

# The journal coverage of bibliometric databases: A comparison of Scopus and Web of Science

Philippe Mongeon and Adèle Paul-Hus

École de bibliothéconomie et des sciences de l'information, Université de Montréal

## Introduction

Bibliometric and scientometric methods have multiple applications in information science, sociology and history of science, research evaluation and scientific policy (Gingras 2014). Large scale bibliometric research was made possible by the creation and development of the Science Citation Index (SCI), which is now part of Web of Science (WoS) along with two other indexes: the Social Science Citation Index (SSCI) and the Arts and Humanities Citation Index (A&HCI). WoS has been the sole tool for citations analysis until the creation of Scopus in 2004.

The validity of bibliometric analyses lies in large part on the databases' representativeness of the scientific activity studied. However, the local orientation of research and the scientific communication practices in Social Sciences as well as in Arts and Humanities make it more difficult for those databases to accurately reflect research activities in these fields (Hick 2011, Nederhof 2006).

Analyzing data from 2004, Archambault et al. (2006) observed an important English-language journals overrepresentation in the WoS coverage compared to Ulrich's database which is considered the most comprehensive worldwide list of periodicals. They concluded that "Thomson Scientific databases cannot be used in isolation to benchmark the output of countries in the [Social Sciences and Humanities]" (p. 329).

Ten years later, we propose to compare in this paper the coverage of both WoS and Scopus in terms of fields, countries and languages, thus verifying if the previously found bias in WoS databases still exists and if a similar bias can be found in Scopus.

## Methodology

### Journal lists

In Ulrich's periodical database, we found 70 644 unique journals classified as "Academic/Scholarly". We downloaded the list of 34 278 journals indexed in Scopus as well as the Web of Science journal list containing 16 957 entries indexed in the Science Citation Index (SCI), the Social Science Citation Index (SSCI) and the Arts and Humanities Citation Index (AHCI).

### Matching

Journals from the Web of Science and the Scopus lists were matched to the Ulrich's list in two steps. In the first step, the journals were matched by means of their ISSN. In the second step, the remaining journals were matched by means of their title and publishing country and these matches were manually verified to eliminate any false

positives. Using this procedure, we were able to match 13 605 and 20 346 from the Web of Science and Scopus lists, respectively, with the Ulrich's list.

### Journals classification

A broad discipline classification of journal titles was done according to Ulrich's subject code. Every journal was classified into one of the following research areas: Health Sciences, Natural Sciences and Engineering, Arts and Humanities and Social Sciences.

## Results

### Coverage by field

Table 1 shows the distribution of journals by field for each database. We see that Health Sciences (HS) and Natural Sciences and Engineering (NSE) are overrepresented in both WoS and Scopus in comparison to Ulrich, Scopus showing a stronger positive bias for HS while WoS is showing a stronger positive bias for NSE. Inversely, Arts and Humanities (A&H) and Social Sciences (SS) are both underrepresented in WoS and Scopus in comparison to Ulrich.

Table 1. Distribution of journals by field

Field	Ulrich	WoS		Scopus	
	% of journals	% of journals	Difference with Ulrich	% of journals	Difference with Ulrich
HS	24,64%	30,56%	24%	34,48%	40%
NSE	28,64%	42,72%	49%	32,89%	15%
AH	13,59%	8,59%	-37%	8,73%	-36%
SS	33,13%	18,12%	-45%	23,90%	-28%

### Coverage by publishing country

Figure 1 shows the difference of relative coverage between the two databases and Ulrich in terms of publishing country. For example, Health Sciences journals published in the United Kingdom are overrepresented by 46% in Scopus and 78% in WoS. In all but a few cases, the same countries are either overrepresented or underrepresented in both databases. The exceptions are France in HS (overrepresented in Scopus and underrepresented in WoS) and Italy, Brazil and Poland in NSE (all underrepresented in Scopus and overrepresented in WoS). In all fields, journals published in the United Kingdom and in the Netherlands are the most strongly overrepresented. This may not be surprising following the fact that some of the most important publishing companies are located in these countries.

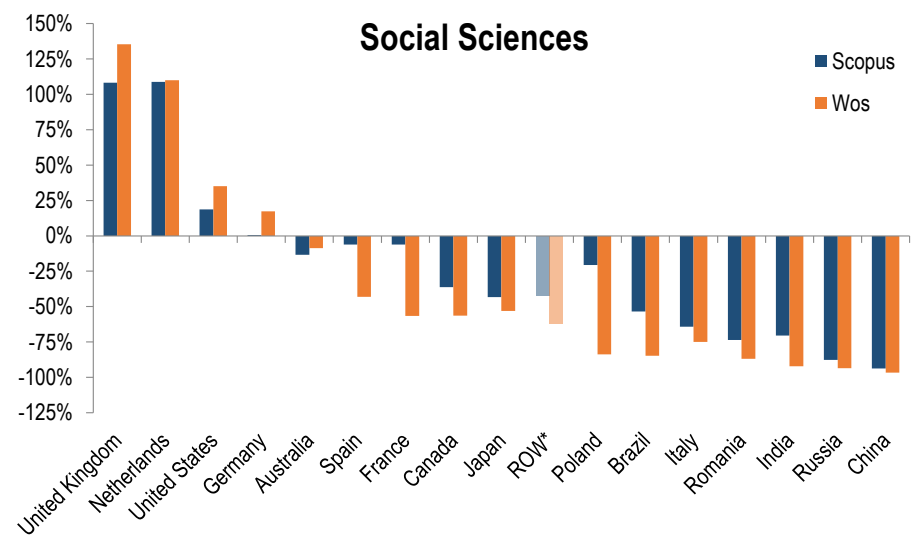
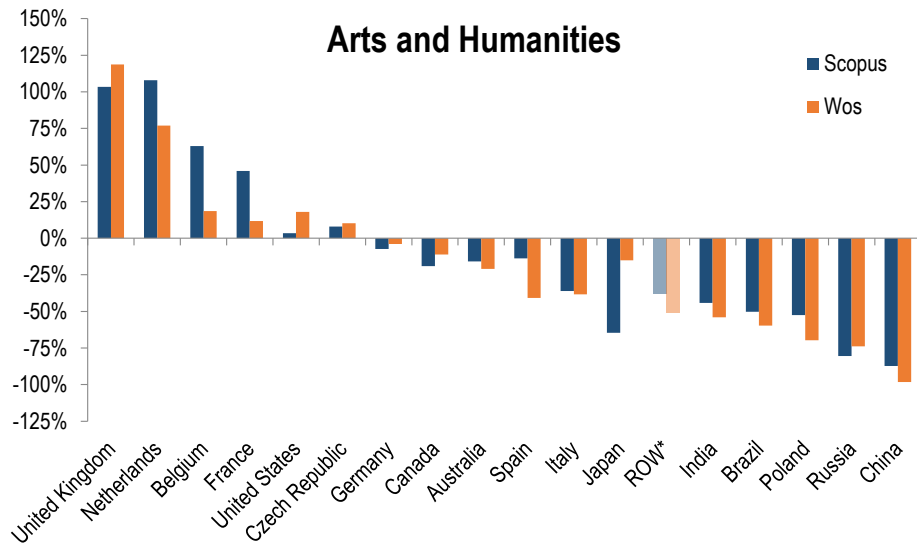
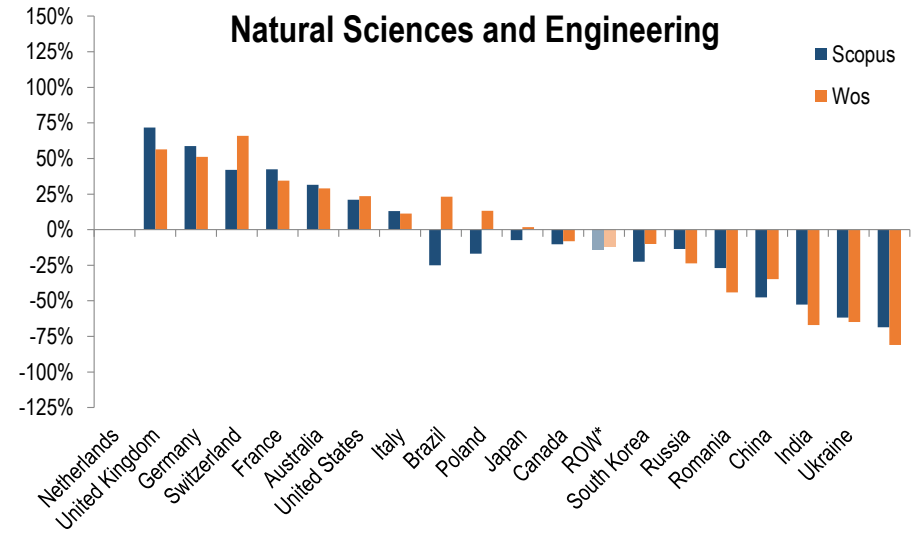
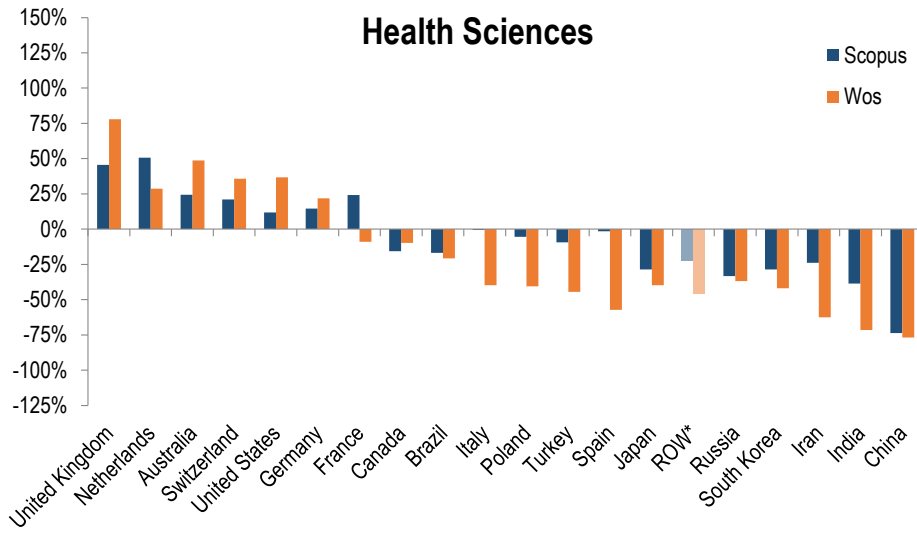


Figure 1. Relative journal coverage by publishing country compared to Ulrich

### **Coverage by language**

Figure 2 shows the difference of relative coverage between both databases and Ulrich in terms of journals language. In Scopus, the majority of languages are underrepresented in all fields. WoS shows a similar trend except for NSE where nine languages are overrepresented. English is the only language that is constantly and strongly overrepresented in the two databases and in all fields.

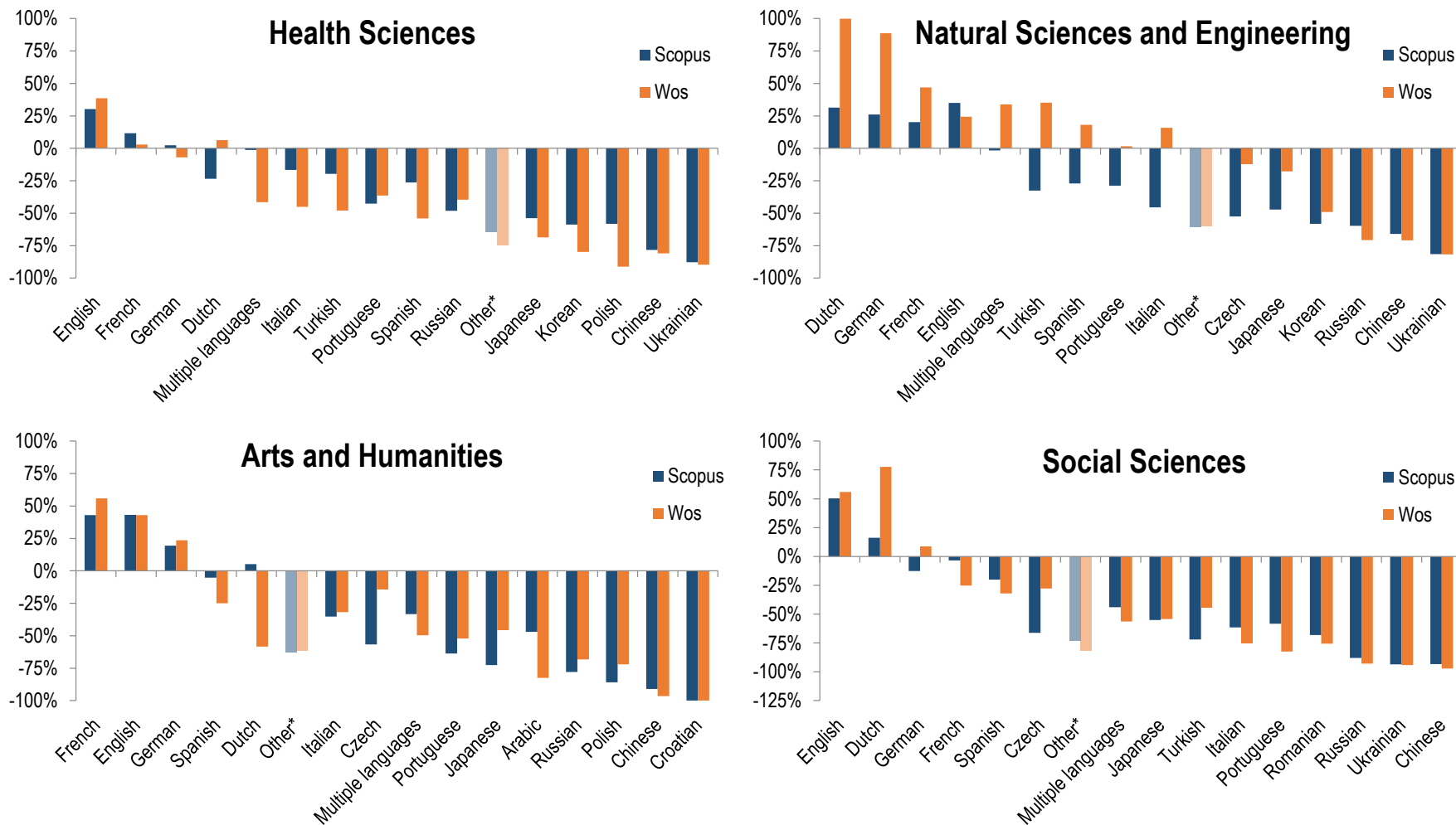


Figure 2. Coverage by language of journals

## Conclusion

Our analysis of the journal coverage in WoS shows that Social Sciences and Humanities are still underrepresented and shows a strong bias toward English-language journals. Despite the greater number of journals included in the Scopus database, its journal coverage presents similar biases. Consequently, using WoS and Scopus for research evaluation introduces biases that favor Natural Sciences and Engineering as well as Health Sciences to the detriment of Social Sciences and Arts and Humanities. Similarly, English-language journals are favored to the detriment of other languages. These important limits should be taken into account when assessing scientific activities.

## Cited references

- Archambault, É., Vignola-Gagné, É., Côté, G., Larivière, V., & Gingras, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, *68*(3), 329–342.
- Gingras, Y. (2014). Les dérives de l'évaluation de la recherche : du bon usage de la bibliométrie. Paris : Raisons d'agir.
- Hicks, D., & Wang, J. (2011). Coverage and overlap of the new social sciences and humanities journal lists. *Journal of the American Society for Information Science and Technology*, *62*(2), 284–294.
- Nederhof, A. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, *66*(1), 81-100.