Presence and visibility of local journals in international scientific databases

Victoria Di Césare[1]* (ORCID: 0000-0001-7098-1914) and

Nicolas Robinson-Garcia[1] (ORCID: 0000-0002-0585-7359)

[1]Unit for Computational Humanities and Social Sciences (U-CHASS), University of Granada (Spain)

*Corresponding author: vdicesare@ugr.es

Abstract

Mainstream databases present significant language, geographical, disciplinary, and thematic biases, particularly against research addressing local needs. Evaluation policies framed around mainstream-based indicators tend to reproduce these biases and, consequently, underestimate local research. We address this issue empirically by merging four databases and conducting a large-scale analysis of local journals across mainstream (WOS and Scopus) and non-mainstream (OpenAlex and DOAJ) knowledge circuits. We aim to (1) determine the representation of local journals in the four databases, (2) assess the visibility of the fields and countries publishing in them, and (3) examine their non-English and open access features. We build a dataset of more than 75,000 OpenAlex journals, which are enriched with WOS, Scopus, and DOAJ variables and characterized as local or global according to a local research framework. Results show that most local journals are poorly represented in the mainstream, regardless of their field, language, or access type. Countries publishing in non-English and open access local journals do not follow clear patterns when examined from different locality perspectives. One conclusion is straightforward: local journals circulate primarily outside the mainstream. From a policy perspective, local research should not be discussed and evaluated solely from the content covered by WOS and Scopus.

Keywords: local journals, scientific databases, visibility, local research, scientometrics.

Introduction

Evaluation policies were introduced to seek greater control and accountability over publicly funded science. Governments around the world developed national systems that, in the pursuit of research excellence, allocate funding based on performance indicators (Hicks, 2012). The measures employed by such systems largely rest on publication quantity, impact, and visibility in Web of Science and Scopus, which best suit powerful and expensive fields like the natural and physical sciences from OECD countries (Hicks, 2013). For other, less powerful fields and regions, this means adapting to a competition regime that rewards publishing in elite English-language journals typically indexed in the mainstream (Chavarro et al., 2017; Vessuri et al., 2014). However, evidence suggests that mainstream databases present significant

1

language, disciplinary, and geographical biases (e.g., Archambault et al., 2006; Chavarro et al., 2018; Moed et al., 1995; Seglen, 1997; van Leeuwen et al., 2001). More importantly, they have been shown to reproduce substantial thematic bias against research addressing local needs in developing countries (Rafols et al., 2015).

In spite of this, many policy instruments rely on mainstream databases and their English-language journals to assess quality and performance, often through quantitative approaches that disadvantage these local perspectives and topics (Ràfols et al., 2016). A notable case is that of China's evaluation system which, until recently, was steered by indicators based on the Science Citation Index, increasing international visibility at the expense of local language and venues (Shu et al., 2022). In Spain, the Impact Factor was established as a quality indicator for allocating individual economic rewards, although it generates negative effects on researchers' publication practices and agendas (Delgado-López-Cózar et al., 2021). Other examples of this mainstream-driven assessment style can be found in countries like Argentina (Beigel et al., 2018), Chile (Gibert Galassi & Espina Bocic, 2022), Indonesia (Fry et al., 2023), Mexico (Sandoval-Romero & Larivière, 2020), Russia (Kosyakov & Guskov, 2022), and South Africa (Tomaselli, 2018). Researchers affiliated there face strong incentives to publish in mainstream journals in order to gain academic prestige and financial rewards that, ultimately, allow them to remain in the system. In sum, evaluation policies framed around mainstream-based indicators tend to reproduce their biases and, thus, underestimate local research capabilities and agendas (Rafols et al., 2015). However, placing value on local research can help foster more contextualized and effective policies, grounded in topics and needs that are often overlooked by mainstream interests and approaches (Marginson & Xu, 2021; Stiglitz, 1999).

In this article, we contribute to the matter from an empirical perspective by conducting a large-scale analysis of local journals across various databases. Since journals reflect the cognitive and social structure of science (Leydesdorff, 1987), we employ them as our object of study to address the local from three different conceptualizations (Di Césare & Robinson-Garcia, 2026b):

(1) Territory, where locality is tied to a site-specific object of study. The journals that comply with this definition are identified through the presence of toponyms in them and labelled as locally situated.

(2) Producers, where locality is grounded on their unique expertise and knowledge base. The journals classified under this conceptualization present geographically concentrated references and are labelled as locally informed.

(3) Recipients, where locality is shaped by the specific audiences it addresses or is relevant to. The journals that respond to this definition receive geographically concentrated citations and are labelled as locally relevant.

We analyze these local journals across mainstream and non-mainstream databases, which embody contrasting representations of science due to their distinct, partial selection criteria (Mongeon & Paul-Hus,

2016). In the first group, we include WOS and Scopus because, despite their biases, they are widely regarded as the dominant circuits for global scientific communication. In the second one, we group OpenAlex and DOAJ, which are two major initiatives dedicated to providing open access to research products worldwide. They have been extensively investigated over the years, in terms of content coverage (De Moya-Anegón et al., 2007; Visser et al., 2021), citation (Martín-Martín et al., 2018), reference (Culbert et al., 2025), language (Céspedes et al., 2025; Vera-Baceta et al., 2019), geographic (Asubiaro et al., 2024; Singh et al., 2021), and open access patterns (Simard et al., 2025). However, much remains to be understood about the presence of local journals across mainstream and non-mainstream databases. On the one hand, features such as language and type of access may influence the inclusion of local journals in these data infrastructures. On the other hand, local journals' indexing status might shape the visibility of knowledge fields and countries whose scientific output is largely published in them. Therefore, our aim is threefold:

(1) To create a merged dataset of four databases (WOS, Scopus, DOAJ, and OpenAlex) that incorporates locality measures to determine the representation of local journals across the mainstream (meaning WOS and Scopus) and non-mainstream (meaning DOAJ and OpenAlex).

(2) To assess the visibility of knowledge fields and countries that publish in local journals within mainstream and non-mainstream knowledge circuits.

(3) To examine the non-English and open access features of local journals across mainstream and non-mainstream databases.

Methodology

We built a large dataset based on 75,694 OpenAlex journals with articles published throughout 2023. This data was extracted from the October 2024 public database stored in Google BigQuery, which is provided by the InSySPo program (Mazoni & Costas, 2024). It was further enriched with Web of Science Master Journal List (MJL), Journal Citation Reports (JCR), Scopus, Scimago Journal & Country Rank (SJR), CWTS Journal Indicators (CJI), and Directory of Open Access Journals (DOAJ) variables. As detailed in Table 1, we selected these additional data sources because of the added value they provide to our existing journal set, both in terms of their production characteristics (e.g., organizations involved, places of origin, licenses, policies) and constructed measures (e.g., citation counts, impact indicators, quartile positioning). When such information was already available in OpenAlex, these other sources allowed for estimating its accuracy and completeness. Moreover, they helped us establish the distinction between mainstream and non-mainstream journals, which was central to addressing our research goals.

Table 1. Selected data sources with total journal counts and provided information.

Data source	Acronym	Total journals	Provided information
OpenAlex	-	75,694	Baseline dataset containing journals' bibliographic data, such as ISSN codes, title, title variants, publisher, country, language, thematic classification, access type, APCs, website.
Web of Science Master Journal List	MJL	22,455	Journals' bibliographic data, such as ISSN codes, title, publisher, country, language, and thematic classification.
Journal Citation Reports	JCR	21,988	Journals' bibliographic data, such as ISSN codes, title, title variants, publisher, thematic classification, edition. Journals' bibliometric indicators, such as JCI, JIF, rank, quartile, percentile, citations, AIS, Immediacy Index, Eigenfactor, etc.
Scopus	-	29,221	Journals' bibliographic data, such as ISSN codes, title, title variants, publisher, thematic classification, language, coverage, access type, Medline, etc.
Scimago Journal & Country Rank	SJR	28,174	Journals' bibliographic data, such as ISSN codes, title, publisher, country, region, thematic classification, coverage. Journals' bibliometric indicators, such as rank, quartile, h index, references per article, citations, citable article, percent female, SDG, Overton, etc.
CWTS Journal Indicators	CJI	27,879	Journals' bibliographic data, such as ISSN codes, title. Journals' bibliometric indicators, such as SNIP, IPP, and self-citations.
Directory of Open Access Journals	DOAJ	20,955	Journals' bibliographic data, such as ISSN codes, title, title variants, publisher, country, organization, language, thematic classification, access type, APCs, licenses, policies, review process, website, etc.

We developed a three-step workflow to merge the six data sources into OpenAlex journals (Figure 1): (1) matching via ISSN codes, (2) matching by journal name, and (3) matching via articles' DOI. In the first step, we were able to match, on average, 86% of all journals included in the six data sources. An extra 0.75% on average was added in the second step via journal name matching, while 1,000 other journals were matched through articles' DOI in the last step. The final dataset includes the original 75,694 journals from OpenAlex, out of which 35,989 (48%) have enriched information extracted from MJL, JCR, Scopus, SJR, CJI, and DOAJ.
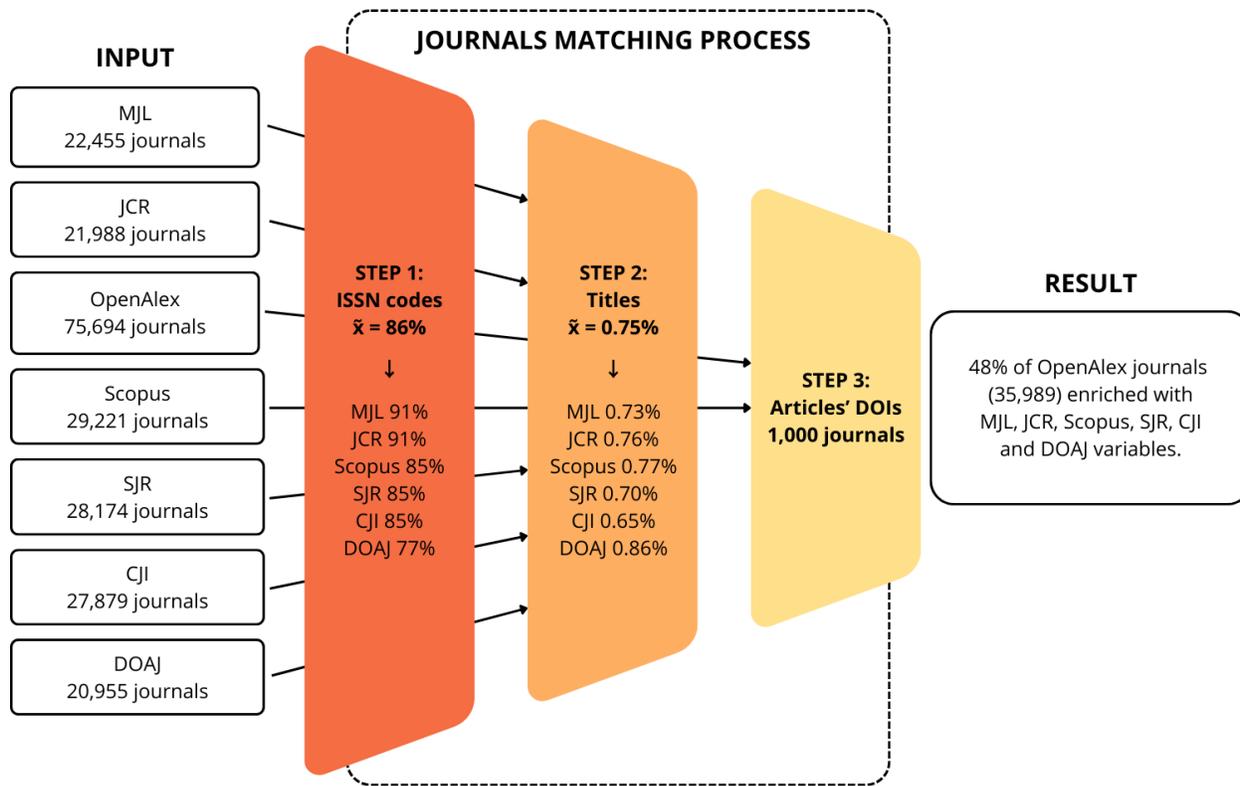
Figure 1. Workflow of the data sources merging process.

Table 2. Selected variables employed in this analysis and extra variables included in the dataset.

Variable	Description	Data source
Employed in this analysis		
OA_source_ID	IDs assigned to all journals within the database. It allowed us to uniquely identify each journal.	OpenAlex
other_IDs	Nested variable with IDs assigned to every MJL, JCR, Scopus, SJR, CJI, and DOAJ journal within this dataset.	Assigned by the authors
ISSN_codes	Nested variable with the print and electronic ISSN codes assigned to each journal. It allowed us to compute the first step of our merging process, which consisted of matching journals through their ISSN codes.	All
journal_name	Nested variable with each journal's title. It allowed us to compute the second step of our merging process, which consisted of matching journals through their titles.	All
OA_domains	Knowledge domains assigned to all journals within the database. It allowed us to produce the results by field.	OpenAlex
language	Nested variable with the languages detected for each journal. It allowed us to classify the journals into English or non-English publishing languages.	OpenAlex, MJL, Scopus, DOAJ

open_access	Nested variable with each journal's type of access. It allowed us to classify the journals into open or closed access.	OpenAlex, Scopus, DOAJ
mains	Whether the journal is indexed in databases other than WOS or Scopus. It allowed us to classify the journals into mainstream or non-mainstream indexing status.	MJL, Scopus
pubs_prop	Proportion of appearance of all publishing countries per journal according to the geographical affiliation of the publishing authors. It allowed us to compute the 2023 national output to produce the results by country.	Di Césare and Robinson-Garcia (2026b)
refs_prop	Proportion of appearance of the maximum referenced country per journal according to the geographical affiliation of the referenced authors. It allowed us to identify the local journals according to the producers conceptualization.	Di Césare and Robinson-Garcia (2026b)
cits_prop	Proportion of appearance of the maximum citing country per journal according to the geographical affiliation of the citing authors. It allowed us to identify the local journals according to the recipients conceptualization.	Di Césare and Robinson-Garcia (2026b)
tops_prop	Proportion of appearance of country-level toponyms mentioned in the articles' titles per journal, selected in the most preponderant languages (95%) of the dataset: English, Indonesian, Spanish, Portuguese, German, French, Arabic, Ukrainian, Turkish, and Russian. It allowed us to identify the local journals according to the territory conceptualization.	Di Césare and Robinson-Garcia (2026b)

Other journal information data

journal_name_variants	Nested variable with journals' title variants.	OpenAlex, JCR, Scopus, DOAJ
publisher, organization	Publishers and organizations involved in the journals' production.	OpenAlex, MJL, JCR, Scopus, SJR, DOAJ
country, region	National and regional origin of the journals.	OpenAlex, MJL, SJR, DOAJ
topics, fields, categories, areas, subjects, keywords	Thematic classification of the journals at different aggregation levels.	OpenAlex, MJL, JCR, Scopus, SJR, DOAJ
coverage	Nested variable with the time period covered by the database for each journal.	Scopus, SJR
licenses, policies	Usage licenses and policies related to deposit, plagiarism, exceptions, etc.	DOAJ
review	Characteristics of the review process and its duration.	DOAJ

APCs	Article Publication Charges and other fees.	OpenAlex, DOAJ
persistent identifiers	Usage and typology of persistent identifiers.	DOAJ
website	Link to each journal's website.	OpenAlex, DOAJ

Journal metrics		
JCI, JIF rank, quartile, percentile	2023 value, rank, quartile, and percentile according to the Journal Citation Indicator and Journal Impact Factor.	JCR
AIS	2023 value and rank by the Article Influence Score.	JCR
eigenfactor	Score and normalized score by the eigenfactor metric.	JCR
SJR rank, best quartile	Value, rank, and best quartile according to the Scimago Journal Rank.	SJR
articles	2023, last 3 years, and overall total articles.	OpenAlex, JCR, SJR
citations	Values related to total citations, citing and cited half-life, percent citable items and open access, self-citations, policy citations via Overton, etc.	OpenAlex, JCR, SJR, CJI
references	Overall total and per article references.	SJR
SNIP, IPP	Values related to the Source Normalized Impact per Paper and Impact per Publication.	CJI
percent female	Percentage of women in each journal's editorial board or authorship.	SJR
SDG	Number of articles per journal related to the UN's Sustainable Development Goals.	SJR

The resulting dataset includes more than 100 variables that provide information about the journals and their bibliometric indicators. We also added six measures of locality computed on the basis of the detected toponyms, affiliations, references, citations, languages, and databases indexing based on the framework developed by Di Césare & Robinson-Garcia (2026b). When such information was present in any of the mentioned sources, we expanded the journals' characterization to reflect their local profile from multiple, complementary perspectives. As a result, more than 90% of OpenAlex journals are now depicted as either local or global from at least one viewpoint. In this analysis, we worked with the variables described in Table 2, which are presented alongside additional information and indicators for each journal.

The R, Python, and SQL code scripts developed to obtain the data, merge the sources, compute the local variables, and produce the results can be found in Di Césare (2025). The final merged dataset can be downloaded from Di Césare & Robinson-Garcia (2026a).

Results

Next, we present a brief characterization of our merged dataset, alongside the results of our local journals analysis. It is conducted at both the field and country levels to assess their visibility across mainstream and non-mainstream knowledge circuits. We also examine the role of language and type of access in order to better understand how non-English and open access publishing interact with the mainstream or non-mainstream indexing status of local journals. The non-English threshold was applied in a strict way, meaning that any journal publishing in English alongside other languages was left out of the non-English subgroup. This section is structured along the above-mentioned conceptualizations of the local: territory, which represents locally situated journals; producers, that captures locally informed journals; and recipients, which groups locally relevant journals. The following results represent only a fraction of what could be analyzed using our large-scale dataset. Hence, we make it freely available for researchers wishing to explore it further in future studies (Di Césare & Robinson-Garcia, 2026a).

Overall dataset description

Figure 2 illustrates the journal overlap between all the data sources in our merged dataset. OpenAlex's set of journals is roughly 3 times bigger than any of the additional sources. MJL and JCR set sizes of around 22,000 journals are similar and present an overlap with OpenAlex of more than 90%. The same happens with Scopus, SJR, and CJI set sizes, of approximately 29,000 journals. However, they have a lower overlap with OpenAlex than that of WOS sources, of around 85%. DOAJ presents both the smallest set size, with just over 20,000 journals, and the least overlap with OpenAlex, at 78%. DOAJ's overlap is even lower with the rest of the data sources (33% on average), hence being the database that leaves the most journals out of the analysis. As WOS products, MJL and JCR have almost complete overlaps. The same happens between Scopus, SJR, and CJI. Therefore, from now onwards, we treat these five different sources as two unified databases, namely WOS and Scopus, which we analyze alongside DOAJ and OpenAlex. In addition to Figure 2, Table A1 (Appendix) details all data sources' set sizes and overlaps.
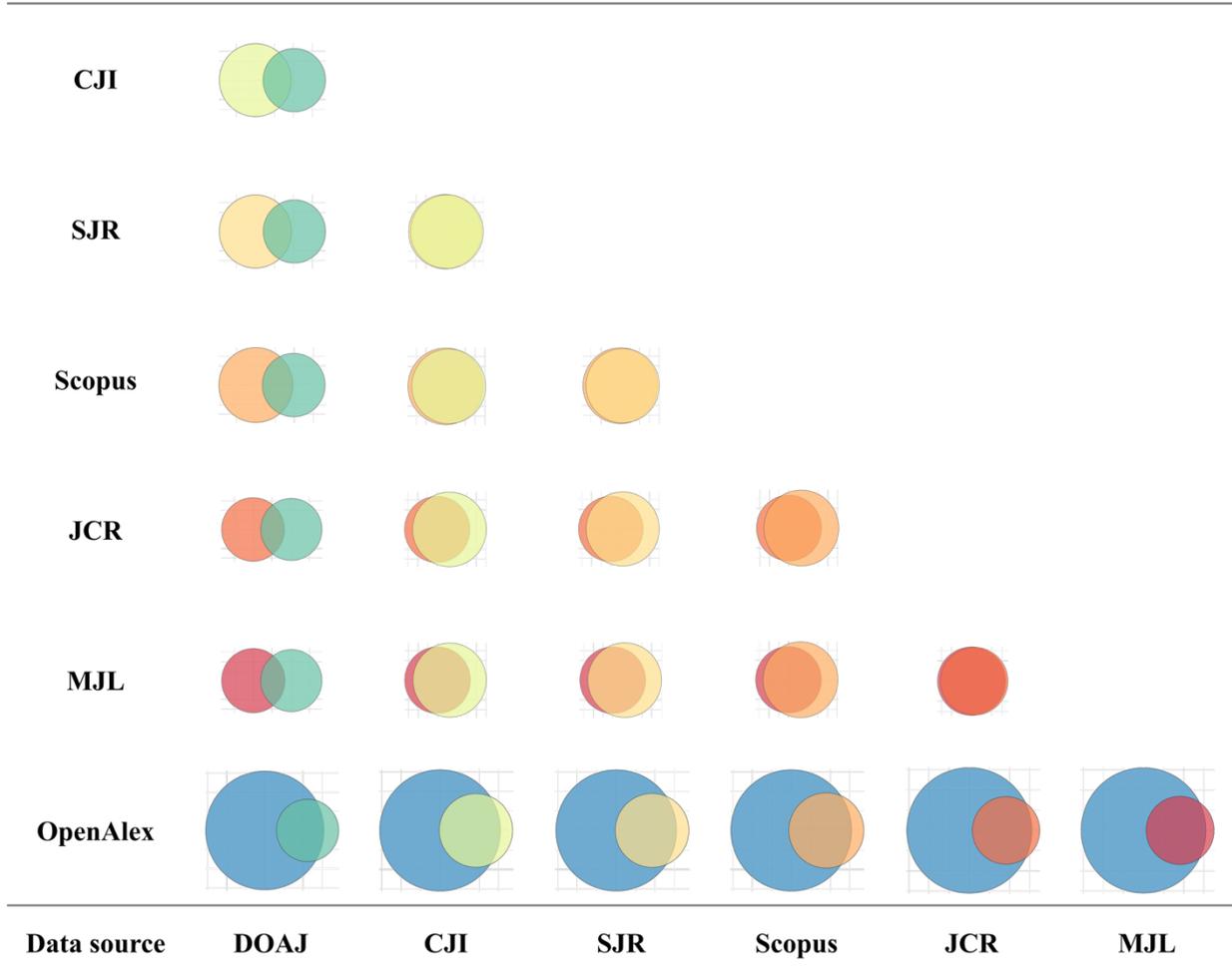
Figure 2. Journal overlap between data sources (OpenAlex, MJL, JCR, Scopus, SJR, CJI, and DOAJ) in our merged dataset.

Table 3 displays the number of journals per database and their distribution by field. It follows OpenAlex's domain characterization and is computed through fractional counting, since a given journal can belong to multiple fields. Table 3 also shows the number of locally situated, informed, and relevant journals, as well as the number of non-English language and open access journals, overall and by field. Scopus provides information on 25,333 journals (34%). They mainly belong to the Social (28%) and Physical Sciences (27%), but are still evenly distributed across all fields. WOS incorporates data on 20,474 journals (27%) with a distribution and ratio between fields similar to Scopus. DOAJ has the fewest journals in common with our main set (22%), but its field proportions resemble those of OpenAlex. Social Sciences journals are the most prevalent (30% on average) across all databases, followed by Physical ($\bar{x}$ = 28%), Health ($\bar{x}$ = 22%), and Life Sciences ($\bar{x}$ = 20%).

Locally situated journals stand out in DOAJ (32%), whereas the locally informed and relevant journals are more frequent in OpenAlex (22% and 20% respectively). Scopus and WOS exhibit a similar

coverage of local journals across the three conceptualizations, with the locally situated accumulating the largest amount ($\bar{x} = 25\%$) and the locally relevant the smallest ($\bar{x} = 4\%$). In terms of language, DOAJ (17%) and OpenAlex (15%) exhibit higher shares of non-English publications compared to Scopus (7%) and WOS (5%). Regarding access, open journals represent almost 40% of OpenAlex's content, and over a third of both Scopus (38%) and WOS's (35%) coverage. Throughout Table 3, the field distribution remains constant. Social Sciences journals are always at the top, except for Scopus and WOS open access journals, where Physical Sciences is the most frequent field.

Table 3. Overall number of journals per database and fractional counting by field.

Database	Field	Number of journals					
		Total	Locally situated	Locally informed	Locally relevant	Non-English	Open Access
OpenAlex	**Overall**	**75,694**	**19,766**	**16,872**	**15,004**	**11,236**	**30,012**
	Health Sciences	16,278	3,987	3,107	2,924	2,231	6,537
	Life Sciences	13,804	3,583	2,334	2,348	1,710	5,724
	Physical Sciences	21,357	5,459	4,801	4,370	3,045	8,462
	Social Sciences	24,255	6,737	6,630	5,362	4,250	9,289
DOAJ	**Overall**	**16,242**	**5,173**	**2,286**	**2,248**	**2,726**	**16,242**
	Health Sciences	3,509	1,049	398	419	542	3,509
	Life Sciences	3,154	997	274	342	447	3,153
	Physical Sciences	4,525	1,431	645	632	750	4,527
	Social Sciences	5,054	1,696	969	855	987	5,053
Scopus	**Overall**	**25,333**	**6,551**	**3,032**	**1,110**	**1,779**	**9,527**
	Health Sciences	5,821	1,398	671	211	362	2,202
	Life Sciences	5,470	1,347	523	187	397	2,123
	Physical Sciences	6,981	1,798	722	282	453	2,662

	Social Sciences	7,061	2,008	1,116	430	567	2,540
	Overall	**20,474**	**5,140**	**2,502**	**767**	**1,309**	**7,067**
WOS	Health Sciences	4,753	1110	568	141	234	1,665
	Life Sciences	4,467	1056	444	114	287	1,597
	Physical Sciences	5,618	1394	598	188	342	1,959
	Social Sciences	5,636	1580	892	324	446	1,846

Table 4 summarizes the share of mainstream and non-mainstream local journals relative to all journals in our dataset, by locality conceptualization and field. Locally situated journals present even values across fields regardless of their indexing status. On average, they represent 26% of the dataset, with mainstream Social Sciences local journals reaching the highest value (28%), and all Health Sciences local journals the lowest (25%). The locally informed and relevant journals display a clearer difference between the mainstream and non-mainstream. These mainstream journals are less present in the dataset, with average values at 11% and 4%, respectively. In both groups, the least represented mainstream local journals belong to the Life Sciences ($\bar{x} = 6$%), while the most represented local journals are non-mainstream Social Sciences ($\bar{x} = 30$%). Locally informed journals from the non-mainstream Social Sciences are the most substantial (30%), whereas locally relevant journals from the mainstream Life Sciences are the smallest group (3%).

Table 4. Percentage of mainstream and non-mainstream local journals relative to all journals, by locality conceptualization and field.

Field	Indexing status	% of journals		
		Locally situated	Locally informed	Locally relevant
Health Sciences	Mainstream	25%	11%	4%
	Non-mainstream	25%	24%	27%
Life Sciences	Mainstream	25%	9%	3%
	Non-mainstream	27%	23%	27%
Physical Sciences	Mainstream	26%	10%	4%
	Non-mainstream	26%	28%	29%
Social Sciences	Mainstream	28%	13%	5%
	Non-mainstream	27%	30%	29%

Locally situated journals

We further examine the locally situated journals, meaning journals where locality is tied to a site-specific object of study, measured through the presence of toponyms. Figure 3A shows the share of mainstream and non-mainstream local journals by field. On the left, the overall results indicate a higher prevalence of non-mainstream local journals across fields, particularly in the Social Sciences (64%). In contrast, Life Sciences exhibits the highest proportion of mainstream local journals (41%), although the shares do not vary substantially between fields. A similar mainstream-non-mainstream ratio is evident in non-English (middle bars) and open access (right-hand bars) local journals. Still, non-English journals have a higher average prevalence outside the mainstream ($\bar{x} = 68\%$). The highest share of non-mainstream local journals corresponds to the non-English Social Sciences (72%), whereas mainstream local journals are most present in the open access Life Sciences (42%).

Figure 3B displays the share of mainstream journals among the locally situated sources in which each country publishes, grouped by region. Africa ($\tilde{x} = 58\%$) presents the lowest median value of all, with countries like Chad (36%) and South Sudan (38%) having the least participation in the mainstream. In contrast, Oceania ($\tilde{x} = 87\%$) publishes its locally situated research almost exclusively in indexed journals. Although most median values are high (between 60% and 75%), all regions show considerable variability between countries, with some of the lowest shares observed in Indonesia (35%) and Turkmenistan (40%).
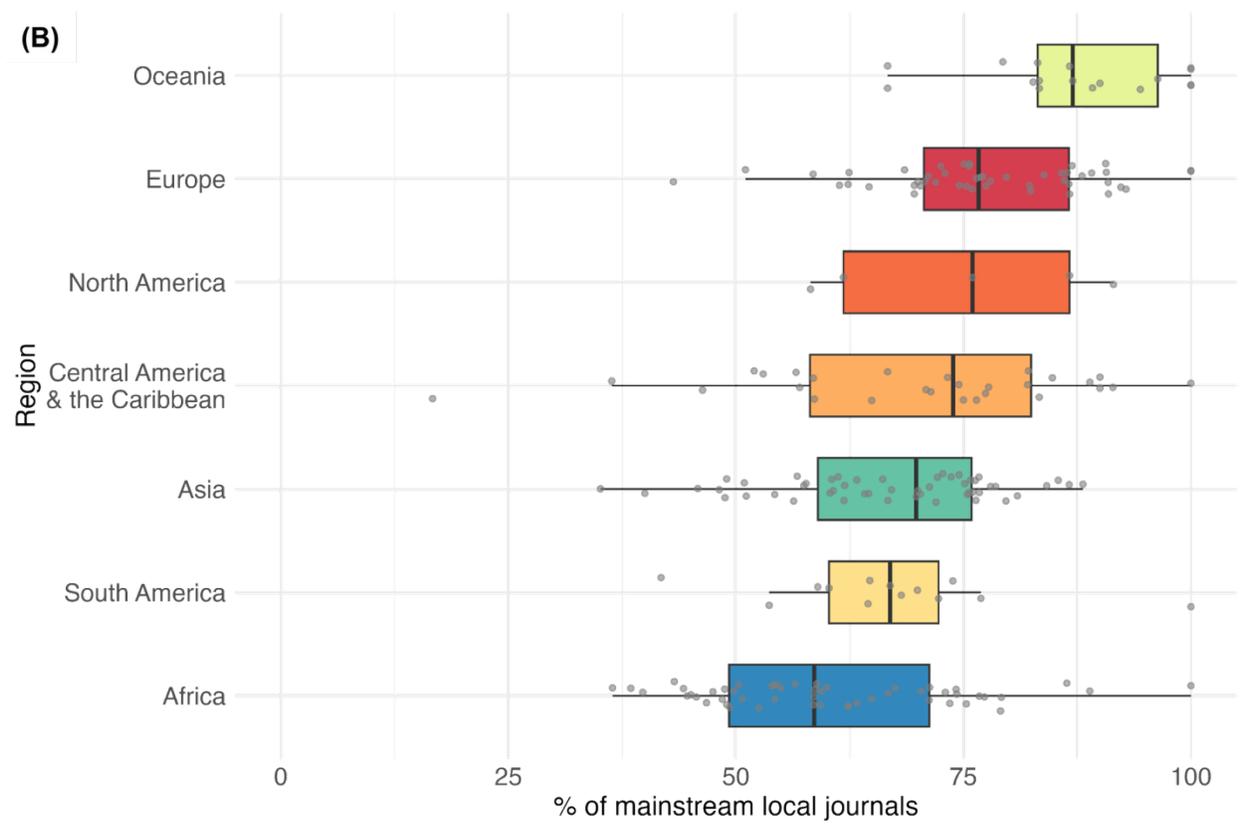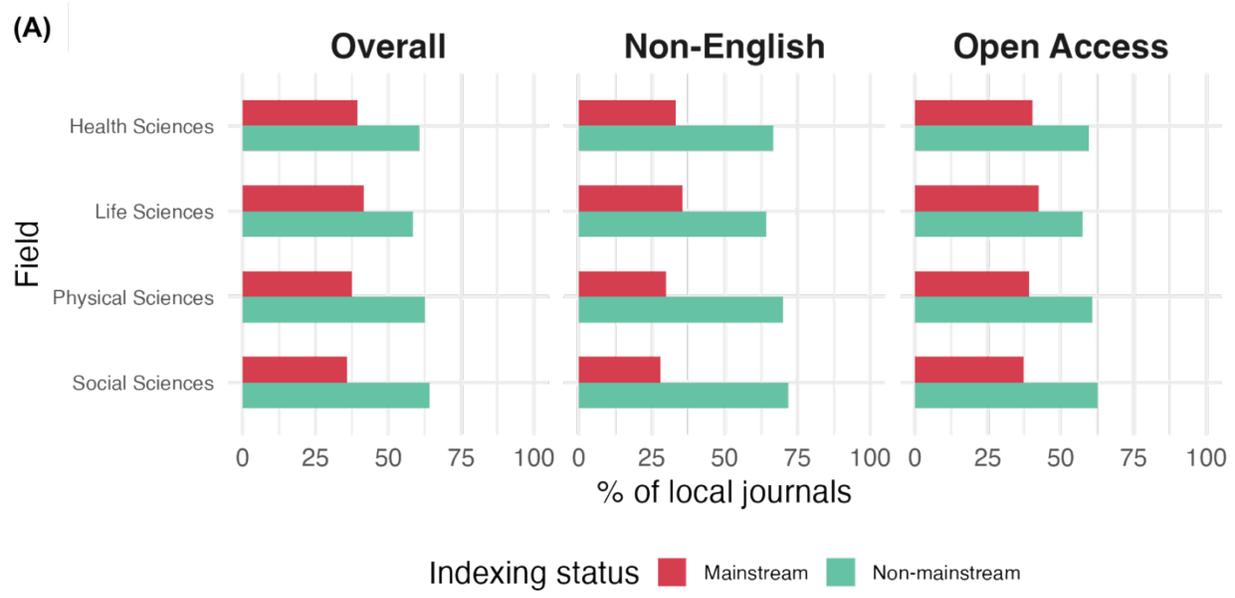
Figure 3. (A) Percentage of mainstream and non-mainstream locally situated journals by field. The left-hand bars represent all journals, while the middle and right-hand bars include only non-English and open access journals. (B) Percentage of mainstream indexing among the locally situated journals each country publishes in, by region.

Figure 4A shows the share of articles published in locally situated journals relative to each country's total output. Sub-Saharan Africa stands out, with particularly high levels of local publications in countries such as Sierra Leone (77%), Niger (73%), Somalia (72%), and the Central African Republic (70%). Figure 4B shows the share of articles published in mainstream locally situated journals by country. Here, the pattern reverses: most African countries publish their high shares of local research articles outside the mainstream. In contrast, countries like Finland (92%), Denmark (91%), the Netherlands (91%), and Australia (91%) publish 3 to 4 times less in locally situated journals, but they do so almost exclusively within the mainstream.

(A)

% of articles in local journals   25   50   75

(B)

% of articles in mainstream local journals   25   50   75   100
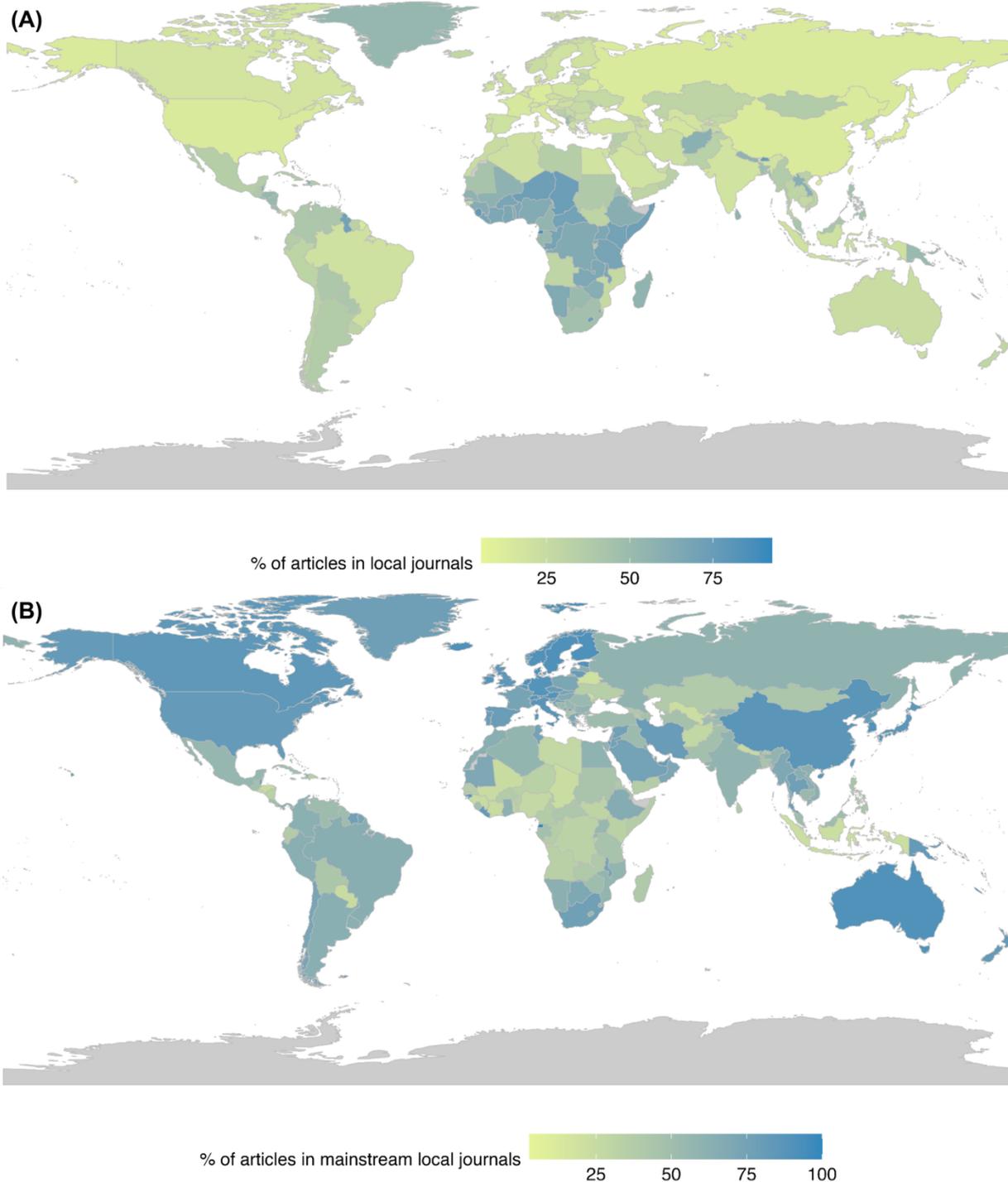
Figure 4. (A) Percentage of articles in locally situated journals by country, relative to their total output.

(B) Percentage of articles in mainstream locally situated journals by country.

In Figure 5, we characterize this mainstream local output based on language and access. Figure 5A shows the distribution by region, highlighting the case of South America. This region publishes on average 31% of its local research in non-English mainstream journals, which is the highest share of all. Central

America and the Caribbean follow closely with 22% on average. Asia displays the opposite scenario, with just 1% on average of its output in non-English mainstream journals and a tight distribution. In Figure 5B, South ($\tilde{x} = 66\%$) and Central America ($\tilde{x} = 60\%$) publish the most articles in open access journals. In fact, all regions but North America ($\tilde{x} = 44\%$) disseminate more than half of their locally situated articles through open access mainstream venues.
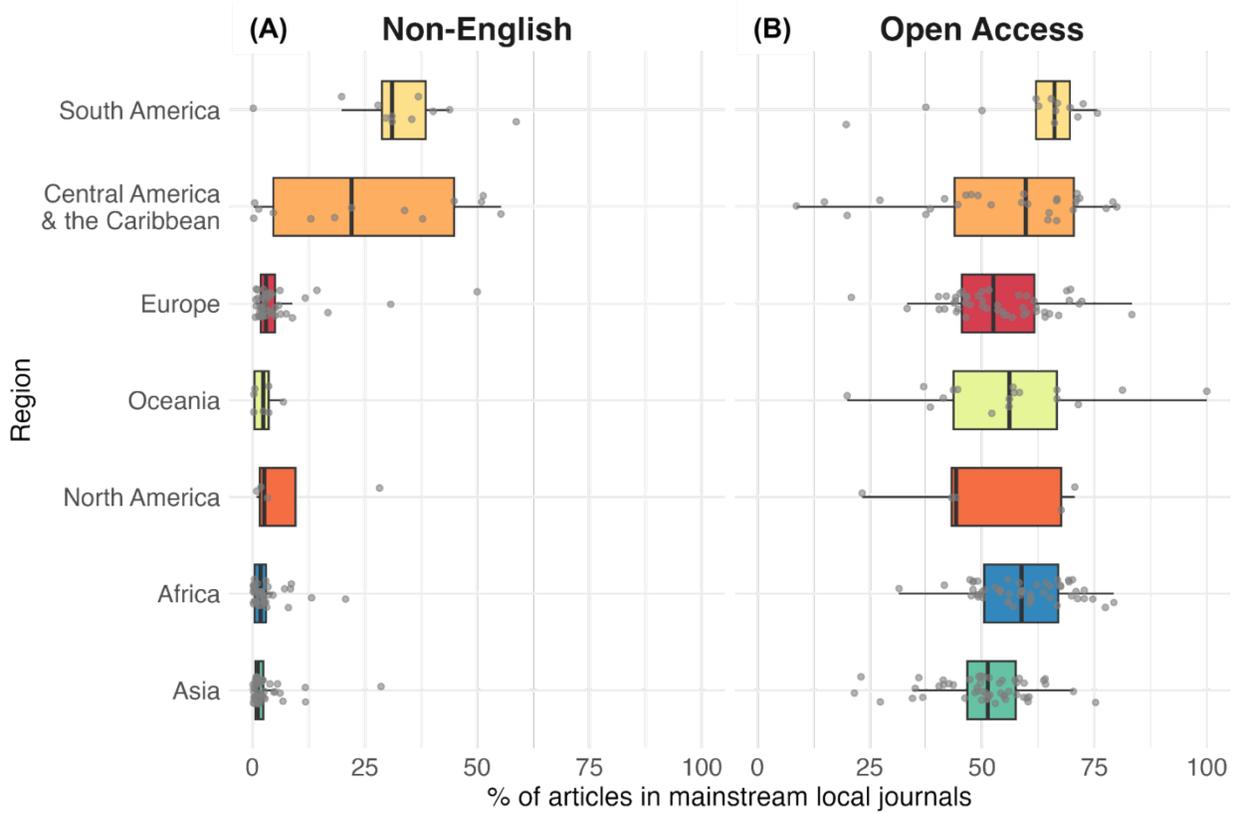

Figure 5. Percentage of articles in (A) non-English and (B) open access mainstream locally situated journals, by region.

Locally informed journals

Next, we delve into locally informed journals, that is journals where locality is grounded on their producers' unique expertise and knowledge base, measured through references. On the left of Figure 6A, there is a predominance of non-mainstream local journals across fields, particularly in the Social (81%) and Physical Sciences (82%). In contrast, mainstream local journals reach their highest share in Life Sciences (24%). Overall, proportions remain consistent across fields. The non-English (middle bars) and open access (right-hand bars) local journals have greater mainstream-non-mainstream differences. Non-mainstream indexing is 12 times more common among non-English journals and 6 times more prevalent among open

access journals. Almost all non-English ($\bar{x}$ = 92%) and open access ($\bar{x}$ = 86%) locally informed journals fall outside the mainstream. Non-English local journals from the Physical Sciences are the most common outside the mainstream (93%), whereas open access local journals from the Life Sciences are the most present in the mainstream (15%).

In Figure 6B, both Africa and South America have the lowest median shares of mainstream locally informed journals, at exactly half the distribution. All other regions publish in more mainstream local journals, with Oceania again being the most extreme case. However, Asia, Africa, and Europe present considerable variability at the country level, with most cases falling in the second and third quartiles. In North America, the lowest point in the distribution corresponds to the United States, as 50% of the local journals in which the US publishes are not indexed in WOS or Scopus.
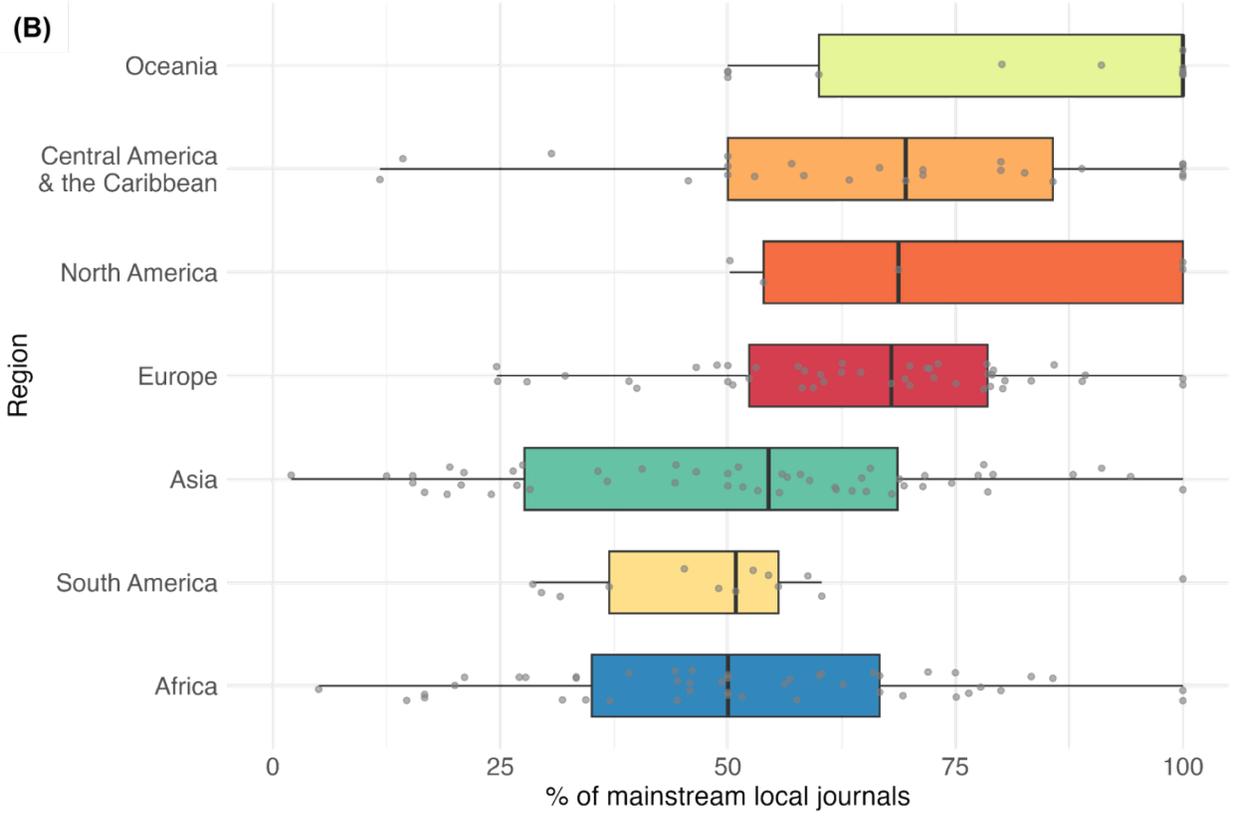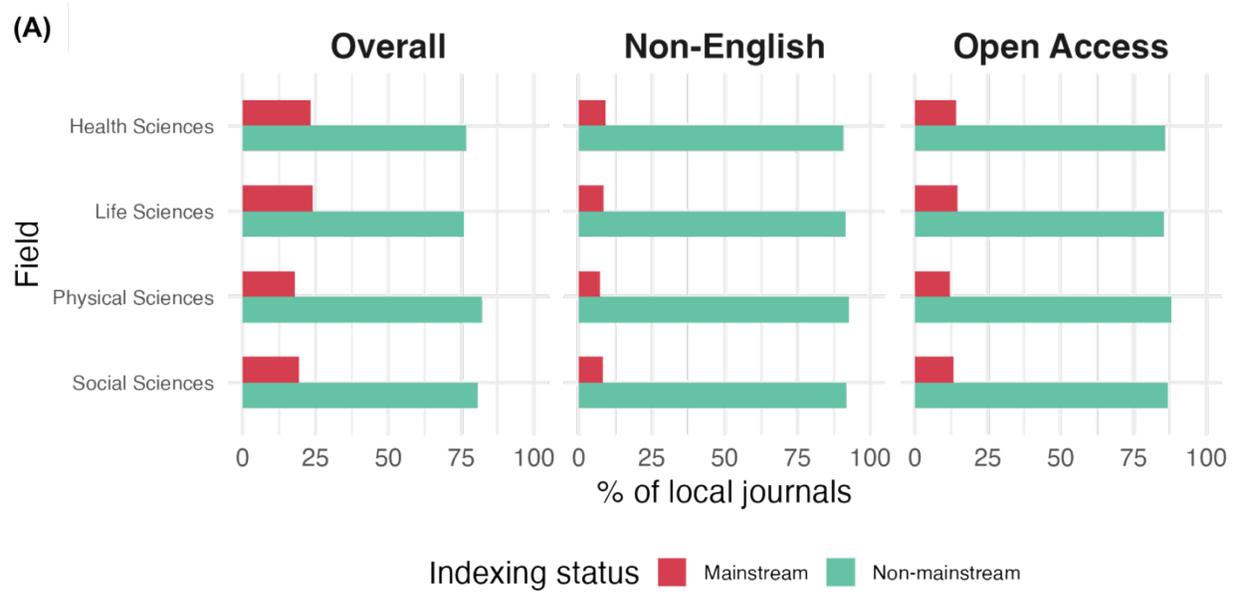
Figure 6. (A) Percentage of mainstream and non-mainstream locally informed journals by field. The left-hand bars represent all journals, while the middle and right-hand bars include only non-English and open access journals. (B) Percentage of mainstream indexing among the locally informed journals each country publishes in, by region.

18

Figure 7 shows the concentration of articles in locally informed journals. In Figure 7A, Indonesia (58%) stands out, with more than half of the country's articles published in local journals. The cases of the United States and Brazil are also notable, since they publish 15% of their considerable 2023 output in locally informed journals. Figure 7B displays a more mixed scenario. While a few countries from different regions show high proportions of local articles in the mainstream (e.g., Israel, Australia, the US, China, and South Africa, averaging 85%), there are widespread areas from South America, Africa, and Asia where non-mainstream dissemination of locally informed research predominates.

(A)

% of articles in local journals   10   20   30   40   50

(B)

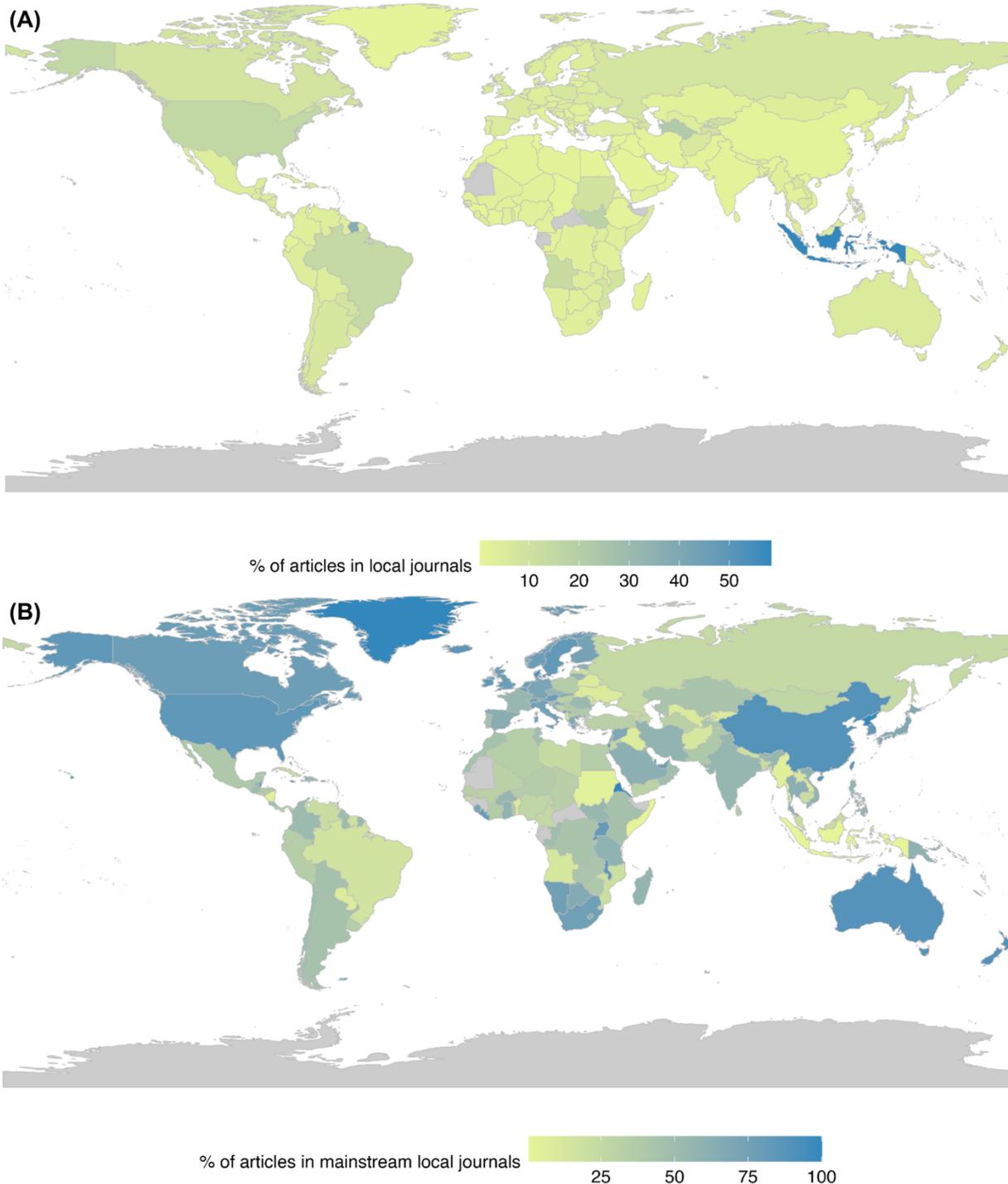% of articles in mainstream local journals   25   50   75   100

Figure 7. (A) Percentage of articles in locally informed journals by country, relative to their total output.

(B) Percentage of articles in mainstream locally informed journals by country.

Figure 8 focuses on the non-English (A) and open access (B) journals within the mainstream locally informed subgroup. In Figure 8A, South America ($\tilde{x}$ = 48%) stands out with almost half its articles published in non-English journals. The opposite happens in North America ($\tilde{x}$ = 4%) and Asia ($\tilde{x}$ = 5%). In

Figure 8B, South America ($\tilde{x}$ = 48%) showcases the highest median value of articles in mainstream open access journals, while Asia and Africa ($\tilde{x}$ = 28%) have the lowest. The other regions indicate that around a third of their articles are published in mainstream locally informed journals with an open type of access.
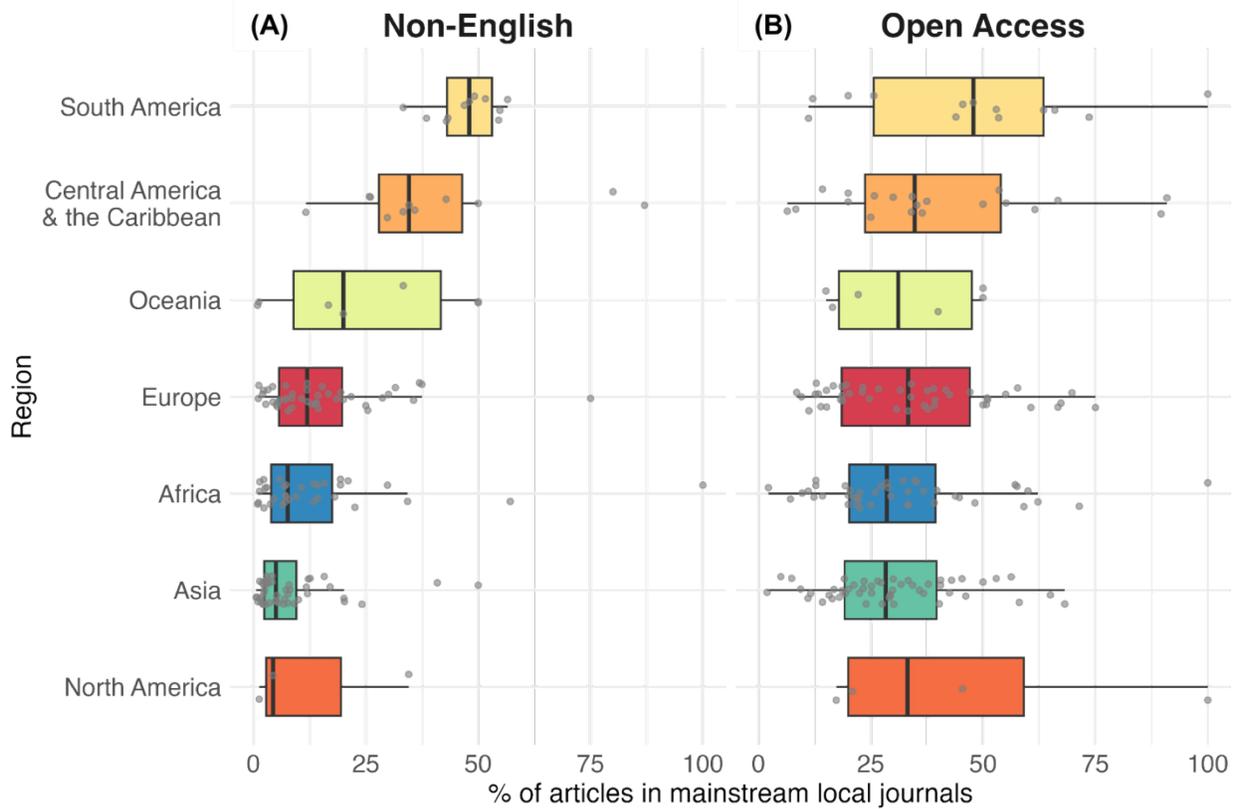

Figure 8. Percentage of articles in (A) non-English and (B) open access mainstream locally informed journals, by region.

Locally relevant journals

Finally, we report locally relevant journals, which are measured through the geographically concentrated citations they receive. On the left of Figure 9A, there is an overall dominance of non-mainstream ($\bar{x}$ = 92%) over mainstream ($\bar{x}$ = 8%) local journals across fields. The Physical Sciences display the lowest shares in the mainstream as well as the highest in the non-mainstream. This pattern at the field level is mirrored in the non-English (middle bars) and open access (right-hand bars) local journals. Within these subsets, non-mainstream local journals present a slight decline ($\bar{x}$ = 89%), yet they still predominate by around 7 times over the mainstream.

In Figure 9B, there are considerably low shares of mainstream indexing among the locally relevant journals where the countries publish. All regions reveal a tendency towards the non-mainstream,

21

particularly Africa ($\tilde{x} = 12\%$) and Asia ($\tilde{x} = 13\%$). In South and North America, large scientific producers such as Brazil (16%) and the United States (20%) have some of the lowest mainstream presence on the continent. The great majority of the locally relevant journals in which they publish are not indexed in WOS or Scopus.
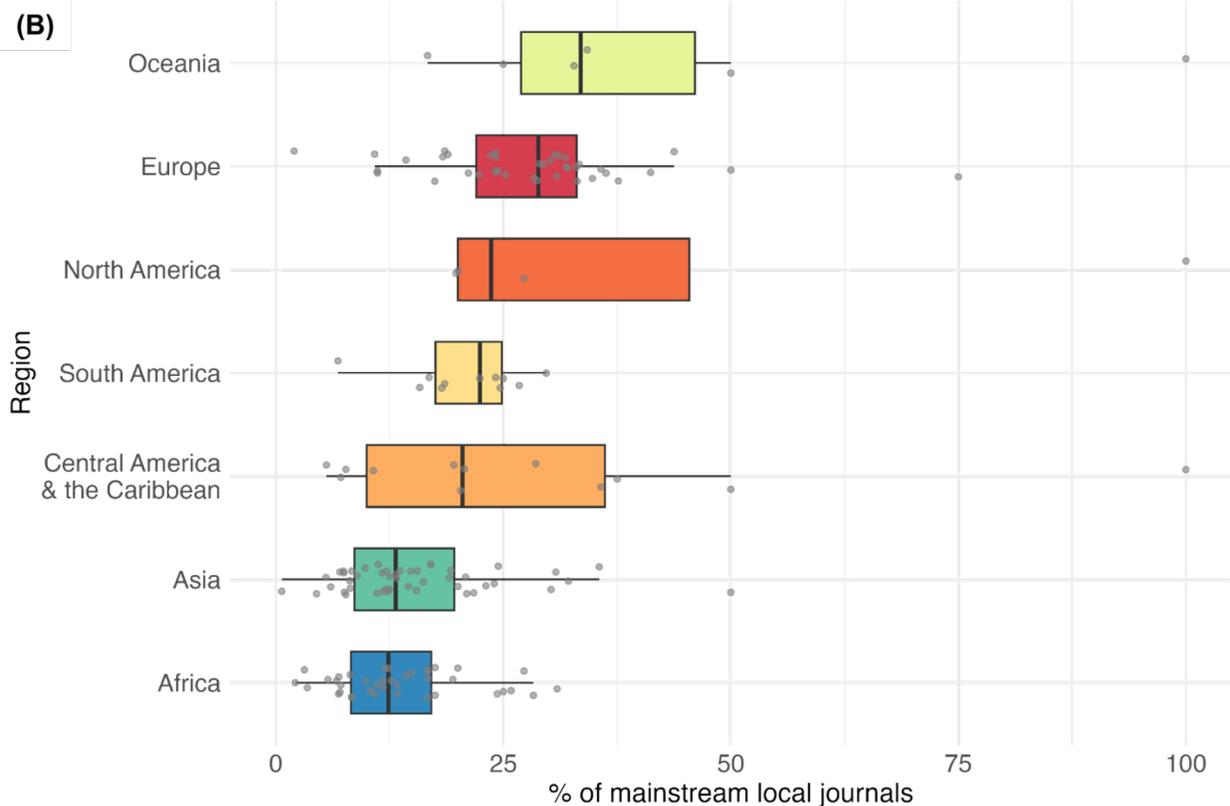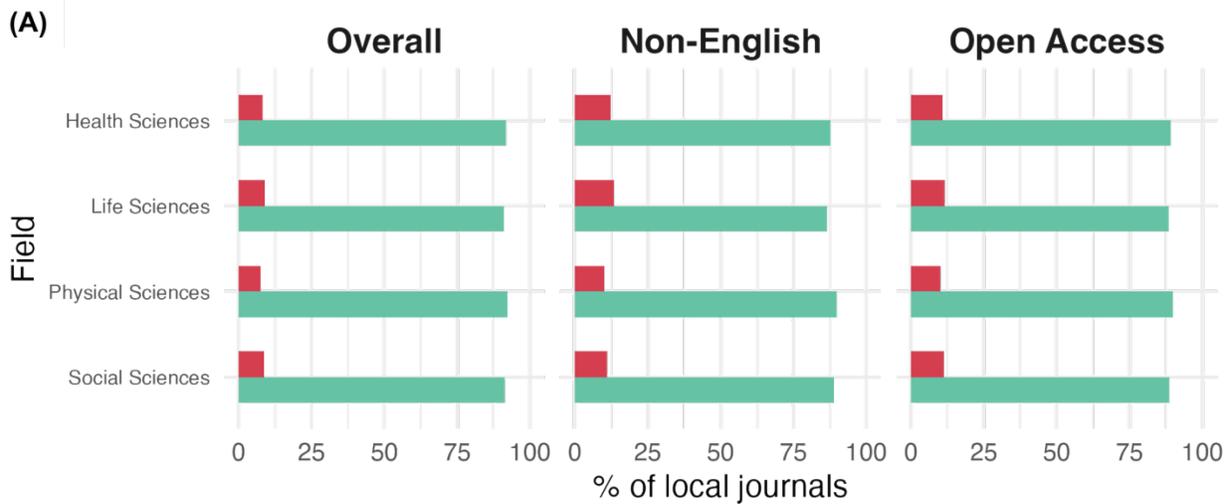
(A)

(B)

Figure 9. (A) Percentage of mainstream and non-mainstream locally relevant journals by field. The left-hand bars represent all journals, while the middle and right-hand bars include only non-English and open access journals. (B) Percentage of mainstream indexing among the locally relevant journals each country publishes in, by region.

Figure 10A exhibits uniform and low values except for a few outstanding cases, like Indonesia (30%) and Nicaragua (20%). Despite such low shares, Figure 10B displays more contrasting results on the journals' indexing status. Apart from a few countries that show an even mainstream-non-mainstream split (e.g., Greenland and Estonia), most reveal a tendency towards publication outside the mainstream. The lowest shares of articles are observed in Africa and Asia. In the Americas, countries such as the United States (30%), Mexico (15%), and Brazil (10%) also stand out due to the limited mainstream dissemination of their local articles. Although their behavior contrasts with that of South American neighbors such as Argentina, Chile, and Bolivia ($\bar{x} = 32\%$).

(A)

% of articles in local journals   10   20

(B)

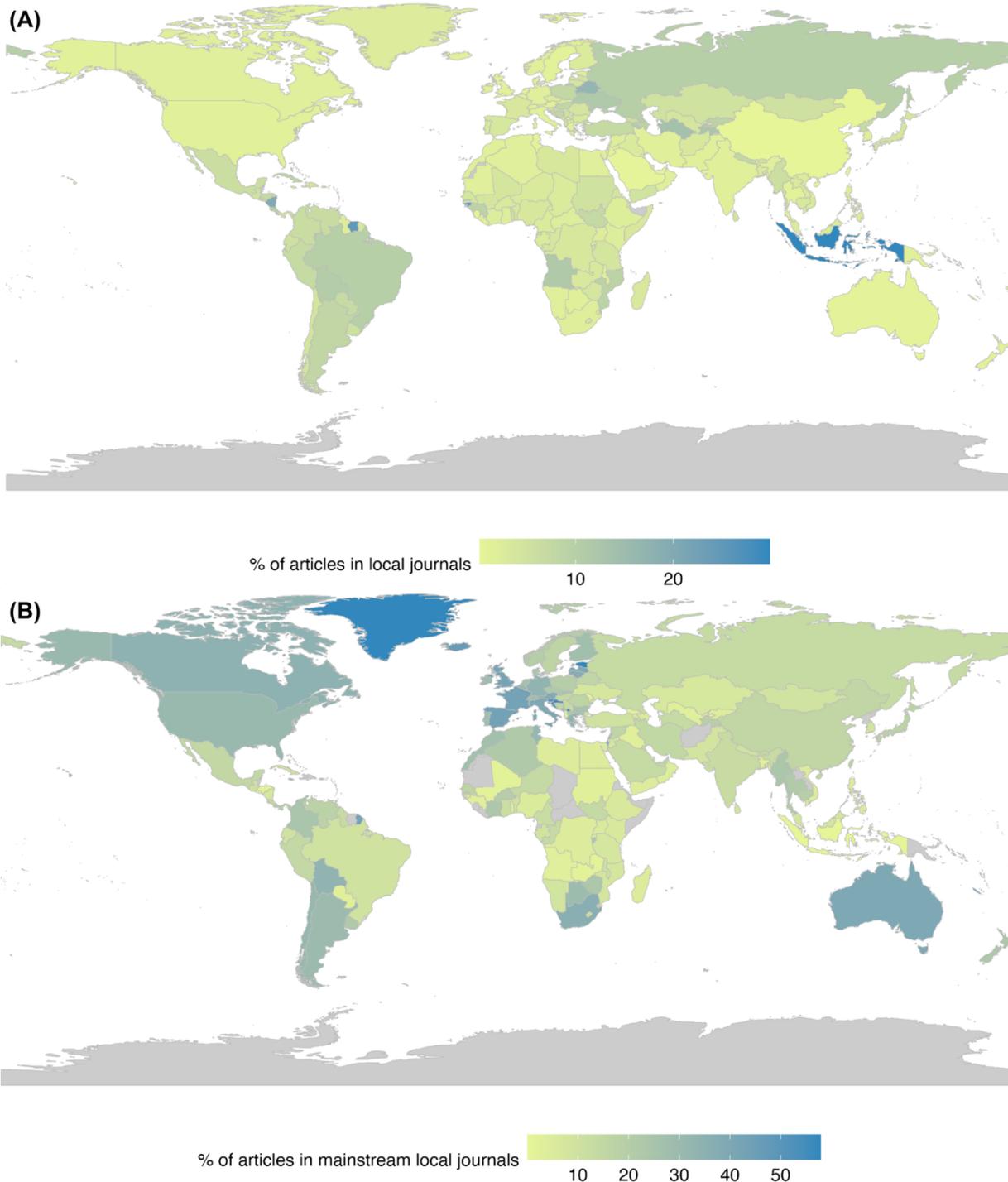% of articles in mainstream local journals   10   20   30   40   50

Figure 10. (A) Percentage of articles in locally relevant journals by country, relative to their total output.

(B) Percentage of articles in mainstream locally relevant journals by country.

Figure 11 exhibits the share of articles in non-English (A) and open access (B) journals within the mainstream locally relevant group, by region. According to Figure 11A, South America ($\tilde{x} = 52\%$) publishes the most in non-English journals and presents a tight country distribution. On the contrary, Asia ($\tilde{x} = 5\%$)

and Africa (x̃ = 8%) have both little participation in this type of journal. In Figure 11B, Central America and the Caribbean (x̃ = 69%) present a considerably high share of articles in these local journals. Median values for South and North America, Europe, and Africa are close to 50%, but with very different behaviors at the country level.
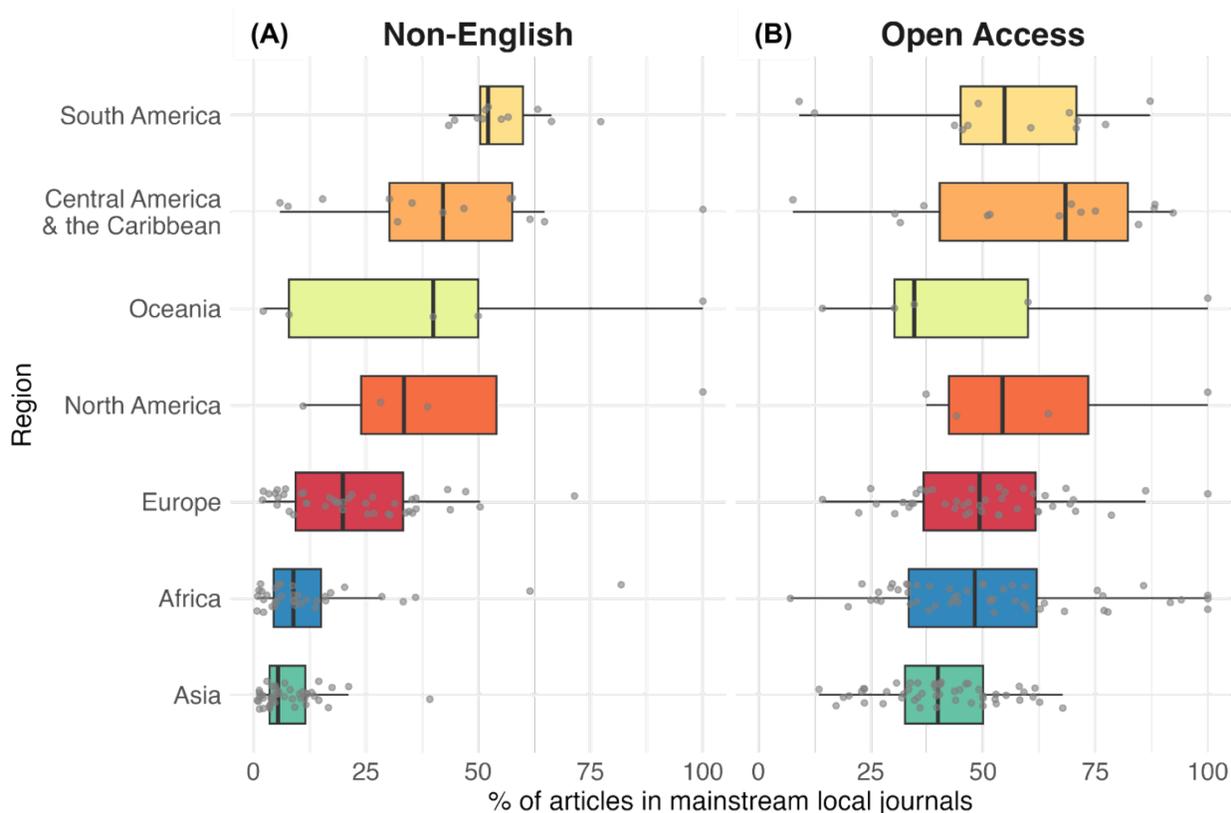

Figure 11. Percentage of articles in (A) non-English and (B) open access (right) mainstream locally relevant journals, by region.

Discussion

In this research article, we conducted a large-scale analysis of local journals across mainstream (WOS and Scopus) and non-mainstream (OpenAlex and DOAJ) databases, which embody contrasting representations of science (Mongeon & Paul-Hus, 2016). Since journals reflect the structure of scientific activity (Leydesdorff, 1987), we employed them as our object of study and addressed their local nature from three different conceptualizations: territory, producers, and recipients (Di Césare & Robinson-Garcia, 2026b). We aimed to (1) determine the representation of local journals in the four merged databases, (2) assess the visibility of the fields and countries that publish in them, and (3) examine their non-English and open access features, all done across mainstream and non-mainstream knowledge circuits.

25

To achieve that, we developed a dataset based on more than 75,000 OpenAlex journals, which were further enriched with WOS, Scopus, and DOAJ information. The resulting merged dataset included more than 100 variables that enhanced almost 50% of OpenAlex journals, and characterized more than 90% of them as either local or global from multiple, complementary perspectives (Di Césare & Robinson-Garcia, 2026b). To the best of our knowledge, such large and complete data is not available in any other consolidated source (Di Césare & Robinson-Garcia, 2026a).

Its overall characterization indicated that OpenAlex was roughly 3 times bigger than any additional source. This coincides with other studies reporting OpenAlex's more extensive coverage of both documents and journals (Alperin et al., 2024; Culbert et al., 2025). Particularly at the article level, Visser et al. (2021) found a similar ratio when comparing Scopus, WOS, and OpenAlex's precursor, Microsoft Academic. In line with Simard et al. (2025), DOAJ's overlap with OpenAlex was greater than that of DOAJ with WOS or Scopus. At the same time, we detected OpenAlex presenting bigger overlaps with both WOS and Scopus. So, ultimately, DOAJ was the data source leaving the most journals out of this analysis.

At the field level, all databases depicted Social Sciences journals as the most prevalent, irrespective of whether they were local, non-English, or open access venues. The WOS and Scopus open access category was the only exception, with Physical Sciences journals being slightly more frequent. Social Sciences have usually been deemed underrepresented in the mainstream (Mongeon & Paul-Hus, 2016), but by analyzing its 2023 active journals, the field stood out. Considering local journals alone, the majority belonged to the non-mainstream regardless of the field, language, and access type. However, Life Sciences local journals were slightly more visible in the mainstream than the rest. Across all fields, the gap between mainstream and non-mainstream indexing was most pronounced in the locally relevant journals. This finding could reinforce the profile of mainstream audiences as geographically dispersed.

At the country and regional levels, our results suggested that Sub-Saharan Africa published the most articles in non-mainstream locally situated journals. In spite of disseminating its English language output through open access, this could imply a double penalty for the region: the lower impact recorded for geolocalized research (Kahalon et al., 2022; Mongeon et al., 2022) is compounded by lower mainstream visibility. Similarly, Asubiaro et al. (2024) found Sub-Saharan African journals to be the most underrepresented in WOS and Scopus. In contrast, Global North countries produced fewer geolocalized articles (Castro Torres & Alburez-Gutierrez, 2022; Di Césare & Robinson-Garcia, 2026b) but disseminated them almost exclusively within the mainstream.

The United States stood out with up to 15% of its substantial 2023 output published in locally informed journals. Interestingly, half of these venues with geographically concentrated references were not indexed in WOS or Scopus, although they did not receive as many US articles as the mainstream local journals. Moreover, less than a fifth of these mainstream local articles were openly disseminated, which

marks the lowest value for North America. This denotes a considerably closed profile for US local research and contrasts with the overall trend towards open access publishing identified by the NSF (National Science Board, 2023).

The least mainstream indexing was detected for the locally relevant journals, especially those publishing African and Asian English language articles. On a similar note, most of Central and South America tended towards non-mainstream publication, which aligns with Latin American journals having a limited presence in WOS and Scopus (Asubiaro et al., 2024). On top of that, their limited mainstream presence was mostly characterized by non-English, open access local articles. Major producer Brazil stood out in the region because, alongside the US and Mexico, it published considerably more articles in journals with a geographically concentrated audience and a non-mainstream dissemination circuit. Moreover, almost half of them were non-English articles. All these features combined could reveal a knowledge circuit quite unique to this Global South superpower that contrasts with South American neighbors such as Chile and Argentina.

In summary, our results revealed no clear patterns. While Sub-Saharan Africa stood out in non-mainstream locally situated journals with its English language, open access articles, Latin America tended towards non-mainstream locally relevant journals with its primarily non-English, open access local output. But Asian English language articles in non-mainstream locally relevant journals were less open. Europe tended more regularly to the mainstream, whereas the US had a mixed profile depending on the typology of local journal analyzed. Nevertheless, a key finding of this research is that local journals circulate primarily outside the mainstream.

Concluding remarks

The creation of a merged dataset combining mainstream (WOS and Scopus) and non-mainstream (DOAJ and OpenAlex) databases allowed us to determine that most local journals were not well represented in the mainstream, regardless of their field, language, or access type. The most pronounced differences were identified at the country and regional levels, where representativity in or outside the mainstream varied according to the conceptualization of locality applied. Additionally, countries publishing in non-English and open access local journals did not produce clear patterns when observed from different locality perspectives. These contrasting, layered findings provide additional insights to reinforce that no individual approach entirely captures the local in every context (Di Césare & Robinson-Garcia, 2026b). However, one conclusion is straightforward: local journals circulate primarily outside the mainstream. Therefore, from a policy perspective, local research should not be discussed and evaluated exclusively from the content covered by WOS and Scopus.

We acknowledge that this research is not free from limitations. Measurements at the country level might be influenced by their size, and extending the time period under analysis could, to some degree, reshape the snapshot we described. Additionally, some of our methodological decisions could be contested, such as the definition of non-English journals, the thresholds applied to label the local, or the selection of data sources. Nonetheless, we see potential in this proposal to advance knowledge related to local journals, especially when addressed from the intersection of locality approaches (Di Césare et al., 2025). Future research could investigate the costs, in terms of authors' impact and recognition, of publishing in local journals, or the evolution of these journals' thematic profiles when they acquire mainstream visibility. Our openly shared dataset also provides opportunities to delve deeper into the behavior of specific countries and regions.

Acknowledgments

Funding information

Authors contributions

Victoria Di Césare: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft
Nicolas Robinson-Garcia: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

References

Alperin, J. P., Portenoy, J., Demes, K., & Larivière, V. (2024). *An analysis of the suitability of OpenAlex for bibliometric analyses*. 28th International Conference on Science, Technology and Innovation Indicators (STI).

Archambault, É., Vignola-Gagné, É., Côté, G., Larivière, V., & Gingrasb, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, *68*(3), 329–342. https://doi.org/10.1007/s11192-006-0115-z

Asubiaro, T., Onaolapo, S., & Mills, D. (2024). Regional disparities in Web of Science and Scopus journal coverage. *Scientometrics*, *129*(3), 1469–1491. https://doi.org/10.1007/s11192-024-04948-x

Beigel, F., Gallardo, O., & Bekerman, F. (2018). Institutional Expansion and Scientific Development in the Periphery: The Structural Heterogeneity of Argentina's Academic Field. *Minerva*, *56*(3), 305–331. https://doi.org/10.1007/s11024-017-9340-2

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. https://doi.org/10.1073/pnas.2119373119

Céspedes, L., Kozlowski, D., Pradier, C., Sainte-Marie, M. H., Shokida, N. S., Benz, P., Poitras, C., Ninkov, A. B., Ebrahimy, S., Ayeni, P., Filali, S., Li, B., & Larivière, V. (2025). Evaluating the linguistic coverage of OpenAlex: An assessment of metadata accuracy and completeness. *Journal of the Association for Information Science and Technology*, asi.24979. https://doi.org/10.1002/asi.24979

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. https://doi.org/10.1093/reseval/rvy001

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. https://doi.org/10.1016/j.respol.2017.08.002

Culbert, J. H., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2025). Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. *Scientometrics*, *130*(4), 2475–2492. https://doi.org/10.1007/s11192-025-05293-3

De Moya-Anegón, F., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Corera-Álvarez, E., Muñoz-Fernández, F. J., González-Molina, A., & Herrero-Solana, V. (2007). Coverage analysis of Scopus: A journal metric approach. *Scientometrics*, *73*(1), 53–78. https://doi.org/10.1007/s11192-007-1681-4

Delgado-López-Cózar, E., Ràfols, I., & Abadal, E. (2021). Letter: A call for a radical change in research evaluation in Spain. *El profesional de la información*, e300309. https://doi.org/10.3145/epi.2021.may.09

Di Césare, V., Costas, R., & Robinson-Garcia, N. (2025). *Testing the knowledge bridging function of non-mainstream journals in scholarly communication*. 29th Annual International Conference on Science and Technology Indicators (STI). https://doi.org/10.5281/zenodo.17276539

Di Césare, V. (2025). *OpenAlex journals dataset* (Version 1.0.0) [R, Python]. https://github.com/vdicesare/OpenAlex-journals-dataset

Di Césare, V., & Robinson-Garcia, N. (2026a). *Dataset of the manuscript 'Presence and visibility of local journals in international databases worldwide'* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.18298992

Di Césare, V., & Robinson-Garcia, N. (2026b). *What is local research? Towards a multidimensional framework linking theory and methods*. Zenodo. https://doi.org/10.5281/zenodo.14033472

Fry, C. V., Lynham, J., & Tran, S. (2023). Ranking researchers: Evidence from Indonesia. *Research Policy*, *52*(5), 104753. https://doi.org/10.1016/j.respol.2023.104753

Gibert Galassi, J., & Espina Bocic, P. (2022). A panoramic view to the evolution of three scientific communities in chilean academia 2012-2022. *Perfiles Económicos*, *13*, 87–107.

Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, *41*(2), 251–

261. https://doi.org/10.1016/j.respol.2011.09.007

Hicks, D. (2013). One size doesn't fit all: On the co-evolution of national evaluation systems and social science publishing. *Confero: Essays on Education, Philosophy and Politics*, *1*(1), 67–90. https://doi.org/10.3384/confero13v1121207b

Kahalon, R., Klein, V., Ksenofontov, I., Ullrich, J., & Wright, S. C. (2022). Mentioning the Sample's Country in the Article's Title Leads to Bias in Research Evaluation. *Social Psychological and Personality Science*, *13*(2), 352–361.

Kosyakov, D., & Guskov, A. (2022). Reasons and consequences of changes in Russian research assessment policies. *Scientometrics*, *127*(8), 4609–4630. https://doi.org/10.1007/s11192-022-04469-5

Leydesdorff, L. (1987). Various methods for the mapping of science. *Scientometrics*, *11*(5–6), 295–324. https://doi.org/10.1007/BF02279351

Marginson, S., & Xu, X. (2021). *Moving beyond centre-periphery science: Towards an ecology of knowledge*. University of Oxford.

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, *12*(4), 1160–1177. https://doi.org/10.1016/j.joi.2018.09.002

Mazoni, A., & Costas, R. (2024, June 6). Towards the democratisation of open research information for scientometrics and science policy: The Campinas experience. *Leiden Madtrics*. https://www.leidenmadtrics.nl/articles/towards-the-democratisation-of-open-research-information-for-scientometrics-and-science-policy-the-campinas-experience

Moed, H. F., De Bruin, R. E., & Van Leeuwen, Th. N. (1995). New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. *Scientometrics*, *33*(3), 381–422. https://doi.org/10.1007/BF02017338

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5

Mongeon, P., Paul-Hus, A., Henkel, M., & Larivière, V. (2022). *On the impact of geo-contextualized and local research in the global North and South*. 26th International Conference on Science, Technology and Innovation Indicators. https://doi.org/10.5281/ZENODO.6956978

National Science Board. (2023). *Publications Output: U.S. Trends and International Comparisons* (No. NSB-2023-33; Science and Engineering Indicators 2024, p. 64). National Science Foundation. https://ncses.nsf.gov/pubs/nsb202333/

Rafols, I., Ciarli, T., & Chavarro, D. (2015). *Under-reporting research relevant to local needs in the South: Database biases in the representation of knowledge on rice*. 13th Globelics International Conference.

Ràfols, I., Molas-Gallart, J., Chavarro, D., & Robinson-García, N. (2016). *On the dominance of quantitative evaluation in 'peripheral' countries: Auditing research with technologies of distance*. SSRN. https://doi.org/10.2139/ssrn.2818335

Sandoval-Romero, V., & Larivière, V. (2020). The national system of researchers in Mexico: Implications of publication incentives for researchers in social sciences. *Scientometrics*, *122*(1), 99–126. https://doi.org/10.1007/s11192-019-03285-8

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ*, *314*(7079), 497–497. https://doi.org/10.1136/bmj.314.7079.497

Shu, F., Liu, S., & Larivière, V. (2022). China's Research Evaluation Reform: What are the Consequences for Global Science? *Minerva*, *60*(3), 329–347. https://doi.org/10.1007/s11024-022-09468-7

Simard, M.-A., Basson, I., Hare, M., Larivière, V., & Mongeon, P. (2025). Examining the geographic and linguistic coverage of gold and diamond open access journals in OpenAlex, Scopus, and Web of Science. *Quantitative Science Studies*, *6*, 732–752. https://doi.org/10.1162/qss.a.1

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, *126*(6), 5113–5142. https://doi.org/10.1007/s11192-021-03948-5

Stiglitz, J. (1999). *Scan globally, reinvent locally: Knowledge infrastructure and the localization of*

knowledge. First Global Development Network Conference.

Tomaselli, K. G. (2018). Perverse incentives and the political economy of South African academic journal publishing. *South African Journal of Science*, *114*(11/12). https://doi.org/10.17159/sajs.2018/4341

van Leeuwen, T. N., Moed, H. F., Tijssen, R. J. W., Visser, M. S., & Van Raan, A. F. J. (2001). Language biases in the coverage of the Science Citation Index and its consequencesfor international comparisons of national research performance. *Scientometrics*, *51*(1), 335–346.

Vera-Baceta, M.-A., Thelwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics*, *121*(3), 1803–1813. https://doi.org/10.1007/s11192-019-03264-z

Vessuri, H., Guédon, J.-C., & Cetto, A. M. (2014). Excellence or quality? Impact of the current competition regime on science and scientific publishing in Latin America and its implications for development. *Current Sociology*, *62*(5), 647–665. https://doi.org/10.1177/0011392113512839

Visser, M., Van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, *2*(1), 20–41. https://doi.org/10.1162/qss_a_00112

Appendix

Table A1. Data sources' overall set sizes and percentage overlaps between them.

Data source	DOAJ	CJI	SJR	Scopus	JCR	MJL	OpenAlex
OpenAlex	16,242	23,759	23,963	24,997	20,038	20,419	75,694
	78%	85%	85%	86%	91%	91%	
MJL	5,964	19,345	19,533	19,839	21,191	22,455	
	28%	69%	69%	68%	96%		
JCR	5,728	19,097	19,278	19,472	21,988		
	27%	68%	68%	67%			
Scopus	8,230	27,051	27,518	29,221			
	39%	97%	98%				
SJR	7,448	27,246	28,174				
	36%	98%					
CJI	7,406	27,879					
	35%						
DOAJ	20,955						