

Workshop, 12 May 2014, Paris

“Guidelines on good practices for quantitative assessments of research”

Workshop organised by the *Observatoire des Sciences et des Techniques (OST)*, in collaboration with *CWTS, OST-UQAM, SPRU, NIFU, INGENIO* and *OcyT*.

Summary Report

This workshop was dedicated to discuss what type of guidelines on good practices for quantitative assessments of research should be recommended to funding agencies, public research organisations and higher education institutions. It aimed to set up an agenda for a network of institutions willing to elaborate a “manual” (similar to Unesco/OECD manuals) that helps users by offering them standards for accountability and expert advice on good scientometric practices. The workshop was mainly concerned with indicators’ relative usefulness and robustness for diverse purposes and contexts, for example evaluation of scientific impact at individual or organisational level, strategic positioning of various research institutions, or research concerning societal challenges.

The workshop organizers hope to start a process involving many experts, intermediary S&T organisations and users all over the world, in order to build shared recommendations for various uses of standard scientometric indicators.

The **scientific committee** consisted of:

Ismael Rafols, INGENIO & SPRU

Paul Wouters & Sarah de Rijcke, CWTS

Ghislaine Filliatreau, OST

Yves Gingras, OST-UQAM

Gunnar Sivertsen, NIFU

The workshop had **participants from the following organisations:**

NIFU (NO), CWTS - University of Leiden (NL), OECD (Intl.), Digital Science (UK), ORCID (Intl.), EMBO (Intl), OST (FR), ERCEA (EC), OST-UQAM (CA), Euro-CRIS (Intl), PSI (UK), IFQ (DE), Royal Academy of Arts and Sciences (NL), INGENIO (ES), SPRU - University of Sussex (UK), INRA (FR), Science Europe (UK), Mines-ParisTech (FR), University of Utrecht (NL)/ Academic Medical Centre Utrecht (NL).

The problem of standards and quality criteria for scientometric performance indicators was approached from a wide variety of perspectives. The workshop was opened by Ghislaine Filliatreau and Ismael Rafols. They explained the motivation of the initiative in the context of the emerging European and international science policies on accountability and information systems. **Paul Wouters** presented key points discussed on quality standards for research assessment in the past and sketched the main conceptual issues involved, such as the relationship between the goals of the assessment and the indicators used, the role of intermediaries, and the political aspects of evaluation for policy.

Frank Miedema informed the participants about the international debate on the management of science, the special issue of the Economist (“How Science Goes Wrong”) and the Dutch initiative Science in Transition. He also explained why there is widespread dissatisfaction with the current system of research evaluation (including both scientometrics and traditional peer review).

Jack Spaapen presented the new national protocol for research evaluation in the Netherlands, which is an example of international trends in research evaluation practices. He informed the workshop that this protocol gives more weight to societal relevance than in the past. Moreover, the number of publications is no longer a criterion in itself --this shift is a response to anxiety that production of papers has become a career goal in itself for many researchers in the Netherlands (rather than a means to communicate results).

Yves Gingras discussed the characteristics of indicators and their effects at the level of the scientific system as a whole. He went into great detail about the conceptual and technical requirements that indicators should meet, and too often fail to meet.

The second part of the workshop consisted of roundtable discussion on the standardisation process and the content of possible guidelines on the use of quantitative assessments. The roundtable discussion engaged a number of intermediary organizations that play an increasingly important role between the bibliometric experts on the one hand and the “end-user” on the other. These intermediaries were represented in the panel by **Laure Haak** (ORCID), **Jonathan Adams** (Digital Science), **Michele Garfinkel** (EMBO), and **Fred Steward** (EASST). The roundtable was followed by a group discussion with all participants focused on the next steps that need to be taken.

The workshop noted that it is not the first time that the scientometric community (in its widest sense) discusses the need for standards. Already in 1996 Wolfgang Glänzel argued for the need for standards in bibliometric research and technology in the journal *Scientometrics*. Since then a number of steps have been taken leading to quality improvement of evaluative bibliometric analyses. However, it is necessary to go beyond the purely technical issues because many political, ethical and normative problems are equally important.

During workshop the following reasons were presented as important drivers behind the creation of an international consensus about quality standards:

- bibliometrics is increasingly being used in research assessment, in various forms;
- data and indicators for evaluation have become widely available;
- an increasing number of databases are linked to bibliometric analyses, although the reliability and validity of these analyses are not always well presented;
- although the bibliometric community has achieved some consensus regarding the key indicators and methods, many methodological issues in evaluative bibliometrics still need to be solved;
- the bibliometric knowledge base is not easily accessible to the general users;
- the general user is often confronted with a bewildering number of options with respect to data sets and indicators and it is often unclear which choices are best;

- the ethical and political responsibility for research evaluations, including their implications for scientific and scholarly careers, is often ignored or only partially acknowledged;
- the ethical and political responsibility for research evaluations is distributed over various parties (the evaluators, the research institute, the individual researchers, the peer reviewers, the metrics experts) and this increases the need for guidance with respect to good evaluation practices.

These arguments for guidance on good evaluation practices are urgent:

- “end users” demand more clarity about the best way to assess quality and impact from bibliometric experts;
- most bibliometric research is focused on creating more diversity and complexity in the area of indicators and data sets, whereas the general user would prefer some pruning and simplification;
- the bibliometric community is very open about its methods, but it is mainly organized as a scientific community without a well-organized professional channel for quality control of its services (although individual bibliometric service providers variously offer more or less explicit guidelines);
- the bibliometric community does not yet have a Code of Conduct with respect to its evaluation services.

The workshop also noted that three types of standards are relevant:

- standards regarding the organization and quality of the data sets that are used for research evaluation;
- indicator standards for performance measurement and assessment;
- standards for the practices of evaluation (covering both quantitative and qualitative methods and procedures), which regulate issues of transparency and accountability.

Each of these three types of standards has technical, organizational, and ethical dimensions that need to be taken into account.

The workshop noted that data that is being used for research assessment are key ingredients for the publicly funded research and education system, covering universities as well as specialized research institutes. At the same time, some of these data is being delivered by for-profit, commercial entities. These companies can play an important and productive role in the further development of science and scholarship. At the same time, this does entail that these data sets, such as those included in the commercial citation indexes, need to meet minimum requirements regarding transparency with respect to the underlying data (often delivered by the universities themselves) as well as with respect to the methods used to compute key indicators. This transparency is currently often lacking (for example, most global university rankings do not meet these requirements). The workshop participants therefore decided to start writing a common position paper alerting to the need for accessibility and transparency of the meta-data of the scientific and scholarly system in its broadest sense.

No agreement was reached regarding the potential contents, participants or process that might lead to some "guidelines on good practices on quantitative assessments of research". It was noted that in

the UK (via Hefce) and in Germany (with iFQ participation) there is an initiative along there are ongoing public processes towards guidelines and standardisation.

Participants agreed to maintain a collective agenda with conferences, workshops and other meetings where particular aspects of the emerging standards for research assessment will be discussed and worked out in more detail. Different types of initiatives were discussed at the workshop and deemed important:

- technology assessment exercises “USA-style” regarding performance indicators and evaluation technologies;
- exploration of the measurement of societal impact or relevance for assessments;
- studies on how bibliometric evaluation works at the level of the individual researcher (see the ACUMEN project);
- the development of researcher profiles in current social media, Current Research Information Systems, and commercial services;
- a more detailed and fine-grained analysis of how assessments work out at the workbench (comparative ethnography and sociology);
- exchange of experiences with new forms of expert review including non-academic stakeholders.

The next upcoming activity is a plenary session in the Science and Technology Indicators Conference, on September 3rd, 2014 at Leiden (see: <http://sti2014.cwts.nl/Program>)

Paul Wouters & Sarah de Rijcke (CWTS), Ismael Rafols (Ingenio), Ghislaine Filliatreau (OST)