Towards more inclusive S&T indicators: a review on efforts to improve STI measurements in 'peripheral' spaces

Ismael Rafols[1,2,3], Jordi Molas-Gallart[1], Richard Woolley and Diego Chavarro

[1] Ingenio (CSIC-UPV), Universitat Politècnica de València, València
[2] SPRU (Science Policy Research Unit), University of Sussex, Brighton
[2] OST (Observatoire des Sciences et des Techniques), HCERES, Paris

Submission to OECD Blue Sky Conference, 14th of March 2016

Short abstract

S&T systems all over the world are routinely monitored and assessed with indicators that were created to measure the natural sciences. These S&T indicators are often inappropriate in other contexts: conventional S&T databases may not include data from low and middle income countries, less prevailing disciplines, and research addressing the problems of socially excluded groups. We review effort being made to create data and indicators that better reflect research activities and contributions in these "peripheral" spaces.

Introduction

In many countries there is the perception that science could be better at helping address societal problems, such as climate change or obesity. Yet, paradoxically, some of the research activities that may help solve societal issues are not well reflected in the S&T indicators typically used by policy-makers. This is a serious problem because the way research is monitored influences what is valued in assessments and planning, and, consequently, the research agendas that are prioritized (de Rijcke, et al. 2015).

Why is there this gap or misalignment between conventional S&T indicators and some relevant research activities? Conventional S&T indicators work reasonably well to capture scholarly contribution and some forms of innovation in areas that have constituted the centre of science, i.e. the natural sciences in rich countries. For example, in the US one can monitor contributions to molecular biology via citations, or commercial innovations in biotechnology by patents.

However, conventional S&T indicators are very problematic in "peripheral" spaces. These "peripheries" can be thought in geographical, cognitive or social dimensions. In geographical terms, developing countries have long been described as "the" periphery. Southern and Eastern European regions can also be conceived as EU peripheries, and less developed regions of a country as dependent on the cities where the financial and political power is located. The relative invisibility of certain disciplines or topics can be interpreted as a signature of cognitive periphery. This is the case of the social sciences and the humanities compared to medicine, or of epidemiology within health sciences. Research for marginalised

social groups may also be seen as less “central” than research aligned with the interests of dominants institutions.

S&T indicators aim to capture relevant properties (e.g. innovative activity) by means of data or measures (e.g. patents). To do so, we need to have a model (a “theory”) of why the measure represents the desired property (Keith Smith, 2005). For example, the use of patents counts (measure) as an indicator of innovative activity (property) is based on the assumption (theoretical model) that a patent is a trace of an innovation with potential socioeconomic value.

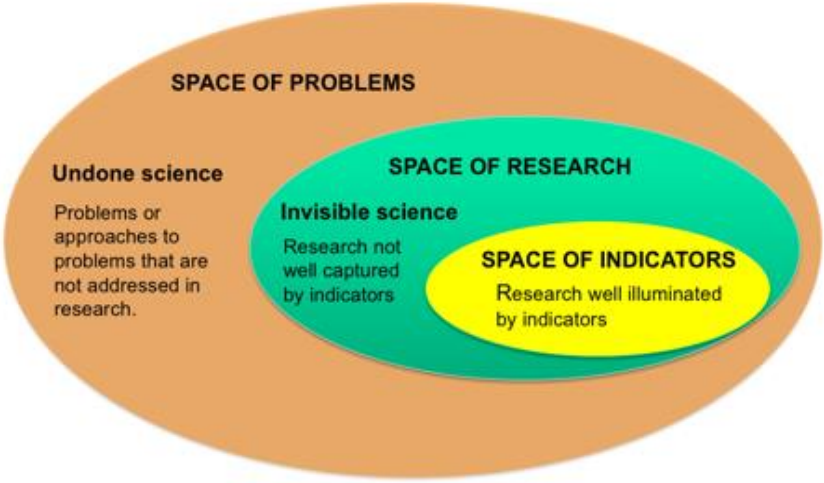

Figure 1. Schematic representation of the limited coverage of indicators.

The problem is that in developing countries, in “softer” disciplines, or in many social innovations, the assumptions behind conventionally S&T indicators do not hold. In other words, in “peripheral” spaces the models often break down and the indicators lose their meaning: patents say little of innovation in Southern Europe or new cultivation practices; citation counts cannot be used to assess the value of an article on Valencian history.

When indicators are used beyond their range of validity, they provide a distorted perspective of the STI system. When this distortion occurs, the use of indicators may cause more harm than good. Decision-makers may be influenced by the research that is illuminated by existing indicators and forget about the science that is not visible. This is a major issue, for example, in bibliometric analyses. The amount of publications from developing countries that are not covered in the conventional databases (Web of Science and Scopus) is very high. In the case of rice research, for example, the share of Indian and Chinese publications in WoS and Scopus is half the share of the observed in the more comprehensive CAB Abstract database (specialised in agriculture). For the US, the situation is the reverse. Developing countries are thus heavily under-represented, and developed countries are over-represented in the conventional databases.

The uneven representation of databases is likely to induce serious biases in the standard bibliometric indicators, which are used for research assessment. Critical voices have been raised on the potential effects that these biases may have in shifting research contents away

from locally relevant research. Last year, the Leiden Manifesto advised to develop metrics using local data so as to "protect excellence in locally relevant research" (Hicks et al, 2015). However, global science reports by organisations such as the UNESCO[1] or the Royal Society[2] continue to use Web of Science and Scopus' data.

Let us see some examples of analytical dimensions where indicators are problematic.

Basic vs. applied vs. research: Applied studies tend to cite fundamental studies more than the reverse. Even within a given scientific field as defined by conventional classifications such as Web of Science Categories, applied research tends to be significantly less cited than fundamental research (van Eck et al., 2013).

Interdisciplinary research: Interdisciplinary research can be thought of as peripheral to the extent that it is published in areas outside the disciplinary cores. It turns out that interdisciplinary research tends to be published in journals with lower rating in journal rankings and, within a field, with journals with a lower Journal Impact Factor (Rafols et al., 2012). As a results interdisciplinary research tends to be in a disadvantage when using this type of journal-based indicators (with citation indicators, the effect may vary as it depends on relative citation rates between fields that are being cross-fertilised).

Language: Language has long been known to be a major problem for performance measures, given that non-English articles tend to be much less cited. Van Leeuwen et al. (2001) showed that the inclusion or not of non-English publications in the analysis of citation impact has a major influence in the outcomes of indicators. Van Raan et al. (2011) showed that this also had major effects in university rankings. Vasconcelos et al. (2008) showed that language proficiency is highly correlated with citation impact and h-index of researchers. This means that for the purposes of comparison, non-English publication should be excluded in most analysis.

Gender: In many fields of science, women tend to publish less or in less visible journals and accrue less citations than men. For example, Brooks et al. (2014) reported that women on average receive lower scores according to journal ratings lists. However, various studies have consistently found that women tend to have higher degree of interdisciplinary research (e.g. Leahey, 2007). Hence, the effect of gender on performance depends on the indicators choice: if publications and citations are taken as a measure of the value of a contribution, the indicators will tend to disadvantage female researchers; if indicators of interdisciplinarity were used, they might be advantaged.

Efforts to improve STI measurement

How can indicators be developed in such a way that they can capture S&T activities in a broader set of spaces? What policy processes are needed for these new indicators to become adopted in policy? The difficulty of answering these questions lies in the multiplicity of purposes and contexts in which S&T indicators need to be deployed. As suggested above,

[1] http://en.unesco.org/unesco_science_report
[2] https://royalsociety.org/topics-policy/projects/knowledge-networks-nations/

for indicators to be valid, a theoretical model is needed linking the research mission to be captured with the specific data to be measured. Given the variety of contexts, a responsible approach to metrics has to embrace diversity, as proposed by Hefce's report *The metric tide* (Wilsdon et al., 2015)[3].

This contribution will review efforts in data collection or indicator design that reduce bias and improve the measurement. While these examples are drawn from science indicators, we believe they illustrate problems and potential solutions in measurement of technology and innovation as well. We review four approaches: (i) broadening database coverage; (ii) normalization of indicators to account for field or sector differences; (iii) design of new indicators to capture scientific processes; (iv) gathering new datasets such as social media (altmetrics).

Database coverage: The database most widely used bibliometric analysis, Web of Science (WoS), was developed under the assumption that the science that matters most is that published in top journals. However, for understanding research capabilities in developing countries or regions, it is important to capture as well scientific publications aimed at regional audiences in local languages. This is even more relevant in a context of fast growth of piublications in developing countries. To address these demands, Elsevier's strategy has been to make Scopus a more comprehensive database with some 20,000 journals (Leydesdorff et al., 2014), while WoS strategy has been to keep a journal core of some 10,000 journals in the Journal Citation Reports, while introducing complementary databases, such as a Korean or Russian Citation Index. In parallel, specialised databases in certain regions (e.g. Scielo and Redalyc[4] in Iberoamerica) or domain (e.g. CABI in global health and agriculture) have being flourishing with the aim of providing a more comprehensive coverage in certain issue. Case studies comparing CAB Abstracts in agriculture with WoS and Scopus for rice research suggest that WoS and Scopus cover are still major differences in country and topic coverage (Ciarli et al., 2015). The advantages and weaknesses of the different approaches will be discussed.

Normalization of indicators by field: Although there is a consensus among analysts about the need to normalise performance indicators such as citations per paper by field, the appropriate unit of analysis remains *ad hoc,* sometimes carried at with 10-20 disciplines over all science (e.g. in HighlyCited), sometimes at some 200-300 fields (e.g. in WoS Categories) and in some important cases with more than 1,000 subdiscipilnary categories (e.g. in the Leiden University Ranking). New methods (e.g. citing-side normalisation) were proposed to improve measures but have not been adopted. We will briefly review the consequences of using different normalisation methods in relation to "peripheral" spaces.

Indicators to capture scientific processes: Conventional S&T analysis have focused on indicators of output (productivity and scientific impact). However, a shift towards in interest on S&T towards societal contributions and human development places more emphasis on research processes that help create 'healthy' S&T system in a longer term perspective. This

[3] https://responsiblemetrics.org/
[4] http://www.scielo.org, http://www.redalyc.org

new emphasis has led to measuring research processes such collaboration or interdisciplinarity, e.g. in Caroline Wagner's work. However, these measures may not be easily interpreted as simple monotonic indicators (i.e. the more, the better), but need to rely on network-based interpretations where centre-periphery relations are prominent.

New data sources: Traces of scientific communication in webpages and social media (webmetrics and altmetrics) may allow to capture research contribution so far hidden, in particular, in societal dimensions. The unstructured nature of the data, however, is a major handicap in terms of offering a reliable, standardised tool for monitoring or evaluation. Disciplinary and topic differences in coverage may have a major effect in the relative visibility of research in social media. We will review methods develop until now in order to use webmetrics and social media data in such a way that they become meaningful information to inform policy.

Conclusion and policy relevance

Conventional STI indicators, e.g. those based on bibliometric data, have been significantly problematic in some "peripheral" spaces such as social science and humanities and developing countries. In this article we describe efforts to improve their coverage, design and use.

As a result of the expanding uses of S&T in geographical and social spaces, our review suggests that rather than a universal set of indicators useful for all contexts, a specific set of indicators and databases will be appropriate under specific conditions. While this diversity of information may enrich analytics, it is raises major challenges for the purpose of comparability. A smart use of indicators will require a more explicit statement of relationship between the set of indicators used and their contexts and policy goals where they are applied.