Research Evaluation, 28(1), 2019, 7–22
doi: 10.1093/reseval/rvy030
Advance Access Publication Date: 14 November 2018
Special Issue

S&T indicators in the wild: Contextualization and participation for responsible metrics

Ismael Ràfols[1,2,3,*]

[1]Ingenio (CSIC-UPV), Universitat Politècnica de València, València, Spain, [2]Centre for Science and Technology Studies (CWTS), University of Leiden, Leiden, Netherlands, and [3]SPRU (Science Policy Research Unit), University of Sussex, Brighton, UK

*Corresponding author: Email: i.rafols@ingenio.upv.es

Abstract

The use of indicators in research policy and evaluation is widely perceived as problematic. Responding to demands for explicit normative framings in STI governance, I propose an agenda for transforming the place and role of indicators in policy. Given that expert advice should not separate knowledge formation from decision-making under conditions of uncertainty and lack of value consensus, I argue that current scientometrics is too focused on technical issues, too reductionist, and too isolated from the contexts and values of its use. Using Callon's analytical framework of 'secluded research' vs. 'research in the wild', I propose three moves for improving design and use of science, technology, and innovation (STI) indicators. First, to continue ongoing trends towards pluralizing the data sources, processing and visualization techniques, and expand the research communities involved in scientometrics. Second, to develop forms of quantitative evidence that can be contextualized with the participation of a more diverse set of stakeholders. Third, to open up the policy framings implicit in measurement, and use quantitative analyses to reveal more balanced perspectives of existing and alternative STI options. I conclude by arguing that these shifts are necessary to preserve epistemic diversity and pluralism in the face of ongoing managerial push for standardization via 'platforms' run by commercial oligopolies.

Key words: indicators, contextualization, responsible metrics, evaluation.

1. Introduction: A call to action

The use of science, technology, and innovation (STI) indicators[1] in research management and evaluation is widely perceived as being problematic. Some analysts even suggest that nowadays the use of indicators in evaluation can be more harmful than beneficial. As a result, recently various initiatives have asked for a responsible use of metrics, such as the *San Francisco Declaration of Research Assessment* (DORA 2013), the *Leiden Manifesto* (Hicks et al. 2015), or *The MetricTide* (Wilsdon et al. 2015).

In the face of the limitations, misuses, and abuses of conventional indicators, policy demand for new indicators has surged—asking, for example, for indicators of societal impact, of open science, of responsible research and innovation (Monsonís-Payá, García-Melón and Lozano 2017; MoRRI 2018), or reproducibility (Munafò et al. 2017). This thirst for new indicators is illustrated by the attention paid to the promises of altmetrics as indicators of societal impact, in spite of its poor conceptualization (Haustein, Bowman and Costas 2015) and lack of evidence regarding its relationship to impact (Robinson-Garcia et al. 2017).

Here, I will argue that trying to improve indicators' use only by adding more indicators is misguided. Having more indicators can be helpful if they represent valuable research activities not captured by conventional indicators. However, the problems with current use of quantitative evidence does not lie in the indicators themselves, but in the role that STI indicators play in STI governance—in how they are positioned in policy processes, and in the way they are used to inform STI strategy and research evaluation.

Rémi Barré, one of the leading proponents of indicators in Europe since the 1980s,[2] both as scholar and practitioner, presented in his keynote at the S&T Indicators conference 2017 a compelling argument of 'what went wrong with indicators' (Barré 2018, in this issue of *Research Evaluation*). His argument is that, although indicators were developed in the spirit of enlightenment as tools that would inform decision-making, they have become 'ignorance producing devices', in the sense that they are used as horse blinkers that reduce the issues taken into consideration in STI policy. The reasons for this narrowing are that 'design and use have been shaped and produced by the scientifico-industrial complex—the dominant forces', in such a way that 'mainstream S&T indicators have become the instruments of insertion of the science system into the neo-liberal economic paradigm and practices'.

Barré's despair about the evolution of STI indicators is shared. Other key pioneers in the adoption and use of scientometrics for science policy such as Ben Martin (Martin, Nightingale and Ràfols 2014; Martin 2016), Diana Hicks (Hicks et al. 2015), or Wolfgang Glänzel (Glänzel and Schoepflin 1994; Wouters et al. 2013) have also expressed concerns.[3] Barré's critique is consistent with many analyses of the construction and effects of indicators under managerialism, or what has often been called the 'neoliberal academy'—meaning the application of New Public Management (NPM) frameworks in higher education and R&D management (Weingart 2005; Gläser and Laudel 2007; Burrows 2012; Halffman and Radder 2015; Fochler and De Rijcke 2017).

This article takes as its point of departure Barré's analysis (2018) and his call for action: 'the collective responsibility of the S&T indicators scientific community is to call for an ending of the culturally produced ignorance drift and to pave the way for new designs and uses based on its founding "science in democracy" problematique'. Barré's call comes at a moment in which the research establishment and elite has become aware of the negative effects of current bibliometric evaluation (DORA was an initiative of 'top' scientists and editors) and is seeking alternatives to conventional indicators. Pressures to change current use of indicators also come from reform movements within the science system that see current evaluation as hindering the adoption of new practices sanctioned by visions of 'Open Science', 'Responsible Research and Innovation', or more productive interactions with society (as hoped by policy calls to deliver 'impact'). These movements are part of a broader agenda on democratizing STI, and making STI an engine of progressive social transformations (Stirling 2015; Schot and Steinmueller 2018).

Adopting this agenda for a pluralistic and democratic governance of STI, I will enquire into how quantitative analyses can help in making STI policy more open to debate and participation, to make it more sustainable and equitable in its outcomes. I propose a framework for re-positioning STI indicators in research policy in such a way that they help reconnect decision-making with practitioners and stakeholders within and around research.

The argument is that scientometrics has been framed in purely technical terms, paying scant attention to its context of use. Research policy takes place under conditions of high uncertainty and lack of value consensus, and in these circumstances knowledge formation and its use in decision-making are inevitably entangled (Pielke 2007; Stirling 2007). Following Callon, Lascoumes, and Barthe's (2001) framing of the research process, I propose that the development of STI indicators should take place not only in 'secluded spaces' such a scientometric labs but with the participation of stakeholders so as to take in consideration their contexts, both in terms of relevant social spaces and values. 'Indicators in the wild' would be the metaphor for this contextualized and participatory work of designing and using quantitative evidence for taking action in the social wilderness.

The article starts with a parable (Section 2), and it presents the incipient re-framing of STI policies in terms of societal goals (see Figure 1). Next, it introduces Pielke and Stirling's frameworks on scientific advice in policy in the face of uncertainty and lack of value consensus (Section 4). Then it introduces Callon, Lascoumes, and Barthe's (2001) notions of 'secluded research' and 'research in the wild' (Section 5). In Section 6, I claim that current scientometrics fits with the notion of 'secluded research'. In Section 7, I sketch a vision for 'indicators in the wild', i.e. the development and use of indicators through participation in the contexts of application, meeting the conditions of value pluralism and the demands for addressing societal goals. The closing section discusses the implications of the agenda proposed in the face of transformations towards a political economy of quantitative evaluation increasingly shaped by commercial information infrastructures.

2. The parable of Prussian scientific forestry

In the beginning of the book *Seeing like a state*, the anthropologist James Scott presents the invention of scientific forestry in Prussia as a parable of the forms of knowledge that state bureaucracies use to render the world legible and transform it—though not always with the expected results (Scott 1998).

Scott explains that forests were viewed by European early modern states as a source of revenue. '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' (Scott 1998). However, since forests were ecologically and socially diverse in 18th-century Europe and inefficient for timber production, the Prussian state embarked on a plan to measure and map forests only through the amount of timber produced with the aim of improving productivity.

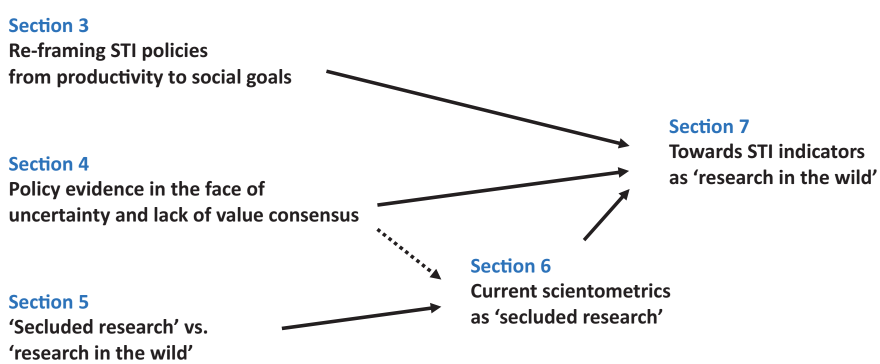

Figure 1. Scheme of the argument presented in the article.

The outcome was the invention of scientific forestry: a method of management that modelled the forest according to the vision of timber metrics used. Old, diverse, disordered, and unplanned trees were razed and replaced by forests of Norway spruce trees, all with the same age and similar height, planted in grid patterns to facilitate exploitation. Scientific forestry was initially very successful. In the first generation, it saw a massive increase in timber yield and revenue. It was thus adopted and became hegemonic across the modern world. There were losers from the monoculture: people who used forests for purposes that fell outside the vision of scientific forestry such as grazing, hunting, collection of herbal medicines, fruits or mushrooms, or use of other raw materials, such as twigs for bedding.

However, the simplification of a complex ecological system into a timber-producing machine eventually came back to haunt the enterprise. The second and later generations of forests managed with scientific forestry saw an alarming decrease of yields due to nutrient depletion, followed by disasters such as pests and falls by storms due to monoculture and uniformity. A new word, *Waldsterben* (forest death), described these unexpected outcomes of scientific forestry. To remediate *Waldsterben*, Prussian engineers invented 'forest hygiene', which aimed to restore the health of the forest by re-introducing some degree of diversity with ant colonies, spiders, and tree-nesting birds. 'Restoration forestry' tried to re-create what engineers had previously destroyed.

> The metaphorical value of this brief account (...) is that it illustrates the dangers of dismembering an exceptionally complex and poorly understood set of relations and processes in order to isolate a single element of instrumental value. (...) Everything that interfered with the efficient production of the key commodity was implacably eliminated. Everything that seemed unrelated to efficient production was ignored. (...) Utilitarian simplification in the forest was an effective way of maximizing wood production in the short and intermediate term. Ultimately, however, its emphasis on yield and paper profits, its relatively short time horizon, and, above all, the vast array of consequences it had resolutely bracketed came back to haunt it.
> Scott (1998)

Two themes of this story are worth highlighting. First, that scientific forestry started with measurements and mappings which turned out to become performative: 'these state fictions transformed the reality they presumed to observe [making it similar to the mappings], although never so thoroughly as to precisely fit the grid' (Scott 1998). Second, that the effect of a narrowing vision was a reduction of their diversity and contextualization, which harmed local stakeholders and eventually hurt the socioecological system.

3. Re-framing STI policies from productivity to social goals

The development of scientific forestry illustrates some of the beliefs of the enlightenment (or modernity) about science, expertise, and the state, which (still?) dominate policy action. One might naively describe public STI policy as rooted in three assumptions:

- First, the assumption that STI eventually lead to positive outcomes, which means the more STI knowledge is created, the better.
- Second, that the state bureaucracy is benevolent in supporting STI, and thus it aims to foster general well-being.
- Third, that scientists and experts responsible for developing and applying STI serve the public good rather than particular interests.

From these assumptions stem the belief, possibly still held by a majority in modern societies, that STIs are a force for the 'good' and deserve public support. Within this picture, we can think of scientometrics as a form of scientific expertise aimed at helping state bureaucracies to improve STI systems since the 1970s—which we should remember were (and in part, still are) run behind closed doors by the old boys' networks of the scientific elite (Mulkay 1976; Travis and Collins 1991). This view of STI for the 'public good' and of scientometrics for the 'good of STI' may have been a reasonable representation for some fields at some points in time, for example when indicators were used to challenge the scientific establishment (Martin and Irvine 1983), particularly in low-trust countries with a tradition of nepotism (Ràfols et al. 2016).

Yet, perhaps paradoxically, in recent decades we have seen both the diffusion of indicators as a means of the STI governance under NPM, and the erosion of the uncritical belief in the benefits of STI. I will argue the place of STI indicators in policy has to change to address these two developments.

3.1 The use of STI indicators for NPM

As explained by Barré (2018), although initially developed within 'welfare' ideals, the quick diffusion of indicators by state bureaucracies since the 1980s is associated with the adoption of NPM, both in discursive and procedural terms (Dahler-Larsen 2011; Burrows 2012). This meant embracing a discourse on 'performance' narrowly focused on efficiency and productivity, which, through its restrictive vision led to management procedures based on few indicators, may be leading to 'task reduction' and 'goal displacement' (De Rijcke et al. 2016) and may be reducing epistemic diversity (Schmidt et al. 2006).[4]

While this emphasis on efficiency resulted from the neo-liberal ideology of the end of 20th century, it is worth remembering that it is a belief shared by centralized or/and authoritarian states either in East (communist regimes were one of the historical cradles of scientometrics; see Wouters 1999: 84) or West (e.g. in Latin and Spanish bureaucracies; Vessuri, Guédon and Cetto 2014; and apartheid South Africa).[5]

In any case, professional scientometrics, though predicated as a science of science, has been dependent on state bureaucracies—irrespective of whether its particular institutional arrangements were through a public agency (as in France), university-based consultancies (e.g. in The Netherlands or Flanders), or private consultancies (in the USA and the UK). This dependency has favoured an alignment between state framings of STI and the analytical frameworks of the STI indicators' communities.

3.2. Questioning the interests and values shaping STI

Yet the 1970s and 1980s also saw the development of civic movements and academic communities that challenged the belief systems about science. Since then, the benefits and downsides of STI have been openly debated and contested. The practices and outcomes of STI have been shown to be shaped by social factors (Bijker, Hughes and Pinch 1987; Williams and Edge 1996); in some instances, the state and its experts have been shown to support STI policies that favoured particular political interests, specific industries, and privileged rather than disenfranchized social groups (Frickel et al. 2010;

Hess, 2016); and sometimes, STI has been shown to cause unexpected harmful events—such as thalidomide, car pollution, or nuclear meltdowns. In STI, as in other arenas, policies tend to reflect the interests in the political economy—and STI priorities are shaped and constrained by leading industries, military goals, and the interest of researchers or other concerned groups.

In summary, although STI has sometimes led to great benefits, it has also led to negative consequences, as perceived by some stakeholders—and these 'negative' outcomes emerged both out of the uncertainty of innovation, and from the shaping of STI by powerful actors pursuing particular interests. In fact, it has been argued that in 'late modernity', some of the biggest risks or challenges facing humanity such as climate change are the result of STI (Beck 1992).

The new awareness of the potential harm caused by innovation (and specifically innovation supported by the state and endorsed by expert advisers who are both influenced by hegemonic stakeholders) challenges the three assumptions outlined above: namely, that STI advice by experts is aimed at the public good, that the state is benevolent when it comes to STI, and that STI is a force for good. In short, there is a now consensus in STI studies that more science and more innovation are not necessarily for the better for many citizens (Schot and Steinmueller 2018). Let me stress this point because it questions the foundation of most policy uses of STI indicators—which revolve around linear views in terms of increasing STI production (more papers), scientific impact (more citations), societal impact (more patents), or more interactions between research and economic actors (more licences, more industrial projects).

The uncertain and contested nature of STI, i.e. that it may sometimes lead to questionable or 'negative' outcomes, has been documented by many Science and Technology Studies (STS) over the past four decades (Bijker, Hughes and Pinch 1987; Williams and Edge 1996; Stirling and Scoones 2009), and frameworks have been developed to think on STI policy appraisal with explicit normative commitments on sustainability and social justice (Stirling et al. 2007; Leach, Stirling and Scoones 2010). Some of these critical notions have entered STI policy since the 2000s, for example in the attention to public engagement in policies for nanotechnology, and particularly for health and environmental research—although much STI policy is still dominated by linear STI push rationales.

Schot and Steinmueller (2018) provide a useful conceptual analysis of three main framings of STI policy since the Second World War. According to Schot and Steinmueller (2018, p. 2) policy is now partly shifting to 'a questioning of how to use science and technology policy for meeting social needs' and is increasing addressing 'the issues of sustainable and inclusive societies at a more fundamental level than previous framings or their associated ideologies and practices'. Evidence of the breadth and depth of this shift away from the traditional narrow focus of economic growth is yet anecdotal. Examples of new STI policies towards needs and demands may be the EC framework on societal challenges and missions (Mazzucato 2017) and the recent prominence of Sustainable Development Goals in setting STI priorities (Giovannini et al. 2015).

3.3. Towards quantitative evidence for alternative framings of STI

This incipient (though still small) shift to alternative, explicitly normative framings of STI policy, has led to demands for indicators that incorporate considerations of value, for example indicators about societal impact (Robinson-Garcia et al. 2017), open science, eco-innovation (Kemp 2010), or about responsible research and innovation (Monsonís-Payá et al. 2017, MoRRI, 2018). The proposal of this article is that to escape from current narrow focus on productivity and efficiency under NPM, and to address explicitly normative framings, their position in policy and the practices on STI indicators must change.

The following sections outline first, the entanglement of normative and technical issues, and second, the need to engage with stakeholders in different contexts—in the wild, outdoors—in indicator making and use. In doing so, rather than thinking of scientometrics as 'science of science', we frame it as a postnormal science (Funtowicz and Ravetz 1993), a form of scholarship that needs to be situated and listening to social contexts, i.e. to social spaces and values.

4. Evidence for policy in the face of uncertainty and lack of value consensus

STI indicators are a particular form of expert knowledge that is used to inform STI policy. As quantitative forms of evidence, STI indicators' scholarship can be understood as 'scientometrics' (the metrology of science), or as a 'science of science' (Lane 2010), i.e. as a 'scientific' form of expert knowledge. We can therefore use the frameworks developed for the use of scientific expertise in policy (e.g. use of toxicology in food regulation, or fluid physics in water management), to the use of STI indicators in policy. In short, we can think of STI indicators as scientific expertise used in appraisal of STI policies or evaluations in the same way that we think of chemical measurement is used in policy appraisal for chemical regulation. This allows us to rely on an extensive literature in 'expert advice to policy' (Doubleday and Wilsdon 2013).

In the spirit of enlightenment, at some points in the past two centuries there was the hope that scientific advice would be able to guide policy decisions. While scientific knowledge has proved valuable and has improved many decisions, it has been argued that it can only provide socially robust advice when two conditions are met: when there is social consensus on the values that matter, and when uncertainty is perceived as low. For example, there can be consensus to ban the use of chemicals that are shown to produce diseases or environmental damage (e.g. asbestos or chlorofluorocarbons) when there are alternative chemicals to use—which means that relative values awarded to health, the environment, or economic growth do not enter in conflict (Hess 2016).

However, uncertainty and lack of value consensus (or its outcome, value ambiguity) are often present when scientific knowledge is mobilized to support and legitimize policy decisions (see Figure 2) (Stirling et al. 2007: 9; Pielke 2007). In these conditions, expert knowledge can be useful, but it does not suffice: decisions have to rely on judgement (which assumptions make more sense in the face of the uncertainty?), and prioritization of values (health or wealth?). In practice, both judgement and values are likely to associated with social and economic interests (whose health? whose wealth?). This is the case of many environmental and health arenas, from genetically modified crops to xenotransplantation to climate change, which have been the focus of public controversies, and which are visibly shaped by power relations.

Pielke (2007) argues that under uncertainty and lack of value consensus, it is not possible to disentangle the processes of knowledge formation (the construction of scientific expertise) from decision-making, because the conclusions reached depend on the assumptions embraced and the values prioritized (which tend to

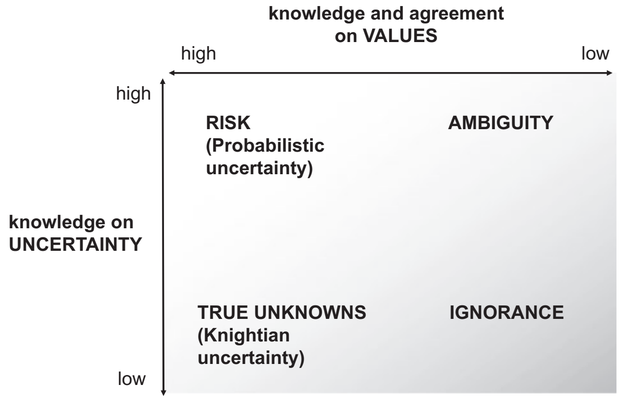

Figure 2. Types of incomplete knowledge. The vertical axis represents knowledge on uncertainty. The horizontal axis degree of value consensus. Adapted from Stirling and Scoones (2009).

reflect particular interests). In these conditions, he proposes the expert to be an 'honest broker', i.e. to provide plural and conditional advice to policy, being transparent about how different assumptions and values may lead to different decisions (Stirling et al. 2007). Funtowicz and Ravetz (1993) 'postnormal science' framework about the limits of scientific knowledge for policy advice builds on similar arguments.

In principle, one may think that research evaluation is not arena of public controversies. Instead, it may seem a technical activity with few discussions on uncertainty in either peer review or metrics. One might believe that there is a strong consensus on the values guiding evaluation: 'scientific quality' or 'excellence', and more recently 'societal impact'. Yet an examination of research evaluation processes reveals that the apparent consensus on 'excellence' or 'impact' is based in ambiguity, given the disparate interpretation or operationalization of these concepts under different conditions and contexts (Molas-Gallart and Castro-Martínez 2007). And the choice of peers or indicators responds to power relations which decisively shape the evaluation criteria and strongly influence the outcomes (Barré 2010). While it may be expedient for hegemonic STI actors to believe that research evaluation builds on experts' agreement, lack of controversy possibly stems from the inability of potentially dissenting actors to participate—given their lack of legitimacy, influence, or other forms of power—rather than consensus.

The widespread use of rankings based on indicators such as number of publications, patents, or citation counts suggests that STI indicators are being used as though they had little uncertainty and value ambiguity. However, when looked through Pielke's or Stirling's frameworks, we realize that indicators tend to hide many sources of uncertainty, and that disagreement in values may lead to different yet equally legitimate operationalizations or even divergent conceptualizations of the same properties.

4.1 Sources of uncertainty in scientometrics

Uncertainty in scientometrics stems from limitations inherent in knowledge about data and its processing. First, traces such as publications, keywords, and citations appear with stochastic patterns (i.e. with probability distributions). Therefore, at low levels of aggregation, particularly at the level of individual researchers, one cannot robustly differentiate meaningfully the 'performance' of different researchers (Wouters et al. 2013). Even at much higher levels of aggregation, the Leiden University ranking shows that dozens of universities are crammed at the tails of the citation distribution within the same range of 'interval of stability'.[6] Schneider and van Leeuwen (2014) explain that some policy reports assumed the differences observed to be meaningful to highlight the success of their policy instrument, although they are not statistically robust.

This form of uncertainty (technically 'risk') can be dealt with using statistical techniques that require conceptual assumptions based on contextual understandings—but which are open to interpretation (see Stirling 1998: 14–16; Waltman 2016a). Without tackling this statistical robustness, use of indicators may be problematic, as highlighted in the *Leiden Manifesto* (Hicks et al. 2015)—and in some instances disaggregate description rather than indicators may help users to grasp lack of robustness.

A second, deeper type of uncertainty stems from 'true unknowns' about the data, also called 'Knightian uncertainty' after the economist Alfred Knight (Stirling 1998: 15). In scientometrics, this type of uncertainty can stem, for example, from coverage of databases or the classifications used, which may generate unexpected biases given the very unequal distribution of some variables such as citations (Waltman 2016b) across fields, languages, and countries (Hicks 1999; Van Leeuwen et al. 2001; Chavarro 2017).

In short, understanding our knowledge about uncertainty is important because rankings and comparisons rely on assuming that observed differences in performance between two individuals or organizations are meaningful. Contexts (spaces but also values of stakeholders) are important to judge how to treat these various uncertainties to make choices that will have effects on indicators—which is why indicator-making and decision-making cannot be separated.

4.2 Sources of value disagreement in scientometrics

In many research assessments, the lack of value consensus will result from divergent views about the dimensions of research or the validity of the models to be taken into account (Molas-Gallart and Ràfols 2018). Conventional STI indicators have focused on measuring mainly production and impact without considering the social consequences. As discussed in Section 3, this is changing. As directionality in sociotechnical terms becomes an issue, STI policy needs indicators that are explicit about normative goals such as eco-innovation or preventive approaches to health. The intertwining of value choices and technical data issues may become more transparent during the development of these indicators to openly espouse some values to pursue some societal goals (e.g. equity in health).

In summary, in the case of STI policy there are high uncertainties and lack of agreement about what values should be prioritized. Pluralizing perspectives in indicators is a step forward for opening up appraisal so as to include diverse assumptions and values (Ràfols et al. 2012a, 2012b). In the next section, I will argue that conventional STI indicators have been developed secluded from the social, value-laden contexts. To become 'responsible metrics', they need to be built in their context of use with the participation of stakeholders.

5. Secluded research vs. research in the wild

5.1 The three translations of the research process

Michel Callon, Pierre Lascoumes, and Yannick Barthe (Callon, Lascoumes, and Barthe 2001, 2009) proposed a model of research

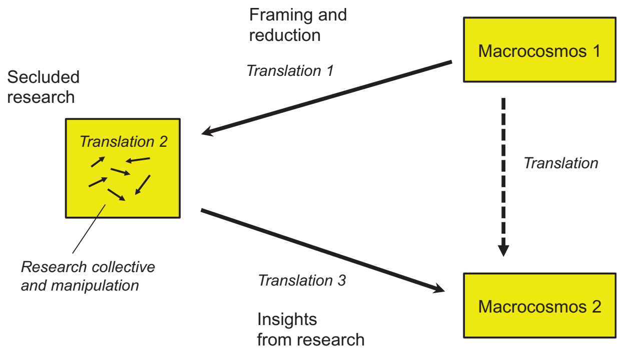

Figure 3. Callon's three stages of translation. Based on Callon, Lascoumes, and Barthe (2009: 69).

based on the notion of translation. As outlined in Figure 3, the process of translation of the unintelligible world to an understandable world involves three stages of translation (Callon, Lascoumes, and Barthe 2009: 48):

> The first is that of the reduction of the big world (the macrocosm) to the small world (the microcosm) of the laboratory. The second stage [secluded research] is that of the formation and setting to work of a restricted research group that, relying on a strong concentration of instruments and abilities, devises and explores simplified objects. The third stage is that of the always perilous return to the big world: Will the knowledge and machines produced in the confined space of the laboratory be able to survive and live in this world?

The idea is that research works through a process of reducing complexity (first translation) and isolating the lab from the macrocosmos (second translation) before applying the insights back in the macrocosmos (third translation). This isolation, what Callon calls 'secluded research', has been and remains important for creating stable experimental contexts that allow regularities to appear and new technologies to be built (forms of knowledge)—which later may (or may not) be found to hold when tried out back into the messiness of the macroscomos, which is in turn transformed by the knowledge produced.

The transformation by research of the macrocosmos from one configuration to a new configuration means that research is not just a way of looking into the world, but that it actively changes it—and therefore it has a political nature in the sense that it has the power to make choices (even if they may seem apparently technical) which will have effects in collective life. Going back to forestry, we can reflect on how focusing observation and measurement only on timber yield created the analytical conditions for 'cartesian' tree monocultures.

Now, given that three processes of translation are carried out by experts, research 'is a machine for changing the life of laypersons, but without really involving them in the conception and implementation of this change' (Callon, Lascoumes, and Barthe 2009: 70). In other words, important political choices are made through research that are not based in democratic mandates.

5.2 Research in the wild

To incorporate the knowledge and interests of laypersons, thus overcoming the isolation of expert knowledge, Callon proposes to carry out 'research in the wild' (in the original French, 'recherche à plein air', carrying somehow gentler undertones). Research in the wild is presented as conducted with the participation of stakeholders in non-purified and non-isolated environments, with awareness of the values and contexts of use, and thus infused with reflexivity on the consequences of the knowledge created in the macrocosmos. The participation of patients and their families in research on a particular neuromuscular disease is presented as an example of research in the wild.

It should stress that the goal is not to dismantle 'secluded research', but to get it to cooperate with 'research in the wild' so as to democratize science and innovation. This broader agenda of 'democratizing science' through participation of stakeholders in different stages of the research process is becoming increasingly adopted in some areas, such as health and the environment (e.g. James Lind Alliance on health priority setting), or in EC programmes seeking 'co-creation'.

Specifically, three moves are proposed for developing research in the wild (Callon, Lascoumes, and Barthe 2009: 124–5):

a. the adaptation of the results to specific contexts of use, when insights and technologies are shifted from the laboratories to the macrocosmos (in Translation 3);
b. the expansion and opening of the 'research collective' that carries out experiments and data manipulations (in Translation 2); and
c. the identification and framing of problems, the moment in which the macrocosmos' complexity is reduced (in Translation 1).

6. Conventional scientometrics as secluded research

Let us now return to STI indicators. Given that we can in principle think of scientometrics as scientific expertise embedded in policy practice, we can thus ask ourselves how this form of knowledge is positioned in the scheme outlined in Figure 3.

Now, what do we understand by scientometrics? We may wish to think of scientometrics as it is carried out in professional research evaluation by specialized university departments, such as CWTS or ECOOM, and consultancies such as Science Metrix or SciTech Strategies. This professional field supports policy makers in specific decision-making by providing empirical evidence based on rigorous use of data in the messy world of evaluation contexts. This is a type of scientometrics that has some (but limited) aspects of 'research in the wild': it is sometimes carried out in dialogue with some of the users (usually policy makers), partly adapted to contexts and sometimes involves discussions or interviews with selected stakeholders.

However, while this type of scientometrics was perhaps dominant until the 1990s, the diffusion of STI indicators, facilitated by easier access to databases and NPM demands, spread scientometrics beyond its small professional evaluation community. In parallel, the field as a whole shifted towards more standardized practices in information science aimed at policy and higher education management. In terms of Callon's framework in Figure 3, I believe that this type of scientometrics (dominant nowadays) fits with the notion of secluded research, i.e. a research field using a highly reduced description of the macroscosm (in this case, the STI systems) in terms of a limited number of data on publications, patents, spin-off counts, and related statistics.

The research collective in Translation 2 is circumscribed to researchers working with a few private databases and the manipulations restricted to a few variables in these databases. Most secluded scientometrics seldom relies on theory (Gläser and Laudel 2007: 113), and the theoretical framings are mainly about the communication practices that leave traces in the databases (citations, co-authorships, patent filing, or citing), not about other scientific practices. In particular, until recently no attention had been paid to the interactions between STI indicators, their institutionalization and use in evaluation, and their effects on research practices (Wouters 2014; De Rijcke et al. 2016; Kaltenbrunner and De Rijcke 2017)—an omission that has led to problematic consequences in Translation 3. The research collective seldom includes qualitative research to provide contrasting evidence. In the few cases that it was conducted, it sometimes led to converging results (Martin and Irvine 1983), and sometimes to contradictory evidence (Hicks 1987), suggesting an excessive reduction of complexity.

Problem framings, alternative perspectives or values are not discussed within this secluded space, in Translation 2. Instead underlying framings are given in Translation 1, generally by policy makers or companies funding the study. Not surprisingly, the analytical lenses of scientometrics (e.g. coverage databases and classification) tend to respond to the needs of those commissioning the study. This means that STI indicators use framings that tend to reflect and support discourses and research agendas of dominant stakeholders in STI policy. This dependence of scientometrics analysis on external funding has long been highlighted as one of its weaknesses (Glänzel and Schoepflin 1994). For example, the coverage of the most widely used database, Web of Science, reflects US research profiles with a strong coverage of biomedical research and poor coverage of non-English journals (Chavarro 2017). Another case in point are scientometric studies of nanotechnology in the 2000s, which often reified it as an 'emergent field', although fine-grained publication and patent clustering would have revealed it as consisting of disparate, and sometimes non-converging, epistemic areas (Schummer 2004).

Finally, in Translation 3, the return of insights back to the macrocosmos, the institutionalization of STI indicators in evaluation has meant that indicators are often used without care for its contextual validity (Gläser and Laudel 2007). Pioneers in the policy use of scientometrics such as Rip (1997) and Noyons (2001) proposed that validation should take place not only in quantitative terms (i.e. in internal mathematical logic within secluded research) but also through validation by field experts, and by users—which at the time were thought of as research managers, but we may now also think as relevant stakeholders. However, in many scientometric studies, the contextualization of the results is dispensed with one sentence explaining that the indicators were 'validated by experts' without providing further details.

Thus, the Translation 3 from the STI scholarship back to the macrocosmos raises two serious concerns regarding their social robustness. The first concern is whether STI indicators provided are perceived as valid or meaningful by stakeholders', given their specific contexts and values.

Without contextualization, indicators are inadequate, and the analysis can be harmful (Gläser and Laudel 2007; Waltman and Van Eck 2016). And given the political use of evaluation, there can often be dissenting understandings regarding appropriate contextualization.

The second concern is that, while many efforts have gone into discussing the descriptive validity of indicators [e.g. citation theories (Leydesdorff 1998; Wouters 2014) or patents (Meyer 2000), potential performative effects were not studied until recently (Weingart 2005; De Rijcke et al. 2016).[7] As put by Wouters (2017: 111):

> By purporting to measure reality, [research evaluations] effectively transform reality. (...) Peer review and performance metrics do not so much record pre-existing quality, reputation or excellence, but literally enact them. Outside of quality assessment practices, 'quality' does not actually exist. Given the strong interactions between quantitative indicators and qualitative judgment, then, 'quality' in all its versions (excellence, impact, top, etc.) is a hybrid result of the research evaluation exercises that have blossomed at all levels in academia since the 1980s.

In summary, scientometrics successfully developed as secluded research, as a form of expert knowledge that has achieved technical brilliance within confined databases and their quantitative manipulations (Translation 2). It has flourished thanks to its instrumental utility to support policy in reducing complexity and in narrowly framing science policy problems (Translation 1). However, in conducting Translation 3, which reshapes the macrocosmos, the use of STI indicators has shown to be problematic and lack social robustness. In my view, this lack of robustness is due to the separation of the technical expertise from the problem framings and the value-laden contexts of applications.

The secluded nature of scientometrics has led the field to make the same mistakes as scientific forestry. Reductionism and lack of contextualisation were useful to foster 'efficiency' under narrow political interests of productivist NPM agendas—but due to this narrowing of vision and neglect for context, it has not paid attention to its own problematic performative effects.

In the face of these problems, *The Metric Tide* report (Wilsdon et al. 2015) proposed that research evaluation should adopt general principles for 'responsible metrics': robustness, humility, transparency, diversity, and reflexivity. In the following section, I will argue that the notion of 'indicators in the wild' can help us in rethinking the role that STI indicators play in STI governance, identifying steps towards meeting the principles of responsible metrics.

7. Developing STI indicators

Let us revisit the arguments made so far (see Figure 1 for an illustration). First, I argued that science policy is being shifted towards societal goals and that this raises demands for new indicators or alternative type of quantitative evidence (Section 3). Second, I reviewed scholarship proposing that in many STI issues, where there is uncertainty and lack of value agreement, advice for policy should not separate technical knowledge from decision-making (Section 4). Third, I proposed that scientometrics fits with the notion of 'secluded research' (Sections 5 and 6), i.e. research that is carried out without much interaction with stakeholders and the contexts in which it will be used.

In this section, I argue that to adopt the notion of responsible metrics, and to respond to new policy demands of addressing societal goals, STI indicators should follow the practices of what Callon called 'research in the wild'. Let me emphasize that this is not necessarily about new or different indicators, but about indicators playing a different role in policy appraisal.

In terms of responsible metrics, the need for 'indicators in the wild' comes from the great diversity of communities in science and innovation, with their disparate practices, institutions, social relations, values, and objectives (Galison and Stump 1996; Gläser et al. 2010). A case in point is the evaluation of impact case studies in the Research Excellence Framework in the UK. The narratives of societal impact were so extremely diverse in their pathways and societal contributions that the use of indicators is very hard to support, even if some policy makers might prefer the justification provided by 'objective' metrics.

If STI policy is to shift towards the pursuit of societal goals, 'indicators in the wild' are needed to understand the interplay between scientific choices, values, and social configurations in specific contexts. Not all, but at least some quantitative analysts will have to produce 'rich descriptions' of their topics of study and to engage in deep dialogue with stakeholders. This does not mean asking for an impossible amount of detail or falling into relativism. It means that comparisons need to take into account relevant contexts, including attention to both specific spaces and their values, to allow for plural views. This also may mean to facilitate the exploration by stakeholders of the rationales for the methodological choices (which carry a normative consequences) used to build the quantitative descriptions. Plural and conditional advice will result from this reciprocal engagement between analysts (and their methods) and stakeholders (and their contexts). Qualitative research evaluation can offer many examples of adaptation to specific contexts (Crow and Bozeman 2005).

Let us now explore the construction of 'indicators in the wild' by opening up secluded research in successive stages. Following Stirling (2007), and in order of increasing difficulty, I propose that indicators in the wild can be (and are already being) developed through:

1. Expansion of the research collective of quantitative evidence (Translation 2)

2. Adaptation of indicators into contexts for plural and conditional use (Translation 3)

3. Opening up problem framing by mapping options (Translation 1)

7.1 Translation 2: Expansion of the research collective

Having agreed that one of the problems of scientometrics is its excessive reduction of STI systems to a small number of variables in a small set of databases, the expansion of the databases, visualization tools, and the type of researchers involved (the research collective) is an obvious first step towards enriching this field of scholarship.

7.1.1 *Data sources*

Data sources have been already increasing diversified. We have seen first a shift from a quasi-monopoly (Web of Science) to a quasi-duopoly (with Scopus since 2005), except for regional (e.g. Redalyc and Scielo in Ibero-America) and specialized databases (e.g. CAB Abstracts in agriculture). This situation is changing with a new series of more comprehensive open-access databases becoming available (including *crossref.org*, *1findr.com*, *lens.org*, and *dimensions.ai*, among others). These databases are useful for providing contrasting results. They would be even more useful if they were transparent about their technical, financial, and legal conditions: for instance, who and how decides the inclusion of journals, or how classifications are made. These understandings would facilitate the adaptation of analyses to marginalized countries and fields, and the scrutiny of biases. The use of classifications, ontologies, or thesauri may also lead to new perspectives on the same data (Leydesdorff, Rotolo and Ràfols 2012).

Another technical expansion is the shift towards new data sources, in part following big data approaches of using non-structured data. In the case of publications, the *Altmetrics* manifesto (Priem et al. 2010) suggested that social media data could be useful to measure 'impact'. While the meaning and usefulness of purely counting mentions in *Twitter* or *Facebook* remains highly contested (Haustein, Bowman and Costas 2015; Robinson-Garcia et al. 2017), social media data might be valuable for more elaborate analysis such as network approaches (Robinson-Garcia, van Leeuwen and Ràfols 2018). Similarly, in the case of innovation measures, social media data are being used to map innovation activity and networks. NESTA has been particularly active in this dimension (Mateos-García 2018).[8] Other approaches include carrying out semantic analyses from text scrapes from websites (e.g. on charities, Klavans and Boyack 2014; on technological companies, Gök, Waterworth and Shapira 2015), or combining STI with other sectoral statistics such as health (Evans, Shim and Ioannidis 2014; Ràfols and Yegros 2018) or agriculture (Ciarli and Ràfols 2018).

7.1.2 *Processing and visualization techniques*

New data algorithms and visualisations tools are allowing a fundamental transformation of how analysts present and how users relate to quantitative evidence. Visualization techniques can show complex quantitative data in intuitive ways. Interactive methods allow the users to change classifications or analytical perspectives, and to obtain new analyses dynamically (on-the-fly) (Van Eck and Waltman 2017; Fortunato et al. 2018). Zooming-in and pop-up of information functions allow to dig into the details of some specific parts of the quantitative evidence, and to go back-and-forth in the aggregation of categories (Latour et al. 2012; Venturini et al. 2017).

These techniques may in principle greatly facilitate users to re-open critically and re-do the analyses to fit with their own perspectives and interests. In practice, this deluge of choices may overwhelm users. Also, new visualizations are still highly stylized representations of complex dimensions—and thus may mislead users by creating unstable descriptions and artefacts. Therefore, the success of these new methods will depend on the cooperation in Translation 3, below, between computational experts guiding stakeholders through the new techniques into the details of the data algorithms, and stakeholders guiding experts in their contextual spaces (Waltman and Van Eck 2016).

7.1.3 *Research communities*

Another transformative change is the emergence of social studies on the use of STI indicators. Until recently, only a few scholars such as Godin (2002) and Wouters (1999) had carried out empirical qualitative studies on the history and sociology of scientometrics and its role on science policy. Yet since the 2010s, there is a growing community of scholars scrutinizing the uses and institutionalization of indicator-based evaluation, following the type of work that Porter (1996) and Desrosières (2002) had carried out on statistics in other areas of public policy (Rottenburg et al. 2015). See, for example, work by Pontille and Torny (2015), the special issue by Fochler and De Rijcke (2017), or by Thomas Heinze and collaborators (Jappe, Pithan and Heinze 2018). These contributions are important for understanding the position and possibilities of indicators work in STI systems. Other valuable insights can come from social scientists working on critical data studies or digital humanities (Marres and Gerlitz 2016; Venturini et al. 2017).

In summary, in the past decade there is a significant broadening of the research collective of STI indicators, certainly in terms of the data and tools and partly in terms of the research and professional communities involved—which may reflect policy demands.

7.2 Translation 3: Contextualization and participation for plural and conditional use

In relation to Translation 3, the use of indicators back into the world, it is worth remembering Barré's (2004 2010) emphasis on their 'political nature': the choice and interpretation of indicators are value-laden and political processes. The value of indicators lies in that they facilitate debate in specific contexts, 'enabling collective learning' in policy arenas. Responsible metrics is about providing scientometric analyses that can be scrutinized, adapted, and interpreted by stakeholders of specific contexts, so as to facilitate discussion about research strategy and prioritization. In my view, this involves two parallel processes: contextualization and participation. They go hand in hand because contextualization not only applies to geographic, social, and epistemic conditions of the evaluation but also to value preferences of the stakeholders involved. Participation is thus useful because stakeholders can contribute an understanding of their contexts (and whether they see them properly captured by data algorithms), and of their preferences, shaped by values and interests.

The pluralization of the research collective through enrichment of databases, more fine-grained topic classifications (Gläser, Glänzel and Scharnhorst 2017; Klavans and Boyack 2017), seductive visualizations (Chen and Leydesdorff 2014), and new statistical methods to reveal uncertainty (Schneider 2015) are developments that might favour a context-sensitive use of indicators. Similarly, the increasing importance of democratization and public engagement in research governance (e.g. in RRI discourse) might favour participation (Stirling 2015).

However, the political economy of indicators does not help in fostering contextualization. Gläser and Laudel (2007: 119) explained how the position of scientometrics between science–policy 'patrons', commercial marketing, and specific managerial demands militated against contextual adaptations ('modalities' in Gläser's terms) and favoured its spread as 'a universal tool that measures research quality'.

Progress towards a better adaptation of indicators to specific contexts and participation has been fairly limited. This is, in a way, surprising, given that many research constituencies see indicators as unstable, incorrect, or irrelevant (Molas-Gallart and Ràfols 2018) and particularly given the various types of discrimination they may foster (Vessuri, Guédon and Cetto 2014).[9]

Yet there are some interesting cases of indicator contextualization. For example, the University Medical Centre of Utrecht (UMCU) embarked on a wide consultation process so as to produce a portfolio of indicators representing UMCU's values regarding the balance between research inspired by curiosity and research driven by clinical needs (Benedictus and Miedema 2016: 454):

> This led to a suite of semi-qualitative indicators that include conventional outcome measurements, evaluations of leadership and citizenship across UMC Utrecht and other communities, as well as assessments of structure and process, such as how research questions are formed and results disseminated. We think that these shifts will reduce waste increase impact, and attract researchers geared for collaborations with each other and with society at large.

Another example is evaluation of development policy. This is a field that had already shown that for indicators of properties such as poverty, 'counting' is dependent on 'who is counting', i.e. on the contexts and values of the stakeholders (Chambers 1995). Not surprisingly, thus, evaluation of 'research-for-development' has been particularly sensitive to the use of locally meaningful indicators (Douthwaite et al. 2007; Lebel and McLean 2018), given that conventional STI indicators are perceived as 'sending negative signals to scientists willing to conduct research on contextualized agendas, particularly those negotiated with non-scientists' (Bianco, Gras and Sutz 2016: 399).

From these experiences, it follows that indicator design may also need to change to help users choose and interpret the forms of quantitative evidence relevant for their specific conditions, and which facilitate constructive discussions. Traditional ranking of a unidimensional indicator provides unitary and prescriptive advice, which prompts the closing down of decision-making (Ràfols et al. 2012b). Instead, for opening up the debate, it is useful to build forms of evidence that allow users:

1. dig into the underlying data and algorithms and to see what is behind the numbers (e.g. disaggregating categories)
2. explore robustness of descriptions (e.g. showing uncertainties),
3. show contrasting dimensions and options (e.g. via science maps),
4. help users reflect on the relation between options against their values and interest (e.g. by highlighting in a science map the options chosen by stakeholders with explicit interests)

Some simple methods can help in this work towards pluralization. For example, as illustrated in Figure 4, unpacking one composite indicator (here the European Innovation Scoreboard performance) into some of its constituent dimensions allows to see that that countries with similar performance may have very different strengths in innovation (Grupp and Schubert 2010).

More complex, novel interactive visualizations offer great possibilities to support the production of quantitative evidence in a plural and conditional manner that is useful for these participatory processes of contextualization.

7.3 Translation 1: Opening up problem framing: Mapping diverse options

According to Callon, Lascoumes, and Barthe (2009), the first stage in translation, about problem framing and complexity reduction (see Figure 3), is the most difficult step towards developing 'research in the wild'. Expectations of pluralization at this stage may be even lower for scientometrics, given its dependence on commercial or institutional databases, and on external funding from policy. However, the shift towards making explicit societal goals in STI policy, as discussed in Section 3, may open a window of opportunity.

Formulation of STI policy in terms that relate to improvements in well-being, (e.g. as stated in SDGs) poses many practical challenges—indicators for monitoring and evaluation being one of them. A focus on societal goals defines unambiguously its expected outcomes (e.g. SDGs goal 3.3. 'end the epidemic of AIDS'), but it allows to considers a variety of STI options (education, prevention, vaccines…) to pursue these goals. Nevertheless, when these goals are framed in STI policies, there is a tendency for incumbent voices to prioritize previously supported or mainstream STI approaches, at the expense of alternatives (Stirling 2018: 1):

> Trajectories in research and innovation are often conditioned not only by declared aims (like mitigating the above global challenges), but also by more private and proximate interests. As a result, a variety of social and technical responses may be available to address these global challenges that are more favourable than *status quo* directions for change, but which may remain unduly under-supported. (…) [A] number of alternative 'socio-technical' pathways for change may often be clearly more generally

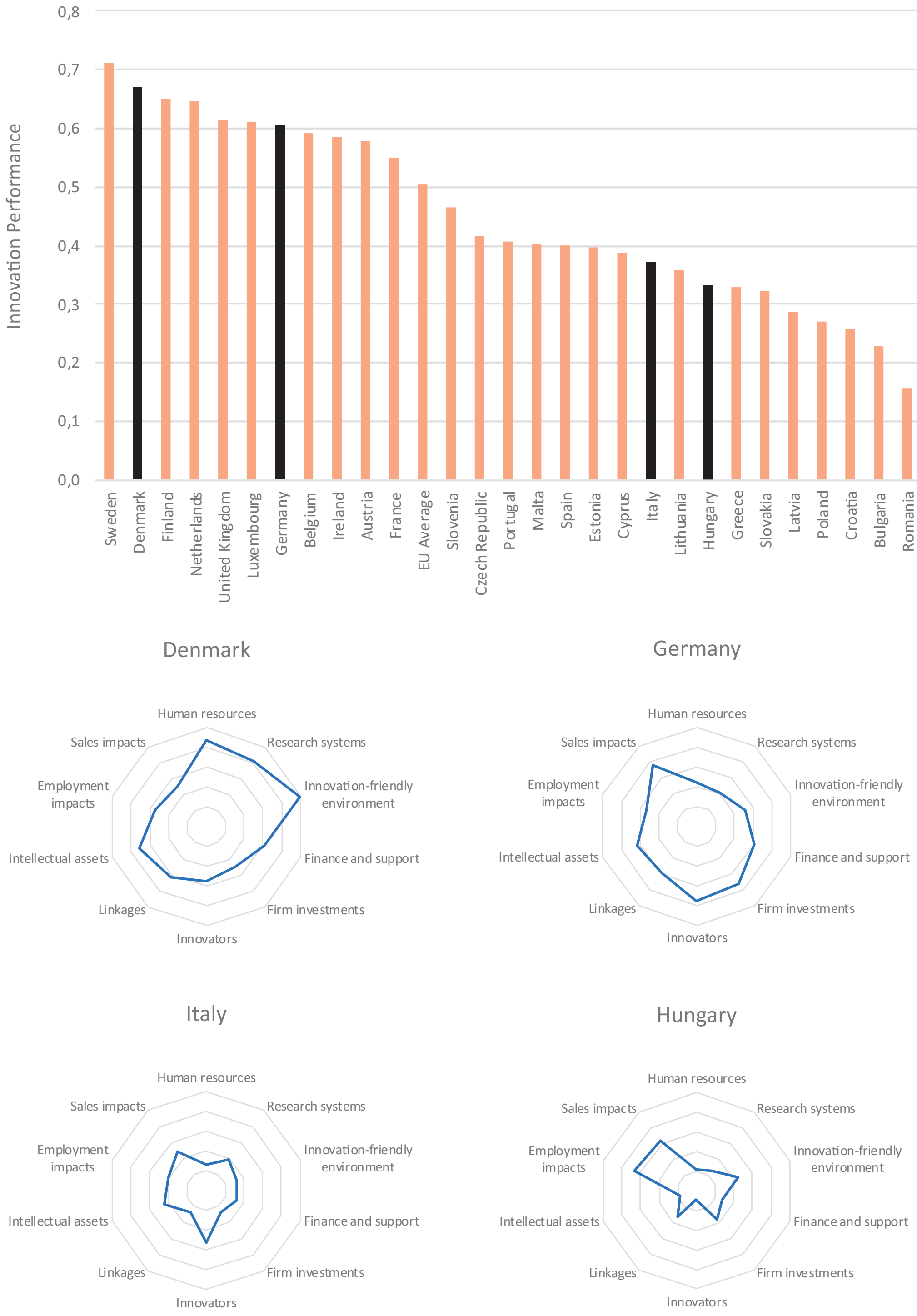

Figure 4. Unpacking indicators: comparison between a unidimensional ranking and radar charts. The European Innovation Scoreboard provides a composite indicator of innovation 'performance' by aggregating various analytical dimensions. A simple radar chart makes explicit the very different profiles by country across dimensions, allowing interpretation. For example, see the disparate strengths in innovation for Denmark and Germany, or for Italy and Hungary, in spite of their similar 'innovation performances'.
Data source: European Innovation Scoreboard (2017). Based in Grupp and Schubert (2010).

desirable than established trajectories. The central point is, that even where these alternative pathways are scientifically realistic, technically practicable, economically feasible and socially viable, incumbency can prevent them becoming historically realisable.

A key contribution of 'indicators in the wild' would be making visible a diversity of STI options, in particular the options that are not prioritised, i.e. relatively 'undone' (Frickel et al. 2010). For example, if the goal is to improve health in relation to obesity, quantitative studies can show that STI can contribute through various agendas: by improving the understanding of the social determinants, fostering changes in life style, proposing medical treatments, or through discoveries based on basic biology (Cassi et al. 2017). In this context, a valuable battery of indicators would consider the trends in these different directions and explore how organizations serving different missions focus on different issues (e.g. in avian flu research, Wallace and Ràfols 2018) or how different countries pursue disparate trajectories as illustrated in Figure 5 (e.g. in rice research, Ciarli and Ràfols 2018).

The aim is to make explicit efforts to counteract what Barré (2018) calls the production of ignorance through indicators that is achieved by focussing the attention of evaluation on narrow and sometimes irrelevant aspects of research (e.g. international visibility, when health is at stake). The framing in terms of diverse options will later allow, in Translation 3, indicator users to adapt the indicators and mapping to their perceived needs and values.

8. An agenda for 'responsible metrics' in a changing political economy

In this essay I have developed a critique of current STI indicators scholarship and proposed one agenda towards its pluralization. Given that policy advice, in the face of uncertainty and lack of value consensus, should not separate knowledge formation from decision-making (Pielke, 2007), I claim that current scientometrics is too focused on technical issues. An analysis of its institutionalization and practices shows that STI indicators are too reductionist and developed in isolation from the contexts of application. Their widespread use is often 'guided by the immediate interests of science policy and planning' (Glänzel and Schoepflin 1994: 376), even when they lack analytical validity or social robustness.

In the face of these practices, there have been calls for 'responsible metrics' (Wilsdon et al. 2015, 2017). I have argued here that Callon's agenda for democratizing expertise with 'research in the wild' can be a useful analytical perspective both for improving scholarly scientometrics and for facilitating a more plural and progressive use of quantitative evidence in STI management and policy. I propose three lines of action:

1. To continue ongoing efforts to expand the data sources, processing and visualization techniques used, and research communities (particularly those in qualitative modes) developing scientometrics.
2. To engage with a more diverse set of stakeholders and develop forms of quantitative evidence that facilitate their contextualization through participation, allowing for plural and conditional advice.
3. To open up analytical framings, so as to foster a more even attention to diverse STI options—particularly in STI policies aiming at societal goals.

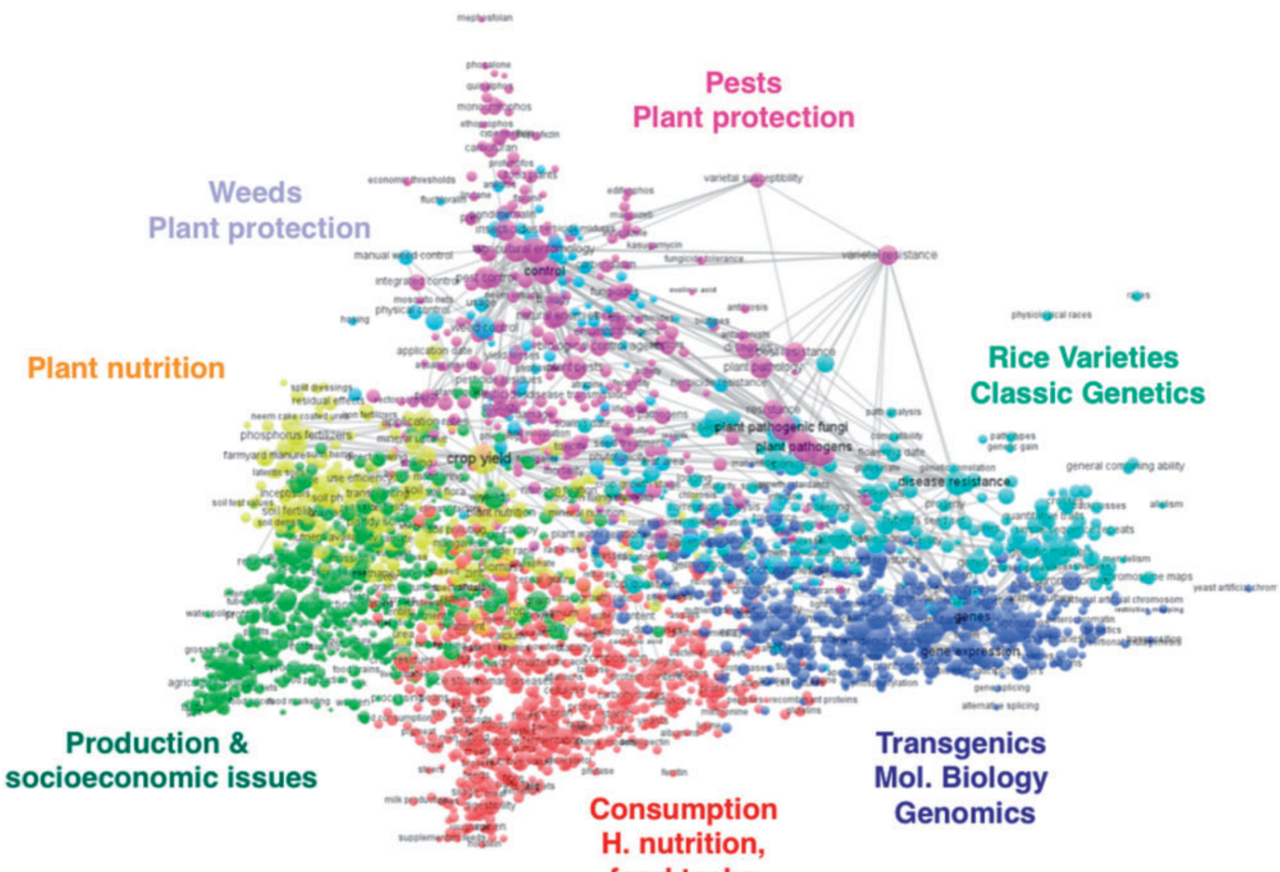

Figure 5. Map of research options in rice research. This is a form of quantitative evidence that highlights the diverse thematic options published on rice research. Based on co-occurrence of specialised descriptors in the database CAB Abstracts.
Source: Ciarli and Ràfols (2018).

Similar discussions on indicator use are occurring in other sectoral areas—in particular the debate on how to overcome simple indicators with undesirable performative effects and to design indicators that facilitate democratic participation (Dahler-Larsen 2011; Lehtonen 2012). Gross domestic product measurement is a prominent and illustrative example:

> Participatory research programmes, combined with active civil society organizations, promote a need for debates and more democratic decision-making processes. Expertise can no longer be limited to top-down application of scientific knowledge and indicators should reflect and contribute to this democratic move. How can and do indicators get involved into this democratic move? How is this enlarged participation of actors into the definition and shaping of indicators changing the modalities of their construction?
>
> *Mismeasuring our lives* by Stiglitz, Sen, and Fitoussi (2010)

For this transformation from technical indicators to contextually meaningful indicators to occur, scholarship in STI indicators needs to be more thoroughly engaged with other research fields in particular sociology of science, STS, economics of innovation, and innovation management. This does not mean rejecting or withdrawing from the traditional focus on statistical patterns, but rather to complement them with theories, methods, and empirical insights from neighbouring fields.

Since the communities of sociology of science and quantitative studies of science split in the early 1980s, there have been recurrent calls for dialogue (Leydesdorff 1989; Woolgar 1991; Leydesdorff and Wouters 1996), with little success (Gläser and Laudel 2007: 113). New communities around *Altmetrics* (Priem et al. 2010) and computational 'science of science' (Fortunato et al. 2018) maintain or even heighten the seclusion of 'data-driven research'.[10] The gap between scientometrics (ISSI conference), economics of innovation (Schumpeter conference), research policy (Eu-SPRI conference), or innovation management (DRUID conference) may not be as stark but is still notorious.

The lack of engagement of scientometrics with contextual and participatory approaches in spite of problematic uses (DORA 2013) may be seen as stemming from its weakness in the political economy (Glänzel and Schoepflin 1994).[11] First, scientometrics has not succeeded at institutionalizing its professional branch and thus lacks mechanisms to enforce codes of conduct or quality standards to practitioners (Jappe, Pithan and Heinze 2018). Second, its income depends on external funding or contract research, which comes from policy and management, rather than the civil society organisations or non-elite researchers which would benefit from wider stakeholder participation. Finally, most relevant databases are owned by private companies which constrain the uses. Under these conditions, with no professional authority nor strong public support, it seems difficult for STI indicators communities to take on an agenda for democratization of their knowledge.

A decade ago, as a response to poor use, misuses and abuses of STI indicadors, there was still hope that *public* institutions might support '*the professionalisation of bibliometric evaluations* by securing the independence of the profession from both its source of data and its customers, and by strengthening 'disinterested' academic work' (Gläser and Laudel 2007: 119). Instead, and against the early visions of open science (Bilder, Lin and Neylon 2015), we have witnessed an increasing concentration of scientific information infrastructure and consultancy work in the hands of large commercial oligopolies (Larivière, Haustein and Mongeon 2015). Through a web of subsidiaries, these companies have begun to provide monitoring and evaluation services to NPM demands through 'platforms' (Jappe, Pithan and Heinze 2018) that capture information traces at different points of the research pipeline (from project to 'societal impact') and aim at integrating them in standardized Current Research Information Systems.

This re-configuration of the political economy of 'S&T metrics' along the lines of 'platform capitalism' (Mirowski 2018) may lead to a transformation of research evaluation and monitoring governance, in which commercial and/or centralized information infrastructures might have further influence in producing even more 'uniformizing' perspectives on STI systems. This transformation of evaluation governance poses serious dangers, in the same way that unchallenged hegemony by Facebook or Google is now seen as posing a threat to pluralism and democracy after the Cambridge Analytica's scandal.

This shift in the political economy of indicators towards 'platform capitalism' gives powerful reasons to communities working on STI indicators to embrace this agenda for 'responsible metrics'– stepping beyond secluded databases and critically engaging with policy makers and stakeholders, conducting 'research in the wild'.

First, in instrumental terms: in the face of the potential standardization of research evaluation, based on commercially available metrics, small institutes and consultancies have and can further develop the know-how to deliver contextual indicators with stakeholder engagement. In contrast, large commercial services of research evaluation are less flexible. Second, in substantive terms: 'STI indicators in the wild' provide insights that are more sensitive to stakeholders needs, contexts, and values, and are that more socially robust thanks to its inclusion of plural perspectives—and thus they fulfil the new aspirations of STI policy to contribute to human well-being. And third, in normative terms: 'indicators in the wild' is 'the right thing to do', since it contributes to the democratizing of science.

We are perhaps at a turning point in research evaluation governance. On the one hand, a shift in the political economy of research with pressures from infrastructure developers and commercial companies towards further managerialism and standardization of evaluation into 'platforms'. On the other hand, the realization by important stakeholders in STI that indicators cannot be simply replaced or expanded, but that their role in research governance needs to move towards sensitivity to epistemic diversity, pluralism, and social robustness. If research genuinely aims to contribute to human well-being, it is time to listen to the call for democratizing science with a new approach to STI indicators.

Notes

1. I will use the term STI indicators and scientometrics as synonyms. STI indicators tend to have a broader meaning clearly including innovation activities. Scientometrics is often limited to metrics of science. The purpose of using them as synonyms is to convey that though science and innovation indicators are different, they cover overlapping spaces and face similar challenges.
2. Rémi Barré was the director of the French Observatoire des Sciences and Téchniques (OST), which is the main provider of scientometric analyses for the French government. He also held important positions in European science policy.
3. Ben Martin is a professor and former director of SPRU (Science Policy Research Unit) at the University of Sussex and editor of *Research Policy.* Diana Hicks is a professor and

former chair of the School of Public Policy at Georgia Tech, and an editor of *Resarch Evaluation*. Wolfang Glänzel is a professor at KU Leuven, the director of ECOOM (R&D Monitoring Centre), and an editor-in-chief of *Scientometrics*.

4. The discourse of efficiency was also apparently prominent in the first liberal globalization, the heyday of imperialism at the end of the 19th century, as famously put by Joseph Conrad's character Marlow in the beginning of *The heart of darkness:* 'What saves us is efficiency—the devotion to efficiency' (Conrad 1899, 2006).
5. I am grateful to Luke Georghiou for pointing to me the South African case.
6. For example, in the Leiden Ranking, if you take the proportion of 10% top cited, and put the mouse near the measure, you will be able to visualize the 'stability interval'. According to this stability, interval dozens of universities have citation performances that are statistically difficult to differentiate except for a few highly cited papers (see http://www.leidenranking.com/ranking/2018/list).
7. Gläser and Laudel (2016) and Gläser (2017) have argued that the specific causal effects of indicators are unclear. However, I think it is reasonable to asume that indicators have performative effects of some sort, even if causal relations are difficult to establish. Anecdotal evidence from case studies shows that indicators have indeed influenced evaluation practices and researchers' behaviours (de Rijcke et al. 2016; Kaltenbrunner and de Rijcke 2017).
8. https://www.nesta.org.uk/feature/innovation-methods/innovation-mapping/
9. It would be interesting to explore the parallels between the discriminations that social media algorithms produce (in terms of race, gender, country, etc.) as a result of in-built biases (Niklas and Peña Gangadharan 2018), and the workings of STI databases and indicators.
10. The smaller community around 'meta-research' is also very centred in large scale systematic analyses, though with a pragmatic focus on health and reproducibility (Ioannidis 2018).
11. I thank Jochen Gläser for suggesting the points made in this paragraph.

Acknowledgements

This article builds on work and insights from many colleagues at Ingenio, CWTS, and SPRU. Here the author draws in particular from And Stirling and Tommaso Ciarli at SPRU on democratization and diversity in research; from Jordi Molas-Gallart and Richard Woolley on evaluation governance; from Sarah De Rijcke, Ludo Waltman, and Paul Wouters on responsible metrics. The author also thank Rigas Arvanitis and Philippe Larédo for the invitation to give the keynote lecture in the STI conference 2017. Last but the least, the author is very grateful to Rémi Barré, Jochen Gläser, Chris Newfield, and the editors of *Research Evaluation* for fruitful comments.

References

Barré, R. (2004). 'S&T Indicators for Policy Making in a Changing Science-Society Relationship', in Moed H. F., Glänzel W., and Schmoch U. (eds) *Handbook of Quantitative Science and Technology Research*, pp. 115–31. Dordrecht: Springer.

Barré, R. (2010) 'Towards Socially Robust S&T Indicators: Indicators as Debatable Devices, Enabling Collective Learning', *Research Evaluation*, 19/3: 227–31.

Barré, R. (2018) Les indicateurs sont morts – vive les indicateurs! Towards a political economy of S&T indicators: a critical overview of the past 35 years. *Research Evaluation*, forthcoming.

Beck, U. (1992) 'From Industrial Society to the Risk Society: Questions of Survival, Social Structure and Ecological Enlightenment', *Theory, Culture and Society*, 9/1: 97–123.

Benedictus, R., and Miedema, F. (2016) 'Redefine Excellence: Fix Incentives to Fix Science', *Nature*, 538/7626: 453–5.

Bianco, M., Gras, N., and Sutz, J. (2016) 'Academic Evaluation: Universal Instrument? Tool for Development?', *Minerva*, 54/4: 399–421.

Bijker, W. E., Hughes, T. P., and Pinch, T. J. eds (1987) *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: MIT Press.

Bilder, G., Lin, J., and Neylon, C. (2015) Principles for Open Scholarly Infrastructure-v1. <http://dx.doi.org/10.6084/m9.figshare.1314859> accessed 19 July 2018.

Burrows, R. (2012) 'Living with the h-Index? Metric Assemblages in the Contemporary Academy', *The Sociological Review*, 60/2: 355–72.

Callon, M. L., Lascoumes, P., and Barthe, Y. (2001) *Agir Dans un Monde Uncertain. Essai Sur la Démocratie Technique*. Paris: Éditions Points.

Callon, M., Lascoumes, P., and Barthe, Y. (2009) *Acting in An Uncertain World: An Essay on Technical Democracy*. Cambridge, MA, and London, UK: MIT Press.

Cassi, L. et al. (2017) 'Improving Fitness: Mapping Research Priorities against Societal Needs on Obesity', *Journal of Informetrics*, 11/4: 1095–113.

Chambers, R. (1995) 'Poverty and Livelihoods: Whose Reality Counts?', *Environment and Urbanization*, 7/1: 173–204.

Chavarro, D. (2017) 'Universalism and particularism: explaining the emergence and growth of regional journal indexing systems', Doctoral thesis (PhD), University of Sussex. <http://sro.sussex.ac.uk/66409/> accessed 18 July 2018.

Chen, C., and Leydesdorff, L. (2014) 'Patterns of Connections and Movements in Dual-Map Overlays: A New Method of Publication Portfolio Analysis', *Journal of the Association for Information Science and Technology*, 65/2: 334–51.

Ciarli, T., and Ràfols, I. (2018) The Relation between Research Priorities and Societal Demands: The Case of Rice. *Research Policy*, forthcoming. <https://papers.ssrn.com/sol3/papers.cfm? abstract_id=3093285> accessed 18 July 2018.

Conrad, J. (2006) (original 1899) *The Heart of Darkness*. California: Project Guthenberg Ebook. <https://www.gutenberg.org/ebooks/526> accessed 24 August 2018.

Crow, M., and Bozeman, B. (2005) *Limited by Design: R & D Laboratories in the US National Innovation System*. New York, NY: Columbia University Press.

Dahler-Larsen, P. (2011). *The Evaluation Society*. Palo Alto: Stanford University Press.

De Rijcke, S. et al. (2016) 'Evaluation Practices and Effects of Indicator Use—a Literature Review', *Research Evaluation*, 25/2: 161–9.

Desrosières, A. (2002) *The Politics of Large Numbers: A History of Statistical Reasoning*. Cambridge, MA: Harvard University Press.

DORA (2013) *San Francisco Declaration on Research Assessment*. <http://www.ascb.org/dora/> accessed 24 August 2018.

Doubleday, R., and Wilsdon, J. (eds) (2013) *Future Directions for Scientific Advice in Whitehall*. London: Alliance for Useful Evidence & Cambridge Centre for Science and Policy.

Douthwaite, B. et al. (2007) 'Participatory impact pathways analysis: a practical application of program theory in research-for-development', *Canadian Journal of Program Evaluation*, 22/2: 127–59.

Evans, J. A., Shim, J. M., and Ioannidis, J. P. (2014) 'Attention to Local Health Burden and the Global Disparity of Health Research', *PLoS One*, 9/4: e90147.

Fochler, M., and De Rijcke, S. (2017) 'Implicated in the Indicator Game? An Experimental Debate', *Engaging Science, Technology, and Society*, 3: 21–40.

Fortunato, S. et al. (2018) ‘ Science of Science’, *Science*, 359/6379: eaao0185.

Frickel, S. et al. (2010) ‘Undone Science: Charting Social Movement and Civil Society Challenges to Research Agenda Setting’, *Science, Technology, and Human Values*, 35/4: 444–73.

Funtowicz, S. O., and Ravetz, J. R. (1993) ‘Science for the Post-Normal Age’, *Futures*, 25/7: 739–55.

Galison, P. L., and Stump, D. J. (1996) *The Disunity of Science: Boundaries, Contexts, and Power*. Palo Alto, CA: Stanford University Press.

Giovannini, E. et al. (2015) ‘The Role of Science, Technology and Innovation Policies to Foster the Implementation of the Sustainable Development Goals’, Report of the Expert Group “Follow-up to Rio+20, notably the SDGs.” European Commission.

Glänzel, W., and Schoepflin, U. (1994) ‘Little Scientometrics, Big Scientometrics... and beyond?’, *Scientometrics*, 30/2–3: 375–84.

Gläser, J., and Laudel, G. (2007) ‘The Social Construction of Bibliometric Evaluations’, in Whitley R. and Gläser J. (eds), *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, pp. 101–23. Dordrecht: Springer.

Gläser, J. et al. (2010) ‘The Limits of Universality: How Field-Specific Epistemic Conditions Affect Authority Relations and Their Consequences’, in Whitley R., Gläser J., and Engwal L. (eds) *Reconfiguring Knowledge Production. Changing Authority Relationships in the Sciences and Their Consequences for Intellectual Innovation*, pp. 291–324. Oxford: Oxford University Press.

Gläser, J., and Laudel, G. (2016) ‘Governing Science: How Science Policy Shapes Research Content’, *European Journal of Sociology/Archives Européennes de Sociologie*, 57/1: 117–68.

Gläser, J. (2017) ‘A Fight on Epistemological Quicksand: Comment on the Dispute between Van Den Besselaar et al. and Butler’, *Journal of Informetrics*, 11/3: 927–32.

Gläser, J., Glänzel, W., and Scharnhorst, A. (2017) ‘Same Data—Different Results? towards a Comparative Approach to the Identification of Thematic Structures in Science’, *Scientometrics*, 111/2: 981–98.

Godin, B. (2002) ‘Outline for a History of Science Measurement’, *Science, Technology, & Human Values*, 27/1: 3–27.

Gök, A., Waterworth, A., and Shapira, P. (2015) ‘Use of Web Mining in Studying Innovation’, *Scientometrics*, 102/1: 653–71.

Grupp, H., and Schubert, T. (2010) ‘Review and New Evidence on Composite Innovation Indicators for Evaluating National Performance’, *Research Policy*, 39/1: 67–78.

Halffman, W., and Radder, H. (2015) ‘The Academic Manifesto: From an Occupied to a Public University’, *Minerva*, 53/2: 165–87.

Haustein, S., Bowman, T. D., and Costas, R. (2015) ‘Interpreting “Altmetrics”: Viewing Acts on Social Media through the Lens of Citation and Social Theories’, in Sugimoto C. R. (ed.), *Theories of Informetrics and Scholarly Communication: A Festschrift in Honor of Blaise Cronin*, pp. 372–406. Germany: De Gruyter.

Hess, D. J. (2016) *Undone Science: Social Movements, Mobilized Publics, and Industrial Transitions*. Boston, MA: MIT Press.

Hicks, D. (1987) ‘Limitations of co-Citation Analysis as a Tool for Science Policy’, *Social Studies of Science*, 17/2: 295–316.

Hicks, D. (1999) ‘The Difficulty of Achieving Full Coverage of International Social Science Literature and the Bibliometric Consequences’, *Scientometrics*, 44/2: 193–215.

Hicks, D. et al. (2015) ‘The Leiden Manifesto for Research Metrics’, *Nature*, 520: 429–31.

Ioannidis, J. P. (2018) ‘Meta-Research: Why Research on Research Matters’, *PLoS Biology*, 16/3: e2005468.

Jappe, A., Pithan, D., and Heinze, T. (2018) ‘Does Bibliometric Research Confer Legitimacy to Research Assessment Practice? a Sociological Study of Reputational Control, 1972-2016’, *PLoS One*, 13/6: e0199031.

Kemp, R. (2010) ‘Eco-Innovation: Definition, Measurement and Open Research Issues’, *Economia Politica*, 27/3: 397–420.

Kaltenbrunner, W., and De Rijcke, S. (2017) ‘Quantifying ‘Output’for Evaluation: Administrative Knowledge Politics and Changing Epistemic Cultures in Dutch Law Faculties’, *Science and Public Policy*, 44/2: 284–93.

Klavans, R., and Boyack, K. W. (2014) ‘Mapping Altruism’, *Journal of Informetrics*, 8/2: 431–47.

Klavans, R., and Boyack, K. W. (2017) ‘Which Type of Citation Analysis Generates the Most Accurate Taxonomy of Scientific and Technical Knowledge?’, *Journal of the Association for Information Science and Technology*, 68/4: 984–98.

Larivière, V., Haustein, S., and Mongeon, P. (2015) ‘The Oligopoly of Academic Publishers in the Digital Era’, *PLoS One*, 10/6: e0127502.

Lane, J. (2010) ‘Let’s Make Science Metrics More Scientific’, *Nature*, 464/7288: 488.

Latour, B. et al. (2012) ‘The Whole Is Always Smaller than Its Parts’–a Digital Test of Gabriel Tardes’ Monads’, *The British Journal of Sociology*, 63/4: 590–615.

Leach, M., Stirling, A. C., and Scoones, I. (2010) *Dynamic Sustainabilities: Technology, Environment, Social Justice*. London: Routledge.

Lebel, J., and McLean, R. (2018) ‘A Better Measure of Research from the Global South’, *Nature*, 559/7712: 23–6.

Lehtonen, M. (2012) ‘Indicators as an Appraisal Technology: Framework for Analysing the Policy Influence of the UK Energy Sector Indicators’, in von Raggamby A. and Rubik F. (eds). *Sustainable Development, Evaluation and Policy-Making. Theory, Practise and Quality Assurance*. Cheltenham, UK: Edward Elgar.

Leydesdorff, L. (1989) ‘The Relations between Qualitative Theory and Scientometric Methods in Science and Technology Studies: Introduction to the Topical Issue’, *Scientometrics*, 15/5–6: 333–47.

Leydesdorff, L. A., and Wouters, P. F. (1996) ‘Quantitative Measuring and Qualitative Understanding: Is It Possible to Bridge the Divide in STS?’, *EASST Review*, 15/23: 20–4.

Leydesdorff, L. (1998) ‘Theories of Citation?’, *Scientometrics*, 43/1: 5–25.

Leydesdorff, L., Rotolo, D., and Ràfols, I. (2012) ‘Bibliometric Perspectives on Medical Innovation Using the Medical Subject Headings of PubMed’, *Journal of the American Society for Information Science and Technology*, 63/11: 2239–53.

Marres, N., and Gerlitz, C. (2016) ‘Interface Methods: Renegotiating Relations between Digital Social Research, STS and Sociology’, *The Sociological Review*, 64/1: 21–46.

Martin, B. R., and Irvine, J. (1983) ‘Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy’, *Research Policy*, 12/2: 61–90.

Martin, B., Nightingale, P., and Ràfols, I. (2014) Response to the call for evidence to the independent review of the role of metrics in research assessment. <https://pdfs.semanticscholar.org/44fb/3d1aaeff2f848efd3db3c2311 7c744cdb65e.pdf> accessed 18 July 2018.

Martin, B. R. (2016) ‘What’s Happening to Our Universities?’, *Prometheus*, 34/1: 7–24.

Mateos-García, J. (2018) How to be More Bee-Like (and Go From Richer Innovation Maps to Better Innovation Policies). <https://www.nesta.org.uk/blog/how-to-be-more-bee-like-and-go-from-richer-innovation-maps-to-better-innovation-policies/> accessed 26 August 2018.

Mazzucato, M. (2017) Mission-oriented Innovation Policy: Challenges and Opportunities. UCL Institute for Innovation and Public Purpose. <https://www.ucl.ac.uk/bartlett/public-purpose/publications/2018/jan/mission-oriented-innovation-policy-challenges-and-opportunities> accessed 18 July 2018.

Meyer, M. (2000) ‘Does Science Push Technology? Patents Citing Scientific Literature’, *Research Policy*, 29/3: 409–34.

Mirowski, P. (2018) ‘The Future (s) of Open Science’, *Social Studies of Science*, 48/2: 171–203.

Molas-Gallart, J., and Castro-Martínez, E. (2007) ‘Ambiguity and Conflict in the Development of ‘Third Mission’indicators’, *Research Evaluation*, 16/4: 321–30.

Molas-Gallart, J., and Ràfols, I. (2018) ‘Why Bibliometric Indicators Break down: Unstable Parameters, Incorrect Models and Irrelevant Properties’, *Biblioteconomia i Documentació*, 40. https://dx.doi.org/10.1344/BiD2018.40.23.

Monsonís-Payá, I., García-Melón, M., and Lozano, J. F. (2017) ‘Indicators for Responsible Research and Innovation: A Methodological Proposal for Context-Based Weighting’, *Sustainability*, 9/12: 2168.

MoRRI (2018) The *Evolution* of Responsible Research and Innovation in Europe: The MoRRI *Indicators Report*. Brussels: European Commission. <http://www.technopolis-group.com/wp-content/uploads/2018/02/D4.3_Revised_20022018_clean.pdf> accessed 18 July 2018.

Mulkay, M. (1976) 'The Mediating Role of the Scientific Elite', *Social Studies of Science*, 6/3–4: 445–70.

Munafò, M. R. et al. (2017) 'A Manifesto for Reproducible Science', *Nature Human Behaviour*, 1/1: 0021.

Niklas, J., and Peña Gangadharan, S. P. (2018) Data-driven discrimination: a new challenge for civil society. LSE Impact Blog. <http://blogs.lse.ac.uk/impactofsocialsciences/2018/07/10/data-driven-discrimination-a-new-challenge-for-civil-society/> accessed 19 July 2018.

Noyons, E. (2001) 'Bibliometric Mapping of Science in a Policy Context', *Scientometrics*, 50/1: 83–98.

Pielke, R. A. (2007) *The Honest Broker. Making Sense of Science in Policy and Politics*. Cambridge: Cambridge University Press.

Pontille, D., and Torny, D. (2015) 'From Manuscript Evaluation to Article Valuation: The Changing Technologies of Journal Peer Review', *Human Studies*, 38/1: 57–79.

Porter, T. M. (1996) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princenton, US: Princeton University Press.

Priem, J. et al. (2010) Altmetrics: A Manifesto. <http://altmetrics.org/manifesto/> accessed 26 August 2018.

Rip, A. (1997) 'Qualitative Conditions of Scientometrics: The New Challenges', *Scientometrics*, 38/1: 7–26.

Ràfols, I. et al. (2012a) 'How Journal Rankings Can Suppress Interdisciplinary Research: A Comparison between Innovation Studies and Business & Management', *Research Policy*, 41/7: 1262–82.

Ràfols, I. et al. (2012b) 'Towards Indicators for 'Opening Up' Science and Technology Policy', in Proceedings of 17th International Conference on Science and Technology Indicators, Vol. 2, pp. 663–74. <2012.sticonference.org/Proceedings/vol2/Rafols_Towards_675.pdf> accessed 1 May 2018.

Ràfols, I. et al. (2016) On the dominance of quantitative evaluation in 'peripheral' countries: Auditing research with technologies of distance. <https://papers.ssrn.com/sol3/papers.cfm? abstract_id=2818335> accessed 19 July 2018.

Ràfols, I., and Yegros, A. (2018) *Is Research Responding to Health Needs?* Social Observatory of La Caixa Foundation (refereed online publication), March 2018. https://observatoriosociallacaixa.org/en/-/responde-la-investigacion-a-las-necesidades-de-salud> accessed 18 July 2018.

Robinson-Garcia, N. et al. (2017) 'The Unbearable Emptiness of Tweeting—about Journal Articles', *PLoS One*, 12/8: e0183551.

Robinson-Garcia, N., van Leeuwen, T. N., and Ràfols, I. (2018) 'Using Almetrics for Contextualised Mapping of Societal Impact: From Hits to Networks', *Science and Public Policy*, 45/6: 815–26.

Rottenburg, R. et al. (2015) *The World of Indicators: The Making of Governmental Knowledge through Quantification*. Cambridge: Cambridge University Press.

Schmidt, M. et al. (2006) 'A Methodological Study for Measuring the Diversity of Science', in *Proceedings International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting, Nancy, France*. <http://eprints.rclis.org/7578/1/Schmidt2006methodoloc%23C73A1.pdf> accessed 18 July 2018.

Schneider, J. W., and van Leeuwen, T. N. (2014) 'Analysing Robustness and Uncertainty Levels of Bibliometric Performance Statistics Supporting Science Policy. A Case Study Evaluating Danish Postdoctoral Funding', *Research Evaluation*, 23/4: 285–97.

Schneider, J. W. (2015) 'Null Hypothesis Significance Tests. A Mix-up of Two Different Theories: The Basis for Widespread Confusion and Numerous Misinterpretations', *Scientometrics*, 102/1: 411–32.

Schot, J., and Steinmueller, E. (2018) 'Three Frames for Innovation Policy: R&D, Systems of Innovation and Transformative Change', *Research Policy*, 47/9: 1554–67.

Schummer, J. (2004) 'Multidisciplinarity, Interdisciplinarity, and Patterns of Research Collaboration in Nanoscience and Nanotechnology', *Scientometrics*, 59/3: 425–65.

Scott, J. C. (1998) *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven and London: Yale University Press.

Stiglitz, J. E., Sen, A., and Fitoussi, J. P. (2010) *Mismeasuring Our Lives: Why GDP Doesn't Add up*. New York: The New Press.

Stirling, A. (1998) 'On the Economics and Analysis of Diversity', SPRU Electronic Working Paper. <http://www.sussex.ac.uk/Units/spru/publications/imprint/sewps/sewp28/sewp28.pdf> accessed 27 August 2018.

Stirling, A. et al. (2007) 'Empowering Designs: Towards More Progressive Social Appraisal of Sustainability', STEPS Working Paper 3. <https://opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/2473/Empowering Designs.pdf? sequence=1> accessed 18 July 2018.

Stirling, A. C., and Scoones, I. (2009) 'From Risk Assessment to Knowledge Mapping: Science, Precaution, and Participation in Disease Ecology', *Ecology and Society*, 14/2.

Stirling, A. (2015) 'Towards Innovation Democracy? Participation, Responsibility and Precaution in the Politics of Science and Technology', STEPS Working Paper, Vol. 78. <http://sro.sussex.ac.uk/39718/1/Innovation-Democracy.pdf> accessed 18 July 2018.

Stirling, A. (2018) How deep is incumbency? Part 1: A 'configuring fields' approach to the distribution and orientation of power in socio-technical change, unpublished.

Travis, G. D. L., and Collins, H. M. (1991) 'New Light on Old Boys: Cognitive and Institutional Particularism in the Peer Review System', *Science, Technology, and Human Values*, 16/3: 322–41.

Van Eck, N. J., and Waltman, L. (2017) 'Citation-Based Clustering of Publications Using CitNetExplorer and VOSviewer', *Scientometrics*, 111/2: 1053–70.

van Leeuwen, T. N. et al. (2001) 'Language Biases in the Coverage of the Science Citation Index and Its Consequences for International Comparisons of National Research Performance', *Scientometrics*, 51/1: 335–46.

Venturini, T. et al. (2017) 'An Unexpected Journey: A Few Lessons from Sciences Po Médialab's Experience', *Big Data and Society*, 4/2.

Vessuri, H., Guédon, J. C., and Cetto, A. M. (2014) 'Excellence or Quality? Impact of the Current Competition Regime on Science and Scientific Publishing in Latin America and Its Implications for Development'', *Current Sociology*, 62/5: 647–65.

Wallace, M. L., and Ràfols, I. (2018) 'Institutional Shaping of Research Priorities: A Case Study on Avian Influenza', *Research Policy*. 47/10: 1975–89.

Waltman, L. (2016a) 'Conceptual Difficulties in the Use of Statistical Inference in Citation Analysis', *Journal of Informetrics*, 10: 1249–52.

Waltman, L. (2016b) 'A Review of the Literature on Citation Impact Indicators', *Journal of Informetrics*, 10/2: 365–91.

Waltman, L., and Van Eck, N. J. (2016) 'The Need for Contextualized Scientometric Analysis: An Opinion Paper', in I. Ràfols et al. (eds) Proceedings of the 21 ST International Conference on Science and Technology Indicator, pp. 541–9. València: Universitat Politècnica de València. <http://ocs.editorial.upv.es/index.php/STI2016/STI2016/paper/viewFile/4543/2327> accessed 28 August 2018.

Weingart, P. (2005) 'Impact of Bibliometrics upon the Science System: Inadvertent Consequences?', *Scientometrics*, 62/1: 117–31.

Williams, R., and Edge, D. (1996) 'The Social Shaping of Technology', *Research Policy*, 25/6: 865–99.

Wilsdon, J. et al. (2015) *Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. London: HEFCE.

Wilsdon, J. et al. (2017) 'Next-Generation Metrics: Reponsible Metrics and Evaluation for Open Science', Report of the European Commission Expert Group on Altmetrics, Brussels. <https://openaccess.leidenuniv.nl/bitstream/handle/1887/58254/report.pdf? sequence=1> accessed 18 July 2018.

Woolgar, S. (1991) 'Beyond the Citation Debate: Towards a Sociology of Measurement Technologies and Their Use in Science Policy', *Science and Public Policy*, 18/5: 319–26.

Wouters, P. F. (1999) 'The Citation Culture', Doctoral dissertation, Universiteit van Amsterdam. <http://garfield.library.upenn.edu/wouters/wouters.pdf> accessed 18 July 2018.

Wouters, P. et al. (2013) 'The Dilemmas of Performance Indicators of Individual Researchers–an Urgent Debate in Bibliometrics', *ISSI Newsletter*, 9/3: 48–53.

Wouters, P. (2014) 'The Citation: From Culture to Infrastructure', in Cronin B. and Sugimoto C. (eds) *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, pp. 47–66. Cambridge, MA: MIT Press.

Wouters, P. (2017) 'Bridging the Evaluation Gap', *Engaging Science, Technology, and Society*, 3: 108–18.