Measuring back: bibliodiversity and the Journal Impact Factor™ brand, a case study of IF-journals included in the 2021 Journal Citations Report™

Little attention has been devoted to whether the impact factor (IF) can be considered a responsible metric in light of bibliodiversity. This article critically engages with this question in measuring the following variables of IF journals included in the 2021 Journal Citation Reports™ and examining their distribution: publishing models (hybrid, open access (OA) with or without fees, subscription), world regions, language(s) of publication, subject categories, publishers and the prices of article processing charges (APCs) if any. Our results show that the quest for prestige or perceived quality through the IF brand poses serious threats to bibliodiversity. The IF brand can indeed hardly be considered a responsible metric insofar as it perpetuates publishing concentration, maintains a domination of the Global North and its attendant artificial image of mega producer of scholarly content, does not promote linguistic diversity, and does not incentivize fair and equitable open access by entrenching fee-based OA delivery options with rather high APCs.

Keywords
bibliodiversity; impact factor; research assessment; metrics; bibliometry; open access

Introduction
The impact factor (IF) is a bibliometric indicator that has been the subject of much criticism, especially in the current state of advocacy for more responsible metrics. Various key reform-oriented texts concerned with the advancement of research assessment such as the San Francisco Declaration on Research Assessment,1 the Leiden Manifesto,2 the Metric Tide3 or the Agreement on Reforming Research Assessment,4 for example, have called to stop using the IF in research evaluation, notably because of its calculation-related deficiencies such as skewed citation distribution.5 These and other texts have also alerted about the normative uses of the IF and its adverse effects on the research ecosystem. These concerns and criticisms include unfair comparisons between fields with different citations practices,6 perverse incentives,7 including citation gaming,8 the amplification of journal status endowment through article processing charges (APCs) inflation9 and the use of the IF as a proxy for perceived quality or excellence,10 despite evidence to the contrary.11
Despite the prevalence of the IF in the research ecosystem, including its influence in research assessment and rewards systems, research has so far failed to propose a comprehensive analysis of bibliodiversity as it can relate to the IF brand. Bibliodiversity is key in maintaining a diverse, inclusive and equitable research landscape that can ‘accommodate the different workflows, languages, publication outputs, and research topics that support the needs and epistemic pluralism of different research communities’. A lack of bibliodiversity can result in homogenized and hegemonic trends in the ways research is carried out, made visible and produced or disseminated. This impacts who can access, read and benefit from research.

Taking its cue from this characterization of bibliodiversity as a multifaceted and structuring concept of scholarly communications, this article examines the bibliodiversity of the IF brand in measuring particular variables of IF journals included in the 2021 Journal Citation Reports™ (JCR™, n = 12,391), namely: publishing models (hybrid, open access with or without fees, subscription), world regions, language(s) of publication, subject categories, publishers and the prices of APCs if any.

Measuring these variables and examining their distribution can provide valuable insights into the IF brand, in particular how it might be said to reject bibliodiversity in reinforcing ‘monopoly, monoculture, and high prices’ in scholarly communications. In turn, these insights may help various stakeholders in scholarly communications, research assessment, information literacy and open science to further critically engage with the IF and how its brand may actually jeopardize the development of a healthy, diverse and sustainable scholarly communications ecosystem.

Literature review

Bibliodiversity is a term that can refer to a variety of realities and practices. The International Alliance of Independent Publishers defines bibliodiversity as ‘cultural diversity applied to the world of books’. In practice, this is envisioned as ‘a complex, self-sustaining system of storytelling, writing, publishing, and other kinds of production of oral and written literature’ which contributes to more socio-cultural diversity and equity. For libraries, providing access to a diversity of outputs and formats is a core principle embedded in both advocacy actions and collection development strategies. In the world of scholarly communications, the concept is even more wide-ranging.

In this context, bibliodiversity indeed pertains to editorial methods, discoverability practices, publishing trends and infrastructures and issues of affordability and sustainability, in particular how the latter can relate to the politics of open access. The 2017 Jussieu Call for Open Science and Bibliodiversity, for instance, argues that developing a scholarly publishing ecosystem by privileging a fee-based open access model may ‘slow if not check the advent of bibliodiversity’. Similarly, Monica Berger describes fee-based open access as a colonizing force and enterprise in the landscape of scholarly publishing. A similar view is shared by Shearer and Becerril-García who contend that bibliodiversity can contribute to the decolonization of scholarly communications. In their respective articles, Berger as well as Shearer and Becerril-García point to various Latin American initiatives of community-owned and non-commercial infrastructures or publishing platforms such as Latindex, Redalyc and SciELO, all of which have enhanced the creation, discoverability, prevalence and significance of locally-produced research. In so doing, these initiatives have shown a commitment to the ‘sustainable, anti-colonial ethos of bibliodiversity’, that is, they constitute a viable alternative to the domination of international English-language journals published by a handful of large commercial publishers, which have increasingly shifted to pay-to-publish open access options methods in recent years.
Bibliodiversity thus runs throughout all the production and dissemination mechanisms and practices of the scholarly communications ecosystem. But this wide scope makes it particularly difficult to propose a comprehensive analysis of how bibliodiversity is articulated in a given corpus. This is why studies usually only examine the state of one main bibliodiversity-related variable in particular bibliographic databases, journals’ lists, or subsets of thereof in focusing on, for example, coverage or scope limitations as they can relate to subjects, geographic representation, linguistic diversity, publisher concentration, or questions of affordability and models and how they tie in with the politics of open access.

This article complements the above-mentioned research in adopting a more holistic approach to bibliodiversity as it can be applied to the IF brand by measuring the previously mentioned bibliodiversity data points of journals included in the 2021 JCR™. In so doing, it provides a blueprint for a reproducible method of ‘measuring back’ which echoes the decolonial ethos of bibliodiversity and can be reused for other journals’ lists or publishing brands.

Materials and methods

The initial dataset used in this study was created by collating data from the 2021 Science Citation Index Expanded (SCIE) and Social Science Citation Index (SSCI) JCR™ lists that are freely available on Clarivate’s website (data downloaded on July 29, 2021). The following information for individual journal entries was obtained from these lists: journal title, ISSN(s), publisher, publisher’s address, publisher country, language(s) of publication, the Web of Science Categories (WoS) and the index(es) that the journal is part of (SCIE, SSCI, or both).

This dataset was then enriched with information regarding publishing models and APC prices, if any, using freely available datasets or lists from publishers and other sources, namely: Walt Crawford’s dataset ‘Gold Open Access 6: 2015–2020’ (GOA6), a journal metadata file produced by the Directory of Open Access Journals (DOAJ; downloaded on August 13, 2021) and APC price lists from various publishers. Information regarding journal models and possible APC was also manually checked and directly retrieved from publishers and journals’ websites when they could not be found in previously mentioned lists.

Detail regarding data standardization and structuration methods is provided in Appendix A for the following variables: publishing models, APC prices, subject categories, publishers’ names and ensembles, world regions and language(s) of publication.

Results

Figures 1 to 4 show the diversity of the IF brand per publishers’ ensembles and one of the following variables: world regions, subjects, language(s) of publications or publishing model.

Figure 1 shows little geographic diversity and indicates high levels of publisher concentration, especially in Western Europe and North America. These two regions produce 85% (n = 10,485) of all IF journals included in the 2021 JCR™. All other world regions are under-represented, especially Eastern and Southern Africa and West and Central Africa, which respectively account for 0.45% (n = 56) and 0.03% (n = 4) of all journals. Figure 1 also shows the global reach of the Oligopoly (i.e. Elsevier, Sage, Springer Nature, Taylor & Francis and Wiley) as it is present in all world regions, albeit in varying degrees. The top 25 publishers, which consist of the Oligopoly and the next 20 publishers with the biggest portfolio, account for 75% (n = 9,290) of all journals. Western Europe and North America publish 90% (n = 8,318) of journals produced by these 25 publishers. In contrast, other world regions show lower levels of publisher concentration as they present bigger, albeit variable shares of journals tagged in the Other publishers category, ranging from 34 (n = 41) in the Middle East and North Africa to 89% (n = 212) in Latin America and the Caribbean. (Please note, due to rounding, some totals may not correspond with the sum of separate figures.)
Figure 1. Distribution of journals per world region and publishers’ ensemble

Figure 2 indicates high levels of publisher concentration for the top 25 publishers in all subject categories. Across all subjects, the Oligopoly represents 59% (n = 7,286) while the next 20 publishers with the biggest portfolio account for 16% (n = 2,004) of all journals. In six subject categories, the Oligopoly’s shares of journals reach levels higher than the overall average representation: Chemistry (61%; n = 366), Geosciences (61%; n = 417), Medical Sciences (59%; n = 1,847), Professional fields (63%; n = 656), Psychology (71%; n = 330) and Social Sciences (63.5%; n = 606). When adding the next 20 publishers with the biggest portfolio to the Oligopoly category, levels of publishers concentration reach levels higher than their combined overall average in six subject categories: Astronomy (82%; n = 54), Chemistry (84%; n = 500), Computer sciences (78%; n = 383), Medical sciences (77%; n = 2,435), Professional fields (78%; n = 809), Psychology (86%; n = 402) and Social sciences (80%; n = 760).

Figure 3 shows that English prevails as the only language of publication for the majority of journals (89% of all journals; n = 11,077) and that the top 25 publishers publish the majority of English-only journals (79%, n = 8,699). In contrast, the category of Other publishers somewhat levels the playing field for more linguistic diversity as publishers that do not belong to the top 25 publishers produce more than 50% of journals in all language categories except English- and German-only titles: Other language(s) (87%; n = 109), Spanish (82%; n = 73), French (59%; n = 37) and Multilingual including English (50%; n = 469).

'English prevails as the only language of publication for the majority of journals'
The proportions of journals per language category fluctuate very little according to subjects or world regions. It is nevertheless worth noting that the least linguistically diverse category is that of Computer sciences, of which 97% (n = 477) of titles are English-only journals. Worth noting as well is the fact that the most linguistically diverse world region is Latin America and the Caribbean, with only 32% (n = 76) of English-only journals.

Figure 4 indicates that the hybrid publishing model is dominant (68% of all journals; n = 8,422) and that the top 25 publishers privilege hybrid and fully OA-APC journals. Together, these 25 publishers produce 90% (n = 7,592) of all hybrid journals and 60% (n = 1,115) of all full OA fee-based journals. In contrast, the category of Other publishers is more diverse in terms of publishing models; it accounts for 80% (n = 608) of all OA journals that do not require fees and represents 60% (n = 610) of all journals using a subscription model. This category of Other publishers is also the one with the most journals for which the possible presence of fees could not be identified (97%; n = 295).
contrast, other regions show more diverse publishing models, albeit in varying degrees. The subscription model is prevalent in Eastern Europe and Central Asia (48%; n = 173), while in Latin America and the Caribbean, the most common model is that of OA without fees (42%; n = 100). Proportionally speaking, Western Europe and North America also have the smallest shares of OA journals that do not require fees, respectively 5% (n = 310) and 2% (n = 96).

Figures 6 to 8 below focus on APC-related information for hybrid and APC-based OA journals. Figure 6 shows a statistical summary of APC prices for hybrid and OA-APC journals per publishers’ ensembles, with box plots showing the first quartile, median and the third quartile of the values. For hybrid journals, the median APC price is US$3,000 for the Oligopoly, US$3,255 for the next 20 publishers, and US$2,380 for the category of Other publishers. The difference in median APC prices for OA-APC journals between the Other publishers and the two other categories of publishers’ ensembles is even more marked: US$2,290 for the Oligopoly, US$1,916 for the next 20 publishers, and US$684 for the Other publishers.
Data shows that the median prices of APCs do not vary much across languages but fluctuate much more per world region, especially for OA-APC journals. Median APC prices of hybrid journals approximate US$3,000 in all world regions, except in Eastern Europe and Central Asia. In contrast, median APC prices of OA-APC journals are below US$1,000 in all world regions but three: East Asia and Pacific (US$1,100), North America (US$2,000) and Western Europe (US$2,000).

Figure 7 shows the number of hybrid journals per APC-price band and publishers’ ensemble. This figure indicates that the top 25 publishers produce more hybrid journals in higher APC price ranges than the category of Other publishers. This is especially true of journals in APC-price bands higher than the overall median price of hybrid journals (i.e. US$3,000), where the top 25 publishers produce more than 90% of journals in all price bands.

Figure 8 shows the number of OA-APC journals per APC-price band and publishers’ ensemble. This figure indicates that the category of Other publishers produces more OA-APC journals in lower price bands, with more than 50% of all of the journals in each price band below US$1,201–1,400. In contrast, the top 25 publishers produce more than 75% of all journals in price bands higher than the overall median price of OA-APC journals (i.e. US$2,000), exceptions being the US$3,801–4,000 and US$4,601–4800 price bands.
Discussion

The IF brand indicates a continued domination of the Oligopoly (Elsevier, Sage, Springer Nature, Taylor & Francis and Wiley) in the scholarly publications system, with even higher shares than the ones reported in a 2015 study of publisher concentration in the WoS content, which averaged slightly above 50% in 2013. These variations can in part result from differences in methodology and scope. While Larivière et al.’s study examined documents throughout the entire WoS content, our study looks at publisher concentration at the journal level and only within a subset of the WoS content (i.e. the SCIE and/or SSCI indexes).

Next to the Oligopoly, our results also show that the IF brand favours well-established and historical publishers and scholarly societies (ACS, APA, BMJ, CUP, De Gruyter, Emerald, OUP, Wolters Kluwer), or more recent publishers with scaling-up capacities and infrastructures that specifically target international audiences (Bentham, Frontiers, MDPI, IOP Publishing, World Scientific Publishing). This is most likely to be due to several factors, including historical reasons, discoverability, indexing and visibility issues, as well as the curation process and inclusion criteria used by the WoS, which notably require that all titles and abstracts must be translated into English to be included in one of its indexes. In any case, encouraging publishing in IF journals contributes to the market consolidation of particular commercial publishers that are mainly located in the Global North.

Similarly, promoting IF journals as publishing venues is bound to hamper the development of multilingual or non-English content in research given the very high shares of IF-journals accepting articles in English (>90%). This result is in line with a previous study showing the proportions of documents in English at 95% in the WoS database. To compare in terms of numbers of journals, a study analysing the 25,671 active journals employing the open-source publishing platform Open Journal Systems (OJS) reports a proportion of journals using English as a main language at 50%. This comparison shows the imbalance of English-only journals across the whole journals publishing sector as represented by the WoS, while pointing to a rescaling possibility of the use of languages other than English in scholarly communications. Of course, the use of a lingua franca such as English can be beneficial for reasons of global dissemination. But it should not prevail to the detriment of research in other languages, which can, as the Helsinki Initiative on Multilingualism in Scholarly Communication reminds us, benefit society ‘beyond academia’, particularly when it is related to issues of ‘heritage, culture, and society’. In fact, a ‘balanced multilingualism’ in research can keep ‘locally relevant research alive’ as well as potentially ‘create localized impacts’.

Several reasons can be advanced as to why the IF brand does not currently achieve this balanced multilingualism by privileging English. Firstly, the IF was historically created and ‘specifically to cater to the needs of US librarians’. It later developed through ‘reshap[ing] international science in favour of both the US and the English language’. Secondly, much of the globalization and internationalization of research since the second half of the 20th century onwards has made English increasingly hypervisible. Thirdly, the increasing use of English-centered databases such as Scopus and WoS in research assessment may influence the publishing agenda of researchers, who ‘may choose to move away from locally relevant research toward decontextualized approaches of interest to English-language audiences’.

The very low share of no-fees OA journals observed here (6.1%) is particularly questionable and worrying when compared to the proportion of APC-based OA journals (15%; n = 1,849). This ratio of OA journals with or without fees indeed sharply contrasts with that of the DOAJ, which features more than 12,959 no-fees OA journals and only 5,886 OA journals requiring APCs (as of January 13th, 2023). Even if all OA-journals for which the presence of fees could not be confirmed (n = 303) were considered to be diamond OA journals, the share of OA-APC journals would still by far surpass the share of diamond OA journals in the present study.

‘encouraging publishing in IF journals contributes to the market consolidation of particular commercial publishers that are mainly located in the Global North’

‘the increasing use of English-centered databases such as Scopus and WoS in research assessment may influence the publishing agenda of researchers’
The IF brand can thus be said to privilege a minority model (OA-APC) over a fairer, more equitable and more frequent OA publishing model without fees, i.e. diamond journals, even if said model undoubtedly has its own challenges. This can be in part explained by the prevalence of big commercial publishers of the Global North (cf. Figure 5). Even so, the share of OA-APC journals reported here entrenches a publishing model supporting exclusionary politics of publishing, which discriminate against researchers from low-income countries, independent scientists, or other scholars unable to finance APCs. This is particularly true when said APC prices, especially as they are practised by the top 25 publishers and journals of the Global North (cf. Figure 7 and Figure 8), surpass by far the actual costs to publish scholarly articles. This APC barrier, in turn, adds another bias in the international research landscape which yet again privileges particular voices such as senior and male academics in contributing to the hypervisible scholarly record.

Finally, the prevalence of the OA-APC model manifest in the IF brand also raises questions of sustainability since it has been demonstrated that there exists a positive relationship between APC prices and journals with a high IF. The IF-brand also de-incentivizes fair, equitable and sustainable open access publishing by indirectly amplifying the hybrid publishing model, which most journals rely on according to our data (cf. Figure 4). Again, this may be due to the significant numbers of big commercial publishers within the IF list, as the top 25 publishers indeed publish more than 90% of all hybrid journals. Whatever the reasons behind the number of hybrid journals, their sheer number contributes to the same exclusionary politics as journals using an OA-APC model, even more so given that the APC levels of hybrid journals reported here (cf. Figure 6 and Figure 7) confirm previous analyses showing that they are more expensive than those practised by OA-APC journals.

The significant shares of OA-APC and hybrid journals observed in the IF brand (cf. Figure 4) should also raise further sustainability and equity concerns as the number of articles delivered through these methods has increased over the last few years. In his yearly studies of gold open access output based on DOAJ data, Crawford has shown that the minority of OA-APC journals included in the DOAJ produce a majority of the total output of OA articles, i.e. OA articles published in either diamond or fee-based OA journals included in the DOAJ. In 2020, the minority of OA-APC journals produced 65% of the overall output of OA articles, a proportion which increased to 69% in 2021. As for hybrid journals, several studies have shown an uptake of hybrid articles published by the Oligopoly. Laakso and Björk reported a number of hybrid OA articles doubling every year between 2007 and 2013. More recently, Butler et al. demonstrated a ‘slight increase’ of hybrid OA articles ‘from 29% in 2015 to 32.4% in 2018’. This increase of OA articles through the hybrid model should nevertheless be put into perspective as it does not seem to be significant enough to allow journals to transition to full OA in a reasonable time frame, thus prolonging the unsustainability and inequity of the model, notably as it is implemented through so-called transformative agreements.

The fact that journals included in other WoS indexes are soon to receive an IF might possibly slightly reduce the significance of some of the hegemonic and oligopolistic trends discussed above. Journals included in the Arts and Humanities Citation Index (AHCI) might, for example, improve the representation of the Humanities within the IF brand. This index might also improve the linguistic diversity of the IF brand since multilingual publishing ‘is an ongoing practice in many SSH research fields regardless of geographical location, political situation, and/or historical heritage’. The addition of journals from ‘more than 3,000 publishers, many of which are smaller publishers from the developing world’, might also reduce the share of English-only journals with an IF, especially if these publishers show more linguistic diversity, as does the category of Other publishers in this study (cf. Figure 3). These publishers may also contribute to reducing the domination of the Global North (cf. Figure 1) and trends of publisher concentration (cf. Figure 1 and Figure 2). Finally, the assignment of an IF to almost 9,000 journals may increase the share of OA journals by 8% in the JCR™ as per Clarivate’s press release. But it is unknown what proportion of the OA journals will be APC-based.
Conclusion

The quest for prestige or perceived quality through the IF brand poses serious threats to bibliodiversity in scholarly communications. On the whole, our results indeed show that the IF brand perpetuates publishing concentration, maintains a domination of the Global North and its attendant artificial image of mega producer of scholarly content, does not promote linguistic diversity, and does not incentivize fair and equitable open access by entrenching fee-based OA delivery options with rather high APCs.

Given the prevalence of the IF in research assessment and its ties to issues of professional advancement or career development, this problematic state of bibliodiversity should be properly reckoned with by the scholarly community to minimize the adverse effects of journals’ lists on the diversity of the publishing landscape and its future developments. In terms of research, this implies a further and regular examination of how the IF and other journals’ lists used for evaluation tackle bibliodiversity over time. In terms of advocacy and policy-making, this implies that bibliodiversity issues be addressed in research training and research evaluation programmes and reforms. A first step in this decolonial endeavour may be to recognize the increasing ‘intertwining of research assessment and open science’, which notably transpires from the UNESCO recommendation on open science.

Appendix A
This appendix details the strategies used for the structuration and standardization of the following data points used in this study: publishing models, APC prices, unique subject categories, publishers’ names and ensembles, world regions and language(s) of publication.

Publishing models
Different criteria were used and articulated with a Boolean logic to identify the publishing models of journals, especially for OA journals. Four main models were identified: fully fee-based OA journals (OA-APC), full OA journals not requiring fees (OA-no fees), hybrid, and subscription. The criteria used to categorize journals into these models are defined below.

OA-APC:
- journals are included in the DOAJ or GOA6 and mention the presence of fees, OR
- journals are included in APC prices lists of publishers and are explicitly labelled as OA, gold OA, or full-OA journals, OR
- all of the journal’s content can be accessed for free and publishing in said journal can only be achieved in exchange of an APC, regardless of the type of licence that may be used or whether said journal refers to or uses the phrase ‘open access’, OR
- journals have transitioned to full APC-based OA and all of their content since their flipping date is published OA.

OA-no fees:
- journals are included in the DOAJ or GOA6 and explicitly mention the absence of fees, i.e. fees = US$0, OR
- all of the journal’s content can be accessed for free and it is explicitly mentioned that publishing in said journal does not require the payment of an APC, regardless of the licence that may be used or whether said journal refers to or uses the phrase ‘open access’, OR
- journals have transitioned to no-fee OA or have adopted the ‘Subscribe to Open’ model but the content predating this flipping moment may still only be available behind a paywall.

Hybrid:
- subscription journals which offer OA publishing only as an option through which articles ‘are immediately free to read under an open license’ that is granted ‘in exchange for an article processing charge (APC)’, including journals labelling themselves as ‘transformative’, OR
- journals are included in APC price lists of publishers and are explicitly labelled as hybrid or optional OA journals, regardless of the licence that may be used.

Subscription:
- the source’s content is only available through subscription or purchase options; it may contain some bronze OA, i.e. ‘free to read content on the publisher’s page, but without an [sic] clearly identifiable license’ OR
- journals use a delayed or embargoed OA model, i.e. there may be an OA-moving barrier but the latest content can only be accessed through subscription or purchase options.

It was sometimes not possible to ascertain the publishing models of some journals, or their possible use of APC or lack thereof. As a result, two other categories were used: ‘unidentified’ (n = 19) and ‘OA-possible presence of fees unknown’ (n = 304a).

APC prices
In order to ensure valid comparison, APC prices were all collected in US$ for US authors without any taxes or any membership or society discounts for what comes closest to ‘traditional’ research articles in terms of output type, without taking into account potential submission fees or extra fees for coloured pages and/or images, tables and Figures. When different types of licences were offered, prices for a CC BY licence were collected. When APC varied according to length or word limit, the same limit was used as that of Crawford in GOA6, i.e. ten pages or 5,000 words. When APC prices were not mentioned in US$, prices were converted using the exchange rate of the date the said price was collected. When journals only charged authors within the country of publication, those charges were used (converted in US$ if necessary).

The label ‘unidentified’ was used for APC prices of full-OA or hybrid journals whose APCs could not be identified (n = 172).
Unique subject categories

The WoS uses over 250 subject categories to classify its indexed journals. To facilitate data analysis for the present study, Milojević’s reclassification scheme of WoS content into broad subject areas was used.

Using this reclassification scheme brought about some challenges. Firstly, journals can be assigned different WoS categories, which are not explicitly hierarchical, even though some subject categories can be considered as part of other, broader ones.14 Most journals assigned with more than one WoS category point to the same broad area after reclassification, but this was not true in some cases. Nevertheless, the arbitrary decision was made to assign a broad area based on the reclassification of the first WoS category appearing in Clarivate’s data. Secondly, the WoS category scheme contains the ‘Multidisciplinary Sciences’ label that has no equivalent in Milojević’s reclassification scheme because it focuses on article-level reclassification, for which the label ‘Multidisciplinary Sciences’ makes little sense. Since this study focuses on journals, this WoS category was kept as a broad area, thereby updating Milojević’s scheme to 15 unique broad areas. Finally, we found three WoS categories that did not have any match in Milojević’s reclassification scheme, namely ‘Development Studies’, ‘Regional & Urban Planning’ and ‘Quantum Science & Technology’. The ‘Social Sciences’ broad area was used to reclassify the journals tagged with ‘Development Studies’ or ‘Regional & Urban Planning’ as a first WoS category because of their semantic similarity with the WoS category ‘Planning and Development’, which falls into Milojević’s broad area ‘Social Sciences’. Journals assigned with ‘Quantum Science & Technology’ as a first WoS category all contained at least one other WoS category which was used to reclassify those journals with a broad area label from Milojević’s scheme.

Standardized publishers’ names and publishers’ ensembles

Many scholarly and scientific societies, institutions or unions are client organizations of commercial publishers, whose platforms they use to distribute content and sometimes manage publishing processes. In the WoS data used to build our initial dataset, journals are associated with these client organizations or with publishers’ imprints and brands, some of which include many name variants or are abbreviated.

Several strategies were used to clean up and standardize the names of most publishers. Firstly, the list of publisher imprints used in a short course on analysing institutional publishing output15 was used to reclassify data regarding imprints and brands into larger standardized publisher names. Secondly, common abbreviations or name variants were searched and assigned a standardized name. Finally, the label ‘Other’ was assigned to all remaining journals without a standardized name (n = 2,711).


World regions

On the basis of the WoS country data included in the initial dataset, a ‘World Region’ value was assigned to journals using UNICEF’s regional classifications scheme.16

Languages

Six categories of languages of publication were used to facilitate analysis: English, Multilingual including English (i.e. journals accepting articles in English in addition to at least one other language), German, Spanish, French and Other language(s) (i.e journals accepting articles in at least one other language that is not included in the previous categories).

Two challenges had to be faced when dealing with the standardization of data regarding language(s) of publication. Firstly, this data was missing for 88 journals in the WoS data. As a result, data was added manually for these journals after checking their websites. Secondly, the WoS information regarding language(s) of publication is not always consistently structured. While for some journals, multiple languages are enumerated, for some others the label ‘Multi-Language’ is used (n = 715). As a result of this discrepancy, every journal tagged with this ‘Multi-Language’ label was considered to be a journal using English in addition to another language of publication. Though this might induce a slight bias in the data, the rationale behind this decision is based on the fact that when they are enumerated, languages of publication of journals almost always explicitly include English.

Acknowledgments

The authors would like to thank Hollie White for her assistance in the conceptualization of this study and Elisa Grégoire, Sabrina Lekhder and David Tredinnick for their contribution in data collection.

Editor’s Note

The editors of Insights would like to apologise for the unavoidable poor quality of Figure 6. For those who would like more clarity, please see the underlying data the graphic is based on. It is available in the article’s dataset: https://zenodo.org/records/7683744.

Data accessibility statement

The dataset used in this study is openly available on Zenodo https://zenodo.org/records/7683744 (Bardiau, Marjorie and Dony, Christophe 2023).

Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other Insights articles can be accessed here – click on the URL below and then select the ‘full list of industry ALAs’ link: http://www.uksg.org/publications/aa.

Competing interests

The authors have declared no competing interests.

References


23. Berger, “Bibliodiversity at the Centre.”

24. Shearer and Becerril-García, “Decolonizing Scholarly Communications through Bibliodiversity.”


32. Berger, “Bibliodiversity at the Centre.”


35. Berger, “Bibliodiversity at the Centre.”


40. Kulczycki et al., “Multilingual Publishing in the Social Sciences and Humanities.”

41. Archambault and Larivière, “History of the Journal Impact Factor.”

42. Archambault and Larivière.


44. Kulczycki et al., “Multilingual Publishing in the Social Sciences and Humanities.”


53. Laakso and Björk, “Hybrid Open Access—A Longitudinal Study.”


58. Kulczycki et al., “Multilingual Publishing in the Social Sciences and Humanities.”

59. Clarivate, “Clarivate Announces Changes to the 2023 Journal Citation Reports.”

60. Clarivate.


68. Milojević, “Practical Method to Reclassify Web of Science Articles.”
