Conference:
Future of the sciences in the light of authorship laws, researcher’s ethical codes and government evaluations of the quality of research.

Contextualisation for Responsible Metrics in Evaluation

Ismael Ràfols

CWTS (Centre for Science and Technology Studies), University of Leiden

Ingenio (CSIC-UPV), Universitat Politècnica de València

SPRU (Science Policy Research Unit), University of Sussex
Summary

- **Indicator-based evaluation may lead to some negative consequences**
  - Goal displacements: publishing become an end in itself
  - Task reduction: Some research / teaching practices and topics may be marginalised
  - Reduction of diversity (language, topics, etc.)

- **Evaluation has different functions & research different purposes**
  - Monitoring, allocation, learning / improvement
  - Academic knowledge, professional training, solving societal problems

- **Need to develop more contextualised evaluation**
  - Specially for the purpose of reflecting on improving practices
  - Indicators inform, but do not substitute for expert judgement (P1)
  - In some contexts - particularly in SSH for societal impact / local issues / (P2/P3)

- **Only possible in more decentralised, autonomous processes?**
  - Not in large national allocation programs -- but at dept / lab level. (appraisal)
  - There are experiences in various agencies / countries
The parable of Prussian scientific forestry (Seeing like a state, J. Scott)

**Forests in pre-modern Prussia**
- Wild
- Uncontrolled
- Unpredictable
- Inefficient for timber production

**Enlightenment and scientific forestry:**
- Cut the wild forest
- Plant Norway spruce – reduce diversity
- Increase yield and predictability
- Loss of forest activities for peasants: (fruits, hunting, medicinal herbs, mushrooms...)
The parable of Prussian scientific forestry (Seeing like a state, J. Scott)

Monocultures and Forest death

• Nutrient depletion leading to 20-30% production loss in 2nd generation
• Storm felling
• Pests due to loss of ‘services’ of insects, birds and animals.

Restoration forestry or forest hygiene:

• Artificial ant colonies & spiders
• Wooden boxes to provide bird nests
• The dangers of dismembering a complex set of relations and processes to isolate a single element of instrumental value
**Performativity**

“Backed by state power through records, courts, and ultimately coercion, these state fictions transformed the reality they presumed to observe, although never so thoroughly as to precisely fit the grid.”

**Task reduction**

“Exaggerating only slightly, one might say that the crown's interest in forests was resolved through its fiscal lens into a single number: the revenue yield of the timber that might be extracted annually..”
Problems, research, indicators and peripheries

Space of problems

Space of research

Space of STI indicators
Problems, research, indicators and peripheries

Space of problems

Space of research

Research well illuminated by indicators
Problems, research, indicators and peripheries

Space of problems

STI Peripheries:
- research spaces not well captured by indicators
- research well illuminated by indicators
Streetlight effect in indicators: mistaking light with “problems”

Problems and research not well covered is not random:

SSH in UK
Indicators make “peripheral” spaces invisible

Problems and research not well covered is not random:

- Public health in rural Brazil
- Space of problems
  - Space of research
Indicators make “peripheral” spaces invisible

Problems and research not well covered is not random:

Agriculture in Chad

Space of problems

Space of research

Space of problems

Space of research
Streetlight effect in indicators: mistaking light with “problems”
Problems, research, indicators and peripheries

Multiple types of space:

Geographical: regional, “South”

Cognitive: SSH, engineering

Linguistic: non-English

Sectoral: low-tech, agriculture, creative ind.

Social: gender, minorities

STI Peripheries: research spaces not well captured by indicators

Research well illuminated by indicators

Usual multiple peripheries: primary health care of poor neighbourhoods in Salvador de Bahia, written in Portuguese (Cog, Geo, Ling, Soc.)
The streetlight effect and the performative role of indicators

- Use of indicators may have consequences on research system

- Incentive structure: indicators signal to stakeholders what is important.
  - Goal displacement: instead of mission, follow indicators

- Potential suppression of diversity
  - Shift towards English publications (Siversten, 2014)
  - Shift towards more technical / mainstream issues
  - Diversion of research away from local or national issues (Hicks, 2015)
  - Bias toward positive reporting (Ioannidis, 2005)
  - Invisible / undone science (Hess, 1997)

We cannot correct for these biases mathematically → expert values are needed
Goals of evaluation and functions of research

Evaluation (a dictionary definition): ‘To ascertain or fix the value or worth of an object against a certain specified criteria.’

Goals of research evaluation (Molas-Gallart, 2012; Adam, 2018)

- Monitoring or auditing (or control)
- Distribution of resources
- Improving or learning
Goals of evaluation and functions of research

Evaluation (a dictionary definition): ‘To ascertain or fix the value or worth of an object against a certain specified criteria.’

Missions of research (Molas-Gallart et al., 2003)

- Academic contribution
- Education & training
- Societal contributions
Bib. indicators vs. goals of evaluation and functions of research

• **Goals of research evaluation** (Molas-Gallart, 2012; Adam, 2018)
  - Auditing: bib. indicators convenient
  - Distribution of resources: bib. indicators useful for justification but not necessarily for setting strategy
  - Improving or learning: bib. indicators not very helpful

• **Functions of research**
  - Academic contribution: bib. indicators can be appropriate in some fields (natural sciences)
  - Education & training: bib. indicators NOT appropriate
  - Societal impact: bib. indicators NOT appropriate

Bibliometrics ONLY for auditing SOME academic contributions

However, in a bureaucratic system – auditing academia is main goal

Useful for managers but possibly harmful for the research system
Uses and abuses of bibliometric indicators tend to favour the ivory tower

San Francisco Declaration of Research Assessment (DORA)

Don’t use Impact Factors!!

ASCB (2013)

Conventional indicators perceived favour elitist research in the ivory tower.

Barrier to adoption of Open Science

The Metric Tide

UK (Hefce) Report

Use indicators with

- Robustness
- Humility
- Transparency
- Diversity
- Reflexivity

Wilsdon et al. (2015)

Need to contextualise

Indicator support (not replace) judgement.

Hicks et al. (2015)
Principles of the “The Leiden Manifesto”

1. Quantitative evaluation should support qualitative, expert assessment.

2. Measure performance against the research missions of the institution, group or researcher.

3. Protect excellence in locally relevant research.

4. Keep data collection and analytical processes open, transparent and simple.
5. Allow those evaluated to verify data and analysis.

6. Account for variation by field in publication and citation practices.
7. Base assessment of individual researchers on a qualitative judgement of their portfolio.
8. Avoid misplaced concreteness and false precision.

9. Recognize the systemic effects of and indicators.
10. Scrutinize indicators regularly and update them.

Hicks et al. (2015)
Evaluation informed by indicator frameworks

Dimensions to consider in evaluation
• Goal of evaluation
• Research mission
• Level of assessment
• Scientific field and methodological approaches
• Potential stakeholders, audiences and beneficiaries
• Research environment: Human and technical resources

For a given configuration of relevant dimensions, certain indicators will be relevant while others will not.

Attention:
• There cannot be general indicators of research ‘quality’ or OS

Between FULLY TAILORED – UNIVERSAL

EC Expert Group on Indicators for Researchers' Engagement with OS
(P. Wouters (chair), B. Holbrook, M. Jacob, Lynn Kamerlin, A. Oancea, I. Ràfols)
Science as a biosphere with dozens of diverse ecosystems
To evaluate means to value -- to make our values explicit

It the wake of movements towards Open Science and RRI
We need:

Forms of more ‘open’ research:
  • More oriented to public good and close to society
  • More diverse and pluralistic
  • With the participation of social actors

This requires forms of evaluation that

Foster pluralization  ➔ Since S&T are uncertain and there are options
Sensitive to values   ➔ ¿What type of research is socially desirable?
Contextual            ➔ depending of specific socio-ecological spaces

Need to change (complement) the FORMS of evaluation:
  • From centralised large evaluations to small (dept/lab) assessment
  • By experts aimed at learning / improving
  • Focus on the scientific and societal goals posed (not abstract ‘quality’)
What is research for?

If indicators are the answer, what is the question?

“It is sometimes easier to develop quantitative ‘indicators’ of performance than to work out what the program has to accomplish.”

David Roessner (2000)
Evaluar significa valorar -- explicitando nuestros valores

Una investigación:
- más orientada al bienestar social de tod@s.
  CyT más plural y más justa
- más cercana, más abierta a la sociedad
  procesos de evaluación con participación de agentes sociales
- indicadores para informar la deliberacion, abre el debate

¿Qué tipos de evaluación?

Pluralizadora  → el futuro de CyT es abierto, hay opciones diversas
Con valores    → ¿qué tipo de investigación es socialmente deseable?
Contextual     → depende de espacios sociales y naturales específicos.

Propuesta para la bibliometría:
- No se trata de buscar nuevos indicadores sino de *pensar cómo los indicadores pueden contribuir a nuevas formas de evaluación*
Classic values of science

In 1942, Merton wrote about the aspirational values of science

- Communalism (shared knowledge)
- Universalism (all humans can participate)
- Disinterestedness (public good)
- Organised Scepticism (scrutiny & transparency)

...key values behind Open Science

His argument was prompted by use of science in authoritarian regimes (1930s-40s):

His answer: ‘Good’ science blossoms in pluralistic, democratic societies”

- Authoritarian times
  - Suppression of rights and liberties, especially for migrants & minorities

Bureaucratic orders vs. public good
‘Open’ science in ‘closed’ societies?
Evaluation -- about valuing -- about values -- Open science?

- OS in a time of increasing authoritarian govs.
- OS policies in China, Turkey, Russia? ... or some EU countries?

- Gov’t decree to shut down websites without judiciary request
- Police requests to shut down of Git-hub application
  - Application for coordinating demonstrations at Git-hub. Civil disobedience in non-violence actions
A research evaluation process

What are the goals of this research?

What are the criteria to see alignment of activities with goals?

What indicators capture the criteria?

ISRIA Statement, Paula Adam et al. 2018
References


Rottenburg, Richard, Sally E. Merry, Sung-Joon Park, and Johanna Mugler, eds. The world of indicators: The making of governmental knowledge through quantification. Cambridge University Press, 2015.


Manifesto of OCSDNet (Open and Collaborative Science in Development Network)

For all the claimed benefits of OS...
... current model is NOT making science a more inclusive practice.
... many scientists continue to be underrepresented and excluded
... new technologies exclude those with limited digital rights.
... citizens rarely get to shape the research agenda.

Principles
• knowledge commons
• cognitive justice
• situated openness
• right to research
• equitable collaboration
• inclusive infrastructures
• use knowledge as a pathway to sustainable development
Evaluation informed by indicator frameworks

Dimensions to consider in evaluation

• Goal of evaluation (monitoring, allocation, learning)
• Research mission (academic, training, social contribution)
• Level of evaluation (system, institutional, individual)
• Scientific field and methodological approaches
• Potential stakeholders, audiences and beneficiaries
• Research environment:
  • Human and technical resources
  • OS capabilities, infrastructure

Between FULLY TAILORED & UNIVERSAL → Prêt-à-porter