

## VIEWPOINT

# Web of Science and Scopus are not global databases of knowledge

Jonathan P Tennant<sup>†</sup>

*Institute for Globally Distributed Open Research and Education (IGDORE), UK; ORCID 0000-0001-7794-0218*

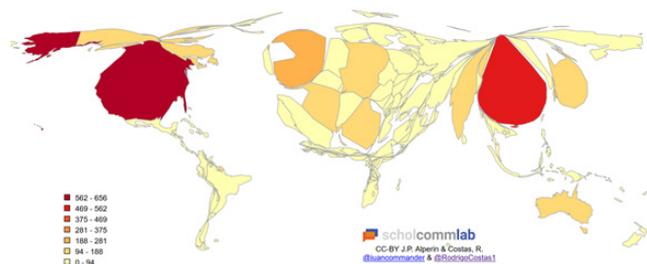
DOI: 10.3897/ese.2020.e51987

### Abstract

Both Web of Science and Scopus are critical components of the current research ecosystem, providing the basis for university and global rankings as well as for bibliometric research. However, both platforms are structurally biased against research produced in non-Western countries, non-English language research, and research from the arts, humanities, and social sciences. This viewpoint emphasizes the damage that these systematic inequities inflict upon global knowledge production systems and the need for research funders to unite to form a more globally representative, non-profit, community-controlled infrastructure for the global pool of research knowledge.

There has never been a more pressing need for science to be working to address the major challenges that we face as a global society, as envisaged by the UN Sustainable Development Goals.<sup>1</sup> The earth's climate is changing catastrophically and irreversibly; we are in the midst of yet another global pandemic<sup>2</sup>; and we are facing resource distribution inequities like never before in the face of a booming global population. These are global challenges that affect us all, and therefore we need to ensure that the science we are using to help address them is truly globally representative.<sup>3,4</sup>

Clarivate Analytics' Web of Science (WoS) and Elsevier's Scopus platforms are synonymous with data on international research. Both are widely considered by the scholarly community to be the two most trusted or authoritative sources of bibliometric data and form the basis for virtually all peer-reviewed knowledge on research across different disciplines. Figure 1 is a cartogram, developed by Juan Pablo Alperin and Rodrigo Costas, which shows the world scaled in proportion to the number of publishing researchers per country. The publication data are from Scopus, and reveal an alarmingly warped version of reality: research from Africa, South America, and major parts of Asia is almost non-existent.



**Figure 1.** Map of the world scaled according to number of published researchers in Scopus.

As the largest of their kind, WoS and Scopus are often hailed as 'global' databases of knowledge and used widely for bibliographic research and academic assessments. This includes the creation of global higher education rankings or their adoption in tenure and promotion guidelines; WoS, for

example, features prominently in the UK Research Excellence Framework 2021 and in international league tables, and bibliographic data from Scopus represent more than 36% of assessment criteria in the popular Times Higher Education world university rankings.

Cameron Neylon and others criticize these rankings, demonstrating that the data sources underpinning them are heavily biased and incomplete.<sup>5,6</sup> This seems strange, given that the current global scholarly ecosystem imposes a critical dependence on them. Neither of the two databases seems to do a fair, precise, or even reasonable job of being unbiased or globally representative; instead, both platforms seem to discriminate against different forms of knowledge, particularly that which does not hail from the English-speaking Western world.

In general, Scopus is larger and geographically broader than WoS; however, Scopus covers only a fraction of journal publishing outside of Europe and North America. This discrimination is especially visible in the case of Asia. As of August 2017, Scopus reports a coverage of over 2000 journals in the Asia-Pacific region, which it boasts as being '230% more than the nearest competitor'. Now, this might seem impressive, until you actually look at the data. In Indonesia alone, now the top country in the world for Open Access (OA) publishing,<sup>7</sup> the national government's Garuda portal currently lists more than 9000 journals published in the country—more than four times what Scopus indexes for the entire Asia-Pacific region. Similarly, in Japan, nearly 3000 journals are currently listed on the national J-Stage platform.

Scopus currently lists 750 journals from Africa and the Middle East, 212% more than the nearest competitor, whereas African Journals Online alone indexes 524 journals; the Directory for Arabian Journals lists 319 journals; the Algerian Scientific Journal Platform indexes 510 journals; and Iraqi Academic Scientific Journals lists 272 OA journals.

In Latin America, we see much the same story. Scopus claims to list about 700 journals from this region, 168% more than the nearest competitor. SciELO, which has been providing open infrastructure for Latin American health science journals since 1997, lists more than 1700 active journals at present. Redalyc also supports more than 550 social sciences journals in Latin America.

In terms of language preference, we see much the same pattern. As Ryan Regier stated in 2018: “For example, Scopus indexes more than 23,000 currently-published scholarly journals and only about 5,000 of these publish in languages other than English. (You can find this data in the Scopus Title List).

“Compare that with the Directory of Open Access Journals (DOAJ), who does a much better job with outreach in non-English countries, of their 11,000 journals, 6,000 publish in languages other than English. They have less than half the journals Scopus does and they still have more non-English Journals!”

A more recent study from 2019 confirmed the extent of this bias: “English dominates both WoS and Scopus (92.64% of the documents indexed in Scopus are in English and this percentage is even higher in the WoS with 95.37% compared to the second language with the highest number of documents in Scopus, Chinese, with 2.76% and the second language in WoS, Spanish, with 1.26%).”<sup>8</sup>

Given that most nations on this planet do not speak English as a first language, unless imposed by colonialism, this clearly represents a hegemonic linguistic practice. This has consequences. First, non-native English-speaking authors may be required to spend part of their research budgets, as well as a significant amount of time and effort, on bringing their work to an ‘acceptable’ standard of English, a problem exacerbated by the exploitative publish-or-perish system. Although there are some clear advantages in homogenizing the language of science, such a practice simultaneously erodes the social and cultural contexts inherent in non-English languages.

Clarivate Analytics have taken some positive steps to broaden the scope of WoS, integrating the SciELO citation index and also through the creation of the Emerging Sources Index (ESI), which has allowed the inclusion of many more international titles into the service. Given the rich history of knowledge production in Latin America, the fact that ESI refers to this as the ‘emerging sources’ seems to display a clear prejudice: they are emergent only in the sense that WoS did not index them earlier, not because the research did not exist or was not well established.

Together, these numbers highlight quite a mismatch between the industry leaders and the global reality, with a number of structural geographic discrepancies. In other parts of the world, this mismatch seems to be less of a problem. In western Europe, Scopus lists 11,000 journals, and 6000 from North America. Other studies have also shown that these services tend to under-represent research from the arts, humanities, and social sciences.<sup>9</sup> Combine this with the linguistic and geographic biases inherent in either database, and it seems difficult to argue that this a quality issue.

One of the reasons for this skew might simply be that journals from the excluded regions are of varying or lower ‘quality’. However, it is well known that legitimate publishing activities from the ‘global south’, such as *Academic Journals*, have an ethnocentric prejudice focused on them. Much of this perspective was created thanks to the infamous, and now largely academically discredited, ‘predatory publishing’ list.<sup>10</sup> However, this remains a pernicious issue at the heart of the global scholarly communication system. First of all, defining the quality of both research and journals is beset with difficulties. A recent consensus definition on what

constitutes a ‘predatory publisher’ did not seek to explicitly delimit legitimate and illegitimate publishing activities, but did highlight the complexity of this situation;<sup>11</sup> by these standards, even the largest publishers such as Elsevier could be classified as predatory.

Irrespective of these problems, the selection criteria for journals and the power imbalance they impose upon the global research community should not be left to commercial third-parties such as WoS or Scopus. It should be the research community, including learned societies and institutes, who makes quality assessments and defines the standards for journals and scholarly communication. WoS and Scopus are both commercial and for-profit services that, irrespective of their methods, have a fiduciary duty and accountability to their shareholders and investors—not a duty to science or to the public. The reality is that the global research community has outsourced the critical functions of acting as custodians for our scholarly ecosystem to a handful of private companies. And not just that, but organizations with an incredible track record of harm to the scholarly community.<sup>12</sup>

Both are shining beacons in the world of ‘platform capitalism’ with business models that not only predate but also outperform those of Google and Facebook in terms of profit.<sup>13</sup> It is perhaps no wonder that these databases are so biased towards research from western Europe and North America. WoS is owned by Clarivate Analytics, based in London, and Scopus by Elsevier, based primarily in Amsterdam and London. Their geographic location is not a coincidence here: they are two of the centres of historical Western colonialism, the aftershocks of which are still being widely felt.

Widespread use of both services continues to reinforce a Western hegemony in global scientific endeavours. The result of this is suppression of global innovation through reducing epistemic diversity in participation in the research process and relegating specific forms of knowledge to the ‘periphery’ as a form of cognitive injustice. Exclusion of non-Western knowledge from these databases dictates what we read, what we value, and we build upon. This in turn discriminates against the contexts of knowledge generation in those places, including invaluable cultural perspectives. For example, in Latin America, there is an incentive policy in higher education that aims to foster increasing internationalization. This is understood as publishing in WoS- and Scopus-indexed journals. A consequence of this is that some researchers no longer focus on ‘local’ topics because they would not get published in ‘international’ journals. Virtually all journals publish important work at different levels and are relevant to different elements of society. That this is suppressed in systematic ways against specific forms of knowledge by WoS and Scopus should be a matter for concern. I do not believe that either entity should be in control of defining what research is deemed significant or not significant, given that they intrinsically have little scientific interest in this matter.

This trend will continue as both Clarivate Analytics and Elsevier continue to extend their control over critical elements of scholarly infrastructure.<sup>14</sup> The fact that Elsevier is also the leading publisher of scholarly content is not a coincidence, and amounts to one of the most significant conflicts of interest in the world.<sup>15</sup> The very fabric of our knowledge society is in the ongoing process of being handed over to for-profit enterprises.

Knowledge infrastructure, academic cultures, and research practices have become subject to the maximization of profits in the interests of a few and at the expense of everyone else.

The first simple step to resolve this problem is to simply stop all research from using either platform, and for users and institutes to stop subscribing to them. By deconstructing or reducing the power of existing faulty elements of the scholarly ecosystem, funds and energy can be liberated and put to use in investing in a more open, non-profit, community-owned global scholarly communication infrastructure, which provides more efficient, effective, and representative information on the global knowledge landscape. This born-digital infrastructure should be truly equitable, comprehensive, and multilingual, facilitating fair participation in knowledge creation.

Decolonizing scholarly communication is not something simple, and the primary focus must be on creating inclusive digital infrastructure that does not replicate the hegemony inherent in the present systems. This requires synchrony in both reducing the power status of those existing systems and simultaneously amplifying other voices that have previously been marginalized to occupy such a space. International research funding bodies have a key responsibility here to unite to help achieve this in the context of the UN Sustainable Development Goals. This promises to address the geopolitical impact that existing systematic discrimination has on knowledge production<sup>16</sup> and to further the inclusion and representation of marginalized research demographics within the global research landscape.

### Acknowledgements

Thank you to Andy Nobes, Bárbara Rivera-López, and Asura Enkhbayar for valuable discussions and input to an earlier version of this article.

### References

- 1 Griggs D, Stafford-Smith M, Gaffney O, Rockström J, Öhman MC, Shyamsundar P, et al. Sustainable development goals for people and planet. *Nature*. 2013 Mar;495(7441):305–7.
- 2 Dunleavy DJ. Coronavirus As Impetus For a Lasting Change in Research Culture [Internet]. SocArXiv; 2020 Mar [cited 2020 Mar 5]. Available from: <https://osf.io/2ryt3>
- 3 Archambault É, Campbell D, Gingras Y, Larivière V. Comparing bibliometric statistics obtained from the Web of Science and Scopus. *Journal of the American Society for Information Science and Technology*. 2009;60(7):1320–6.
- 4 Ciarli T, Rafols I, Llopis O. The under-representation of developing countries in the main bibliometric databases: a comparison of rice studies in the Web of Science, Scopus and CAB Abstracts. *Proceedings of the science and technology indicators conference 2014 Leiden: "Context Counts: Pathways to Master Big and Little Data."* 2014;97–106.
- 5 Huang C-K (Karl), Neylon C, Brookes-Kenworthy C, Hosking R, Montgomery L, Wilson K, et al. Comparison of bibliographic data sources: Implications for the robustness of university rankings. *bioRxiv*. 2019 Sep 1;750075.
- 6 Selten F, Neylon C, Huang C-K, Groth P. A Longitudinal Analysis of University Rankings. *arXiv:190810632 [physics]* [Internet]. 2020 Jan 20 [cited 2020 Mar 12]; Available from: <http://arxiv.org/abs/1908.10632>
- 7 Noorden RV. Indonesia tops open-access publishing charts. *Nature* [Internet]. 2019 May 15 [cited 2020 Mar 12]; Available from: <https://www.nature.com/articles/d41586-019-01536-5>
- 8 Vera-Baceta M-A, Thelwall M, Kousha K. Web of Science and Scopus language coverage. *Scientometrics*. 2019 Dec 1;121(3):1803–13.
- 9 Martín-Martín A, Orduna-Malea E, Delgado López-Cózar E. Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: a multidisciplinary comparison. *Scientometrics*. 2018 Sep 1;116(3):2175–88.
- 10 Eve MP, Priego E. Who is Actually Harmed by Predatory Publishers? 1. 2017 Aug 13;15(2):755–70.
- 11 Grudniewicz A, Moher D, Cobey KD, Bryson GL, Cukier S, Allen K, et al. Predatory journals: no definition, no defence. *Nature*. 2019 Dec;576(7786):210–2.
- 12 Tennant JP. Democratising knowledge: A report on the scholarly publisher Elsevier [Internet]. *Education International*; 2018. Available from: <https://bit.ly/2PPjwRK>
- 13 Mirowski P. The future(s) of open science. *Soc Stud Sci*. 2018 Apr 1;48(2):171–203.
- 14 Posada A, Chen G. Inequality in Knowledge Production: The Integration of Academic Infrastructure by Big Publishers. In: Cha L, Mounie P, editors. *ELPUB 2018* [Internet]. Toronto, Canada; 2018 [cited 2019 Jan 24]. Available from: <https://hal.archives-ouvertes.fr/hal-01816707>
- 15 Tennant J, Brems B. RELX referral to EU competition authority [Internet]. *Zenodo*; 2018 Oct [cited 2018 Dec 12]. Available from: <https://zenodo.org/record/1472045#XBEoWj0IPY>
- 16 Paasi A. Academic Capitalism and the Geopolitics of Knowledge. In: *The Wiley Blackwell Companion to Political Geography* [Internet]. John Wiley & Sons, Ltd; 2015 [cited 2019 Mar 1]. p. 507–23. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118725771.ch37>