

Towards Indicators for ‘Opening Up’ Science and Technology Policy¹

Ismael Rafols,^{*} Tommaso Ciarli,^{**} Patrick van Zwanenberg^{**} and Andy Stirling^{**}

^{*}*i.rafols@sussex.ac.uk*

SPRU–Science and Technology Policy Research, University of Sussex, Brighton
INGENIO (CSIC-UPV), Universitat Politècnica de València, València

^{**}*t.ciarli@sussex.ac.uk; p.f.van-zwanenberg@sussex.ac.uk, a.c.stirling@sussex.ac.uk*
SPRU–Science and Technology Policy Research, University of Sussex, Brighton

Abstract

As a reaction to the simplistic use of scientometric tools in policy, efforts have been made to improve the robustness of measurements by broadening the range of inputs considered in scientometric evaluations. While this ‘broadening out’ of ‘inputs’ in scientometric appraisal is, in our view, commendable, we propose in this paper that a second dimension also needs to be considered. This relates to the extent to which the ‘outputs’ of appraisal ‘open up’ contrasting conceptualisations of the phenomena under scrutiny and consequently allow for more considered and rigorous attention to alternative policy options, both by decision makers and within wider policy debate. We use a recent comparative study on the performance and interdisciplinarity of six organisational units (Rafols & *al.*, 2012) to illustrate the difference between increasing the range of inputs (‘broadening out’) and enhancing the diversity of outputs to policy decision making (‘opening out’). In doing so, we propose that policy appraisal can inform decision making in a more rigorous ‘plural and conditional’ fashion by acknowledging the way in which divergent metrics and associated normative assumptions can yield contrasting understandings of both the phenomena under scrutiny, and of appropriate policy responses.

Introduction

Recent years have seen much critical debate over the simplistic use of scientometric tools for formal or informal appraisal of science and technology (S&T) organisations (e.g. in university rankings) or individuals (e.g. the h-index) (Roessner, 2000; Van Raan, 2004; Weingart, 2005). As a reaction to these critiques, efforts have been made to improve the robustness of measurements by broadening the range of inputs considered in scientometric evaluations. Examples include the inclusion of books and national or regional journals (Martin & *al.*, 2010), or more recently ‘altmetrics’ (i.e. metrics based on alternative data sources, see Priem & *al.*, 2010). In doing so, the S&T indicator and policy communities have reverted to an early conventional wisdom that scientometrics should rely on multiple sources of data that may provide ‘converging partial indicators’ (Martin & Irvine, 1983).

While this ‘broadening out’ of the range of data used as ‘inputs’ in scientometric appraisal is, in our view, commendable (Stirling, 2003), we propose in this paper that a second dimension also needs to be considered. This relates to the extent to which the ‘outputs’ of appraisal ‘open up’

¹ We acknowledge support from the US National Science Foundation (Award #1064146, “Revealing Innovation Pathways: Hybrid Science Maps for Technology Assessment and Foresight”) and by the UK Economic and Social Research Council (Grant #RES-360-25-0076, “Mapping the Dynamics of Emergent Technologies). The findings and observations contained in this paper are those of the authors and do not necessarily reflect the views of the funders.

contrasting conceptualisations of the phenomena under scrutiny and consequently allow for more considered and rigorous attention to alternative policy options, both by decision makers and within wider policy debate (Stirling, 2005; Stirling & *al.*, 2007, p. 54-58; Leach & *al.*, 2010, p. 102-107). We use a recent comparative study on the performance and interdisciplinarity of six organisational units (Rafols & *al.*, 2012) to illustrate the difference between increasing the range of inputs ('broadening out') and enhancing the diversity of outputs to policy decision making ('opening out'). In this way, policy appraisal can inform decision making in a more rigorous 'plural and conditional' fashion—acknowledging the way in which divergent normative assumptions and metrics can yield contrasting understandings of both the phenomena under scrutiny, and of appropriate policy responses (Stirling, 2008).

Conceptual framework: 'Opening up' versus 'broadening out' in policy appraisal

Many S&T indicators have been developed over the past 50 years as means to reveal the 'strengths' and 'weaknesses' of a given country's 'capacity' and 'performance' in science and technology (Godin, 2003). Developments by the OECD and US National Science Board (NSB), were derived from 'a pure accounting framework based on the anticipated economic benefits of science' (Godin, 2007, p. 1388) and hence with a tendency to take an essentialist understanding of scientific excellence and production, influenced by economic concepts such as 'efficiency' and 'effectiveness' (Narin, 1987). Initial scientometric studies were careful to declare methodological limitations, for example stating explicitly that citations were proxies and 'partial and imperfect' measures of impact rather than quality (Martin & Irvine, 1983). But whether cautious or not, the emphasis of scientometric studies has traditionally lain in producing a 'good' measure of a given concept such as 'scientific excellence', rather than in providing contrasting perspectives on what the meaning of 'excellence' is.

In recent years, various parallel developments have begun to challenge this scientometric *status-quo*. First, the pervasive diffusion of simplistic (and very possibly damaging) scientometric measures such as the h-index at various levels of management has renewed the debate over abuse and misuse of indicators (Weingart, 2005). Second, traditional scientometrics is challenged by alternative data sources, like databases from hitherto excluded countries (e.g. Brazil's Scielo), and new web-based indicators such as publication download frequency or popularity in 2.0 websites like *academia.eu* (Priem & *al.*, 2010). Third, new tools have emerged for data visualisation (e.g. Hans Rosling's *Gapminder*), for large network analysis (e.g. Rosvall & Bergstrom, 2008) and, for science mapping (Börner, 2010), which are radically easing the presentation of complex multidimensional quantitative information to non-experts.

Each of these trends is pushing S&T policy towards use of indicators based on more diverse data inputs. These broader portfolios of inputs can in principle make scientometric analyses more robust. However, we contend here that this improved 'breadth' of inputs need not necessarily translate into a more plural and conditional policy process. 'Opening up' is not just about 'more' indicators, nor is it only a matter of 'positioning' or contextualising (Lepori, 2006). It's about the design and use of indicators aimed explicitly at providing plural policy understandings and options. For S&T policy to be 'opened up', indicators used in appraisal need to be re-conceived as 'debatable devices, enabling collective learning' (Barré, 2010, p. 227).

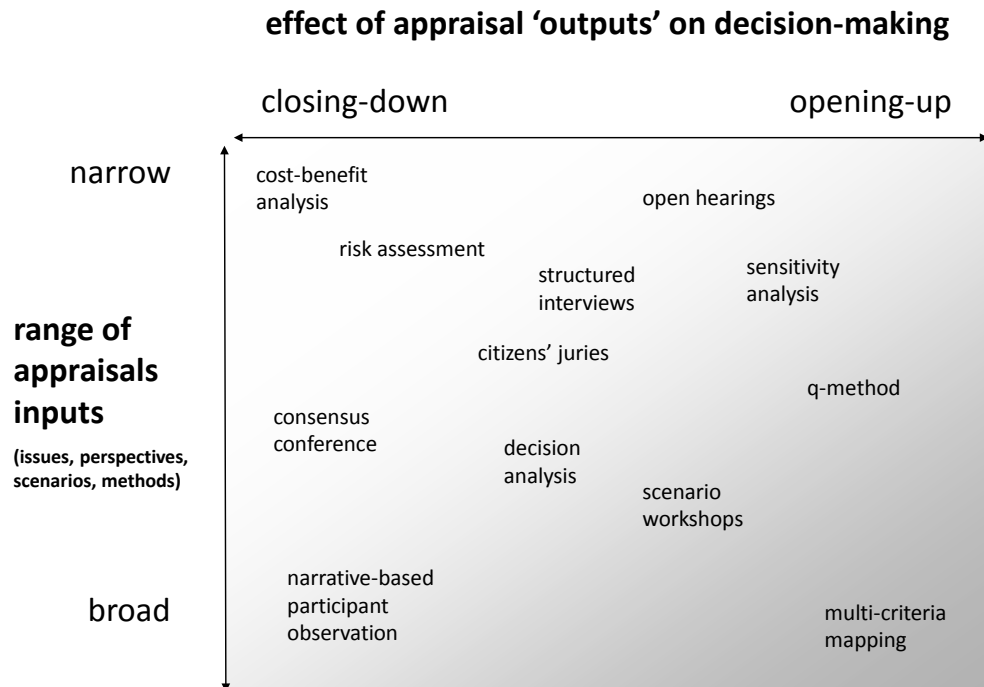


Figure 1. Characteristics of appraisal methods. Source: Stirling & *al.* (2007, p. 57).

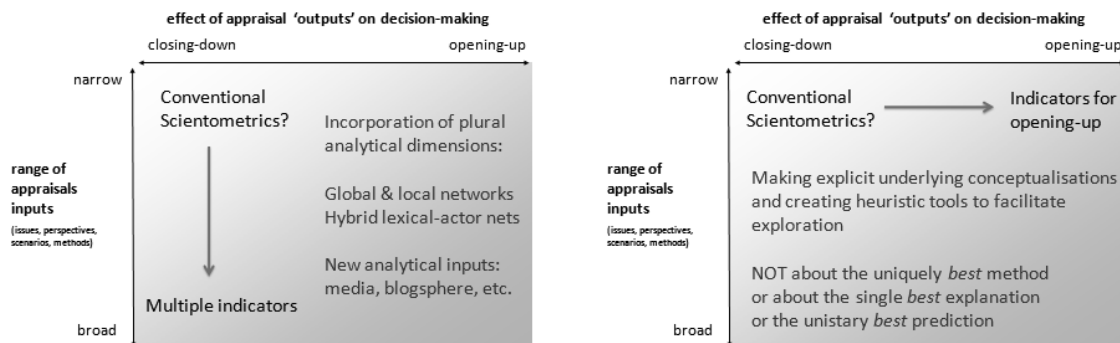


Figure 2. Difference between ‘broadening out’ the range of inputs used in indicators (left) and ‘opening up’ decision making.

In this way, we distinguish two dimensions in any process of policy appraisal, as illustrated in Figure 1. The first dimension, ‘breadth’ refers to the depth, extent and scope with which appraisal includes different types of knowledge that can describe the phenomena under scrutiny (Leach & *al.*, 2010, p. 104). The second dimension, ‘openness’, refers to the degree to which the outputs of appraisal provide plural and conditional interpretations of the phenomena—and thus allow contrasting policy options to be rigorously debated. Unlike analytical tools that ‘close down’ appraisal by establishing an absolute ranking of ‘best’ choices, ‘opening up’ tools allow decision-makers to contrast how under different assumptions the analysis may result in different rankings of options.

Conventional scientometric appraisal is rather narrow: both in the breadth of inputs and the openness of outputs (as illustrated in Figure 2). As with cost-benefit analysis, this narrowness results from measuring performance only in one or two dimensions (e.g. production and efficiency, or number of publications and citations) and focusing disproportionately on artificially singular selections of allegedly ‘best possible’ methodological choices with which to handle empirical data (like normalisation routines or aggregation procedures)—even where equally reasonable alternatives yield disparate output rankings.

Some of the analytical tools in S&T indicators can be relatively broad in terms of the range of inputs. For example, the Shanghai ranking of universities takes into account six different inputs, and the European Innovation Scoreboard includes a total of 25 indicators. However, both tools create a composite index that uses simple weightings to aggregate multiple dimensions into a single scalar. These are broad in inputs but narrow in outputs (as illustrated in the left side of Figure 2). Such scalar scores ‘close down’ debates on performance by univocally establishing which university is ‘best’ or which country is ‘most’ innovative. Such composite indicators have been shown to be potentially misleading as ‘the scope for manipulation of scoreboards by selection, weighing and aggregation is great’ (Grupp & Moggee, 2004, p. 1382).

An obvious way to handle plural input dimensions is to use multidimensional representations, such as ‘spider’ charts (Grupp & Schubert, 2010)—preferably after conceptually and mathematically grounded reduction of dimensions. But in scientometric (and even more so, in bibliometric) analysis, the range of inputs on a given property (productivity or citation impact) is often limited by the nature of data sources. In such cases, can quantitative studies capture and convey diverse outcomes under different analytical assumptions? Our answer is yes. Even when data sources are relatively narrow, there is still scope for opening up (on the right hand side of Figure 2). Even with narrow inputs, tools can be developed that help decision makers scrutinize how different conceptualisations and associated mathematical operationalisations may yield contrasting results (even of exactly the same data). By investigating how different assumptions lead to different methods and rankings, the analyst can provide ‘plural and conditional’ advice—and policy makers can be more reflective and explicit about the normative aspects of their choices.

Opening up measures of interdisciplinarity and performance

Here we will explore and illustrate the process of ‘opening up’, by reviewing a recent bibliometric comparison of performance and interdisciplinarity in six academic organisations (Rafols & *al.*, 2012). Both ‘performance’ and ‘interdisciplinarity’ are complex concepts that can only partially be captured by bibliometric indicators. Indicators in question were derived from only two data sources: generic journal attributes and the references contained in each publication.² Yet in spite of this narrowness of inputs, we show it is possible to conceive of different conceptualisations of interdisciplinarity and performance, and make multiple operationalisations of some of them.

² These data are treated using complementary contextual information such as the classification of journals into disciplinary subject category, and the overall citation patterns across journals in all the web of science.

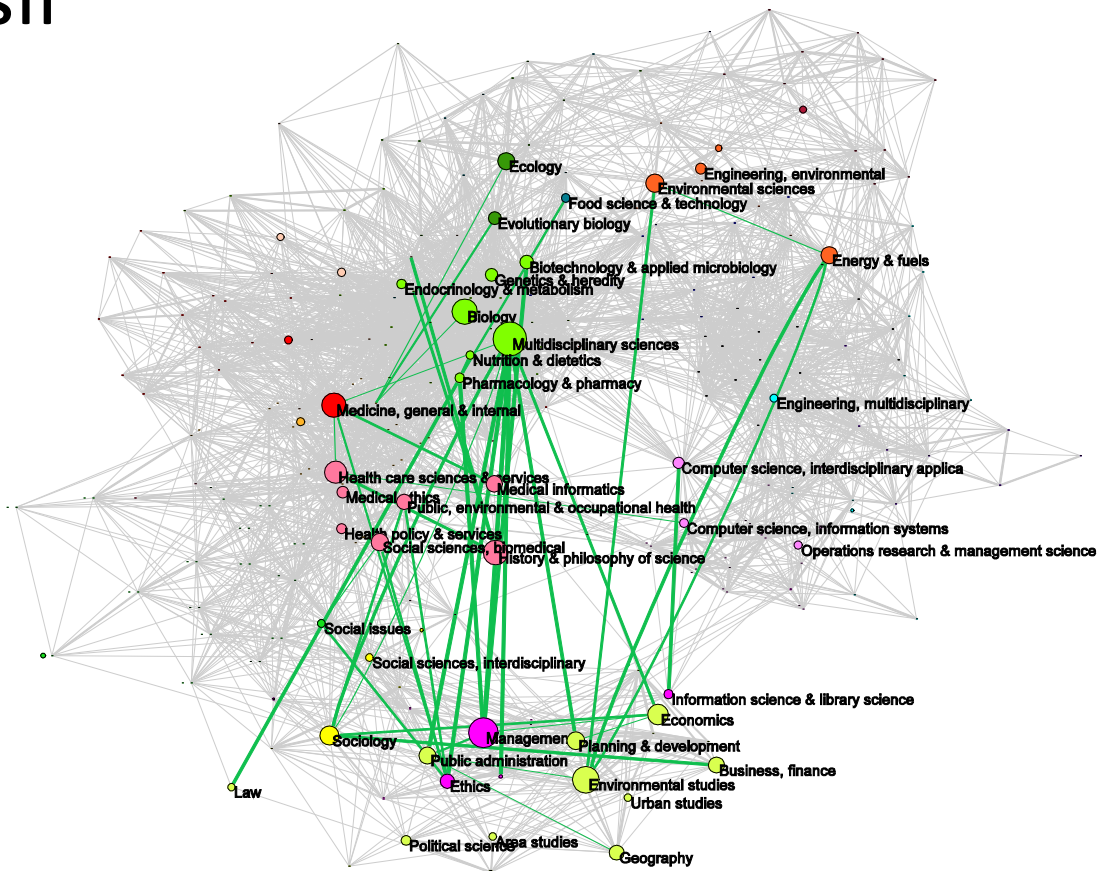


Figure 3. Overlay of number of references on Web of Subject Categories (source) by of the Institute for the Studies of Science Technology and Innovation (ISSTI, University of Edinburgh) on the global map of science. Each node represents a sub-discipline (Subject Category), and node size the number of references. Green links indicate 5-fold above expectation referencing (or citing) between Subject Categories by ISSTI. Grey lines indicate a certain level of similarity between Subject Categories. The degree of superposition in the grey background illustrates the degree of similarity between different areas of science for all 2009 Web of Science data. Diversity of references (as reflected in the spread of nodes over map) and referencing across disparate Subject Categories (the amount of cross-linking) are interpreted as signs of interdisciplinarity. Source: Rafols & *al.* (2012).

Two conceptualisations of interdisciplinarity are shown in Figure 3. On the one hand, we can understand interdisciplinarity as disciplinary *diversity*. Thus diversity measures of the distribution of publications (or references) of a unit across disparate subject categories (as illustrated by the spread of nodes over the map of science) captures the degree to which a unit covers different disciplinary approaches. On the other hand, we can conceptualise interdisciplinarity as the degree of *coherence* in their network of categories where they publish. This aims to capture the degree of cross-fertilisation between disciplines, which would be shown by the extent to which the references of publications criss-cross the map of science (as illustrated by the green lines, which show cases of cross-citation 5-fold above expectation). In the analysis it was found that the most interdisciplinary unit in terms of diversity was not the most coherent—hence there is good reason to differentiate these conceptualisations. Nevertheless, a comparison between three Innovation

Studies (IS) units and three Business and Management units (BM) units showed that under any of the various conceptualisations and operationalisations IS units were more interdisciplinary than BM units. Thus, at this larger scale, the contribution of the opening-up effort was to provide more robust evidence of the difference between IS and BM.

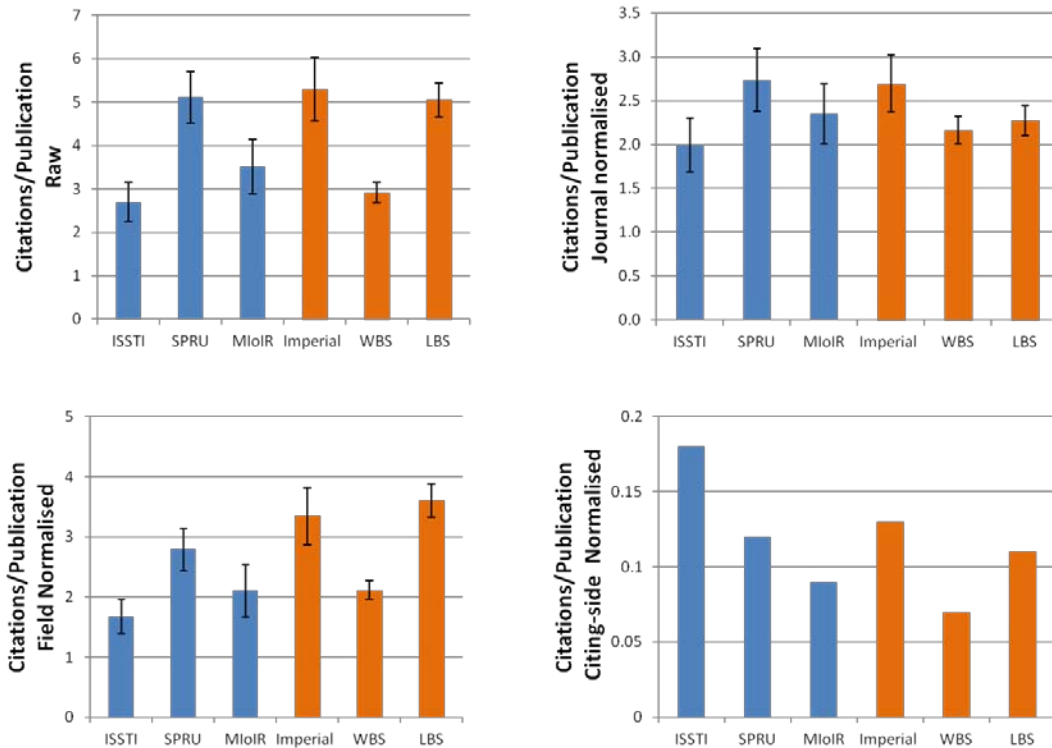


Figure 4. Example of opening-up by using different normalisations to a measure of the average number of citations per publication in a given organisation. Source: Rafols & *al.* (2012).

Conclusions and policy implications

This paper aims to illustrate that even analytical tools as narrow and apparently rigid as scientometric indicators leave room for policy usage that is more explicit about the dependence of analytic outputs on normative assumptions. We have argued that this ‘opening up’ is distinct (and complementary) to the ‘broadening out’ of the range of data inputs.

Indicators in S&T policy and management (as well as in other social spheres) have not only become pervasive as measurement tools, but constitute obvious ‘technologies for governance’ (Davis & *al.*, 2011). Indicators play a performative role, incentivising and thus ‘guiding’ scientists towards particular understandings of ‘good’ performance. ‘Statistical measures tend to replace political debate with technical expertise’ (Merry, 2011, p. S83). Under these circumstances, it becomes imperative to bring out into more open debate the crucial normative choices underlying indicators (Barre, 2010). In short, both broader and more plural forms of S&T indicators and visualisation tools are needed, in order to facilitate the ‘opening up’ of more rigorous and accountable policy appraisal.

References

- Barré, R. (2010) Towards Socially Robust ST Indicators: Indicators as Debatable Devices, Enabling Collective Learning. *Research Evaluation*, 19 (3), 227-231.
- Börner, K. (2010). *Atlas of Science. Vizualizing What We Know*. Cambridge, Mass. and London: MIT Press.
- Davis, K.E., Kingsbury B. & Merry, S.E. (2010). Indicators as a Technology of Global Governance. SSRN. <http://ssrn.com/paper=1583431>
- Grupp, H. & Mogege M.E. (2004). Indicators for National Science and Technology Policy: How Robust Are Composite Indicators? *Research Policy*, 33 (9), 1373-1384.
- Grupp, H. & Schubert, T. (2010). Review and New Evidence on Composite Innovation Indicators for Evaluating National Performance. *Research Policy*, 39 (1), 67-78.
- Leach, M., Scoones, I. & Stirling, A. (2010). *Dynamic Sustainabilities. Technology, Environment, Social Justice*. London and Washington: Earthscan.
- Lepori, B. 2006. Methodologies for the Analysis of Research Funding and Expenditure: From Input to Positioning Indicators. *Research Evaluation*, 15 (2), 133-143.
- Martin, B.R. & Irvine, J. (1983). Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy. *Research Policy*, 12 (2), 61-90.
- Martin, B.R., Tang, P., Morgan, M & al. (2010). Towards a Bibliometric Database for the Social Sciences and Humanities—A European Scoping Project. A report for DFG, ESRC, AHRC, NWO, ANR and ESF. Brighton, UK: SPRU.
- Merry, S.E. (2011). Measuring the World: Indicators, Human Rights, and Global Governance: With CA Comment by John M. Conley. *Current Anthropology*, 52 (S3), S83-S95.
- Narin, F. (1987). Bibliometric Techniques in Teh Evaluation of Research Programs. *Science and Public Policy*, 14 (2), 99-108.
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A Manifesto—Altmetrics.org. <http://altmetrics.org/manifesto/>
- Rafols, I., Porter, A.L. & Leydesdorff, L. (2010). Science Overlay Maps: A New Tool for Research Policy and Library Management. *Journal of the American Society for Information Science and Technology*, 61 (9), 871-1887.
- Rafols, I., Leydesdorff, L. O'Hare, A., Nightingale, P. & Stirling, A. (2012). How Journal Rankings Can Suppress Interdisciplinary Research. A Comparison of Innovation Studies and Business & Management. *Research Policy*. In press. Available at www.interdisciplinaryscience.net
- Rosvall, M. & Bergstrom, C.T. (2008). Maps of Randon Walks on Complex Networks Reveal Community Structure. *Proceedings of the National Academy of Sciences USA*, 105 (4), 1118-1123.
- Roessner, D. (2000). Quantitative and Qualitative Methods and Measures in the Evaluation of Research. *Research Evaluation*, 9 (2), 125-132.
- Stirling, A. (2003). Risk, Uncertainty and Precaution: some instrumental implications from the social sciences. In I. Scoones, M. Leach & F. Berkhout (Eds.). *Negotiating Change: Perspectives in Environmental Social Science* (pp. 33-76). London: Edward Elgar.
- Stirling, A. (2005). Opening Up or Closing Down: Analysis, Participation and Power in the Social Appraisal of Technology. In M. Leach, I. Scoones & B. Wynne (Eds.). *Science and Citizens: Globalization and the Challenge of Engagement* (pp. 218-231). London: Zed.
- Stirling, A. (2008). “Opening Up” and “Closing Down”: Power, Participation, and Pluralism in the Social Appraisal of Technology. *Science, Technology & Human Values*, 33 (2), 262-294.

- Stirling, A., Leach M., Mehta, L., Scoones, I., Smith, A., Stagl, S. & Thompson, J. (2007). Empowering Designs: Towards More Progressive Appraisal of Sustainability. STEPS Centre.
- van Leeuwen, Thed N. (2007). Modelling of Bibliometric Approaches and Importance of Output Verification in Research Performance Assessment. *Research Evaluation*, 16 (2), 93-105.
- van Raan, A.F.J. (2005). Fatal Attraction: Conceptual and Methodological Problems in the Ranking of Universities by Bibliometric Methods. *Scientometrics*, 62 (1), 133-143.
- Weingart, P. (2005). Impact of Bibliometrics Upon the Science System: Inadvertent Consequences? *Scientometrics*, 62 (1), 117-131.