WOS VS. SCOPUS: ON THE RELIABILITY OF SCIENTOMETRICS

Éric Archambault*

Science-Metrix, 1335A avenue du Mont-Royal E, Montréal, Québec, H2J 1Y6, Canada and Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, Montréal (Québec), Canada.
E-mail: eric.archambault@science-metrix.com

David Campbell

Science-Metrix, 1335A avenue du Mont-Royal E, Montréal, Québec, H2J 1Y6, Canada
E-mail: david.campbell@science-metrix.com

Yves Gingras, Vincent Larivière

Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, Case Postale 8888, succ. Centre-Ville, Montréal (Québec), H3C 3P8, Canada.
E-mail: lariviere.vincent@uqam.ca; gingras.yves@uqam.ca
* Corresponding author

Theme 6: Accuracy and reliability of data sources for scientometric studies

Keywords: papers; citation; databases; web of science; scopus.

Background and research question

For more than 40 years, the Institute for Scientific Information (ISI, now part of Thomson-Reuters), produced the only available database making possible citation analysis, the Web of Science (WoS). Now, another company, Reed-Elsevier, has created its own bibliographic database, Scopus, available since 2002. For those who perform bibliometric analysis and comparisons of countries or institutions, the existence of these two major databases raises the important question of the comparability and stability of rankings obtained from different data sources.

Although several studies have compared the WoS and Scopus for Web use [BALL and TUNGER, 2006; BAR-ILAN, 2008; BOSMAN et al., 2006; FALAGAS et al., 2008; JACSO, 2005; MEHO and YANG, 2007; NORRIS and OPPENHEIM, 2007], no study has yet compared them in the context of a bibliometric production environment. This paper provides a comparative analysis of the ranking of countries in terms of the number of papers and the number of received citations. A high correlation between both rankings would suggest that these databases reflect the dynamics of knowledge production at the country level.

Methods

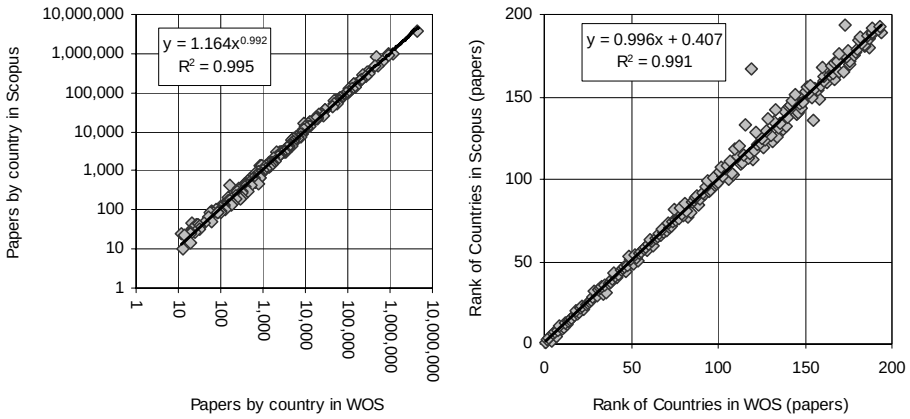
Data for this paper were produced from the WoS and Scopus databases for the 1996–2006 period. This short comparison period is a restriction imposed by Scopus, which does not include references prior to 1996. However, for performance measurement, having the last ten years of data is sufficient in the vast majority of cases. Moreover, our

objective here is not to provide an assessment of countries but rather to compare the results obtained from the two sources in order to evaluate their robustness. All document types were retained when calculating the number of papers and citations by country in both databases, the majority of which are journal articles. The nanotechnology datasets were built by running a keyword (title, abstract, author's keywords) search in each database.

Results

Figure 1 examines how the number of papers per country in Scopus compares to that measured in WoS in absolute terms (1a) and when countries are ranked (1b). The correlations between the measured outputs in both databases are remarkably strong (R^2 of 0.995 and 0.991, respectively). When examining top-ranking countries, Scopus often gives higher grades to Asian countries (e.g. Japan and China each gain two ranks) whereas several European and English-speaking countries cede one rank (e.g. the UK, Germany, France, Canada, the Netherlands). However, except for minor variations, the top-ranking countries have similar ranks when measuring output in either database, and the top 25 countries are identical, the changes never exceeding a difference of two places.

Figure 1 a,b Correlation in number of papers by countries, WoS and Scopus, 1996–2006



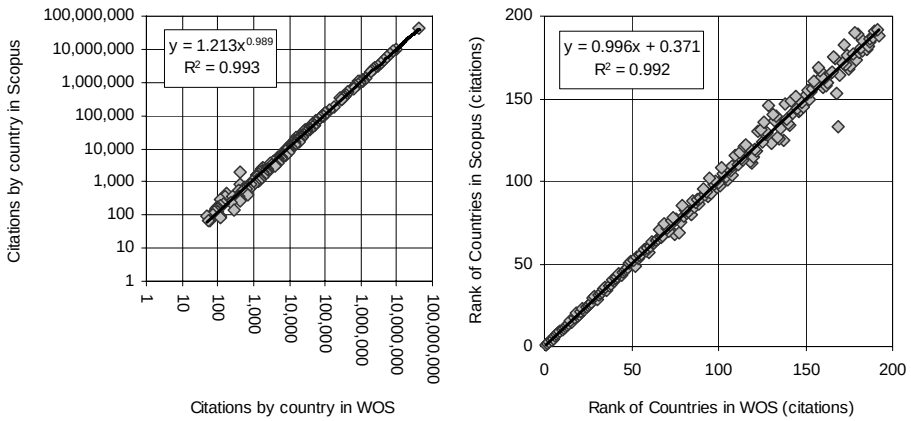
Source: Scopus data compiled by Science-Metrix and WoS data by OST

Figure 2 presents citation data at the country level for both databases. As for papers, citation-level data are strongly correlated in both absolute terms and in terms of rank (R^2 of 0.993 and 0.992, respectively) and the listing of top 25 countries according to citations received is the same regardless of the database used (i.e., there are slight variations in the ranking among top countries but again changes never exceed two ranks).

We also computed how differently these databases measure the output of countries in nanotechnology. This test confirms they produce very similar results. In particular, the correlation coefficient (R^2) for the number of papers and citations is 0.991 and 0.967, respectively. Using rankings instead of absolute numbers of papers and citations, the correlations coefficient become respectively 0.990 and 0.974. For both databases, the top 25 countries are the same in nanotechnology for papers and for citations. A few countries

have somewhat different outputs in the two databases, but these databases produce remarkably similar rankings in terms of number of papers and citations for countries that have at least 100 papers. The variations for these countries never exceed six ranks for papers and seven ranks for citations.

Figure 2 a,b Correlation in citations received by countries, WoS and Scopus, 1996–2006



Source: Scopus data compiled by Science-Metrix and WoS data by OST.

Conclusion

The data presented here provide strong evidence that scientometrics based on bibliometric methods is a sound undertaking at the country level. Despite the fact that the content and coverage of these databases is different in terms of scope and volume of data, the output (papers) and the impact (citations) of countries is highly correlated between WoS and Scopus, even in subsets of the data, such as in nanotechnology. Further research could examine differences at the institutional level as well as in different fields to verify whether these results can be generalized to different scales.

References

BALL, R.; TUNGER, T. (2006), Science indicators revisited – Science Citation Index versus SCOPUS: A bibliometric comparison of both citation databases, *Information Services & Use*, 26: 293–301.

BAR-ILAN, J. (2008), Which h-index? – A comparison of WoS, Scopus and Google Scholar, *Scientometrics*, 74: 257–271.

BOSMAN, J., van MOURIK, I., RASCH, M., SIEVERTS, E., VERHOEFF, H. (2006) *Scopus reviewed and compared*. Universiteitsbibliotheek Utrecht.

JACSO, P. (2005), As we may search – Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases, Current Science, 89(9): 1537-47.

MEHO, L.I.; YANG, K. (2007), Impact of data sources on citation counts and rankings of LIS Faculty: Web of science versus Scopus and Google Scholar, *JASIST*, 58: 2105–2125.

FALAGAS, M.E.,. PITSOUNI, E.I., MALIETZIS, G.A. and PAPPAS, G., (2008), Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses, *FASEB J*, 22, 338–342.

NORRIS, M.; OPPENHEIM, C. (2007), Comparing alternatives to the Web of Science for coverage of the social sciences' literature, *Journal of Informetrics*, 1: 161-169.