ISSI 2023 Proceedings, v2, pp 7–13
DOI: 10.5281/zenodo.8350260 https://cns-iu.github.io/workshops/2023-07-02_issi/

On regional knowledge systems and the limitations of mainstream academic databases: Preliminary findings

Patricia Alonso-Álvarez[1]

[1]*patalons@bib.uc3m.es*
LEMI Research Group, Carlos III University of Madrid, 126 Madrid Str., 28903, Getafe, (Spain)
Research Institute for Higher Education and Science (INAECU), Carlos III University of Madrid, 126 Madrid Str., 28903, Getafe, (Spain)

Abstract
The unequal knowledge production and dissemination structure divide the world's academic system into a dominating mainstream circuit and a peripherical one. Several local and international initiatives have recently emerged to reduce regional inequalities and build alternative transnational publishing circuits. This project studies alternative publishing circuits and their representation in international bibliometric databases. To do so, it focuses on the African publishing environment. Using Ulrich's periodicals database, a regional journal platform (African Journals Online, AJOL hereafter), Scimago Journal Ranking (SJR), and Journal Citation Reports (JCR) data, this paper shows the coverage of the databases of African journals and their relative coverage in terms of countries and research areas. The results show that mainstream databases are biased toward specific countries and research areas, which has important social and academic implications. Building on the preliminary findings presented here, this work will be developed further by including additional variables, improving research area assignation, and considering analysis at the publication level.

Introduction
The unequal knowledge production and circulation structure divide the world's academic system into a dominating mainstream circuit and a peripherical one. Although the main scientific databases claim to include academic journals based on quality standards, it is well known that they are biased toward certain research areas, countries, and languages (Archambault et al., 2009; Mongeon and Paul-Hus, 2016). Focusing on the case of rice research, Rafóls et al. (2015) showed that the Web of Science (WoS) and Scopus are indeed biased toward specific topics and countries. Moreover, Chavarro et al. (2018) found that quality alone could not predict journals' inclusion on WoS, and that discipline, country, and language played a significant role. A recent study by Khanna et al. (2022) estimated that WoS covers only 1.2% of world journals using Open Journal Systems (OJS). As academic journals play an essential role in disseminating knowledge, ignoring journals from specific regions or research areas diminishes the importance of 'local' and 'regional' research outside the knowledge centers –mainly Europe and North America. Rather than being localized to the scientific community, this discussion also has direct political implications as scientific knowledge is strongly linked to social development (see, for instance: Kreimer and Zabala, 2007; Coburn et al., 2022). If researchers are forced to publish in 'international' journals –meaning those included in WoS and Scopus– to gain recognition, it might be the case that local debates remain uncovered (Hanafi, 2011).

Several local and international initiatives have emerged in recent years to reduce regional inequalities and build alternative transnational publishing circuits. The open access movement is one of the core pillars of the shift toward a more universal, accessible, and fair science. Specific projects on open publishing include regional online libraries such as SciELO and Latindex or the Journals Online project. Other initiatives introduce software developments such as OJS, which has been used to decentralize academic publishing worldwide.

This paper studies local and regional journals and their representation in international bibliometric databases. Although previous research has studied this topic, research outside the

Latin American context has been scarce. Thus, I focus on the African publishing environment to show the limitations of mainstream databases. To do so, I study the coverage of the Web of Science and Scopus of African-based research, and I compare it with a local journal aggregator (African Journals Online) and Ulrich's Periodicals Database to cover a broader spectrum of journals –although acknowledging its limitations (Wang et al., 2017). As this is an ongoing project, this paper only explores African-based research at the journal level.

Data and methods

I use the publisher's country to analyze African-based journals and their representation in regional and international academic databases (Mongeon and Paul-Hus, 2016). Therefore, I assume a journal is from a particular country if the publisher is located there.

All data were retrieved from four sources. For international databases, I use both WoS and Scopus. For WoS, I downloaded the JCR 2022 dataset from Clarivariate's webpage and filtered it by region. All African-based journals were retrieved (241 journals). I followed a similar procedure for Scopus: I downloaded all journals belonging to the region 'Africa' from the Scimago Journal Rank (SJR) from 2014 to 2021 (314 journals).

The regional dataset for this research comes from African Journals Online. There are two main reasons for using AJOL data. First, AJOL is "the world's largest and preeminent platform of African-published scholarly journals" (African Journals Online, n.d.). AJOL was launched in 1998 by the International Network for the Availability of Scientific Publications (INASP) in collaboration with the Public Knowledge Project (PKP), and it covered only 50 African journals. In December 2022, the number of covered journals increased to 628 from 37 African countries. The second reason for using AJOL is its sparse use in international academic literature, as most publications continue using mainstream databases. Simple searches in WoS and Scopus illustrate this tendency. Using *TS=(AJOL OR "African Journals Online")* as a search strategy, WoS only retrieves 467 papers, the majority of which are from Medicine and Biomedical Sciences, and only 12 belong to Information and Library Sciences. In Scopus, *TITLE-ABS-KEY ("AJOL" OR "African Journals Online")* retrieves 630 documents. Most of the publications belong to Medicine and areas related to Health Sciences. A manual revision of the retrieved documents from the areas not belonging to these categories –Social Sciences, Multidisciplinary, Environmental Sciences, Arts and Humanities, Computer Sciences, Decision Sciences, Engineering, Decision Sciences, and Business– showed that only 29 documents were not related to Health Sciences. Therefore, AJOL constitutes a large and understudied source of bibliometric data from African countries.

I downloaded all journals' data from the AJOL website using R language. Data were downloaded in December 2022. As this is an ongoing project, all the data will be updated regularly to account for the changes in AJOL coverage. The journal titles, country, and research area were retrieved using *rvest* (Wickham, 2022), an R package for web scrapping. For the ISSN, I downloaded the metadata from all publications in AJOL using *ojsr* (Becerra, 2022), an R package developed to download metadata from OJS sites, and merged them with the journals' dataset to get the ISSN of the journals. Then, the journal data was reviewed manually to account for mistakes in the download.

Ulrich's Periodicals Directory (Ulrich) is one of the most comprehensive databases of academic journals (Grimes and Morris, 2006). Therefore, I use it to explore the broader picture and better understand the coverage of the other three data sources. I manually downloaded journal information from all African countries. As Ulrich has different records for each

journal's print and online versions, I merged the two versions into only one record, including ISSN and eISSN. After removing duplicates, I retrieved 2,923 African-based journals.

All datasets were merged to get a complete dataset with all journals covered by AJOL, Ulrich, SJR, and JCR. The full dataset allowed coverage analysis of the databases. Finally, all downloads and the subsequently merged table were manually revised, and the ISSN Portal was used to resolve inconsistencies between datasets. All manual changes in the databases were recorded in a separate document, so the procedure is traceable and replicable.

The publishers' countries were retrieved from AJOL, Ulrich, SJR, and JCR datasets. In 15 journals, the country recorded in one dataset did not match the country in another. The ISSN Portal was used in these cases to verify the journals' country. In 6 cases, the ISSN portal did not find any results with the reported ISSN, so those journals were removed from the country analysis. One other journal was removed from the analysis as it did not have country data in any of the datasets. Therefore, the analysis includes 3,011 journals.

Research areas were retrieved from the four databases. AJOL, Ulrich, SJR, and JCR all have different research area categories; therefore, I normalized them using the Revised Field of Science and Technology (FOS) Classification in the Frascati Manual (OECD, 2007). I correlated categories in the studied databases with one of the broad category fields of the FOS classification: natural sciences, engineering and technology, medical and health sciences, agricultural sciences, social sciences, and humanities. Those unclear categories (for instance, Ulrich had a category that was 'analysis') were not classified. Journals can belong to more than one category. Further work of this project will try to improve the classification system to include all journals in the databases. The analysis of research areas finally included 2,882 journals.

Results

The first step of the analysis shows the number of journals each database covers and how databases overlap. Figure 1 was generated using *nVennR* (Pérez-Silva et al., 2018), an R package designed to create quasi-proportional Venn and Euler diagrams with any number of sets.

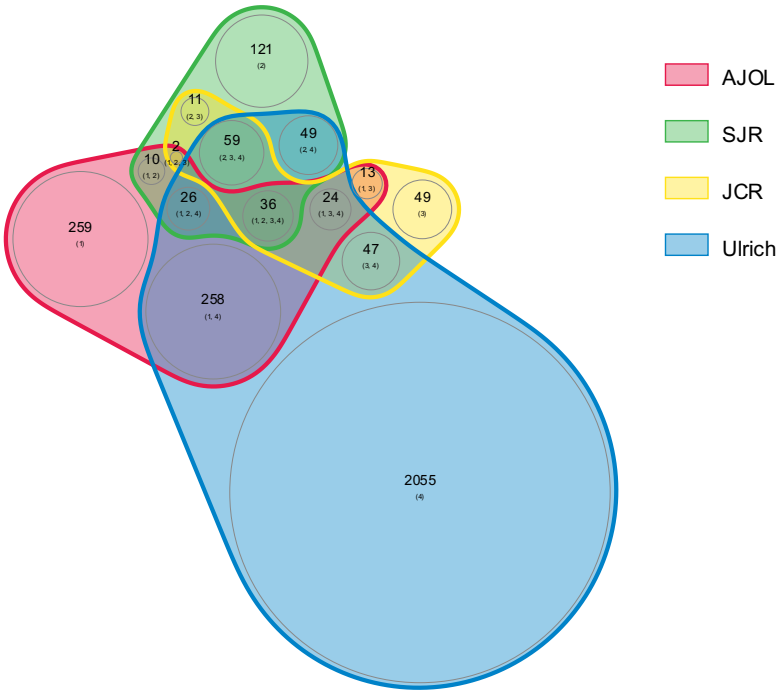

Figure 1. Euler diagram of African journals coverage in AJOL, SJR, JCR, and Ulrich

Figure 1 shows that Ulrich covers most African-based journals and a high number of unique journals (2,055). However, all AJOL, SJR, and JCR include journals not found in Ulrich (284, 144, and 75, respectively). After Ulrich, AJOL is the database with more unique journals (259) and also the second database with the highest total coverage (628). The Figure shows that none of the databases covers all African-based journals, and all are needed to improve the coverage of the dataset. This finding also indicates that it is likely that many journals do not appear in any of the databases, so results have to be interpreted with caution and consider that this is only a part of a broader landscape. JCR, SJR, and AJOL have quality criteria the journals have to fulfill to enter into the database so these along with geographical, language, and other biases could influence the journals they cover. The reason for Ulrich's limited coverage is not clear.
The four maps in Figure 2 show the total number of journals per African country (the biggest map on the left) and AJOL, SJR, and JCR relative coverage of each African country. Ulrich was omitted as it covers almost all journals in the full dataset, so the relative coverage score does not show significant results. The relative coverage is calculated as the proportion between the number of journals belonging to a country covered by a database and the total number of journals belonging to that country covered by the full database. If *c* is a specific country and *Db* is the database we are analyzing, the relative coverage of the country by the database would be defined by the equation:

$$Relative\ coverage_{c,Db} = \frac{Number\ of\ journals_{c,Db}}{Number\ of\ journals_{Db}} \Big/ \frac{Number\ of\ journals_{c,full\ Db}}{Number\ of\ journals_{full\ Db}}$$

Thus, if the relative coverage is between 0 and 1, the country is underrepresented in the specific database when we compare it to the full set. If the relative coverage is above 1, the country is overrepresented compared to the full set. In some cases, there are countries that are highly overrepresented in one database (for instance, South Sudan in the case of AJOL) because there is only one journal from that specific country, and it is included in only one database. It is thus useful to consider both the absolute and relative numbers when doing the analysis.

Overall, the countries hosting most journals are by far Nigeria (1,158), South Africa (550), and Egypt (491). On the other hand, some countries do not appear in any of the databases[1]. AJOL has a relatively high coverage of Eastern, Western and Southern regions, but the coverage of Northern Africa is relatively low. In comparison, SJR and JCR have a relatively high coverage of Northern and South Africa. Especially for JCR, the other countries are very close to 0.

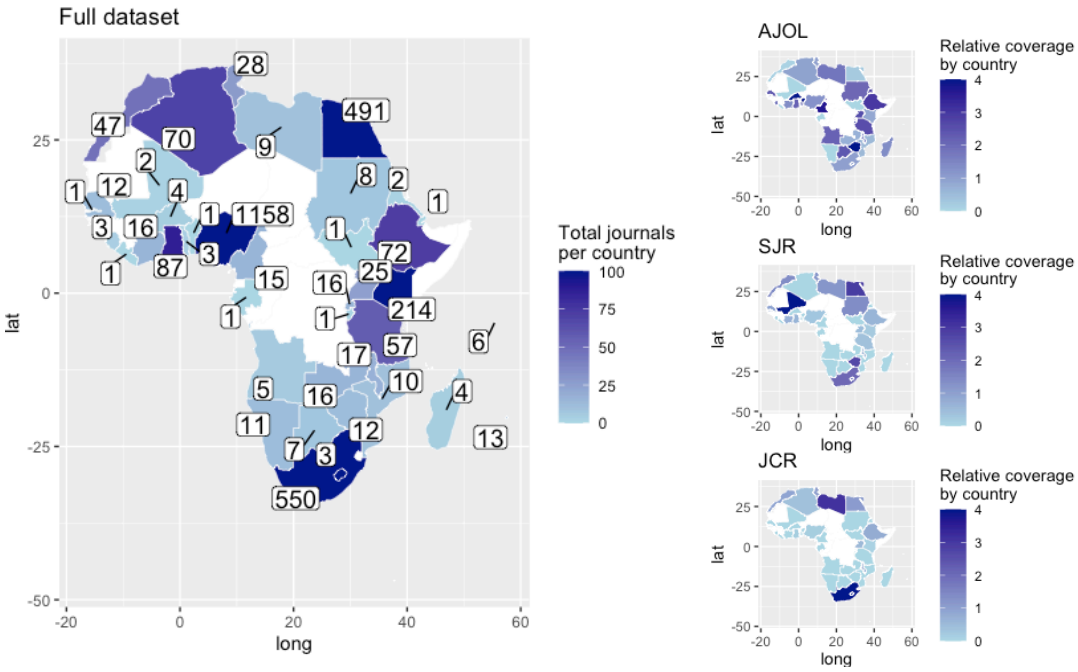

Figure 2. Number of journals per country (left) and relative coverage of African countries by AJOL, SJR, and JCR (right). White spaces represent countries not covered by the datasets.

Figure 3 shows the total number of journals per FOS research area (upper left corner) and the relative coverage of AJOL, SJR, and JCR. Again, Ulrich was not included in the visualization, as all the values were approximately 1. The solid black line shows the 1, so areas above the line are relatively overrepresented, and those below the line are underrepresented. Looking at the total numbers, the majority of the journals belong to Social sciences (1,037), followed by Natural sciences (732) and Medical and health sciences (748). Engineering and technology is the research area with the lowest number of journals (236). The relative coverage plots show that Medical and health sciences are overrepresented in all databases, especially in SJR. Regarding the rest of the areas, some differences arise. For instance, Engineering and technology is highly underrepresented in the case of AJOL but overrepresented in SJR and JCR. The plots show that SJR is generally biased towards more technical disciplines. Surprisingly, JCR's coverage scores are all very close to or above one, which means that its area coverage is similar to the coverage of the full dataset. JCR is also the only database that is biased toward Humanities.

[1]Burundi, Cabo Verde, Central African Republic, Chad, Equatorial Guinea, Guinea, Guinea-Bissau, and Sao Tomé and Principe.

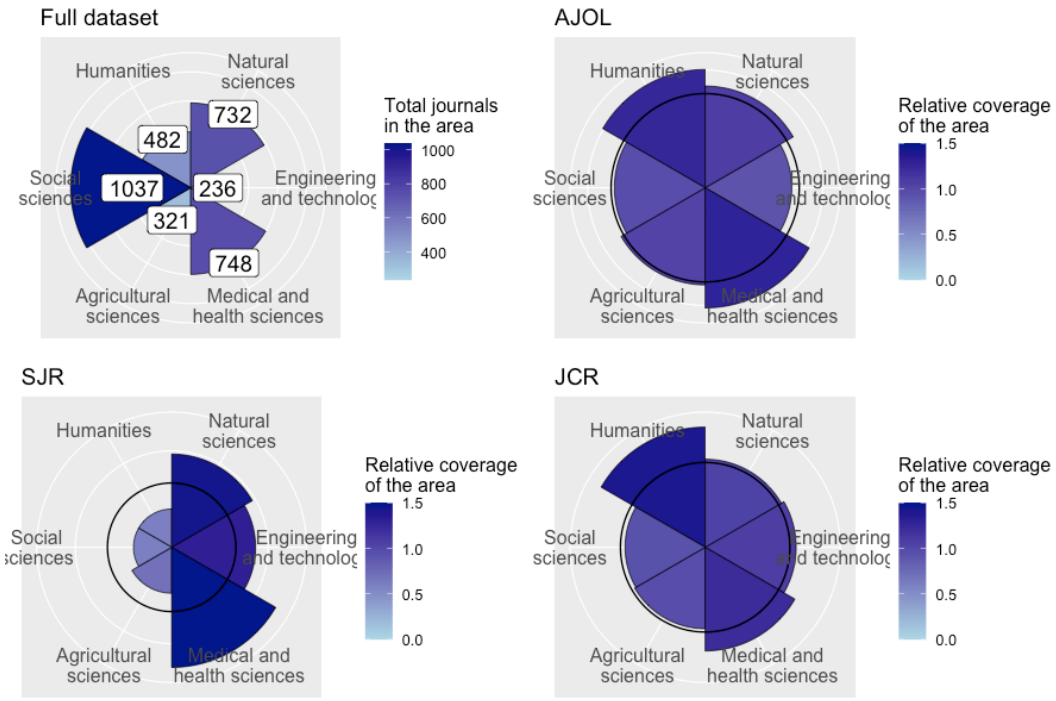

Figure 3. Number of journals by FOS research areas (first plot) and relative coverage in AJOL, SJR, and JCR

Discussion

The analysis contributes to current research on the biases and limitations of academic databases, mainly when covering peripherical world regions. These differences relate to the number of journals included in a specific database and their content in research areas and regional coverage. This has several implications. First, it diminishes the importance of research in peripherical regions. Often, knowledge produced in the Global South is defined as 'too local' to be included in international journals, while this does not happen with localized research in central regions –i.e., Europe and North America– (for instance, Castro-Torres and Alburez-Gutierrez, 2021; and Mongeon et al., 2022). Articles published in international databases are also perceived as higher-quality research by scholars (Chavarro et al, 2017). Finally, academic research not included in international databases often lacks visibility (Tijssen et al., 2006). The second main implication is that coverage biases, especially in mainstream databases, also affect social and political debates, forcing scholars to decide between publishing in international journals that provide recognition and visibility or investigating local problems.

This research in progress paper contributes to the current academic literature on this area and shows the limitations of mainstream databases when covering some regions, especially in the Global South. This ongoing project aims to understand better the role of the different 'circuits' (Beigel, 2014) of knowledge and how they relate and interact. Next steps aim to include other variables that might also offer clues on the coverage of mainstream and non-mainstream databases, such as publication language and journal quality. Further analysis will include publication-level research topics analysis, publication recognition, and visibility.

References

African Journals Online (n.d.). [Main page]. https://www.ajol.info/index.php/ajol

Archambault, É., Campbell, D., Gingras, Y., & Larivière, V. (2009). Comparing bibliometric statistics obtained from the Web of Science and Scopus. *Journal of the American society for information science and technology*, *60*(7), 1320-1326.

Becerra, G. (2022). *Crawler and Data Scraper for Open Journal System ('OJS').* https://CRAN.R-project.org/package=ojsr

Beigel, F. (2014). Publishing from the periphery: Structural heterogeneity and segmented circuits. The evaluation of scientific publications for tenure in Argentina's CONICET. Current sociology, 62(5).

Castro Torres, A. F., & Alburez-Gutierrez, D. (2022). North and South: Naming practices and the hidden dimension of global disparities in knowledge production. *Proceedings of the National Academy of Sciences*, *119*(10), e2119373119.

Chavarro, D., Tang, P., & Ràfols, I. (2017). Why researchers publish in non-mainstream journals: Training, knowledge bridging, and gap filling. *Research policy*, *46*(9), 1666-1680.

Chavarro, D., Ràfols, I., & Tang, P. (2018). To what extent is inclusion in the Web of Science an indicator of journal 'quality'?. *Research evaluation*, *27*(2), 106-118.

Coburn, J., Yaqub, O., & Chataway, J. (2022). Targeting research to address societal needs: What can we learn from 30 years of targeting neglected diseases?.

Grimes, M., & Morris, S. E. (2006). Is accuracy everything? A study of two serials directories. *Reference & User Services Quarterly*, 45-49.

Hanafi, S. (2011). University systems in the Arab East: Publish globally and perish locally vs publish locally and perish globally. *Current Sociology*, *59*(3), 291-309.

Khanna, S., Ball, J., Alperin, J. P., & Willinsky, J. (2022). Recalibrating the Scope of Scholarly Publishing: A Modest Step in a Vast Decolonization Process. SciELO Preprints

Kreimer, P., & Zabala, J. P. (2007). Chagas disease in Argentina: Reciprocal construction of social and scientific problems. *Science, Technology and Society*, *12*(1), 49-72.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, *106*(1), 213-228.

Mongeon, Philippe, Paul-Hus, Adèle, Henkel, Maria, & Larivière, Vincent. (2022, September 7). On the impact of geo-contextualized and local research in the global North and South. *26th International Conference on Science, Technology and Innovation Indicators (STI 2022*), Granada, Spain. https://doi.org/10.5281/zenodo.6956978

OECD (2007). *Revised Field of Science and Technology (FOS) Classification in the Frascati Manual.*

Perez-Silva J, Araujo-Voces M, Quesada V (2018). "nVenn: generalized, quasi-proportional Venn and Euler diagrams." *Bioinformatics*, **34**(13), 2322-2324.

Rafols, I., Ciarli, T., & Chavarro, D. (2015). Under-reporting research relevant to local needs in the global south. Database biases in the representation of knowledge on rice.

Tijssen, R. J., Mouton, J., Van Leeuwen, T. N., & Boshoff, N. (2006). How relevant are local scholarly journals in global science? A case study of South Africa. *Research evaluation*, *15*(3)

Wang, Y., Hu, R., & Liu, M. (2017). The geotemporal demographics of academic journals from 1950 to 2013 according to Ulrich's database. *Journal of Informetrics*, *11*(3), 655-671.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wickham H (2022). *rvest: Easily Harvest (Scrape) Web Pages*. https://rvest.tidyverse.org/, https://github.com/tidyverse/rvest