

OECD, Paris, 25th March 2013

From technical standards to good practices in quantitative assessment of S&T

Ismael Rafols

Ingenio (CSIC-UPV), Universitat Politècnica de València &
SPRU —Science and Technology Policy Research, University of Sussex

Acknowledgements: This presentation has benefited from recent discussions with Ghislaine Filliatreu, Sarah de Rijcke, Paul Wouters and Yves Gingras.

Summary of argument: Beyond technical standards: shift attention from “metrics supply” to “metrics demand”

- Broader discussion about the use of metrics in mgmt/evaluation
- S&T metrics do much more than measure: they signal desired outcomes, produce behavioural changes and have policy effects.
 - Normative effects.
- Behavioural changes and policy effects.
 - Problems are not due to lack of standards –but to evaluative context.
- Technical standards are insufficient.
- Need for guidelines on use of metrics (Barré, 2010) :
 - From justification ... (closing down the debate)
 - ...to deliberation (opening up the debate)

Beyond technical standards: shifting attention from metrics supply to indicators demand

In spite of the title of the session, this not a only a discussion about technical or bibliometric standards.

Let's read again text introducing the session:

It is not only about the '***desirability and feasibility of standards***', in the narrow technical sense,

but also about standards in other senses such as '***the way in which data are recorded in databases***' (let's remember that there is a very serious bias in databases)...

...but also in '***the practices used by public agencies and ministries***' (let's remember that validity and reliability of indicators depends on the specific use)

It is important to put the discussion in context, and talk not only about the technical issues about standards, but more importantly about the **uses of metrics in science policy, management and evaluation.**

Indicators: from “pure” to “socially robust”

Since the uses of indicators are NOT given by technical issues, the process of thinking about good practices should not only be left only to expert scientometricians or statisticians.

Rémi Barré made the distinction between reliable knowledge on metrics (scientometricians -- indicators suppliers) and socially robust knowledge on use of metrics (which involves deliberation between stakeholders --related to the demands of indicator). (Barré, 2010)

This suggests that the governance of scientometrics can have two interrelated tracks:

Technical discussion on metrics (measurement as a "pure" science, detached from policy context)

Discussion on socially robust metrics (about the use) (this is what Roger Pielke calls, the "Honest Broker") (Pielke, 2007)

The effects of scientometric indicators: policy consequences of evaluative scientometrics (I)

- S&T Indicators have a **performative** role". They do not only measure. They don't 'just happen to be used' in science policy. They are a constitutive part of the incentive structure for "disciplining". They signal to stakeholders what is important.
- Indicators have behavioural consequences and policy effects. These effects often results from explicit or implicit focus of indicators, which sometimes have undesirable consequences. Let us see a few examples of the effects of certain focuses.
 - E.g. Focus on production (quantity) has produced spurious increases in publications by means of so-called salami slicing of publications (Butler, 2003)

The effects of scientometric indicators: policy consequences of evaluative scientometrics (II)

- Focus on highly ranked journals and/or journals indexed in main databases might be resulting in:
 - Shift towards English publications with potential loss of cultural/ linguistic diversity.
 - Suppression of diversity within a discipline (e.g. in economics in the UK, Lee, 2007)
 - Suppression of interdisciplinarity (e.g. in management studies, Rafols et. al., 2012)
 - Diversion of research away from local issues (e.g. very little research on Canada by Canadian economists. This effect is particularly worrying in the social science and humanities.
 - Bias toward positive reporting which leads to bias and lack of reliability of research (Ionnidis, 2005)
 - Increase in fraud, lack of integrity (Martin, 2013)

Can these behavioral changes and policy effects be related to lack of technical standards?

- NO (though further research is necessary for documenting these effects).
- The effects mainly result from the policy context (science governance) in HOW indicators are used.
- Indicators are used for justification for closing down the debate (Stirling, 2010)
- Rémi Barré (2010) proposed that "indicators should be debatable devices, enabling collective learning". "being debatable is not a limitation, but the essence of the contribution of indicators to decision-making processes."
- Rémi Barré suggests that the response to the criticisms on S&T indicators (...) is to engage in building them as socially robust and contextualised indicators, that is, building them with the relevant stakeholders."

Conclusion

- Many policy effects created by indicators today do not arise from lack of standardisation, but due to the manner in which indicators are used, as a means for justification of choices, as a way of closing down the debate).
- A technical and managerial view of science is likely to result in this problematic use of indicators. Perhaps technical standards may mitigate damage. Damage mitigation would be a positive step, but it is only mitigation. **It only acts on indicator supply.** Besides standards it will be important to focus on discussing with user implicit conceptual models, limitations and range of validity.
- Instead, we propose a more systemic view of the S&T system, in which the key issue is to **shape indicator demand by fostering the use of indicators for deliberation processes** -- what we have called "opening up". This acts on the **demand side of indicators.**

References

- Barre, R., 2010. Towards socially robust ST indicators: indicators as debatable devices, enabling collective learning. *Research Evaluation* 19, 227–231.
- Butler, L. (2003). Modifying publication practices in response to funding formulas. *Research Evaluation*, 12(1), 39–46.
- Ioannidis, J.P.A., 2005. Why Most Published Research Findings Are False. *PLoS Med* 2, e124.
- Lee, F.S., 2007. The research assessment exercise, the state and the dominance of mainstream economics in British universities. *Cambridge Journal of Economics* 31, 309–325.
- Lepori, B., Barre, R., Fillitreau, G., 2008. New perspectives and challenges for the design and production of S&T indicators. *Research Evaluation* 17, 33–44.
- Martin, B.R., 2013. Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment. *Research Policy* 42, 1005–1014.
- Pielke, R.A., 2007. *The honest broker. Making sense of science in policy and politics*. Cambridge University Press, Cambridge.
- Rafols, I., Ciarli, T., Van Zwanenberg, P., Stirling, A., 2012. Towards indicators for opening up S&T policy. *STI Indicators Conference*, Montreal.
http://sticonference.org/Proceedings/vol2/Rafols_Towards_675.pdf
- Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., Stirling, A., 2012. How journal rankings can suppress interdisciplinarity. The case of innovation studies and business and management. *Research Policy* 41, 1262–1282.
- Stirling, A., 2008. “Opening Up” and “Closing Down”: Power, Participation, and Pluralism in the Social Appraisal of Technology. *Science, Technology & Human Values* 33, 262–294.