

Open access books: A global preference for regional subjects

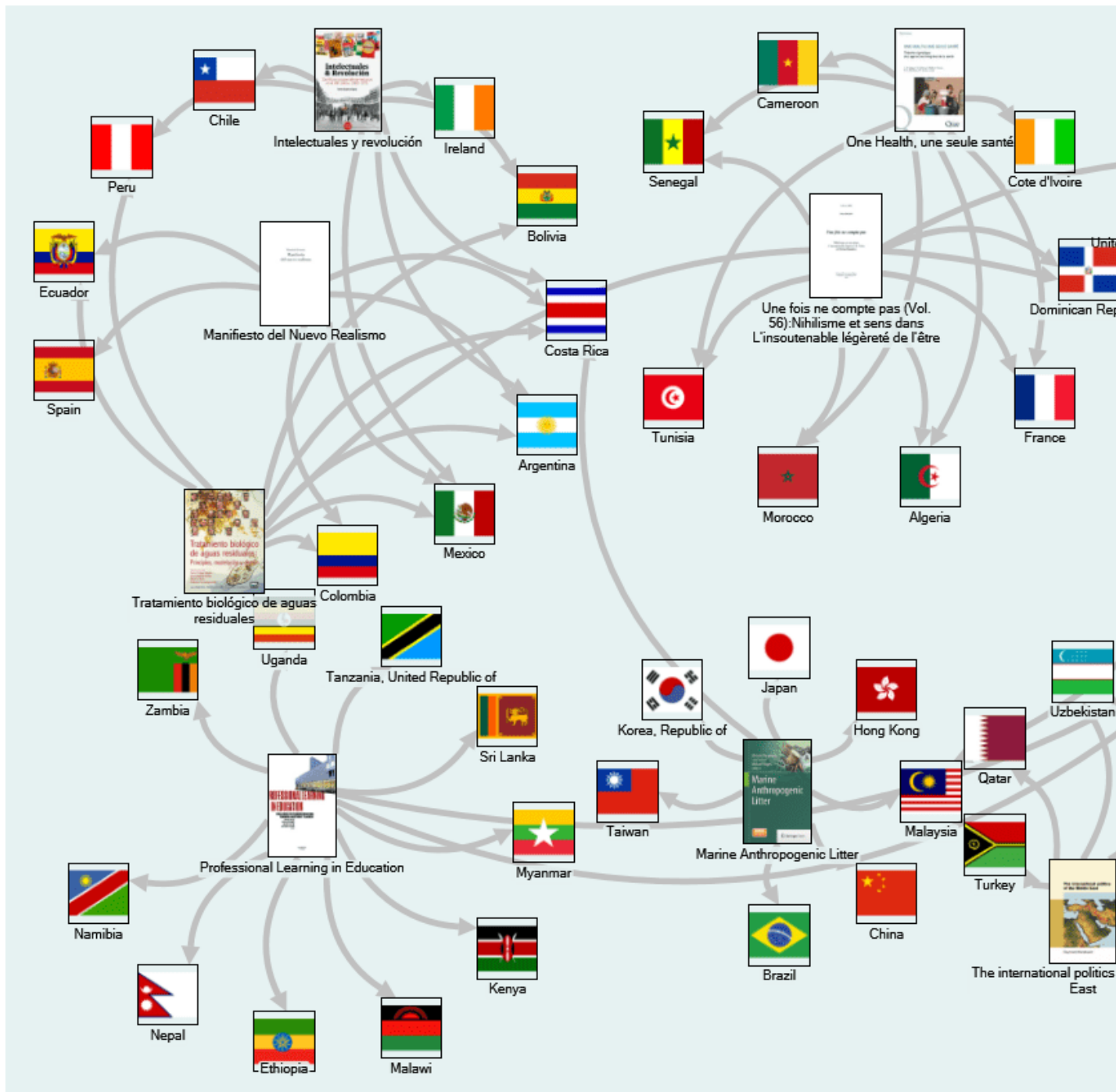
*For many research disciplines English functions as the global language for research. But, how far does this align with patterns of research use globally? Drawing on download evidence from the OAPEN library of open access books, **Ronald Snijder** explores this global demand for open research and finds significant demand for regional research and research published in languages other than English.*

Like everywhere else, internationalisation and globalisation are part of the academic world. As a result, many authors, especially in STEM disciplines publish articles in English. We tend to think of publications in English as aimed at a global audience; publications in other languages will probably have a more regional topic. Authors writing in Dutch, German or Finnish are most likely looking for a different audience than those using English. In the humanities and parts of the social sciences, books instead of articles are the most important medium and many of those are not written in English. In other words: there is more [bibliodiversity](#).

Much research is available on scholarly authors and their choice to publish books in English or in other languages. But what about the readers? Are they also mainly interested in global, English language titles or do they prefer books whose subjects may be closer to home? In an attempt to find an answer to this question I looked at the usage data of the [OAPEN Library](#). It contains over 20,000 open access books and chapters in more than 50 languages. All the publications are available for direct download without any cost or registration requirements. Readers from all over the globe make use of it: in 2021 there were over 11 million downloads.

Looking at the downloads per country, there are large differences. At one end of the spectrum, readers from the United States alone have been responsible for 1.1 million downloads. At the other end, we find eight downloads coming from Antarctica. To create a useful data set, 100 countries with the highest number of downloads were selected. For each of those countries, the ten most popular titles were chosen. Several books are popular in more countries, leading to a set of 710 different titles where 40% have been written in other languages than English.

Some books have a truly global appeal, and this graphic shows the ten most downloaded books in seven countries or more:

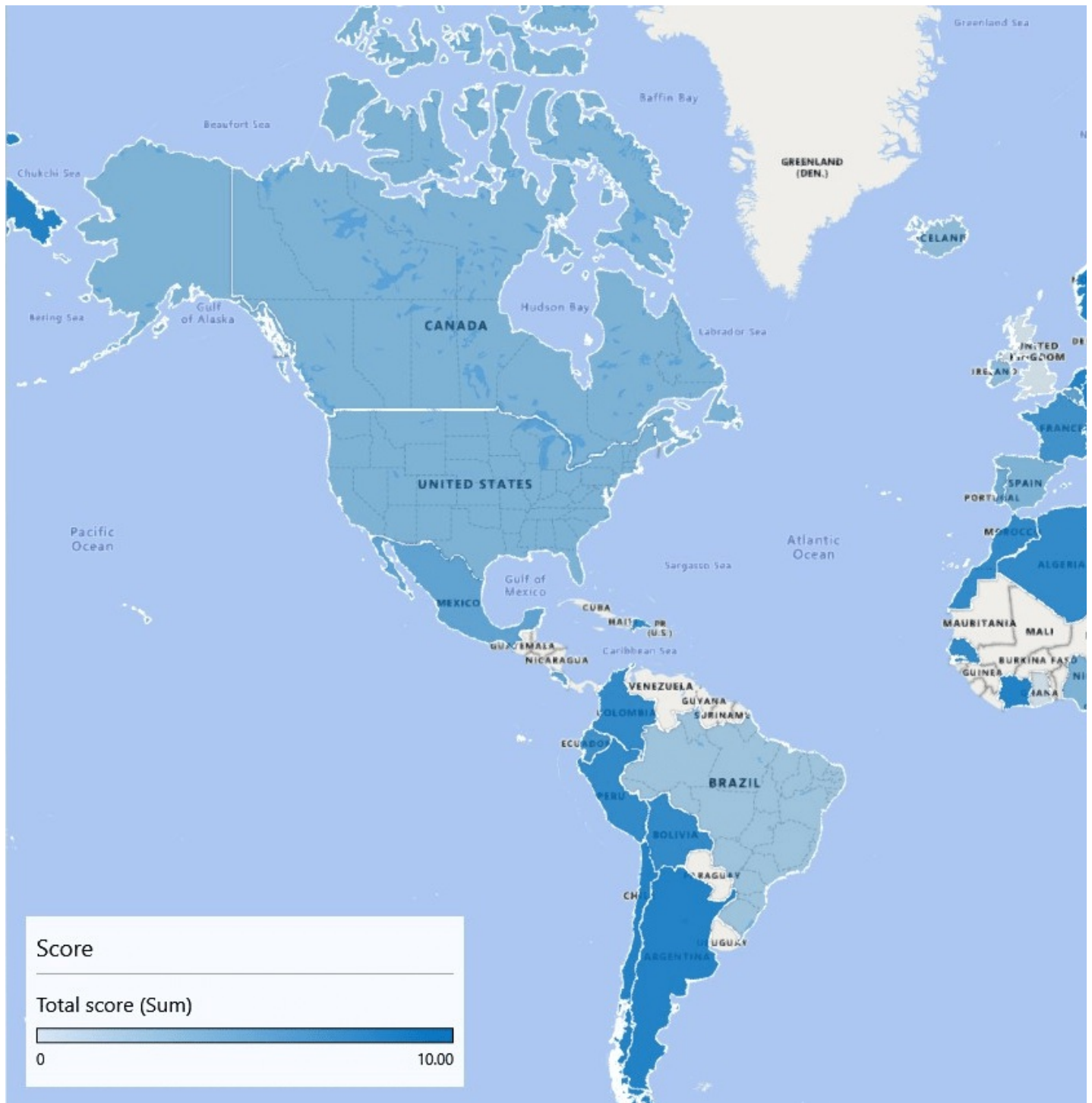


The data can be used to create a score from 0 to 10 for each country. We can add one point for each book published in other languages than English. In many countries, the most popular titles are published in the national language or in a language that is widely used. For instance, in South American countries – with the exception of Brazil – the most popular books are in Spanish. National languages also dominate the top ten of Scandinavian, Western European countries, Russia and Belarus. French is much read in several African countries.

Scoring only books written in other languages does not tell the complete story. It seems rather unlikely that readers in countries where English is the common language are not interested in more local affairs. So, the next step is to determine how many of the English language books are focusing on local instead of global affairs. This was done using text mining techniques. One of the advantages of open access is that the complete text of the books and chapters is completely available and can be read by a computer. An algorithm splits the full text into sets of three consecutive words: trigrams. The next steps are to remove all trigrams that contain commonly used words, and to select trigrams that are specific for that publication. Finally, when an English language book has been downloaded in a certain country, the next phase is to search for the name of that country.

This analysis reveals that the top ten of many countries contains English language books with a local focus. For instance in South Africa, India, Indonesia, Australia or Fiji, at least half of the most popular books fall into this category. The advantage of using a text mining algorithm is that the results are unambiguous and verifiable; it leaves no room for different interpretations. There is also a disadvantage: larger or smaller areas than countries are taken not into account. Also, regional concerns might also be defined by other keywords, such as local traditions or beliefs. As a result, the scores for English are most likely an underestimation.

When all the scores are displayed in one map, it looks like this:



It might sound contradictory, but the results show a global interest in books with a regional focus. Given the 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 – it discusses issues that are relevant for the country of the reader. A systematic focus on the most popular books for many countries helps to identify patterns that may have remained hidden. Using this method, there is no interference of the large number of downloads from countries in the global north. More details – including the data set – can be found at <https://doi.org/10.1629/uksg.580>.

In conclusion, while internationalisation and globalisation are a fact of life, the evidence shows that many people like to read about subjects that are closer to home and preferably in their own language.

The content generated on this blog is for information purposes only. This Article gives the views and opinions of the authors and does not reflect the views and opinions of the Impact of Social Science blog (the blog), nor of the London School of Economics and Political Science. Please review our [comments policy](#) if you have any concerns on posting a comment below.

Image Credit: All images reproduced with permission of the author.
