

Reflections on Tools and Methods for Differentiated Assessments of Individual Scientists, Groups of Scientists and Scientific Journals

Ronald Rousseau^{1,2†}, Xiaolin Zhang^{3,4,5†}

¹Centre for R&D Monitoring (ECOOM) and Department MSI, KU Leuven, 3000 Leuven, Belgium

²University of Antwerp, Faculty of Social Sciences, 2020 Antwerpen, Belgium

³Library and information services, ShanghaiTech University, Shanghai, 201210, China

⁴National Science Library, Chinese Academy of Sciences, Beijing, 100190, China

⁵Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing, 100190, China

Citation: Ronald Rousseau, Xiaolin Zhang (2019).

Reflections on tools and methods for differentiated assessments of individual scientists, groups of scientists and scientific journals. *Journal of Data and Information Science*, 4(3), 1–5

DOI: 10.2478/jdis-2019-0011

Requirements for research assessments

There are huge differences in mission, emphasis, inherent capability, and targeted utilization of research among scientific institutions. Hence, when it comes to assessments, a one-size-fits-all approach cannot meet the goal(s) of these assessments. Probably even larger differences exist between individuals, research teams and departments.

It is up to the research community to come up with objective, sound, reliable, easy to use, easy to understand, scalable, and sustainable methodologies, techniques, and tools for all types of scientific assessments, considering the reality of data availability, quality, and computability. Meeting these needs requires more than just changing to another set of indicators. A better understanding of what are the contributions and impacts for each of different types of research is necessary. Multiple data sources and computational methods may be needed, not just as individual tools but often coherently integrated to reveal pertinent, insightful, and, perhaps, even non-expected results. Moreover, tools for interactive analysis even by non-specialist decision-makers may be called for to support using the combined power of human intuition/experience (peers) and data mining & computational analytics.



[†] Corresponding author: Ronald Rousseau (E-mail: ronald.rousseau@kuleuven.be); Xiaolin Zhang (E-mail: zhangxl@mail.las.ac.cn).

Assessment in a scientific framework

Besides persons, other science related entities are also assessed such as journals, research programs and research infrastructure. Here we provide some examples to illustrate the many contexts in which differentiated assessments are expected.

- (1) Entities of evaluation consisting of persons: researchers, research teams, research institutions. There may be differentiating evaluations for researchers or institutions at different career or development stages. Typical reasons for this type of evaluations include promotion and funding. Universities and scientific institutions may have a legal obligation to perform regular assessments of their research work.
- (2) Types of research or performance: basic or applied research in the natural sciences, basic or applied research in the social sciences and humanities, contributing to interdisciplinary research, advancing clinical medicine, patenting, technique & product development, policy research, social engineering, programming, statistical analyses.
- (3) Level of achievement: creating a new field, leading the field, parallel front runners, following but closing in, following with a distance, losing track, ...
- (4) Organization and environment of research: research facilities; research programs & initiatives; research climate, including the existence of offices and policies against scientific misconduct, gender and minority bias or harassment, agreeable advisor-advisee relations.
- (5) Publication outlets as units of assessment: journals, textbooks, handbooks, ...
- (6) Funding agencies themselves.

In the next section we provide some more details about some of these features without any attempt at completeness.

Remarks on some of these types of assessment exercises

Evaluating research teams should include their composition in terms of size, gender, nationality, sectorial composition, i.e. mixed such as in company-university collaborations or uni-sectoral, and age. Is there a clear team leader (not just on paper) who is accepted and supported by the whole team? How is team (and individual) authorship counted (Sivertsen et al., 2019)?

Concerning journal evaluation we mention Wouters et al. (2019) who call for an expansion of journal indicators to cover all functions of scholarly journals. In their call they explicitly mention: registering, curating, evaluating (peer review; issuing corrections if necessary), disseminating and archiving. Evaluating submissions should include a balanced use of reviewers (in terms of gender, geographic



distribution, specialty). Indicators should be such that they are very impractical to manipulate. They should, moreover, be validated through empirical testing. Wouters et al. (2019) further write that *all stakeholders in the system share responsibility for the appropriate construction and use of indicators*.

In recent years, especially in the context of performance-based funding systems, good progress has been made to evaluate the social sciences, arts and humanities on an equal footing as the natural sciences, engineering and medicine, see e.g. (Sivertsen, 2018; Engels & Guns, 2018).

Although less frequently done, also funding agencies are evaluated in terms of the success of their programs. An early example of a comparison between journal articles published by a selected group of grantees (351 in total) and the general literature on schistosomiasis can be found in (Pao & Goffman, 1990). These colleagues found that this small core of sponsored researchers (those sponsored by the *National Institute of Health (USA)*, the *World Health Organization*, the *Edna McConnell Clark Foundation* and the *Rockefeller Foundation*) produced over a 15 year period one-third of the schistosomiasis literature, with a higher impact per paper for grantees than for the total literature. For a more recent study on the success of a funding agency we mention as an example (Bornmann & Daniel, 2005) who studied the workings of the Boehringer Ingelheim Fonds (B.I.F).

In this issue

This special section of the *Journal of Data and Information Science (JDIS)* includes contributions on journal assessment, the evaluation of artistic research, institutional benchmarking and the evaluation of the Centres of Excellence of the Chinese Academy of Science (CAS), a huge conglomerate of institutes in different fields.

Noyons (2019) proposes the ABC method to characterize journals. Here ABC stands for area-based connectedness to society. In this approach he captures signals connecting research output to society. For journal indicators he implements the following dimensions and corresponding signals: news (papers being mentioned in news items); policy (papers being mentioned in policy documents); industry R&D (industry authorship); technological or commercial application (papers cited in patents) and local scope (papers in local languages, including English language journals with a local interest).

Vanlee and Ysebaert (2019) provide a concrete example of how the quality of artistic research output may be evaluated. Obviously, established evaluation models originating from academia (here understood as non-arts) are not suitable. The authors emphasize the importance of allowing an assessment culture to emerge from



Editorial

practitioners themselves, instead of imposing ill-suited methods borrowed from established scientific evaluation models.

Chang and Liu (2019) illustrate a university evaluation system, by taking ShanghaiTech University as an example. ShanghaiTech is a recently (2013) established research university jointly by the Shanghai Municipal Government and the Chinese Academy of Sciences (CAS). It is purposely organized as a small-scale, internationalized, and first-class research institute aiming at solving globally advanced and difficult scientific challenges. Its research performance should manifest itself not in numbers of papers, average number of citations, or even its h-index, but in competitiveness, breakthroughs, breakaways, and power of leading. So the common ranking schemes do not serve its mission. For this reason the authors, working with the university administration, designed and tested a new benchmarking scheme based on competitiveness and research subject distributions of ShanghaiTech compared to a selective group of first international universities. At the moment this scheme relies on publications of research-oriented departments and is accepted as a regular service for the university.

Finally, Fu, and Li (2019) discuss the evaluation practices of the Centers for Excellence of the Chinese Academy of Sciences (CAS). CAS has been developing its more than 100 research institutions in 4 different categories: Centers of Excellence aiming for internationally first-class basic research; Institutes of Innovation striving for technology breakthroughs with global and national significance; Institutes for Specialized Research, focusing on special, more applied, or locally-oriented areas; and finally, scientific facilities that support the whole research community. Apparently, these categories cannot be assessed by a one-size-fits-all evaluation scheme. The authors focus on the assessment design and practices for the Center of Excellence relying on evaluation panels consisting of local, i.e. Chinese, and international experts.

References

- Bornmann, L., & Daniel, H.D. (2005). Selection of research fellowship recipients by committee peer review. Reliability, fairness and predictive validity of Board of Trustees' decisions. *Scientometrics*, 63(2), 297–320.
- Chang, J., & Liu, J.H. (2019). Methods and practices for institutional benchmarking based on research impact and competitiveness. *Journal of Data and Information Science*, 4(3), 55–72.
- Engels, T.C.E., & Guns, R. (2018). The Flemish performance-based research funding system: A unique variant of the Norwegian model. *Journal of Data and Information Science*, 3(4), 45–60.
- Noyons, E.C. (2019). Measuring societal impact is as complex as ABC. *Journal of Data and Information Science*, 4(3), 6–21.



- Pao, M.L., & Goffman, W. (1990). Quality assessment of schistosomiasis literature. In: *Informetrics* 89/90 (Egghe & Rousseau, Eds.), (229–242). Amsterdam: Elsevier.
- Sivertsen, G. (2018). The Norwegian model in Norway. *Journal of Data and Information Science*, 3(4), 3–19.
- Sivertsen, G., Rousseau, R., & Zhang, L. (2019). Measuring scientific production with modified fractional counting. *Journal of Informetrics*, 13(2), 679–694.
- Vanlee, F., & Ysebaert, W. (2019). Disclosing and evaluating artistic research. *Journal of Data and Information Science*, 4(3), 35–54.
- Wouters, P., Sugimoto, C.R., Larivière, V., McVeigh, M.E., Pulverer, B., de Rijcke, S., & Waltman, L. (2019). Rethink impact factors: Find new ways to judge a journal. *Nature*, 569(7758), 621–623.
- Xu, F., & Li, X.X. (2019). Practice and challenge of international peer review: A case study of research evaluation of CAS centers for excellence. *Journal of Data and Information Science*, 4(3), 22–34.



This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

