inorms

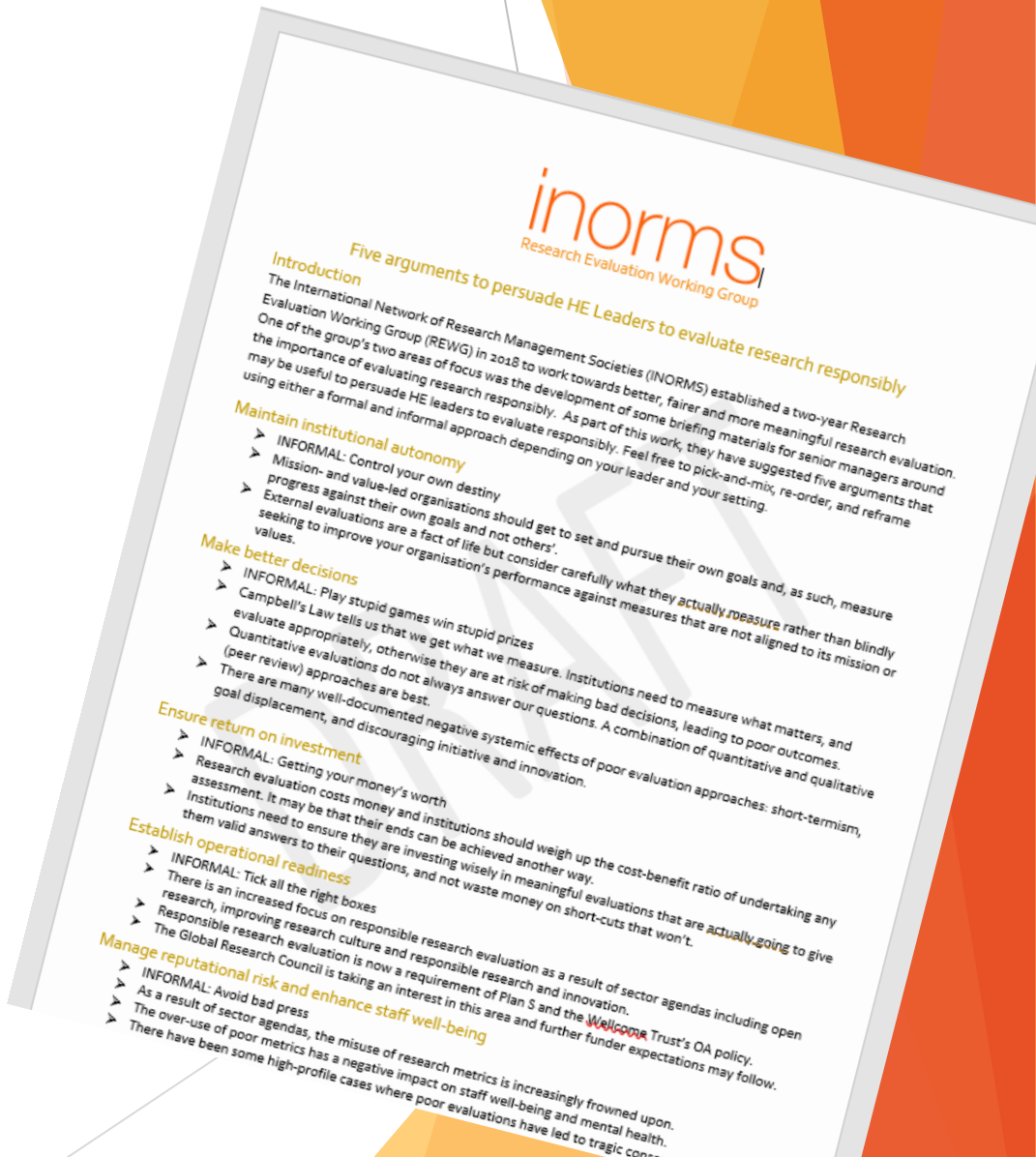
Research Evaluation Working Group

Better decision making through responsible research evaluation

Research Administration
Briefing Material

Why is responsible evaluation important?

- Ensure staff well-being and manage reputational risk
- Make better decisions
- Ensure return on investment
- Establish operational readiness
- Maintain institutional autonomy


inorms

Research Evaluation Working Group

Five arguments to persuade HE Leaders to evaluate research responsibly

Introduction

The International Network of Research Management Societies (INORMS) established a two-year Research Evaluation Working Group (REWG) in 2018 to work towards better, fairer and more meaningful research evaluation. One of the group's two areas of focus was the development of some briefing materials for senior managers around the importance of evaluating research responsibly. As part of this work, they have suggested five arguments that may be useful to persuade HE leaders to evaluate responsibly. Feel free to pick-and-mix, re-order, and reframe using either a formal and informal approach depending on your leader and your setting.

Maintain institutional autonomy

- INFORMAL: Control your own destiny
- Mission- and value-led organisations should get to set and pursue their own goals and, as such, measure progress against their own goals and not others'.
- External evaluations are a fact of life but consider carefully what they actually measure rather than blindly seeking to improve your organisation's performance against measures that are not aligned to its mission or values.

Make better decisions

- INFORMAL: Play stupid games win stupid prizes
- Campbell's Law tells us that we get what we measure. Institutions need to measure what matters, and evaluate appropriately, otherwise they are at risk of making bad decisions, leading to poor outcomes.
- Quantitative evaluations do not always answer our questions. A combination of quantitative and qualitative (peer review) approaches are best.
- There are many well-documented negative systemic effects of poor evaluation approaches: short-termism, goal displacement, and discouraging initiative and innovation.

Ensure return on investment

- INFORMAL: Getting your money's worth
- Research evaluation costs money and institutions should weigh up the cost-benefit ratio of undertaking any assessment. It may be that their ends can be achieved another way.
- Institutions need to ensure they are investing wisely in meaningful evaluations that are actually going to give them valid answers to their questions, and not waste money on short-cuts that won't.

Establish operational readiness

- INFORMAL: Tick all the right boxes
- There is an increased focus on responsible research evaluation as a result of sector agendas including open research, improving research culture and responsible research and innovation.
- Responsible research evaluation is now a requirement of Plan S and the Wellcome Trust's OA policy.
- The Global Research Council is taking an interest in this area and further funder expectations may follow.

Manage reputational risk and enhance staff well-being

- INFORMAL: Avoid bad press
- As a result of sector agendas, the misuse of research metrics is increasingly frowned upon.
- The over-use of poor metrics has a negative impact on staff well-being and mental health.
- There have been some high-profile cases where poor evaluations have led to tragic

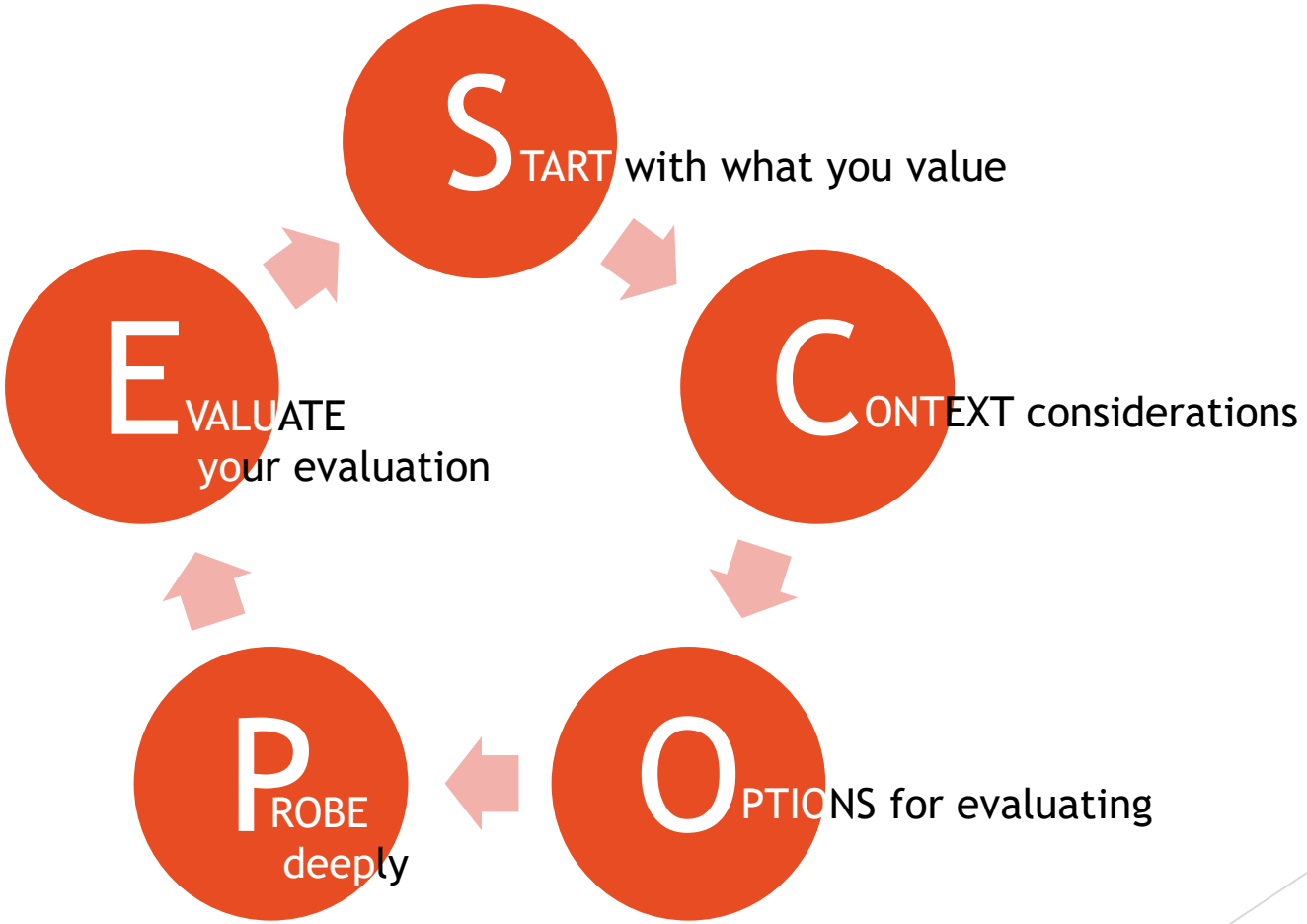
Before starting any evaluation, expertise is needed

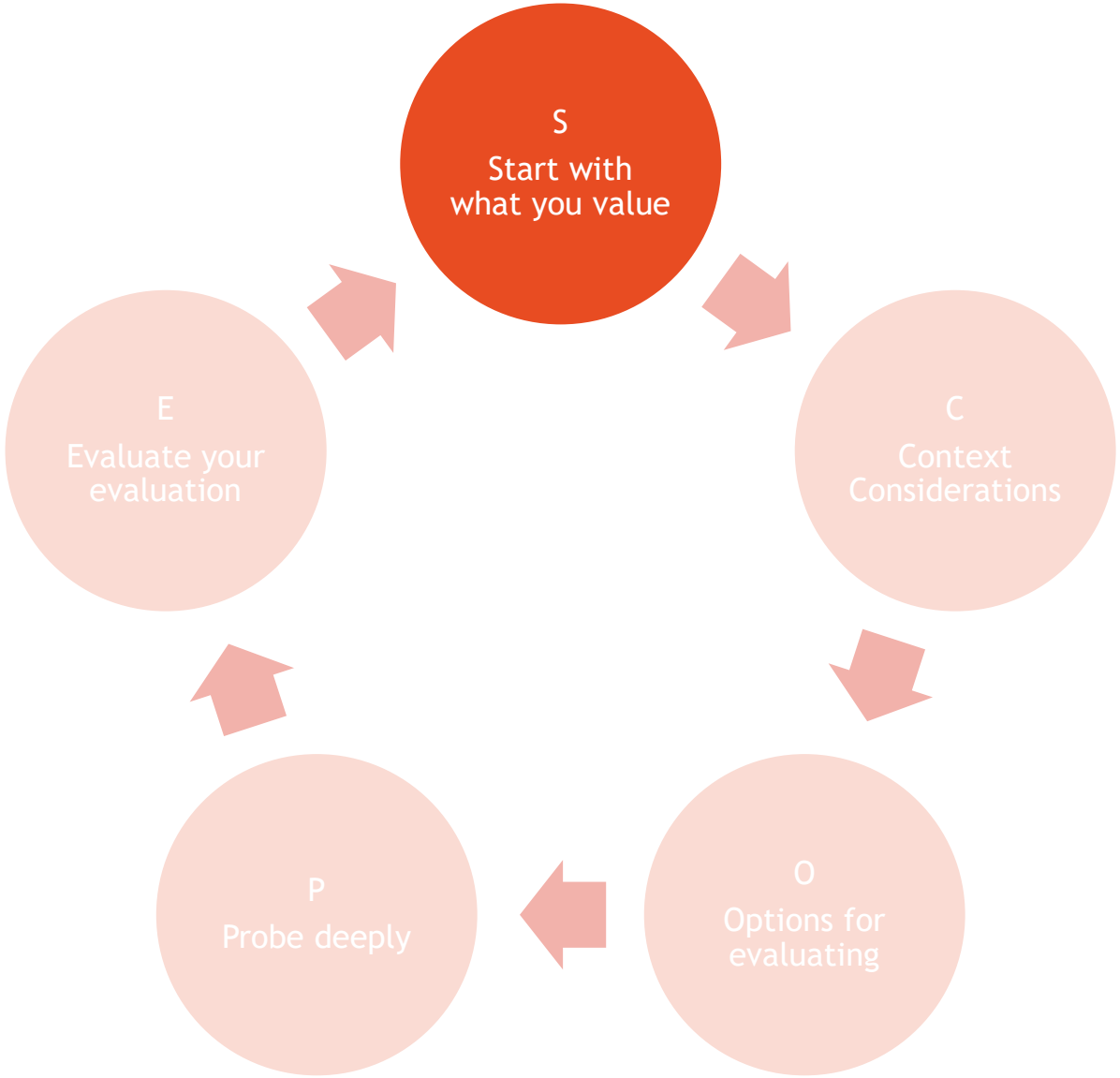
It’s surprising how academic rigour is so often abandoned when it comes to management practice.

- 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.

inorms
Research Evaluation Working Group

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

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

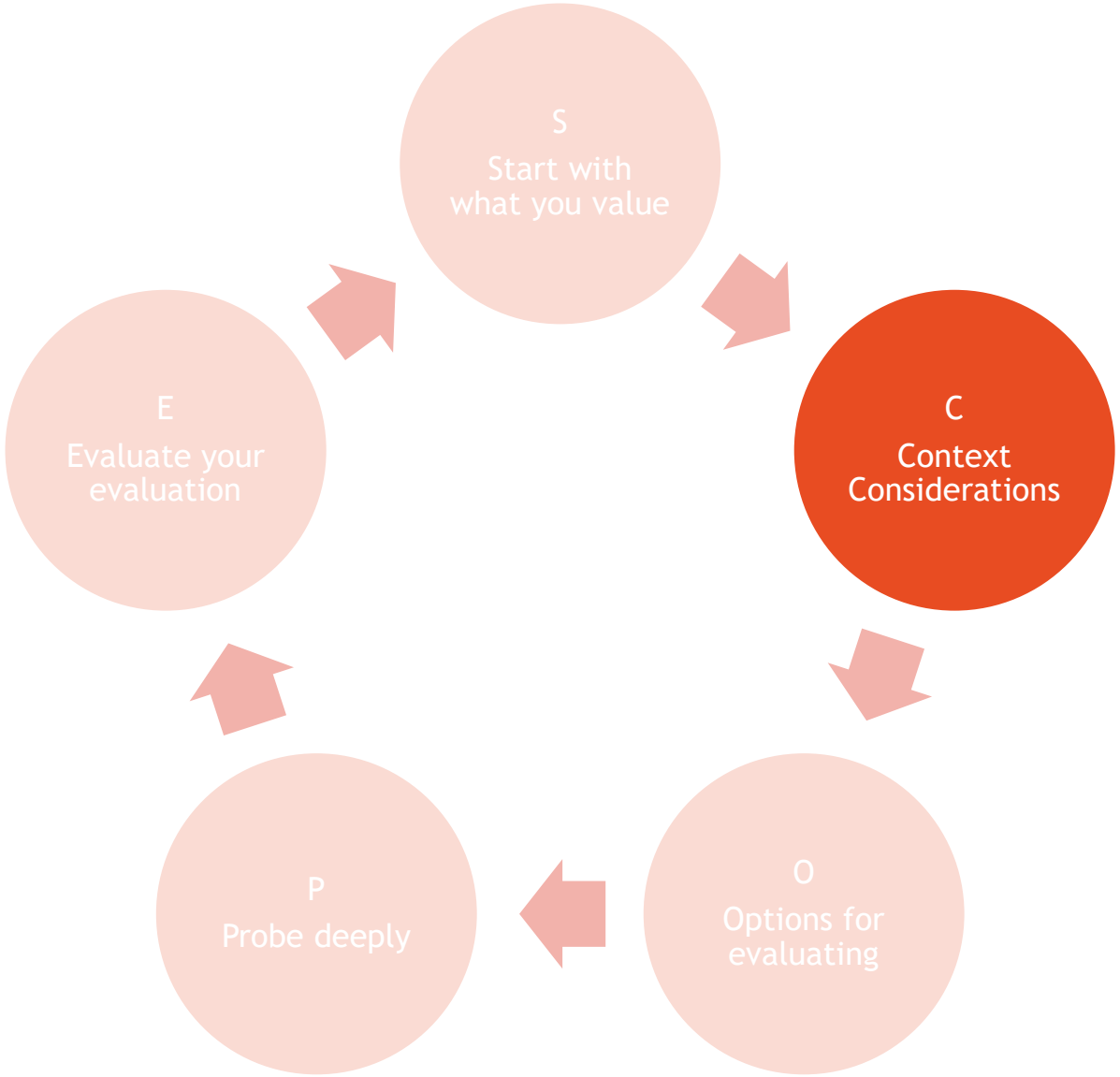
inorms
Research Evaluation Working Group

S-Start with what you value

- HEIs sometimes focus on what OTHERS value (i.e. rankings, national assessment systems, etc.) rather than what THEY value.
- Institutional autonomy is important. Use it or lose it.
- 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.
- So often we value what we think we can measure rather than measure what we value.

inorms
Research Evaluation Working Group

The Streetlight effect:
Measuring by available data not by mission


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

Why are you evaluating?

- **Measure to analyse.** “Science of science” activities that study phenomenon for the sole purpose of understanding them better.
- **Measure to advocate.** “Pick me!” activities. The use of metrics to market an individual, group or university on promotional materials or grant applications.
- **Measure for accountability.** Plotting progress against an objective whether internally or externally set..
- **Measure to acclaim**. The use of indicators to compare one entity with another.
- **Measure to adapt.** The use of indicators to incentivise certain behaviours..
- **Measure to allocate.** 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	HEI	Group	Individual
Analysis	To understand				
Advocacy	To show off				
Accountability	To monitor				
Acclaim	To benchmark				
Adaptation	To incentivise				
Allocation	To reward				

	Low impact
	Medium impact
	High impact

- Risks associated with evaluation activities in various settings

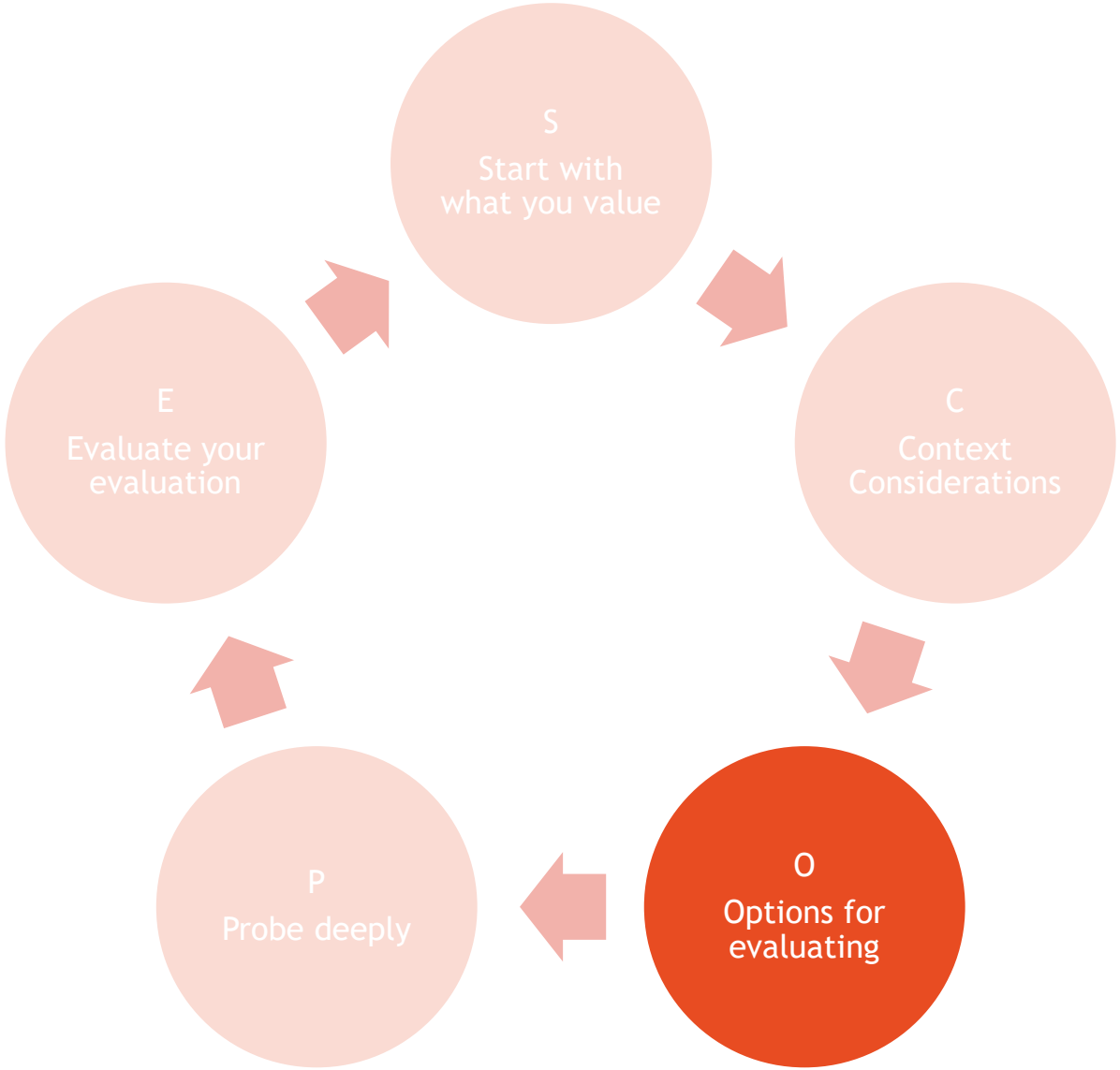
The disparity of scholarly disciplines

- 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
- Metrics are neither objective nor neutral

What works in one context won't work in another

Do you need to evaluate 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)


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

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, excellence, value
- Be careful if using quantitative indicators as a proxy for qualitative things
  - Citations ≠ quality
  - Ranking position ≠ excellence

inorms
Research Evaluation Working Group

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

inorms
Research Evaluation Working Group

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.

inorms
Research Evaluation Working Group

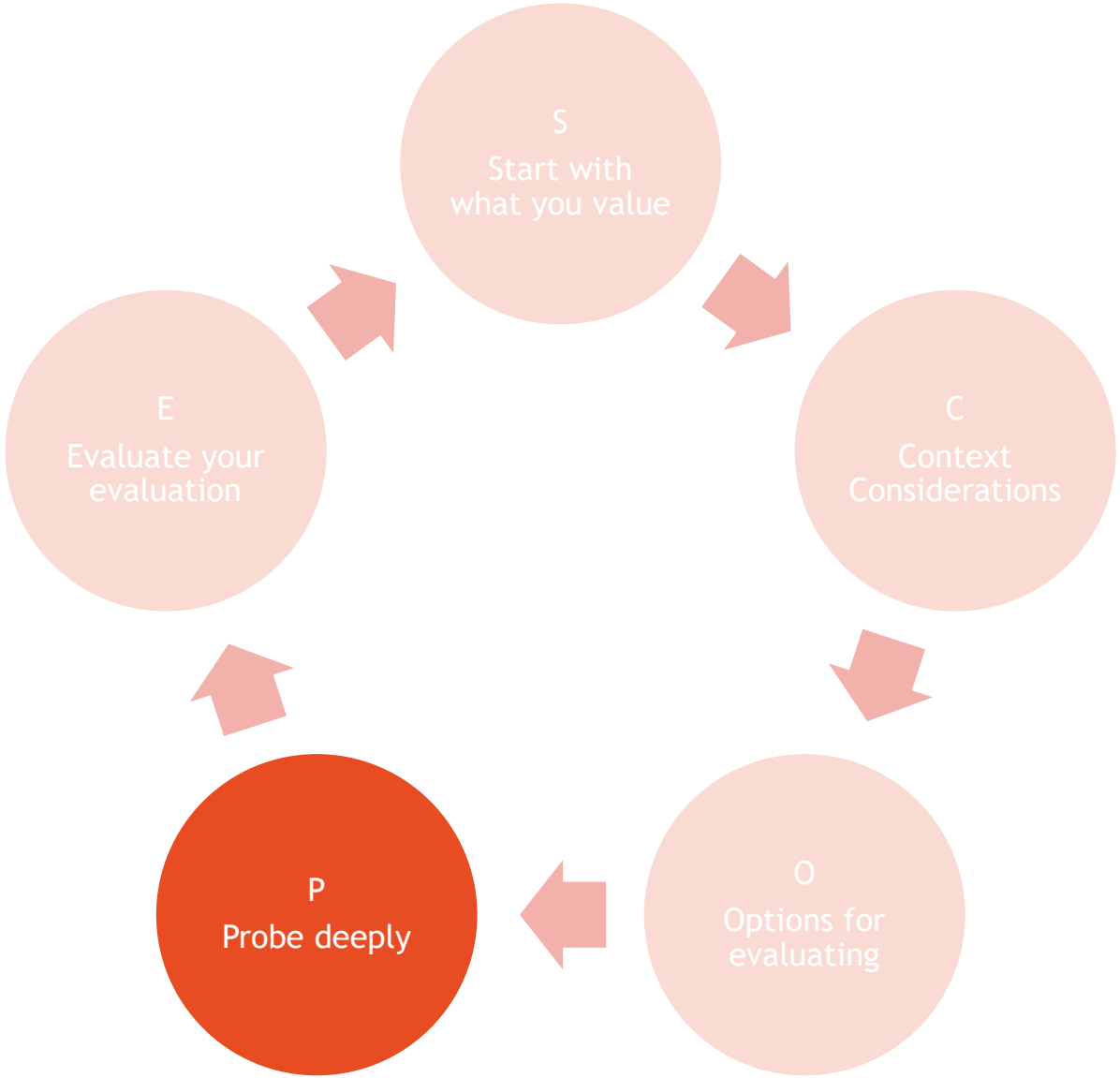
Qualitative measures (peer review)

- 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

inorms
Research Evaluation Working Group

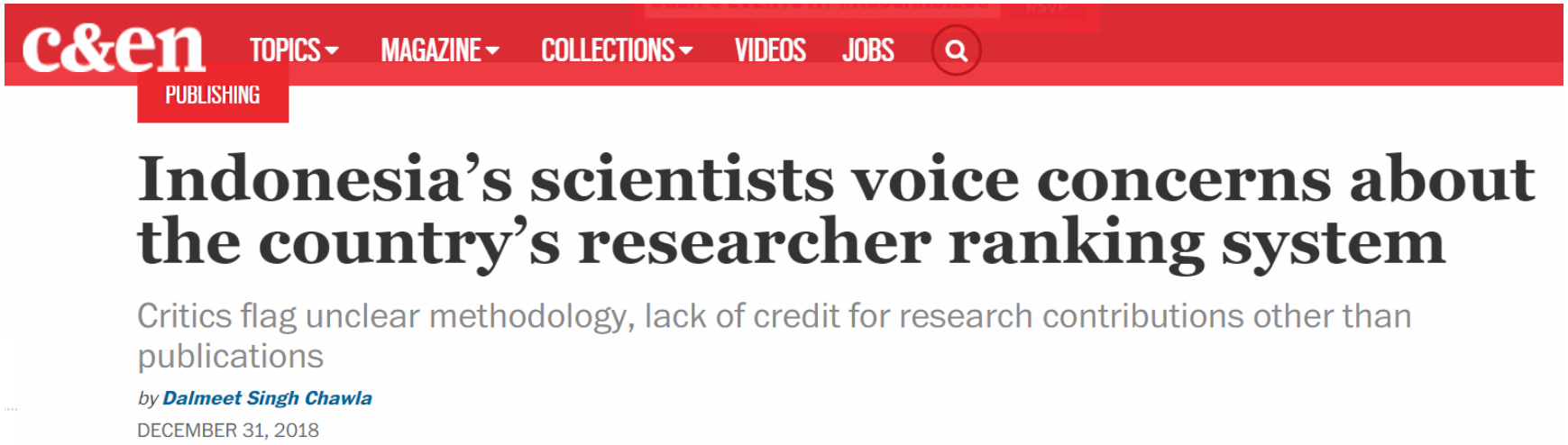
Evaluate WITH the evaluated

- 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' https://www.cwts.nl/blog?article=n-r2u2b4&title=the-evaluative-inquiry-a-new-approach-to-academic-evaluation
  - 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?
  - 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 – 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 propotional 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 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

… 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?

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

Unintended consequences: institutional level

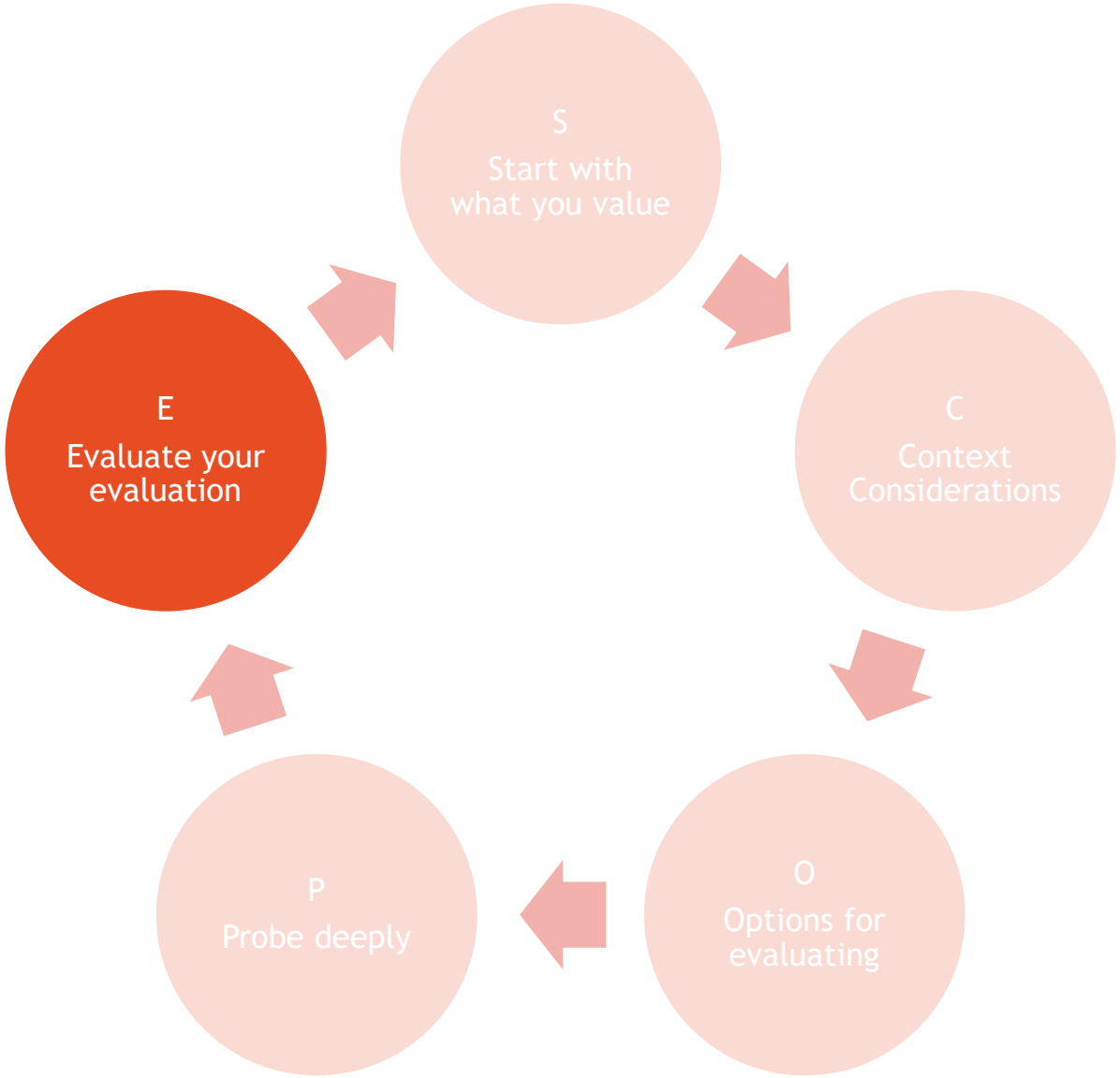
- Poor decision-making
- Neglecting all activities that aren't measured
- Goal displacement, mirroring e.g. government level assessment practice
- Short termism
- Transactional cost of metrics which may not actually add any value to the organisation
- Discourage initiative, innovation and risk-taking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

inorms
Research Evaluation Working Group

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 journals with high impact – research results are published later rather than sooner, the editorial system gets clogged…

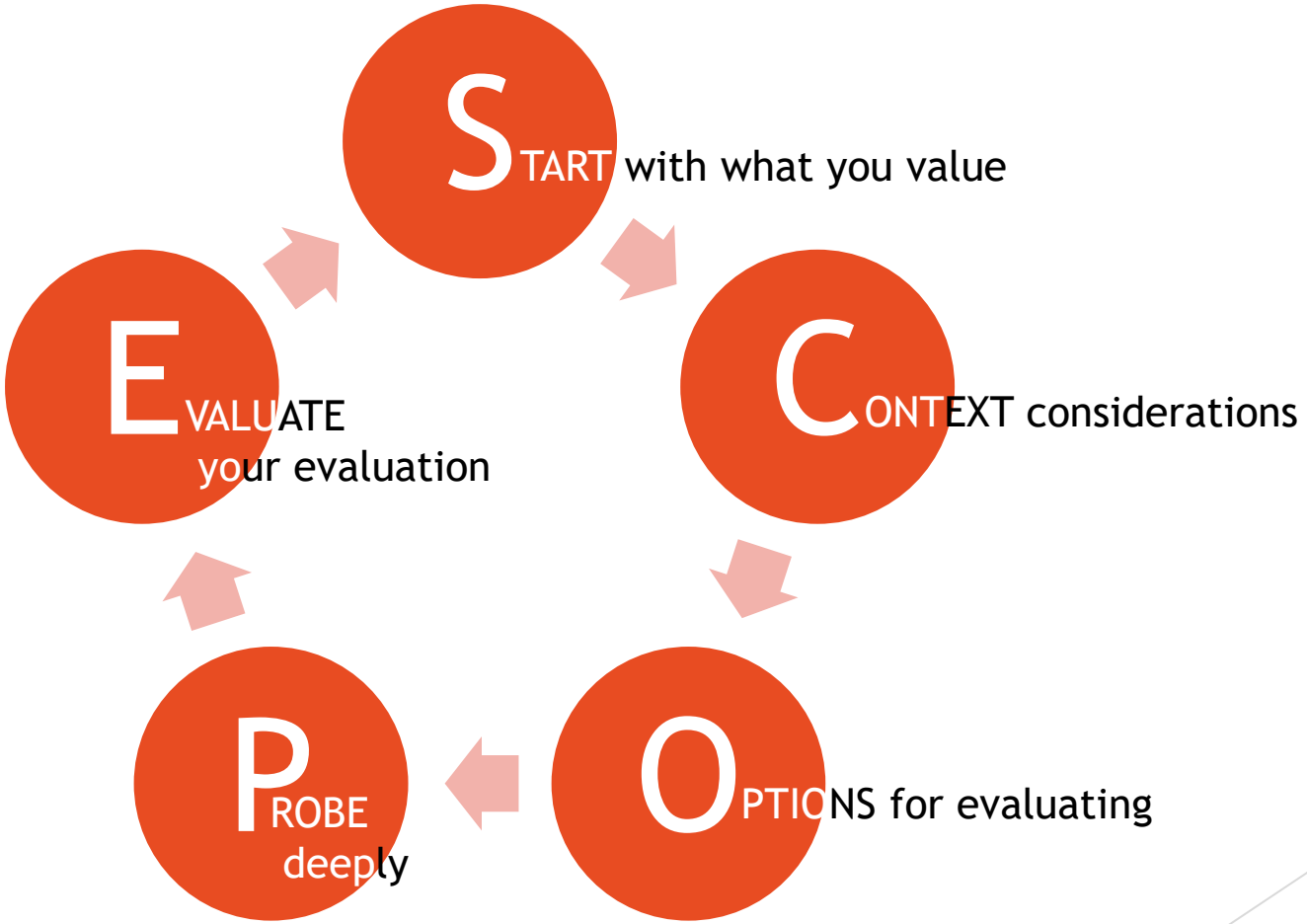
inorms
Research Evaluation Working Group


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

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

SCOPE: 5 stages for doing evaluation responsibly



INORMS Research Evaluation Working Group web pages:

https://inorms.net/activities/research-evaluation-working-group/

Subscribe to the INORMS-RES-EVAL discussion list:

https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=INORMS-RES-EVAL

inorms
Research Evaluation Working Group