

# Bibliometric Evaluation Standards Debate: Introductory Presentation and Panel Discussion

Presentation:

**Daniel Sirtes**                      sirtes@forschungsinfo.de / iFQ Institute for Research Information and Quality Assurance, Berlin (Germany)

**Ludo Waltman**                    waltmanlr@cwts.leidenuniv.nl / Centre for Science and Technology Studies, Leiden University (The Netherlands)

Additional Panel Participants:

**Éric Archambault**                Eric.archambault@science-metrix.com / Science-Metrix Inc., Observatoire des Sciences et des Technologies (OST), Montréal, Québec (Canada)

**Wolfgang Glänzel**                Wolfgang.Glanzel@kuleuven.be / Centre for R&D Monitoring and Dept. MSI, KU Leuven (Belgium)

**Stefan Hornbostel**                hornbostel@forschungsinfo.de / iFQ Institute for Research Information and Quality Assurance, Berlin (Germany)

**Paul Wouters**                      p.f.wouters@cwts.leidenuniv.nl / Centre for Science and Technology Studies, Leiden University (The Netherlands)

## Introduction

The role of bibliometrics in research evaluation is becoming increasingly important, and bibliometric data and statistics are becoming more and more widely available. At the same time, a general consensus on how bibliometrics should be applied in research evaluation contexts and which types of bibliometric indicators should be used has not yet emerged. There is no set of explicitly formulated and commonly accepted ‘good practices’ for the use of bibliometrics in research evaluation, and instead of trying to reach consensus on a standard set of bibliometric indicators that work well for most evaluation purposes, many bibliometricians seem to prefer spending their time on inventing yet another new indicator.

Perhaps the current state of bibliometric research, in which a lot of attention is paid to technical innovations and in which there is less interest in reaching a broad consensus on what does and does not work well, is just a normal and healthy situation for a research field in a stage of quick development and expansion. At the same time, however, there is a serious risk that for many end users of bibliometric analyses and evaluations these controversial discussions are rather unsettling and alienating. University presidents and deans, their support staff, university librarians, research managers at funding agencies, etc. usually do not have the time or the right background in order to keep up with all the developments going on in the field of bibliometrics. In many cases, these people simply expect bibliometricians to tell them how bibliometrics can best be applied to their particular situation, which types of indicators should be used, and so on. All the more, these people expect that bibliometricians tell them a more or less consistent story. Why would one trust bibliometrics as a research evaluation tool if each bibliometrician advises to use

a different set of indicators, possibly leading to quite different outcomes in a research evaluation exercise?

From a practical research evaluation point of view, there appears to be a clear need for an increased level of standardization in bibliometrics. However, standardization could take place in a number of different ways, and some forms of standardization are perhaps more important or more relevant than others. For this reason, we believe that it is useful to make a distinction between three types of bibliometric standardization:

- Standardization of data sources
- Standardization of indicators
- Standardization of good practices

Below, we discuss each of these three types of standardization in more detail.

## **Standardization of data sources**

A bibliometric analysis requires a bibliographic data source. For analyses that are limited in scope, discipline-specific data sources may be used (e.g., PsycINFO), but for large-scale analyses one usually needs a multidisciplinary data source. The most popular multidisciplinary data sources are Thomson Reuters' Web of Science, Elsevier's Scopus, and Google Scholar. However, only the first two are documented and stable enough in their methodologies, so that reproducible results are warranted.

Bibliographic data sources differ from each other in the literature they cover, and therefore analyses based on different data sources normally do not yield the same results. However, even if two analyses are done based on the same nominal data source, they do not necessarily produce identical outcomes. Web of Science, for instance, consists of a number of databases (Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Conference Proceedings Citation Index, Book Citation Index), and depending on which of these databases one includes in an analysis, different results will be obtained. Furthermore, the databases change (even concerning older entries) on a weekly basis. Thus, exactly the same citation scores between studies will be hardly attainable.

In addition, professional bibliometric centers often do not work with the original 'raw' data from a data source such as Web of Science or Scopus, but instead they attempt to enhance the data quality. For instance, they may perform their own citation matching (i.e., the linking of reference strings to the publications being referenced) and they may work with cleaned address data (i.e., address data in which the names of organizations, cities, countries, etc. have been made consistent as much as possible). Obviously, the way in which citation matching, address cleaning, and other related issues are handled affects the outcomes of a bibliometric analysis.

Exclusion and inclusion of parts of the database for different analyses are very common as well. Publication types like serials and proceedings books are sometimes included in publication counts but excluded in citation scores. Typically, not all document types are included in studies. Mostly, only articles, reviews, and letters, the so called 'citable document types', are part of the studies. However, even on these matters there is no consensus.

Another issue is the way in which scientific fields or disciplines are defined. Bibliographic data sources typically offer a classification system that assigns publications to fields. Different data sources provide different classification systems, but even when working with the same data source it is possible to use different classification systems. Some bibliometric centers have for instance developed their own classification systems. It is clear that the use of different classification systems leads to different results in a bibliometric analysis.

Standardization of bibliographic data sources would require agreement to be reached on how to deal with each of the above issues. It is unlikely (and for certain purposes disadvantageous) that all centers will agree to use always the same data source with all the same demarcation criteria. However, guidelines and transparency would constitute important progress..

## **Standardization of indicators**

Especially since the introduction of the *h*-index in 2005, bibliometricians have paid a lot of attention to the development of new indicators, resulting in a large literature in which lots of proposals are reported for new approaches to counting publications and citations. Many indicators have been introduced for evaluating individual researchers, for assessing research institutions and research groups, and for measuring journal impact. A large variety of issues have been studied, ranging from the problem of how to deal with co-authored publications to the problem of weighting citations differently depending on their origin (e.g., PageRank) to the problem of normalizing indicators to correct for field-specific publication and citation practices.

A first step toward standardization may be to reach agreement on a minimum set of core principles that indicators should follow. These principles could for instance relate to the mathematical and statistical properties of indicators (e.g., averages vs. percentiles), the way in which different document types are dealt with and citation windows are chosen, the approach taken to correct for field differences (e.g., cited-side vs. citing-side normalization), and the way in which statistical uncertainty is handled. Reaching consensus on at least the core principles of the design of indicators may be a significant step forward.

## **Standardization of good practices**

In addition to the more technical aspects of bibliometric research evaluation discussed in the previous two sections, standardization may also take place in terms of the development of a commonly accepted set of good practices for the use of bibliometrics in research evaluation. These good practices may for instance relate to:

- Situations in which the use of bibliometrics is considered appropriate and situations in which it is not, with special interest to for instance the use of bibliometrics at the level of individual researchers and the use of bibliometrics in the social sciences and humanities.
- The role bibliometrics should play relative to peer review and other methods of research evaluation.
- The transparency of bibliometric research evaluation..
- The way in which manipulation of bibliometric statistics (e.g., citation cartels, journal self-citations) is handled.
- The degree to which the producers of bibliometric statistics (e.g., professional bibliometric centers) have a responsibility to properly inform and educate the end users of these statistics (e.g., university managers).
- More generally, the way in which the responsibility for the use of bibliometrics in research evaluation is shared between the producers of bibliometric statistics on the one hand and the decision makers that use the statistics on the other hand.

Like in the case of indicator standardization, it may not be feasible to reach consensus on good practices for all aspects of the use of bibliometrics in research evaluation. However, standardization of good practices could start in those areas that are considered most important and in which there is most agreement. At some point, one could imagine that this will result in a commonly agreed code of conduct for the producers and the users of bibliometric statistics.

## **Issues for discussion**

We propose that the following topics will be raised by the panel discussants:

- What are the advantages and disadvantages of bibliometric standardization?
- Given the advantages and disadvantages, should bibliometric standardization be considered as something desirable? If so, which types of standardization are most important, and which are less important?
- To what degree will bibliometric standardization be feasible in practice? Could it still be too early to achieve a high level of standardization?
- How could the process toward bibliometric standardization be organized? What should be the role of bibliometric research centers, and what role should be played by the major commercial parties (i.e., Thomson Reuters and Elsevier)? Which other parties should be involved?

- Do we need a professional code of conduct for providers of bibliometric analyses? If so, how should this code of conduct be designed and implemented?
- What role may recent developments such as the increasing interest in altmetrics and the recently published San Francisco Declaration on Research Assessment play in the process?