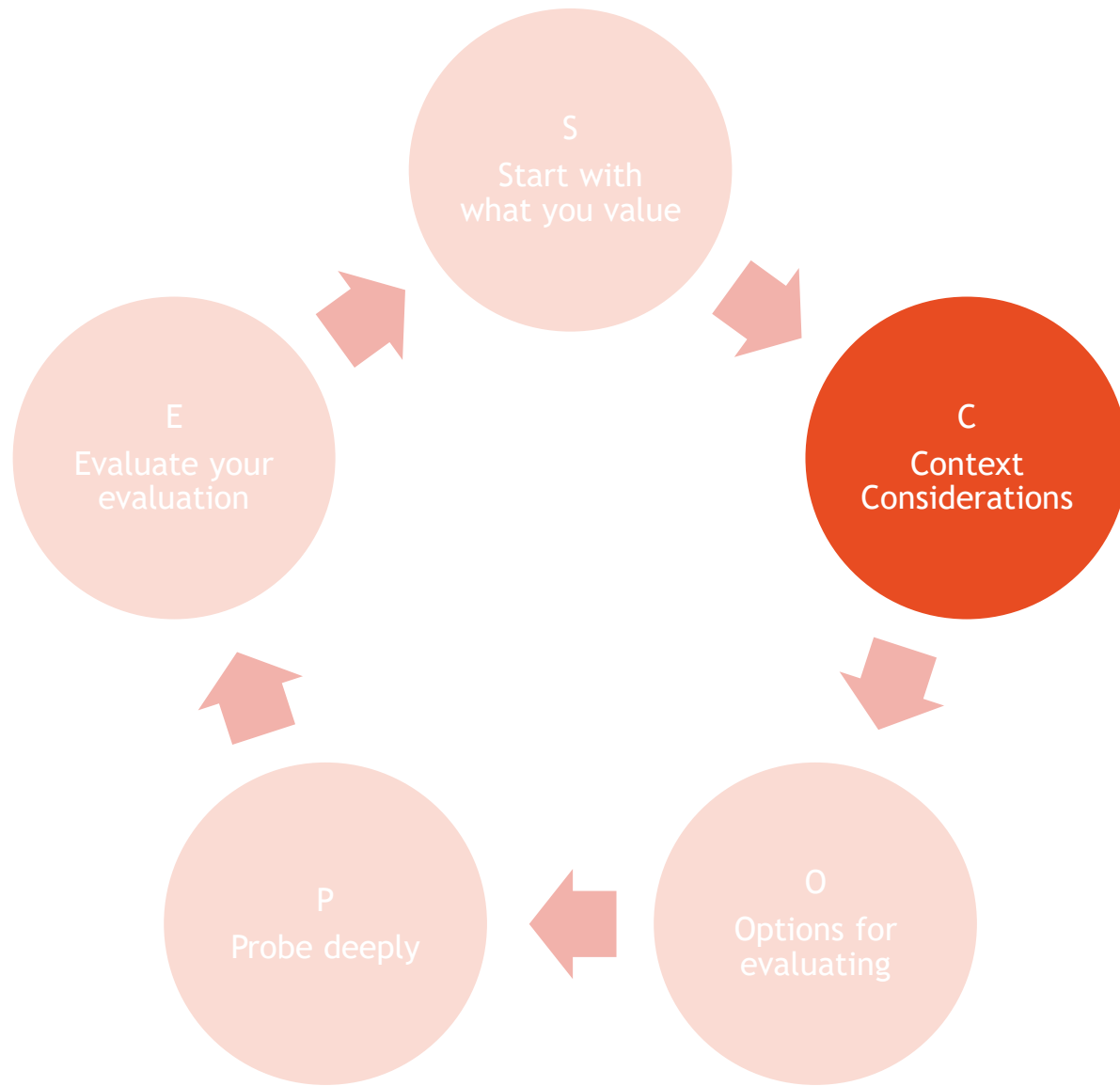
# S-Start with what you value

- ▶ Universities sometimes focus on what OTHERS value (i.e. rankings, national assessment systems, etc.) rather than ask themselves what THEY value.
  - ▶ So often we value what we think we can measure rather than measure what we value.
- ▶ Institutional autonomy is important. Use it or lose it, don't let external drivers dictate how and why you evaluate.
- ▶ If introducing performance indicators, what sort of 'performance' are you looking for?
  - ▶ How do you prioritize your values?
- ▶ Inherent tensions between what you value and staying alive (compliance)
  - ▶ Where those tensions exist, ask yourself what you can do about them.

# The Streetlight effect: Measuring by available data not by mission







# Why are you measuring?

- ▶ **Measure to understand.** “Science of science” activities that study publication patterns and trends for the sole purpose of understanding them better.
- ▶ **Measure to show off.** “Pick me!” activities. The use of metrics to market an individual, group or university on promotional materials or grant applications.
- ▶ **Measure to monitor.** Plotting progress against an objective whether internally or externally set. This may include some comparison activity.
- ▶ **Measure to compare.** The use of indicators to compare one entity with another. University rankings are an example of this.
- ▶ **Measure to incentivise.** The use of indicators to incentivise certain behaviours. The measurement of open access content submitted to REF is an example of this.
- ▶ **Measure to reward.** Any activity that results in some kind of reward for the entity being measured, be this a job, promotion, grant, prize or award of any description.

# Why and what are you measuring? Balancing the risks.

	Country	University	Research Group	Researcher
Analysis	Green	Green	Yellow	Yellow
Advocacy	Green	Green	Yellow	Yellow
Accountability	Green	Yellow	Yellow	Red
Acclaim	Yellow	Red	Red	Red
Adaptation	Yellow	Red	Red	Red
Allocation	Red	Red	Red	Red
Low risk	Green			
Medium risk	Yellow			
High risk	Red			

- ▶ Risks associated with metric use in various settings

# Who are you measuring?

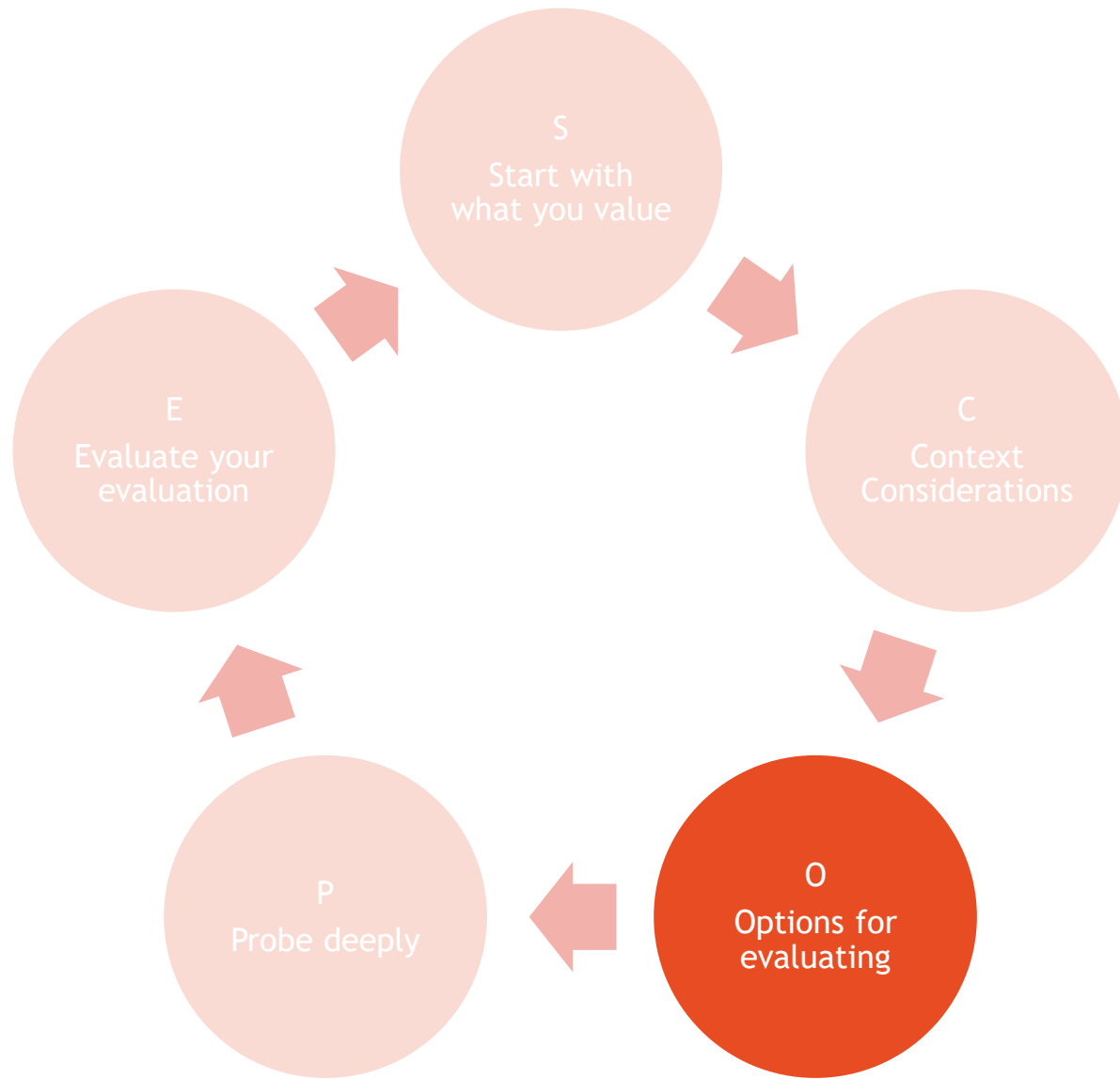
- ▶ Whether a metric is good or bad depends also on who you are measuring → indicators are not evaluators
- ▶ Differences in scientific disciplines, and in the size or structure of units of assessment have to be considered → all indicators are not suitable for all disciplines and metrics are neither objective nor neutral



What works in one context won't work in another

# Do you need to measure at all?

- ▶ “You don’t fatten a pig by weighing it”
- ▶ Measuring is not the only way to incentivise behaviour
- ▶ Changing behaviour requires:
  - ▶ Understanding - Why do I need to do this? (Messaging)
  - ▶ Capability - How do I do this? (Training and support)
  - ▶ Opportunity - Where can I do this? (Tools, services, making it easier to do it than not to do it)
  - ▶ Motivation - What will happen if I do/don’t? (Reward/sanctions)



## Rule of thumb

- ▶ Is your measure a suitable proxy for what you are measuring?
- ▶ Quantitative measures are for quantifiable things...
  - ▶ Citations, publications, money, students
- ▶ Qualitative measures for qualifiable things...
  - ▶ Quality, diversity, excellence, value
- ▶ Be careful if using quantitative indicators as a proxy for qualitative things
  - ▶ Citations  $\neq$  quality
  - ▶ Ranking position  $\neq$  excellence

University search:

Study Level



Subject of interest



Study destination



GO

## Faculty/Student Ratio (20%)

Teaching quality is typically cited by students as the metric of highest importance to them when comparing institutions using a ranking. It is notoriously difficult to measure, but **we have determined that measuring teacher/student ratios is the most effective proxy metric for teaching quality.** It assesses the extent to which institutions are able to provide students with meaningful access to lecturers and tutors, and recognizes that a high number of faculty members per student will reduce the teaching burden on each individual academic.

*Faculty/student Ratio* constitutes 20 percent of an institution's final score.



# Quantitative measures

- ▶ Always imperfect
- ▶ Always looking back
- ▶ Made better by:
  - ▶ Checking validity (does this measure what you value)
  - ▶ Using a ‘basket’ of indicators
  - ▶ Combining with qualitative assessment

# Unpacking some common metrics

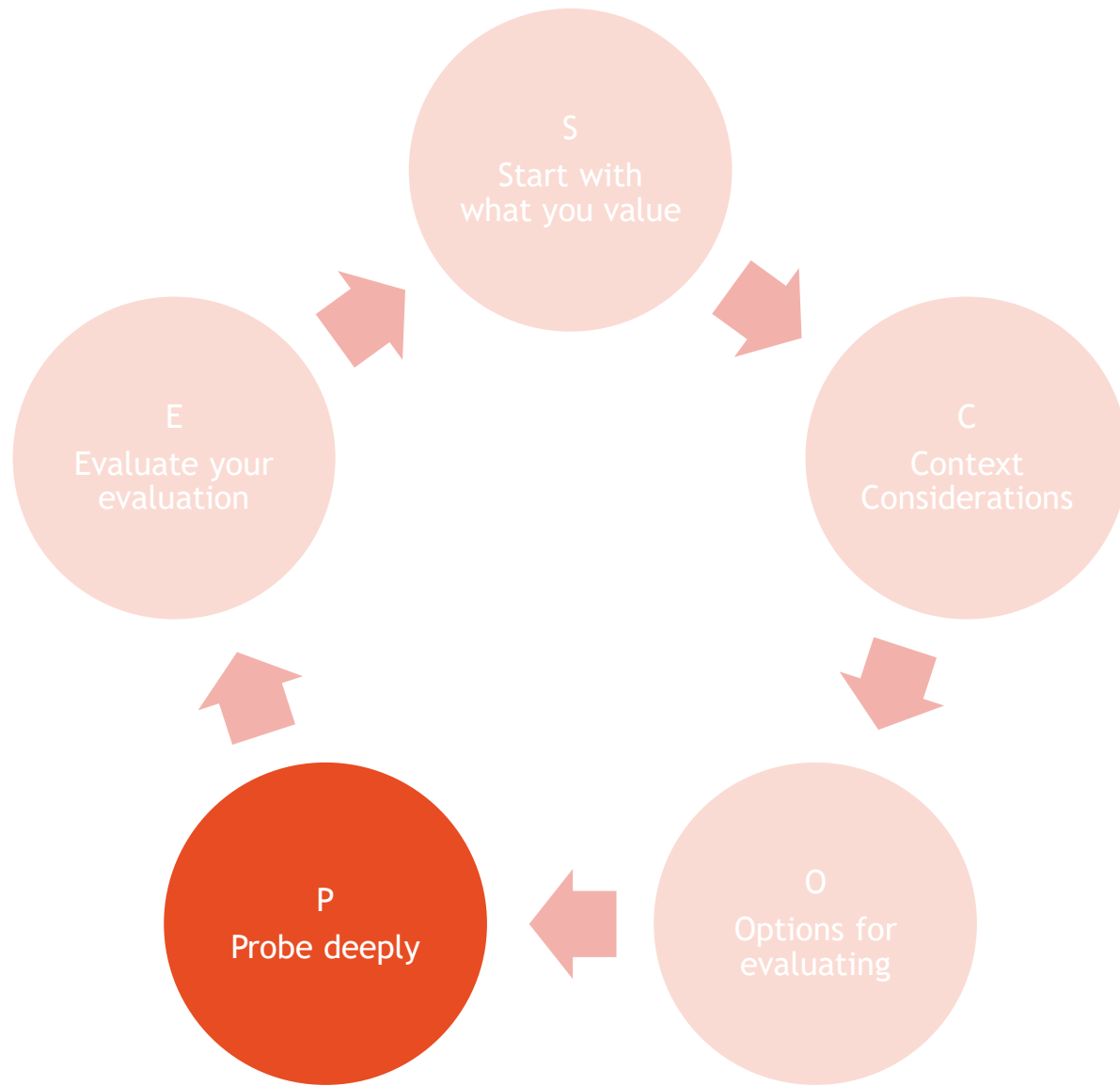
- ▶ Journal Impact Factor
  - ▶ Mean cites-per-paper for a journal over the past two years
  - ▶ Problems:
    - ▶ Citation data is always skewed - should never use the mean
    - ▶ Often used as a metric for researchers or papers. It's a journal metric.
- ▶ H-index
  - ▶ Number of papers (n) with at least n citations
  - ▶ Problems:
    - ▶ Disadvantages early career researchers, and those with non-standard career paths
    - ▶ Only ever increases. (Often called the 'Age-Index')
    - ▶ Correlates with total citations.

# Qualitative measures (peer review)

- ▶ Always imperfect
- ▶ Made better by:
  - ▶ Ensuring appropriate expertise (there is peer review and EXPERT peer review)
  - ▶ Unconscious bias training
  - ▶ Double-blind OR Open review
  - ▶ More than one reviewer
  - ▶ Diverse reviewers
  - ▶ Triangulate with metrics where appropriate

# Evaluate WITH the evaluated

- ▶ Engage with communities under evaluation
  - ▶ Understand what the unit of assessment values, what are their aims
- ▶ Co-produce evaluative approaches where possible
  - ▶ CWTS ‘Evaluative Inquiry’
  - ▶ Consider the scientific fields under evaluation to choose relevant indicators
- ▶ Jointly interpret the results
  - ▶ Openness and transparency increases the legitimacy of evaluation results



# P - Probe deeply

- ▶ Who does this discriminate against?
  - ▶ E.g. early-career researchers, females, non-journal based disciplines
  - ▶ Adjust accordingly
- ▶ How might this be gamed?
  - ▶ Where there is a prize there is a game - intellectual people will optimize
- ▶ What might the unintended consequences be?
  - ▶ On institutional or individual level
- ▶ Does the cost of measuring outweigh the benefit?
  - ▶ The cost, including the workload should be proportional to the aims and anticipated outcomes of evaluation

# Indonesia's scientists voice concerns about the country's researcher ranking system

Critics flag unclear methodology, lack of credit for research contributions other than publications

by *Dalmeet Singh Chawla*

DECEMBER 31, 2018

Critics say the methodology and reasoning behind the metric, known as the **Science and Technology Index** (SINTA), are unclear. SINTA takes into account the number of journal and non-journal articles indexed in the database Scopus, the number of citations these documents accumulate in Scopus and Google Scholar, and researchers' h-index. The h-index is another controversial metric that is designed to measure researchers' productivity and the impact of their publications.

Indonesia recently introduced a new researcher ranking system that uses the volume of articles and citations from GS and Scopus for deciding who are the BEST researchers And for allocating funding to these best researchers...

# Top Ten

**SUHARYO SUMOWIDAGDO**  
Lembaga Ilmu Pengetahuan Indonesia  
NIDN /NIP/NIDK :  
Scopus<sup>®</sup> H-Index : 91 | Google H-Index : 134 |   
Experimental Particle Physics | High Performance Computing

**INDAH SUCI WIDYAHENING**  
Universitas Indonesia  
NIDN /NIP/NIDK : 0311107303  
Scopus<sup>®</sup> H-Index : 7 | Google H-Index : 11 |   
Family Medicine | Primary Health Care | Evidence based Medicine | Community Medicine

**RIYANTO SARNO**  
Institut Teknologi Sepuluh Nopember  
NIDN /NIP/NIDK : 0003085905  
Scopus<sup>®</sup> H-Index : 17 | Google H-Index : 20 |   
Internet of Things | Business Process Management | Process Aware Information Systems | Knowledge Engineering  
Smart Grids

**I GEDE WENTEN**  
Institut Teknologi Bandung  
NIDN /NIP/NIDK : 0015026202  
Scopus<sup>®</sup> H-Index : 21 | Google H-Index : 30 |   
Membrane Technology

**ANWAR MALLONGI**  
Universitas Hasanuddin  
NIDN /NIP/NIDK : 0016087401  
Scopus<sup>®</sup> H-Index : 12 | Google H-Index : 11 |   
Health and Environmental Risks Assessment | Environmental and Health Modeling

**HANUNG ADI NUGROHO**  
Universitas Gadjah Mada  
NIDN /NIP/NIDK : 0024027804  
Scopus<sup>®</sup> H-Index : 10 | Google H-Index : 13 |   
Biomedical signal & image processing & ana | computer vision | medical instrumentation | medical imaging  
statistical pattern

**ACHMAD NIZAR HIDAYANTO**  
Universitas Indonesia  
NIDN /NIP/NIDK : 0024077601  
Scopus<sup>®</sup> H-Index : 8 | Google H-Index : 12 |   
Information Technology | Information System | Business Intelligence | Technology Adoption | Electronic Commerce

**MAURIDHI HERY PURNOMO**  
Insitu: Teknologi Sepuluh Nopember  
NIDN /NIP/NIDK : 0016095811  
Scopus<sup>®</sup> H-Index : 4 | Google H-Index : 17 |   
Artificial intelligence

**MOHAMMAD BASYUNI**  
Universitas Sumatera Utara  
NIDN /NIP/NIDK : 0021047304  
Scopus<sup>®</sup> H-Index : 14 | Google H-Index : 16 |   
Molecular Biotechnology | Plant Lipid Biochemistry | Plant Biology | Bioinformatics

**TOLE SUTIKNO**  
Universitas Ahmad Dahlan  
NIDN /NIP/NIDK : 0512067501  
Scopus<sup>®</sup> H-Index : 13 | Google H-Index : 19 | 

... number of women = 1  
... number of researchers in social sciences and humanities = 0

Are we really confident that these are the best researchers in the country?  
Are we really sure that funding those who do well on this metrics is going to be good for science?



## Unintended consequences: institutional level

- ▶ Poor decision-making
- ▶ Neglecting all activities that aren't measured
- ▶ Goal displacement, mirroring e.g. government level assessment practice
- ▶ Short termism
- ▶ Transactional cost of metrics which may not actually add any value to the organisation
- ▶ Discourage initiative, innovation and risk-taking
- ▶ Negative influence on interdisciplinary research caused by biases against interdisciplinarity

## Unintended consequences: individual level

- ▶ Academic burden- academics leaving the profession - or worse
- ▶ Focusing on research areas that are more visible in e.g. publication databases
- ▶ Narrowing the types of publication according to assessment criteria
- ▶ Focusing on tasks that are favourable in regards to assessments
- ▶ Aiming at publication forums with high impact - research results are published later rather than sooner, the editorial system gets clogged...

## Why the audit culture made me quit

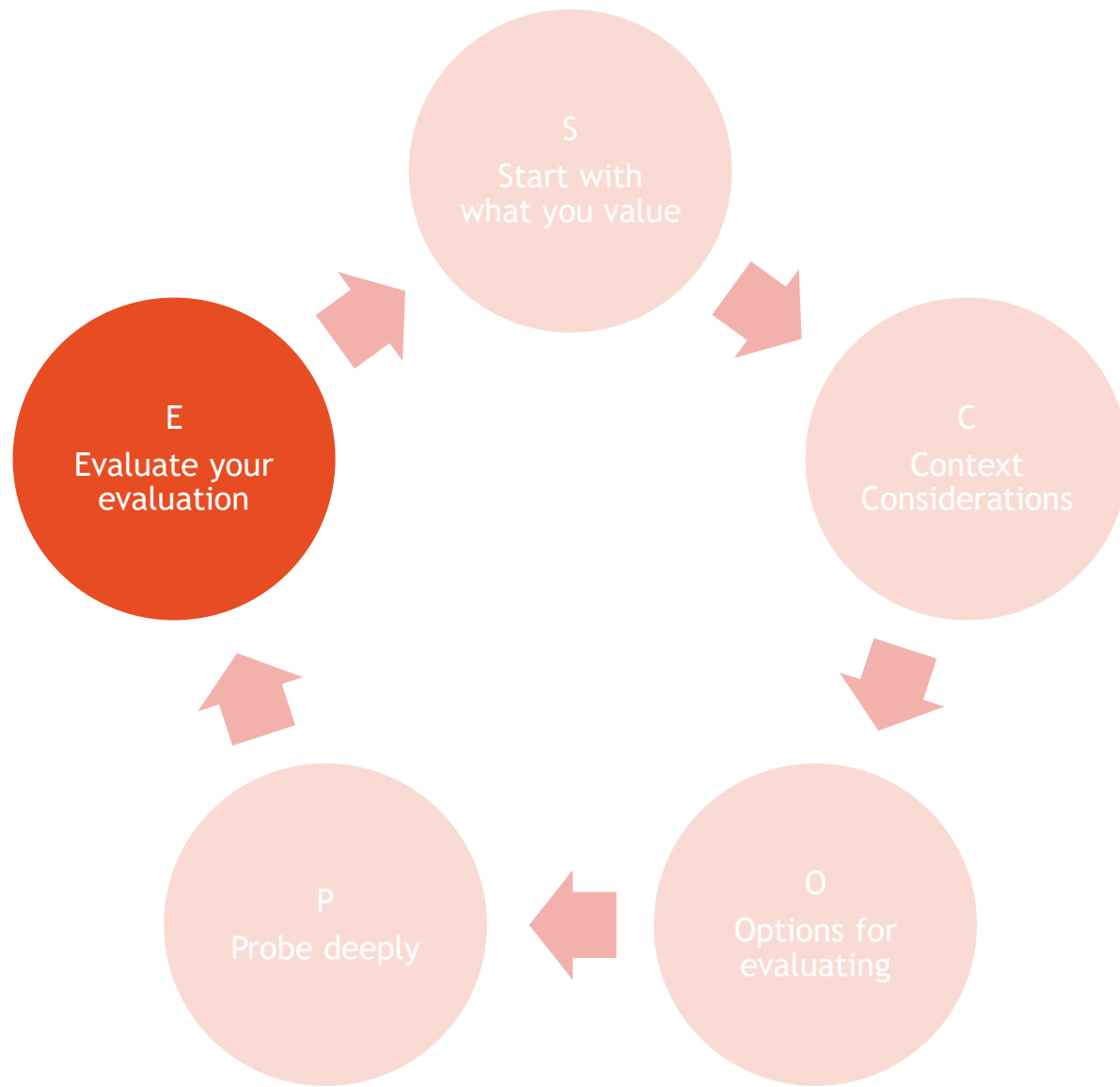
When Liz Morrish opened up to students about the pressures academics are under, the proceedings culminated in her resignation. She reflects on why she chose to leave the neoliberal academy from the outside

March 2, 2017

By  
Twit

In the UK, much of the rush to management by metrics is in response to shifting government incentives and policy changes, which, fed through the mechanism of the research excellence framework, affect institutional priorities, reputations and funding levels. Many of these metrics are quite outside the control of academics. Nevertheless, they have been weaponised as tools of performance management, and the very nature of the scrutiny creates a hostile environment for academic freedom.





# E - Evaluate your evaluation

- ▶ Go back to S - did the evaluation approach bring new insights to what you value?
- ▶ Be open to adjustments, it is always possible to do the evaluation better.
- ▶ The area of evaluation (range of data sources and indicators available as well as institutional missions and strategies) is subject to constant change
  - ▶ Just because an evaluation approach worked previously, does not mean it will work forever
  - ▶ Building in a regular review of evaluation approaches and doing so with units under evaluation is an essential part of the evaluation process

# E - Evaluate your evaluation

- ▶ Keep performance indicators under review
- ▶ Does measuring research actually make the research any better?
  - ▶ While the effect of measurement on researchers might motivate them to adopt better strategies, the long-term effects of over-measurement are almost always negative

# SCOPE: 5 stages for doing evaluation responsibly

