

Big in Japan, Zimbabwe or Brazil – global reach and national preferences for open access books

The perceived effect of internationalization on publishing is that there is a strong focus on global issues written in English only. In academic book publishing – strongly connected to the humanities and social sciences (HSS) – languages other than English play an important role. Non-English academic publications have been linked to regional issues: there is a tension between English as the ‘lingua franca’ enabling a global reach versus local languages that provide a better cultural ‘fit’. This article examines the preference of global readers in a systematic manner, by examining the usage of the open access collection of the OAPEN Library. Based on the ten most downloaded books from 100 countries during a 12-month period, the focus on regional topics is measured using the number of books written in non-English languages and the amount of English language books that mention the country.

The results show a global interest in books with a regional focus. In many countries, the ten most downloaded books are written in non-English languages. Even when English language titles are part of the top ten, many mention regional concerns. The article counters the narrative of the dominance of English as the language of scholarly communication. Instead, it supports the value of bibliodiversity.

Keywords

open access; books; globalization; internationalization; OAPEN Library

Introduction

The debate about globalization and internationalization, and what they mean for the academic world,¹ will be ongoing for quite some time. Internationalization might seem to lead to a publication culture that puts more emphasis on global issues, written in one language – English. This raises the question of whether global interest causes a shift away from regional concerns or not. In an attempt to partially answer this, we will look at the preferences of a worldwide audience regarding open access books.

When the main differences in publication practices between the humanities and social sciences (HSS) and science, technology and medicine (STM) are considered, there are two aspects that almost always get mentioned: the prominence of books and the role of local languages. Not surprisingly, the work of HSS researchers written in languages other than English is most often linked to more regional concerns or a more regional community. Simply put, an author using Dutch will have a different audience in mind than an author writing in English.

The OAPEN Library contains thousands of open access monographs in more than 50 languages.² Launched in 2010, it is managed by the OAPEN Foundation. All books and chapters in the OAPEN Library are available for direct download without any cost or registration requirements for the user. At the end of 2021, the collection consisted of over 19,000 titles. The OAPEN Library is accessed globally, and during 2021 over 11 million downloads were registered.



RONALD SNIJDER

Deputy Director
OAPEN Foundation
The Netherlands

‘Internationalization might seem to lead to a publication culture that puts more emphasis on global issues, written in one language – English’

- 2 While books in English are in the majority – over 60% of the collection – this still leaves a large collection of publications in other languages. If a global audience can freely choose from this collection, will there be a preference for global subjects, or will more regional concerns take precedence? This explorative research looks at the most popular books from 100 countries and tries to determine the level of regional interest.

Literature review

In the literature, non-English academic publications have been linked to regional issues. This is visible in a European context, where Kulczyk et al.³ investigated multilingual publications patterns in the HSS. A recent literature study by Balula and Leão⁴ concludes that a 'balanced multilingualism' is vital for a diverse academic publication landscape. Bibliodiversity (Note 1) as a way to enhance local knowledge is also the subject of Mkhize and Ndimande-Hlongwa.⁵ The authors argue that local African languages and indigenous knowledge systems are indispensable for higher education. The dominance of English versus the role of regional languages is also debated by Flowerdew and Li,⁶ who investigated the publication choices of Chinese HSS scholars. And finally, the study of Argentinian research output conducted by Chinchilla et al.⁷ links Spanish language publications to regional subjects and English language articles to more global issues.

'non-English academic publications have been linked to regional issues'

This short literature overview of academic writing reveals the tension between English as the 'lingua franca' enabling a global reach versus local languages that provide a better cultural 'fit'. Multilingualism could also be framed in terms of centre and periphery.

The literature does not discuss the readers, however. Earlier research on the usage of the OAPEN Library revealed that the readership of non-English books tends to be concentrated in certain countries. For example, the usage of German language books is skewed toward the 'DACH countries' – Germany, Austria and Switzerland.⁸ Furthermore, clusters of German, Dutch and Italian language books tend to be more downloaded by residents from the 'corresponding' countries.⁹ A recent article investigating the effects of open access at the Springer Nature e-books platform also analysed regional effects. The authors concluded that when the publication's title contained either 'Africa' or 'Latin America', this slightly increased the readership from those regions. The publication's language was not considered, possibly because almost all titles of the dataset were written in English.¹⁰

'the tension between English as the 'lingua franca' enabling a global reach versus local languages that provide a better cultural 'fit''

In this article, the preference of global readers is examined in a systematic manner. Based on the ten most downloaded books from 100 countries during a 12-month period, the focus on regional topics is measured in two ways: the number of books written in non-English languages and the amount of English language books that mention the country. In the next section, the methodology and data are further explained. The section following that will explore the results.

'no standardized procedure to measure the regional focus of open access books exists'

Methodology and data

Currently, no standardized procedure to measure the regional focus of open access books exists. Thus, it is important to describe the method used here in sufficient detail. Furthermore, to ensure maximum reproducibility, the method is based on verifiable data that is not subject to interpretation.

Selection of data

Several choices determined the data selection. Firstly, which countries should be examined? To measure global usage, it is important to create a sufficiently large group of countries. However, the number of books that have been downloaded and the total number of downloads strongly differ for each country. Usage data from the OAPEN Library was

3 captured for the period from May 2020 to April 2021. Usage varies greatly from country to country. At one end of the spectrum, people residing in the USA were responsible for over 1.1 million downloads – of thousands of titles. By contrast, during the same period, only eight downloads – of three titles – originated from Antarctica. This issue has been resolved by selecting 100 countries with the highest number of downloads.

The next choice was to determine how many books should be considered. For each country, the ten most popular titles were selected. In the event of there being an equal number of downloads of two books that qualify to be added to the ‘top ten’, both titles were listed. Given the extensive collection of titles available at the OAPEN Library, there is a high possibility that different titles are ‘popular’ in different parts of the world, and we might expect to see a little over 1,000 different titles in our dataset. This is not the case: the set of documents consists of 710 different titles. Of those, 175 titles were part of the top ten in multiple countries. Nine books are among the most popular in seven or more countries.

Measuring web usage, such as the number of books downloaded, is not straightforward. For instance, the large differences between Google Analytics and COUNTER Release 5 show that the choice of metric service has considerable consequences for what is reported.¹¹ When both are used to measure the same events – the downloads of open access books from the OAPEN Library – the variation is striking. The COUNTER Code of Practice aims to measure only ‘genuine, user-driven usage’,¹² and automatically created downloads are discarded. Google Analytics uses a less strict policy, resulting in a difference of almost 60% for the total number of downloads. The usage reported in this article is based on the COUNTER Release 5 standard, as provided by IRUS-UK.¹³

At this point, there is a group of countries, and for each country, there is a set of books. The next step is to determine if those books have a regional focus. Given the discussion on multilingualism, the next question is whether the book is written in a language other than English. When we look at the distribution of titles in English versus titles in other languages, the dataset consists of 60% English language titles. This is in line with the total collection of the OAPEN Library at that period, which consists of 62% English language titles. See Table 1 for more details.

Publication period	Language		Total
	English only	Non-English	
Before 2000	7	20	27
Between 2000 and 2009	93	24	117
Between 2010 and 2019	342	132	474
2020 and beyond	65	27	92
Total	507	203	710

Table 1. Language and publication period of the books in the data set

As will be visible in the results section, the most popular titles of many countries are written in one of the national languages. As the current collection of the OAPEN Library is skewed towards English, this might lead to situations where there are fewer than ten books in the national language. Only selecting countries where the ten most popular books are written in the national language would lead to a result that is biased towards English. Hence the selection of 100 countries.

‘the most popular titles of many countries are written in one of the national languages’

Using text mining techniques on English language books

Most of the titles examined are written in English. For these titles, text mining techniques are deployed to check whether the name of the country is mentioned frequently. The text mining method used in this article¹⁴ is applied for English only. It is based on filtering out all sets of three consecutive words – trigrams – that contain either the commonest English words or terms that are regularly used in academic book publishing. The remaining trigrams are analysed to find the subjects discussed in the book. Of course, the words to be filtered out differ for each language. Unfortunately, these word collections were not available for all the other languages in the dataset.

4 The full text of the English language books was used to find terms that are specific for that document. The algorithm splits the full text of the book or chapter into trigrams. Then it removes all trigrams containing words that are commonly used in everyday language and the trigrams containing terms that are commonly used in (open access) book publishing. When a trigram contains a word – or multiple words – that is commonly used, the whole trigram is discarded.

The goal is to find trigrams that contain words specific to that publication. The remaining trigrams are distinctive to the book or chapter and selecting those that occur most frequently indicates the concepts under discussion. The algorithm is depicted in Figure 1.

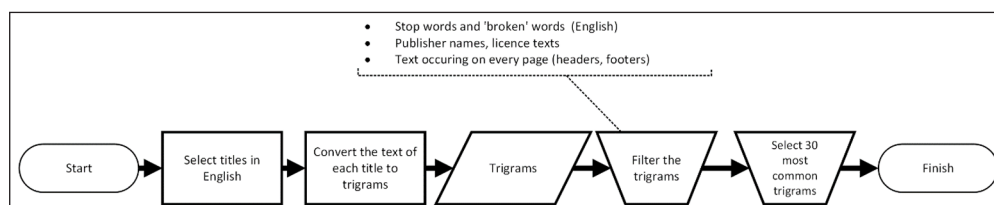


Figure 1. Text mining algorithm

The result is a set of 90 words in the English language (30 trigrams). When an English language book has been downloaded in a certain country, the next step is to search for the name of that country. To allow for multiple variants, the country name is truncated. For example, instead of searching for 'Belgium', the term 'belg' is used. This enables us to find variants such as 'Belgian'. If the truncated term is found, the book is marked as discussing the country. These selected trigrams are part of the dataset.

When the content of a book has been reduced to 90 words, obviously the chance that the country's name is among them is not high. Furthermore, the region is strictly defined as a country, and no larger or smaller areas are considered. Regional concerns might also be defined by other keywords, such as local traditions or beliefs, but this requires the interpretation of what constitutes a regional concern. By sticking to a strict algorithm – reducing the text to 30 trigrams and then searching for the country name – the results are unambiguous and verifiable.

Scoring the books

The language check and the text mining techniques allow us to count how many of the ten most used books per country have a regional focus, either by looking at the book's language or – for English language books – whether the country in question has been mentioned regularly. As a result, we can easily assess for each country how strong the focus on regional issues is and express it as a score from 0 to 10.

'By sticking to a strict algorithm ... the results are unambiguous and verifiable'

Each book that is written in a language other than English and each English language book that mentions the country is counted. The total number of books per country that fall into either of these categories is the score used in this article. Looking at the total scores, it is interesting to see that no usage of regionally focused books could be found for only seven of the 100 countries. In the next section, the results for the 100 countries are described.

The dataset is freely available using the link in the data accessibility statement at the end of the article.

Results: global interest in regional issues

Here we will discuss the results for the different countries listed by continent.

North America

There are scores for four countries in North America with a clear difference between the Dominican Republic and Mexico versus Canada and the United States. For the first two countries, the majority of the top ten books were written in Spanish. In contrast, all books

5 listed for the United States and nine out of the ten books most popular in Canada are English. The one non-English book is partly English, partly French. Three of the English language books mention Canada; in the USA, four popular books mention the country. Figure 2 and Table 2 contain more details.



Figure 2. North America – book scores

Country	National languages	Total downloads of top ten titles	Non-English books	English books mentioning country	Total score
Dominican Republic	Spanish	1,139	8	1	9
Mexico	Spanish	8,322	5	1	6
Canada	English, French	7,139	1	3	4
United States	English	2,6147	0	4	4

Table 2. North America – languages, downloads and book scores

South America

In six countries in South America, the score is eight or higher. The most profound outlier is Brazil, with a score of only two books. This is clearly visible in Figure 3. As was to be expected, the role of English language books is marginal. Only three countries – Peru, Ecuador and Costa Rica – have one English title as part of their scoring list. There is a definite preference for other languages, in this case, Spanish. More details can be found in Table 3.



Figure 3. South America – book scores

Country	National languages	Total downloads of top ten titles	Non-English books	English books mentioning country	Total score
Argentina	Spanish	1,549	10	0	10
Chile	Spanish	3,271	10	0	10
Bolivia	Spanish	989	9	0	9
Colombia	Spanish	3,663	9	0	9
Peru	Spanish	3,366	8	1	9
Ecuador	Spanish, Quechua	2,646	7	1	8
Costa Rica	Spanish	485	5	1	6
Brazil	Portuguese	3,807	2	0	2

Table 3. South America – languages, downloads and book scores

Europe

The dataset contains 37 European countries. We will discuss this relatively large group as three subsections: countries with a score from eight to ten, countries with a score from five to seven and countries with a score of four and lower. Within the first subsection, the role of English is marginal: other languages dominate the top ten almost completely. The most popular books downloaded in France were written in French and German, with only two in English. The preference for books in the national language is clearly visible in the Netherlands (Dutch), Norway (Norwegian), Germany and Austria (German). This is also the case for Switzerland, with seven German and one Italian language titles. In a similar vein, the most popular titles in Italy are written in Italian. The majority of the most downloaded books in Finland are also in the national language. The most popular books in Belarus, Ukraine and the Russian Federation are written in Russian.

The subsection of countries with a score between five and seven shows a slightly different pattern: here the role of non-English language is still important, but this group also contains more English language books that mention the country. For Luxembourg, Portugal and Belgium, half or more of the popular books have been published in their national languages. In this group, we also see combinations of either German or English and – in most cases – the national languages. See for instance Slovenia, Croatia and Estonia. The role of Russian is visible in the top ten of Latvia and Georgia.

The last subsection consists of countries with a score of four or lower. Within Sweden, four of the most popular titles are in the native language, the same holds true for Spain. For Denmark, the number of books in Danish is one. Another distinctive aspect of this group is the role of German. For five countries, the top ten consists of more books in German and – possibly – one or more English language books mentioning the country. This applies to Serbia, Hungary, Romania, Lithuania and Slovakia. In Cyprus and Great Britain, two English language books mention the country. The other European countries display a more diverse choice of languages. Here we see combinations of German, Russian, Chinese, Spanish, Albanian and French plus one or more English language books. The geographical differences are depicted in Figure 4 and in Table 4.

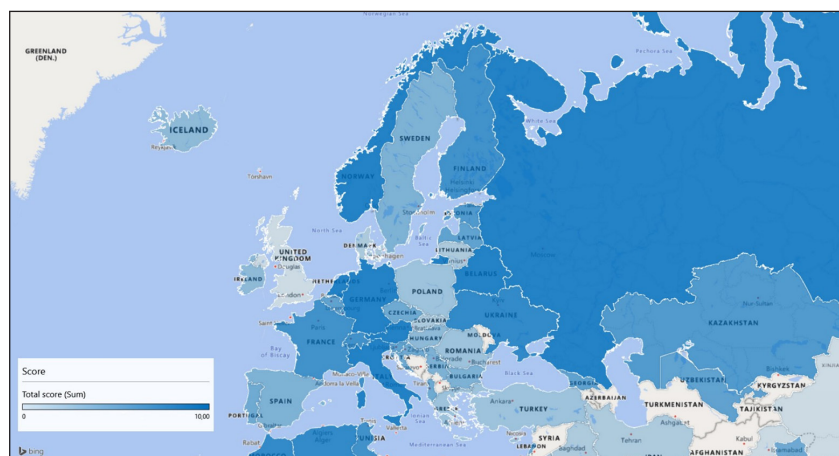


Figure 4. Europe – book scores

Country	National languages	Total downloads of top ten titles	Non-English books	English books mentioning country	Total score
Austria	German	6,536	10	0	10
Belarus	Belarusian, Russian	1,182	10	0	10
Germany	German	28,112	10	0	10
Italy	Italian	17,886	10	0	10
Netherlands	Dutch	18,629	10	0	10
Norway	Norwegian	5,515	10	0	10
Russian Federation	Russian	11,477	10	0	10
Ukraine	Ukrainian	3,364	9	0	9
Finland	Finnish, Swedish	6,482	8	0	8
France	French	10,300	8	0	8
Switzerland	French, German, Italian, Romansh	2,808	8	0	8
Luxembourg	French, German, Luxembourgish	341	7	0	7
Estonia	Estonian, Russian	1,016	1	6	7
Slovenia	Slovene	670	6	0	6
Belgium	Dutch, French, German	6,542	5	1	6
Latvia	Latvian	541	5	1	6
Georgia	Georgian	466	4	2	6
Albania	Albanian	993	5	0	5
Czech Republic	Czech, Slovak	1,181	5	0	5
Portugal	Portuguese	2,759	5	0	5
Croatia	Croatian	618	2	3	5
Spain	Spanish	2,626	4	0	4
Sweden	Swedish	4,739	4	0	4
Bulgaria	Bulgarian	376	3	1	4
Serbia	Serbian	514	2	2	4
Hungary	Hungarian	1,116	3	0	3
Iceland	Icelandic	374	1	2	3
Ireland	Irish, English	4,851	1	2	3
Greece	Greek	1,717	2	0	2
Poland	Polish	2,588	2	0	2
Turkey	Turkish	2,954	2	0	2
Romania	Romanian	865	1	1	2
Cyprus	Greek, Turkish, English	421	0	2	2
United Kingdom	English	23,643	0	2	2
Denmark	Danish	1,967	1	0	1
Lithuania	Lithuanian	518	1	0	1
Slovakia	Slovak	399	1	0	1

Table 4. Europe – languages, downloads and book scores

Africa

In Africa, French plays an important role: by far, most of the non-English books that are part of our dataset are written in this language. For instance, all top ten books downloaded from Cote d'Ivoire are French. The same holds true for the nine books connected to Senegal and the eight books linked to Tunisia, Cameroon and Algeria. However, there are two additional English books mentioning Cameroon, and one Arabic title is linked to Algeria. In Egypt, we see two English titles, one German and one Arabic language title. More details can be found in Table 5.

In contrast, all nine books that mention South Africa are written in English. This is not the only country where English is the main language as far as the most popular books are concerned. We see the same pattern for Ethiopia, Nigeria, Kenya, Malawi and Zimbabwe. And lastly, Namibia and Uganda, who both list two titles mentioning their country.

In Africa, we encounter the first countries from the top 100 where there is no regional connection to any of their most downloaded books: Ghana, Tanzania and Zambia. Figure 5 depicts the book scores and countries that are not part of the data set.

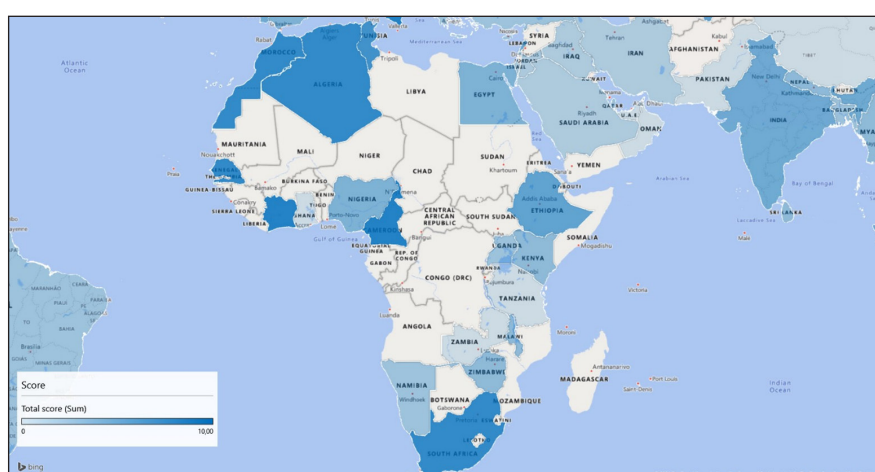


Figure 5. Africa – book scores

Country	National languages	Total downloads of top ten titles	Non-English books	English books mentioning country	Total score
Cote d'Ivoire	French	1,230	10	0	10
Cameroon	English, French	2,686	8	2	10
Algeria	Arabic, Tamazight, French	2,215	9	0	9
Senegal	French	1,032	9	0	9
Morocco	Arabic, Tamazight, French	1,603	7	2	9
South Africa	Afrikaans, English	18,975	0	9	9
Tunisia	Arabic	1,197	8	0	8
Ethiopia	Afar, Amharic, Oromo, Somali, Tigrinya	4,187	0	6	6
Egypt	Arabic	2,274	2	2	4
Nigeria	English	5,681	0	4	4
Kenya	English	3,124	0	3	3
Malawi	English, Chichewa	726	0	3	3
Zimbabwe	English	1,506	0	3	3
Namibia	English, Afrikaans	2,688	0	2	2
Uganda	English, Swahili	1,399	0	2	2
Ghana	English	2,051	0	0	0
Tanzania	English	2,181	0	0	0
Zambia	English	2,333	0	0	0

Table 5. Africa – languages, downloads and book scores

Asia

Contrary to the scores from the other continents, the highest score is seven. The only country reaching this score is Kazakhstan. Together with Uzbekistan – with a score of six – these countries list only titles in Russian. See Table 6 for more details.

Moving slightly further east, we see that another non-English language – Arabic – is amongst the most popular titles in Jordan, Saudi Arabia, Qatar and Iraq. The top ten lists of both Qatar and Iraq feature one English language title with a national focus, while the score for the United Arab Emirates and Oman is zero. In Lebanon, two French and two country-focused English titles can be counted. Israel and Iran do not feature non-English books in their top ten.

For the next group of countries, the most popular titles are all written in English. The number of books with national focus ranges from five for India to four for Sri Lanka and Nepal, three for Myanmar and Vietnam and two for Bangladesh, Thailand and Cambodia. There are no titles in Pakistan’s top ten list that mention the country.

In Japan, the top ten contains one book in Japanese and five English language titles. Most other countries in the region have a lower score. For some countries, only English books are part of the most downloaded titles: Indonesia – with a relatively high score of five titles – together with South Korea, Singapore and Malaysia. A combination of English with another language can be found in Taiwan, Hong Kong and the Philippines. The score for China is zero. See Figure 6 for an overview.



Figure 6. Asia – book scores

Country	National languages	Total downloads of top ten titles	Non-English books	English books mentioning country	Total score
Kazakhstan	Kazakh, Russian	1,989	7	0	7
Uzbekistan	Uzbek	873	6	0	6
Japan	Japanese	2,004	1	5	6
Lebanon	Arabic, English, French	1,139	2	3	5
India	Hindi, English	20,037	0	5	5
Indonesia	Indonesian	5,941	0	5	5
Sri Lanka	Sinhala, Tamil	2,620	0	4	4
Hong Kong	Cantonese, English	2,546	1	2	3
Myanmar	Burmese	971	0	3	3
Nepal	Nepali	542	0	3	3

(Contd.)

Country	National languages	Total downloads of top ten titles	Non-English books	English books mentioning country	Total score
Vietnam	Vietnamese	1,544	0	3	3
Qatar	Arabic	216	2	0	2
Iraq	Arabic, Kurdish	2,354	1	1	2
Philippines	Filipino, English	20,192	1	1	2
Bangladesh	English	1,285	0	2	2
Cambodia	Khmer	672	0	2	2
Malaysia	Malay	2,630	0	2	2
Israel	Hebrew	1,844	0	2	2
Korea, Republic of	Korean	1,520	0	2	2
Thailand	Thai	1,660	0	2	2
Jordan	Arabic	280	1	0	1
Saudi Arabia	Arabic	975	1	0	1
Taiwan	Mandarin Chinese	1,799	1	0	1
Iran, Islamic Republic of	Persian	888	0	1	1
Singapore	English, Malay, Mandarin Chinese, Tamil	1,470	0	1	1
China	Mandarin Chinese	2,813	0	0	0
Oman	Arabic	207	0	0	0
Pakistan	Urdu, English	6,497	0	0	0
United Arab Emirates	Arabic	751	0	0	0

Table 6. Asia – languages, downloads and book scores

Oceania

In Oceania, all books in the top ten are written in English. Fiji’s top ten contains nine titles that mention the country. The score for Australia is seven, and Papua New Guinea lists six books with a national focus. Compared to the other countries, New Zealand’s score of four is moderately low. More details can be found in Table 7 and Figure 7.



Figure 7. Oceania – book scores

Country	National languages	Total downloads of top ten titles	Non-English books	English books mentioning country	Total score
Fiji	English, Fijian, Fiji Hindi	2,377	0	9	9
Australia	English	10,651	0	7	7
Papua New Guinea	English	4,085	0	6	6
New Zealand	English, Māori	2,654	0	4	4

Table 7. Oceania – languages, downloads and book scores

Books, popular in multiple countries

In the Methodology and data section, we mentioned books that are among the most popular in several countries. Of these, the book *Professional Learning in Education*¹⁵ is widely downloaded in twelve countries. This English language title does not mention any of the countries. *How the World Changed Social Media*,¹⁶ also written in English, is listed in ten countries but does not touch upon them. The following books are both listed in eight countries: the Spanish language *Tratamiento biológico de aguas residuales*¹⁷ and *Marine Anthropogenic Litter*.¹⁸ Also, these English language titles have not focused on any of the eight countries. Perhaps not surprisingly, the subjects of these books are important to many people: education, social media, wastewater treatment and marine pollution.

Figure 8 gives an overview of the ten most highly connected books – titles that are popular in seven countries or more. The figure shows publications that are listed in the top ten of several countries at once, for instance *Tratamiento biológico de aguas residuales* and *Manifiesto del Nuevo Realismo*.¹⁹ Both titles are among the top ten of Argentina, Costa Rica and Mexico.

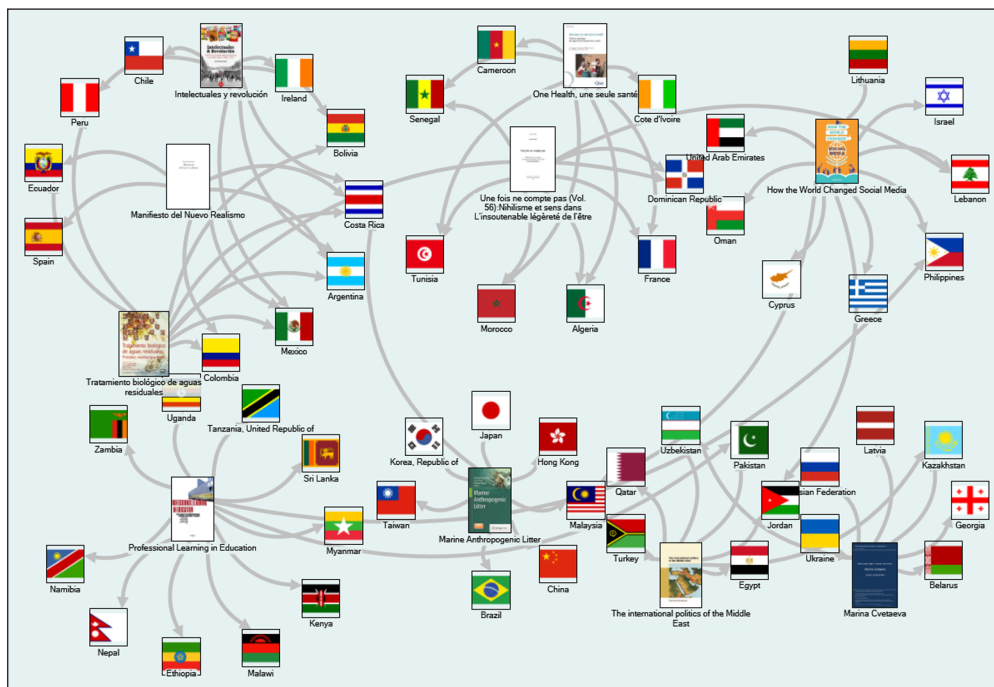


Figure 8. Highly connected books

Conclusion

It might sound contradictory, but the results show a global interest in books with a regional focus. Given a choice from a freely accessible collection, many readers will download a book that either diverts from the clichéd English language monograph or – when the book is written in English – discusses issues that are relevant for the country of

12 the reader. A systematic focus on the most popular books for many countries helps to identify patterns that may have otherwise remained hidden. Using the method described in this article, there is no interference from the large number of downloads from countries in the global north.

The outcomes of this study do not fit in a story of English language publications as the only or the main source of scholarly communication. This is clearly visible in Austria, Belarus, Germany, Italy, the Netherlands, Norway and the Russian Federation, where the top ten most downloaded books are written in the official language of the country. A comparable outcome can be found for Finland: eight of the most downloaded books are written in Finnish, while the total amount of Finnish language books is less than 1% of the OAPEN Library collection. Looking at Cote d'Ivoire, Cameroon, Algeria, Senegal, Morocco and Tunisia, we see a clear preference for French. In these countries, French is either one of the official languages or it is widely used. Spanish is the national language throughout South America – with the exception of Brazil. Again, the use of English is marginal in these countries.

'the results show a global interest in books with a regional focus'

Even when English language titles are part of the top ten, many mention regional concerns. The scores for English are most likely an under-estimate, as it is based on country name only. The method to determine the regional focus does not consider transnational regions, such as the Maghreb. On top of that, books focusing on smaller areas might be excluded. For instance, a monograph whose subject is the city of Jakarta instead of Indonesia would not be marked as 'country focused'. It is interesting to note that in most countries where English is an official or national language, the scores are relatively low. It does not seem likely that in English-speaking countries there will be less interest in topics of national concern. Instead, the methodology based on the text mining results is probably too 'strict'.

'Even when English language titles are part of the top ten, many mention regional concerns'

The results of this study depended on the titles available in the OAPEN collection. The lack of certain subjects or languages especially could have played a role: how much demand has not been met? An even more diverse collection might have led to a different outcome. Another angle could be the distinction between books about HSS subjects versus STM subjects, although the number of titles might differ, and whether there is also a significant difference in usage. Conducting similar research using another open access book platform with a different collection would be welcome. For instance, by analysing the global usage of the SciELO Books platform,²⁰ where 96% of the titles are written in Portuguese or the OpenEdition Books platform,²¹ where 85% of the collection is French.

In conclusion, this article clearly indicates a demand for regionally focused titles. Furthermore, it counters the narrative of the dominance of English as the language of scholarly communication. Instead, it supports the value of bibliodiversity.

Note

1. A definition of bibliodiversity coined by Giménez Toledo et al. reads: 'Bibliodiversity, the diversity of academic content, both at the national and international level is essential for preserving research in a wide range of global and local topics, studied from different epistemic and methodological approaches, inspired by various schools of thought and expressed in a variety of languages.' – Elea Giménez Toledo et al., "Bibliodiversity-What It Is and Why It Is Essential to Creating Situated Knowledge," *Impact of Social Sciences Blog*, 2019. http://eprints.lse.ac.uk/103440/1/impactofsocialsciences_2019_12_05_bibliodiversity_what_it_is_and_why.pdf (accessed 25 May 2022).

Data accessibility statement

The dataset is freely available at <https://doi.org/10.5281/zenodo.6501839>.

Acknowledgements

The author would like to thank his OAPEN colleagues for commenting on a previous version of this article.

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 A&As' link: <http://www.uksg.org/publications#aa>.

Competing interests

The author has declared no competing interests.

References

1. Philip G. Altbach, "Globalisation and the University: Myths and Realities in an Unequal World," *Tertiary Education & Management* 10, no. 1 (2004): 3–25, DOI: <https://doi.org/10.1080/13583883.2004.9967114> (accessed 26 May 2022).
2. OAPEN Foundation, *OAPEN Foundation Stakeholder Report 2021: Key Results and Developments*, The Hague, 2022, <https://oapen.fra1.digitaloceanspaces.com/90b36d1e9131442480030e48076ac263.pdf> (accessed 25 May 2022).
3. Emanuel Kulczycki, et al., "Publication Patterns in the Social Sciences and Humanities: Evidence from Eight European Countries," *Scientometrics* 116, no. 1 (1 July 2018): 463–86, DOI: <https://doi.org/10.1007/s11192-018-2711-0> (accessed 25 May 2022); Emanuel Kulczycki et al., "Multilingual Publishing in the Social Sciences and Humanities: A Seven-Country European Study," *Journal of the Association for Information Science and Technology* 71, no. 11 (2020): 1371–85, DOI: <https://doi.org/10.1002/asi.24336> (accessed 26 May 2022).
4. Ana Balula and Delfim Leão, "Multilingualism within Scholarly Communication in SSH. A Literature Review," *JLIS.It* 12, no. 2 (15 May 2021): 88–98, DOI: <https://doi.org/10.4403/jlis.it-12672> (accessed 25 May 2022).
5. Nhlanhla Mkhize and Nobuhle Ndimande-Hlongwa, "African Languages, Indigenous Knowledge Systems (IKS), and the Transformation of the Humanities and Social Sciences in Higher Education," *Alternation* 21, no. 2 (2014): 10–37, https://www.researchgate.net/profile/Cheryl-Potgieter/publication/328928144_Humanities_Knowledge_Production_and_Transformation/links/5bebe025299bf124fd12257/Humanities-Knowledge-Production-and-Transformation.pdf#page=15 (accessed 25 May 2022).
6. John Flowerdew and Yongyan Li, "English or Chinese? The Trade-off between Local and International Publication among Chinese Academics in the Humanities and Social Sciences," *Journal of Second Language Writing* 18, no. 1 (1 March 2009): 1–16. DOI: <https://doi.org/10.1016/j.jslw.2008.09.005> (accessed 25 May 2022).
7. Zaida Chinchilla-Rodríguez, Sandra Miguel, and Félix de Moya-Anegón, "What Factors Affect the Visibility of Argentinean Publications in Humanities and Social Sciences in Scopus? Some Evidence beyond the Geographic Realm of Research," *Scientometrics* 102, no. 1 (1 January 2015): 789–810, DOI: <https://doi.org/10.1007/s11192-014-1414-4> (accessed 25 May 2022).
8. Ronald Snijder, "Evaluating the Impact of the FWF-E-Book-Library Collection in the OAPEN Library: An Analysis of the 2014 Download Data," *D-Lib Magazine* 21, no. 7/8 (2015), DOI: <https://doi.org/10.1045/july2015-snijder> (accessed 25 May 2022).
9. Ronald Snijder, "Patterns of Information – Clustering Books and Readers in Open Access Libraries," In *The Deliverance of Open Access Books: Examining Usage and Dissemination*, (Amsterdam University Press, 2019) 83–103, DOI: https://doi.org/10.26530/OAPEN_1004809 (accessed 30 May 2022).
10. Cameron Neylon et al., "More Readers in More Places: The Benefits of Open Access for Scholarly Books," *Insights* 34, no. 1 (1 December 2021): 27, DOI: <https://doi.org/10.1629/uksg.558> (accessed 25 May 2022).
11. Ronald Snijder, "Open Access Book Usage Data – How Close Is COUNTER to the Other Kind?," *Insights* 34, no. 1 (28 April 2021): 9, DOI: <https://doi.org/10.1629/uksg.539> (accessed 25 May 2022).
12. "7.0 Processing Rules for Underlying COUNTER Reporting Data," *COUNTER*, accessed 28 October 2020, <https://www.projectcounter.org/code-of-practice-five-sections/7-processing-rules-underlying-counter-reporting-data/> (accessed 25 May 2022).
13. Snijder, "Open Access Book Usage Data – How Close Is COUNTER to the Other Kind?"
14. Ronald Snijder, "Words Algorithm Collection – Finding Closely Related Open Access Books Using Text Mining Techniques," *LIBER Quarterly: The Journal of the Association of European Research Libraries* 31, no. 1 (24 August 2021) DOI: <https://doi.org/10.53377/lq.10938> (accessed 25 May 2022).
15. Antonia Aelterman et al., eds., *Professional Learning in Education* (Academia Press, 2016), <https://library.oapen.org/handle/20.500.12657/31077> (accessed 25 May 2022).
16. Daniel Miller et al., *How the World Changed Social Media* (UCL Press, 2016), DOI: <https://doi.org/10.14324/111.9781910634493> (accessed 25 May 2022).
17. H.A. Hernández García et al., eds., *Tratamiento biológico de aguas residuales: principios, modelación y diseño* (IWA Publishing, 2017), DOI: <https://doi.org/10.2166/9781780409146> (accessed 25 May 2022).
18. Melanie Bergmann, Lars Gutow, and Michael Klages, eds., *Marine Anthropogenic Litter* (Springer Nature, 2015), DOI: <https://doi.org/10.1007/978-3-319-16510-3> (accessed 25 May 2022).
19. Hernández García et al., *Tratamiento biológico de aguas residuales*; Maurizio Ferraris, *Manifiesto del Nuevo Realismo* (Ariadna Ediciones, 2012), DOI: https://doi.org/10.26530/OAPEN_617573 (accessed 25 May 2022).
20. SciELO Livro, <https://books.scielo.org/en/> (accessed 25 May 2022).
21. OpenEdition Books, <https://books.openedition.org/> (accessed 25 May 2022).

Article copyright: © 2022 Ronald Snijder. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use and distribution provided the original author and source are credited.



Corresponding author:

Ronald Snijder

Deputy Director

OAPEN Foundation, NL

E-mail: r.snijder@oapen.org

ORCID ID: 0000-0001-9260-4941

To cite this article:

Snijder R, "Big in Japan, Zimbabwe or Brazil – global reach and national preferences for open access books," *Insights*, 2022, 35: 11, 1–15; DOI: <https://doi.org/10.1629/uksg.580>

Submitted on 11 April 2022

Accepted on 09 May 2022

Published on 29 June 2022

Published by UKSG in association with Ubiquity Press.