DRAFT

inorms

Research Evaluation Working Group

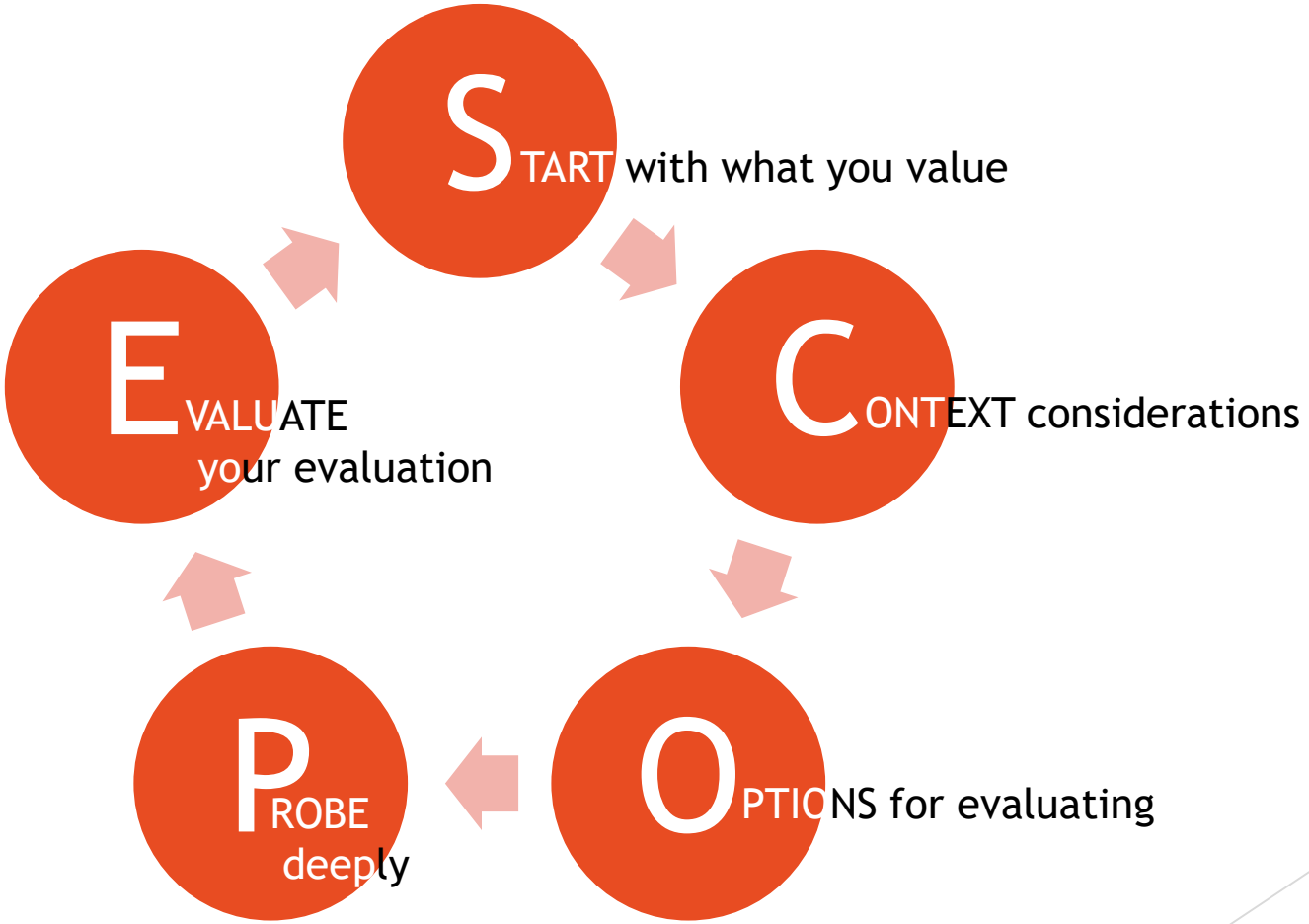
Better decision making through responsible research evaluation

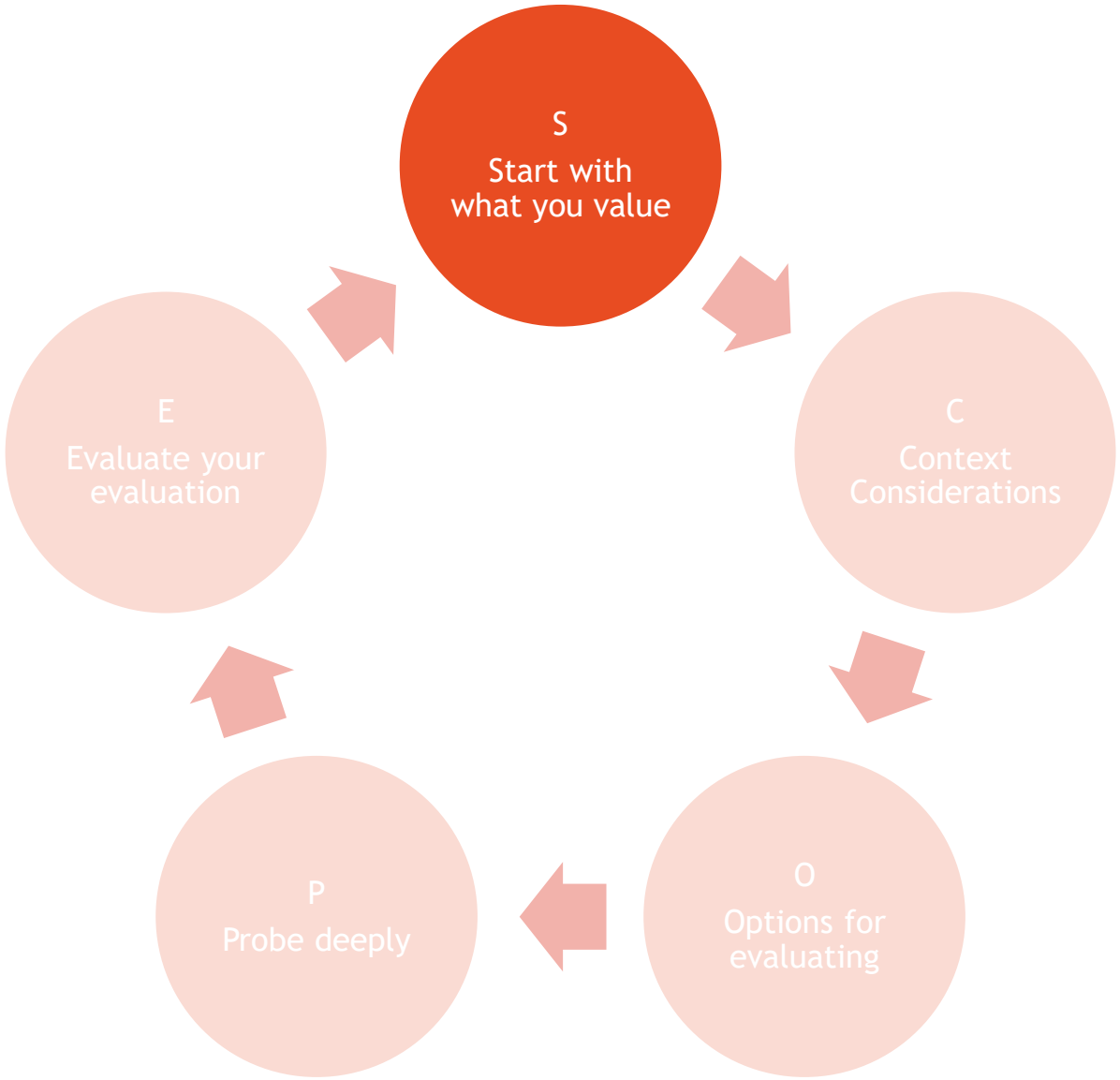
Senior Leader Briefing material

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

SCOPE: 5 stages for doing evaluation responsibly




S
Start with what you value
C
Context Considerations
O
Options for evaluating
P
Probe deeply
E
Evaluate your evaluation

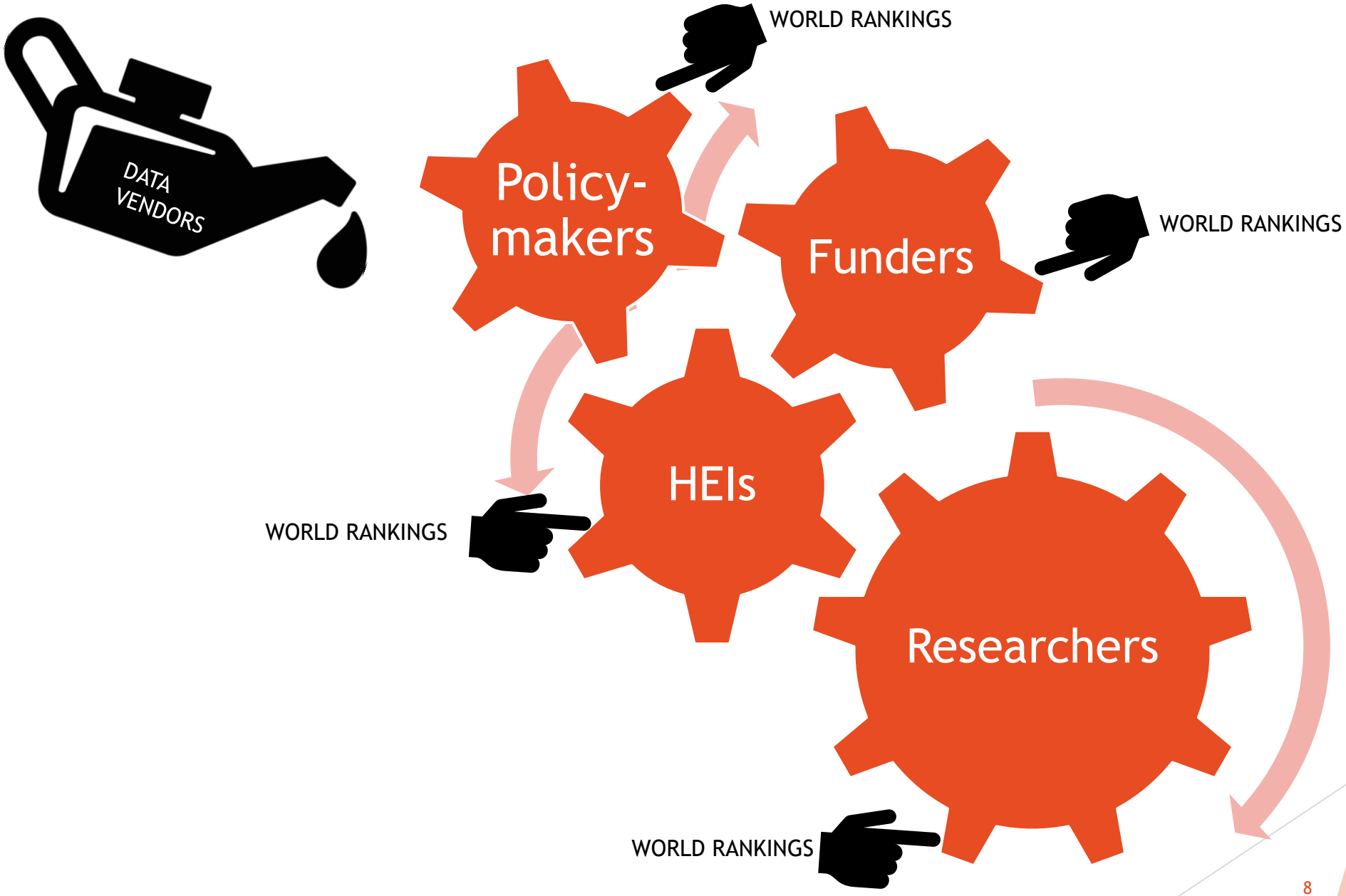
S – Start with what you value

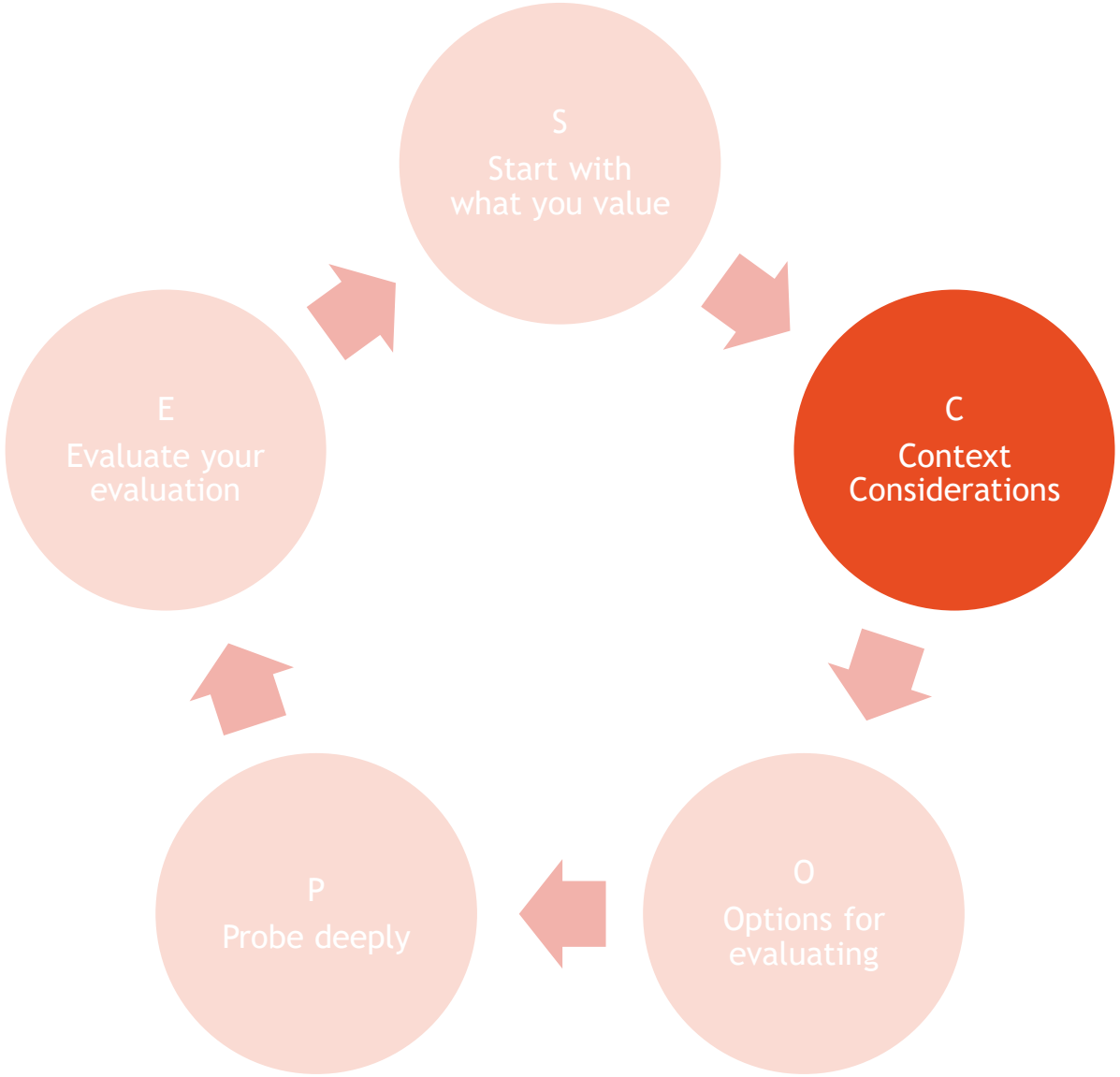
- Evaluation should be guided by an institution's mission and values
  - Not on what can be counted
  - Not on external drivers (such as rankings or national research assessment exercises)
- Allowing external drivers to dictate how and why you evaluate, leads to an outsourcing of those values
  - Keep institutional autonomy instead
  - Maintain distinct institutional character

The Streetlight effect
Measuring by available data not by mission

We are measured, therefore we are

We are measured, therefore we are all the same


DATA VENDORS
WORLD RANKINGS
Policy-makers
Funders
WORLD RANKINGS
WORLD RANKINGS
HEIs
Researchers
WORLD RANKINGS


S
Start with what you value
C
Context Considerations
O
Options for evaluating
P
Probe deeply
E
Evaluate your evaluation

C – Consider the Context

Why are you evaluating?

Who are you evaluating?

What works in one context won’t work in another

Why and what are you measuring? Balancing the risks.

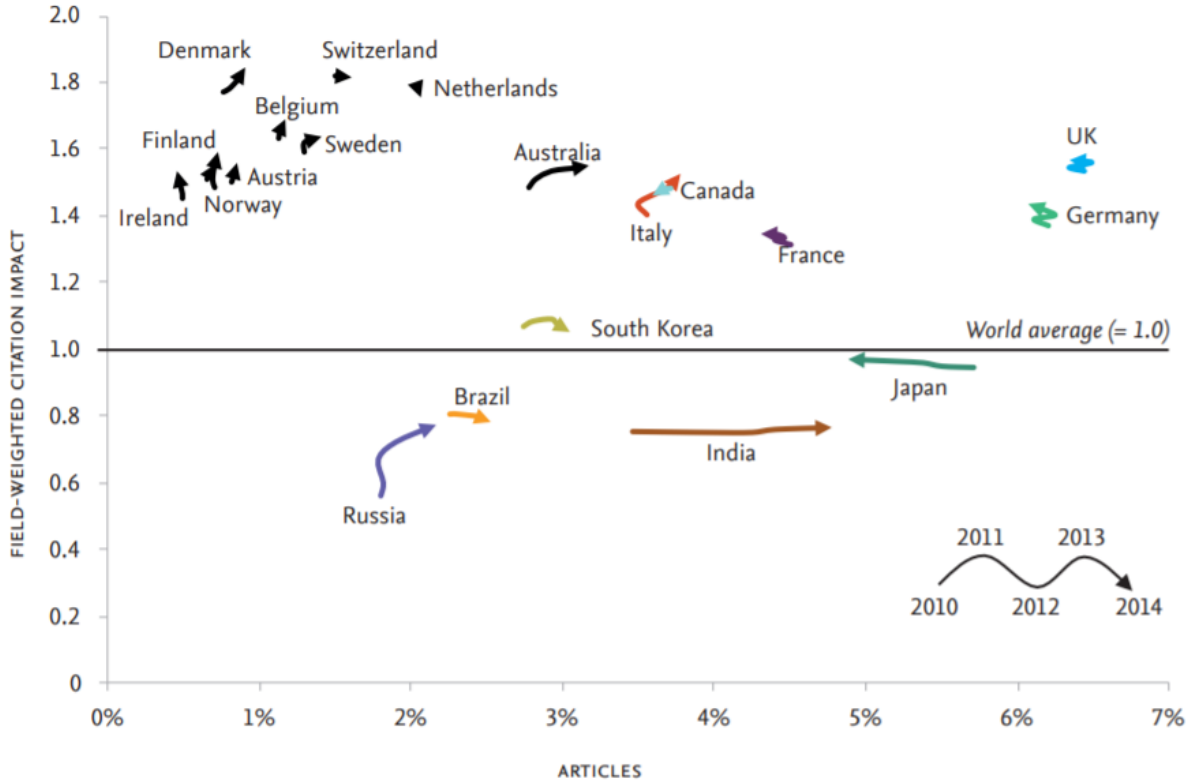
	Country	University	Research Group	Researcher
Analysis				
Advocacy				
Accountability				
Acclaim				
Adaptation				
Allocation				

Low risk
Medium risk
High risk

- Risks associated with metric use in various settings

Use of FWCI in measuring to understand

Panel A(2): *The UK and comparator countries plus top ten countries with the highest field-weighted citation impact in 2014 among OECD countries with at least 5,000 publications in 2014 (excluding the US and China).*



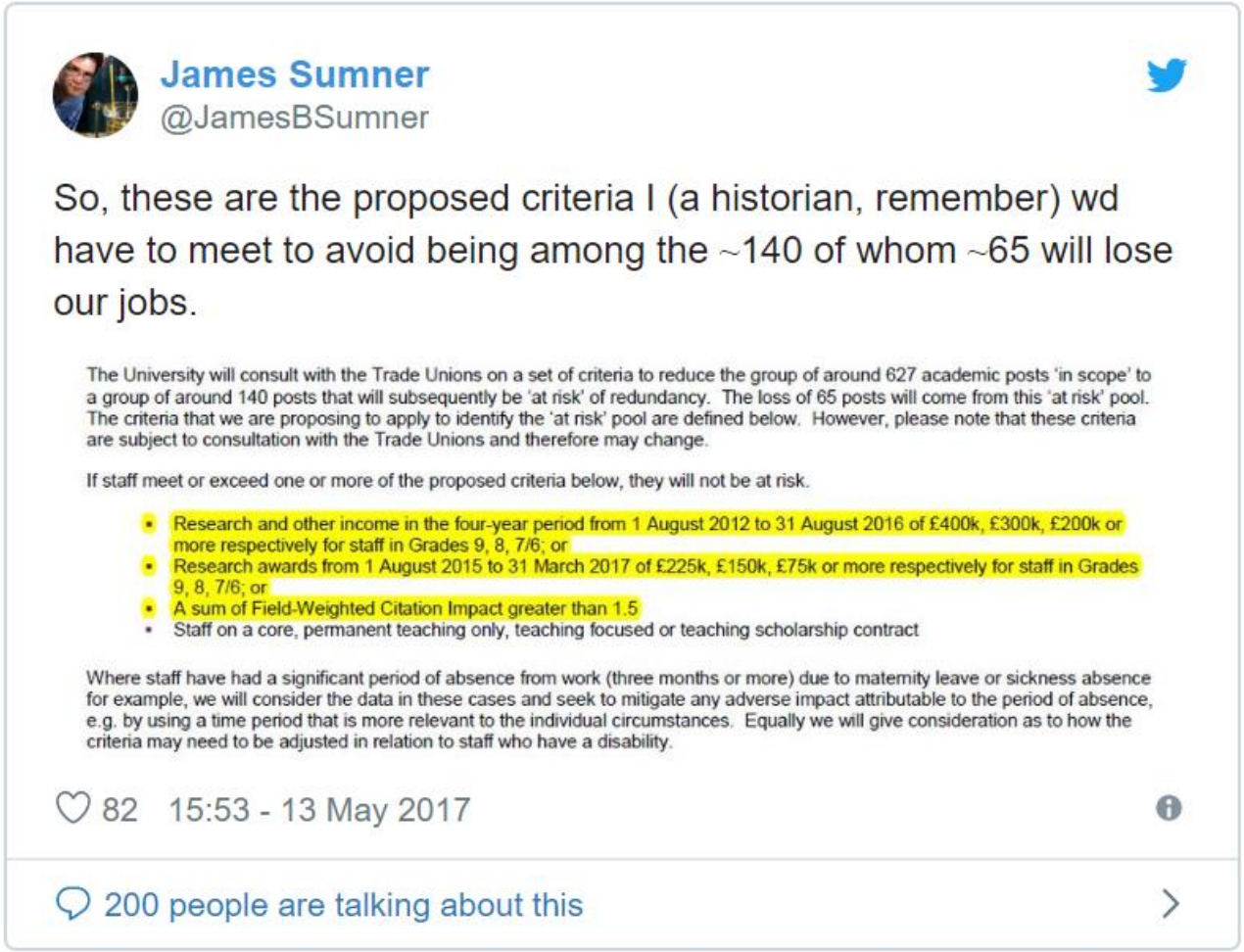
	Country	University	Research Group	Researcher
Analysis				
Advocacy				
Accountability				
Acclaim				
Adaptation				
Allocation				

Low risk	
Medium risk	
High risk	

Measuring to analyse at the level of countries = low risk

International Comparative Performance of UK Research Base – 2016 report on 2011-2014 data
https://www.elsevier.com/__data/assets/pdf_file/0018/507321/ELS-BEIS-Web.pdf

Use of FWCI to identify staff for redundancy…



James Sumner
@JamesBSumner

So, these are the proposed criteria I (a historian, remember) wd have to meet to avoid being among the ~140 of whom ~65 will lose our jobs.

The University will consult with the Trade Unions on a set of criteria to reduce the group of around 627 academic posts 'in scope' to a group of around 140 posts that will subsequently be 'at risk' of redundancy. The loss of 65 posts will come from this 'at risk' pool. The criteria that we are proposing to apply to identify the 'at risk' pool are defined below. However, please note that these criteria are subject to consultation with the Trade Unions and therefore may change.

If staff meet or exceed one or more of the proposed criteria below, they will not be at risk.

- Research and other income in the four-year period from 1 August 2012 to 31 August 2016 of £400k, £300k, £200k or more respectively for staff in Grades 9, 8, 7/6; or
- Research awards from 1 August 2015 to 31 March 2017 of £225k, £150k, £75k or more respectively for staff in Grades 9, 8, 7/6; or
- A sum of Field-Weighted Citation Impact greater than 1.5
- Staff on a core, permanent teaching only, teaching focused or teaching scholarship contract

Where staff have had a significant period of absence from work (three months or more) due to maternity leave or sickness absence for example, we will consider the data in these cases and seek to mitigate any adverse impact attributable to the period of absence, e.g. by using a time period that is more relevant to the individual circumstances. Equally we will give consideration as to how the criteria may need to be adjusted in relation to staff who have a disability.

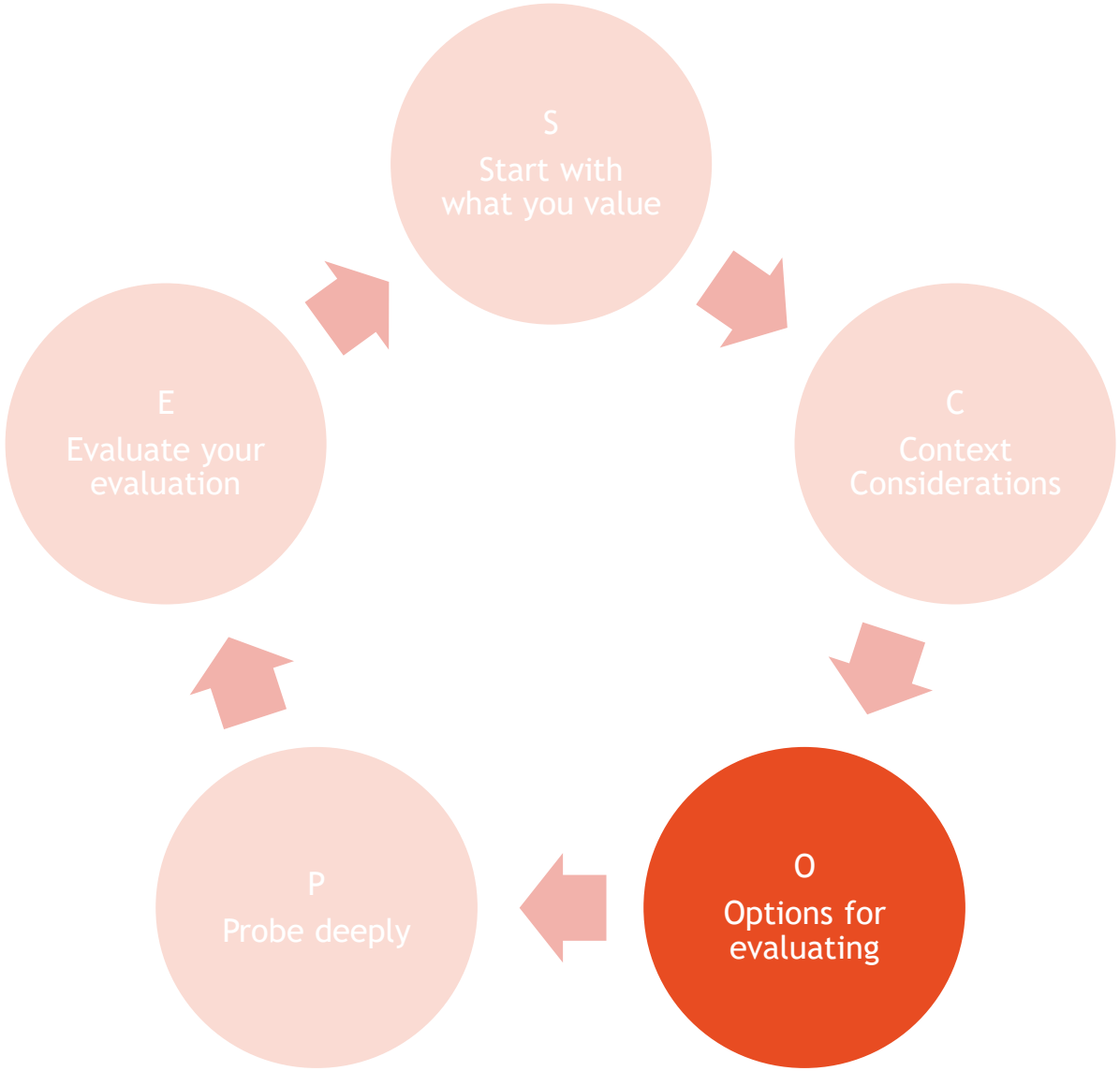
82 15:53 - 13 May 2017

200 people are talking about this

	Country	University	Research Group	Researcher
Analysis				
Advocacy				
Accountability				
Acclaim				
Adaptation				
Allocation				

Low risk	
Medium risk	
High risk	

Measuring to allocate at the level of individuals = high risk

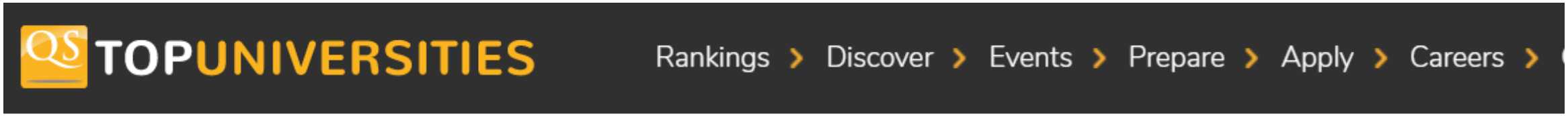

S
Start with what you value
C
Context Considerations
O
Options for evaluating
P
Probe deeply
E
Evaluate your evaluation

Explore all the options

Quantitative - indicators

Qualitative – peer review

Take care when using quantities to indicate qualities



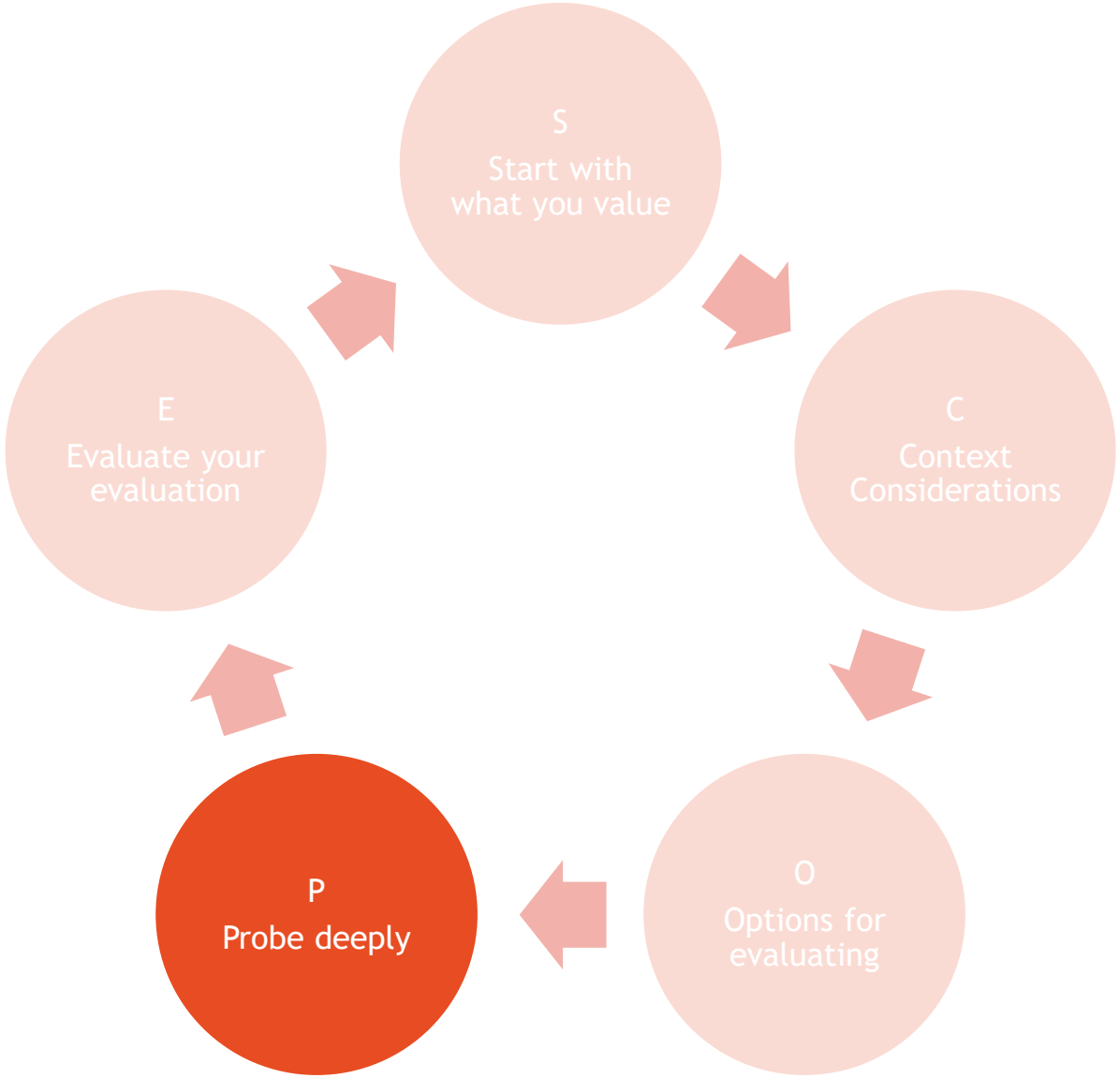
Faculty/Student Ratio (20%)

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.

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


S
Start with what you value
C
Context Considerations
O
Options for evaluating
P
Probe deeply
E
Evaluate your evaluation

P – Probe deeply

- Who does this discriminate against?
- How might this be gamed?
- What might the unintended consequences be?
- Does the cost of measuring outweigh the benefit?

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.

Top Ten

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

52.84
Since 2017

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

43.79
Since 2017

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

40.62
Since 2017

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

38.16
Since 2017

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

31.11
Since 2017

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

30.93
Since 2017

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

30.88
Since 2017

MAURIDHI HERY PURNOMO
Institut Teknologi Sepuluh Nopember
NIDN /NIP/NIDK : 0016095811
Scopus H-Index : 4 | Google H-Index : 17 |
Artificial Intelligence

29.61
Since 2017

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

29.58
Since 2017

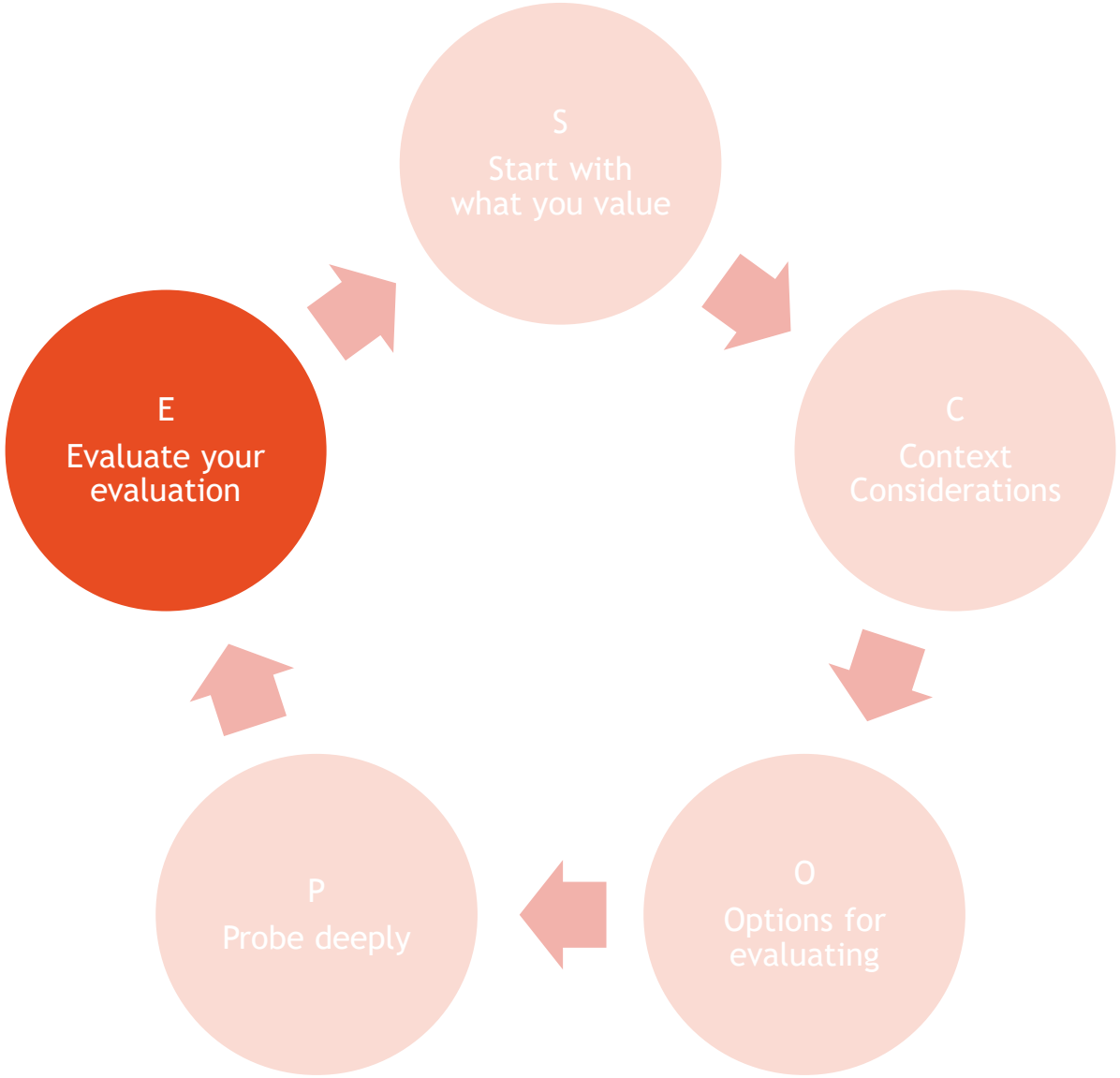
TOLE SUTIKNO
Universitas Ahmad Dahlan
NIDN /NIP/NIDK : 0512067501
Scopus H-Index : 13 | Google H-Index : 19 |

29.24
Since 2017

Number of women = 1
Number of Social Sciences & Humanities researchers = 0

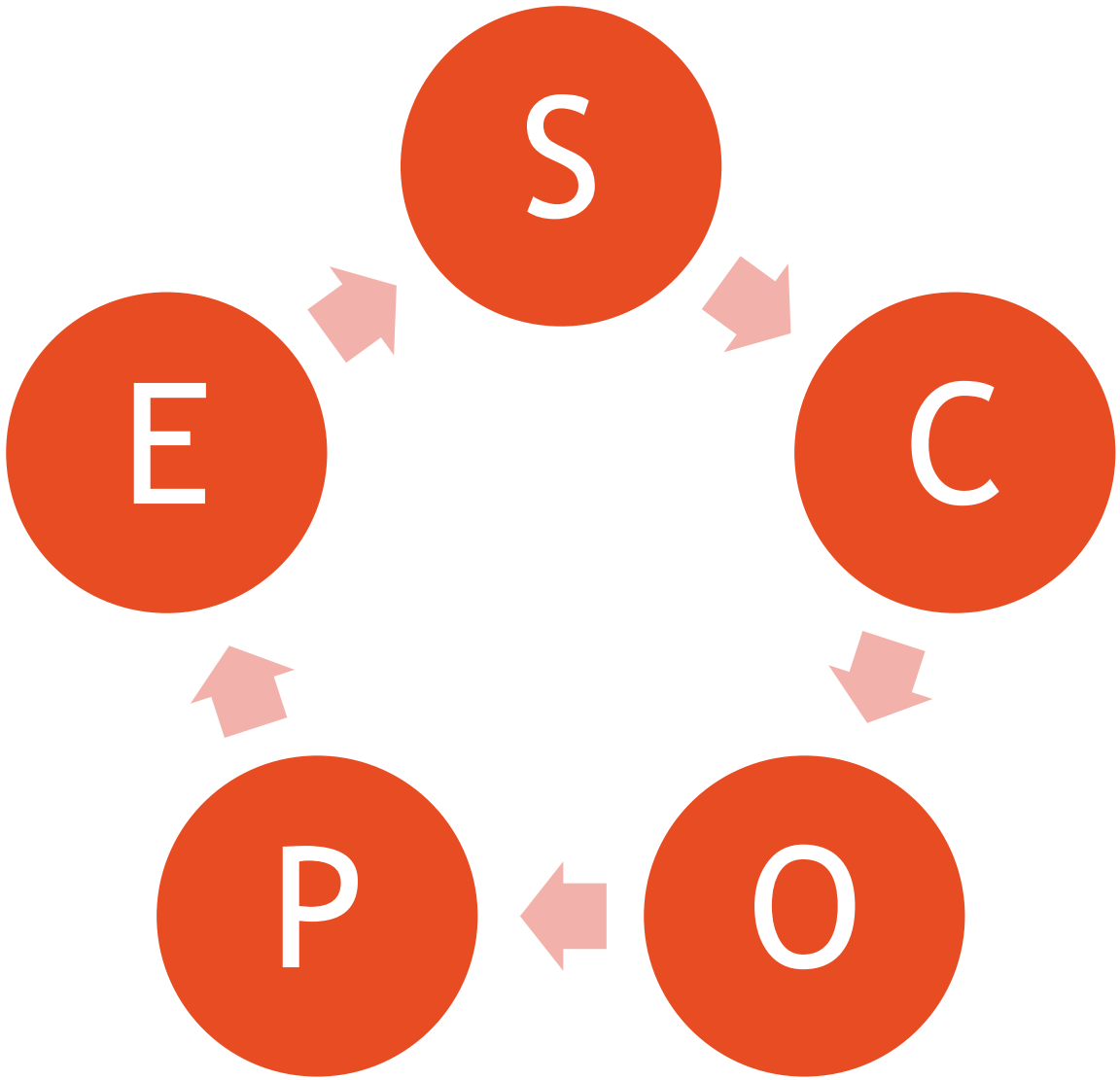
Unintended consequences

- Neglecting all activities that aren't measured
- Goal displacement
- 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


S
Start with what you value
C
Context Considerations
O
Options for evaluating
P
Probe deeply
E
Evaluate your evaluation

E – Evaluate your evaluation

- Did you fulfil the aims of your evaluation?
- Keep performance indicators under review
- Does evaluating research actually make the research any better?
- What will the long-term effects of evaluating be?


S
C
O
P
E