



Andreas Nishikawa-Pacher

October 18th, 2022

The dream of 'editormetrics' – Why a FAIR dataset of journal editors would benefit all researchers

2 comments | 23 shares

Estimated reading time: 6 minutes



*Editors are among the most powerful actors in the scientific community. By deciding which papers (not) to publish, they can influence public discourse and nurture – or obstruct – academic careers. However, there is little available information about aggregate patterns of scholarly journal editorships. This may change soon, as **Andreas Nishikawa-Pacher** writes, thanks to a novel dataset created in collaboration with Kerstin Shoch and Tamara Heck that provides new insights into the landscape of journal editing.*

Perhaps you have heard that some editors of scientific journals misuse their position to favour their own students, thereby circumventing the competitive nature of the scholarly publication system. Or maybe you have read of a **pervasive underrepresentation of women** and minorities in editorial boards. Or you have come across **researchers who prolifically publish in their own journals**. Whilst (hopefully) not the norm, these issues undermine the impartiality of the academic system with its (usually anonymous) peer-review procedures and highlight the important role editors play in shaping the scholarly record.



data about editors are not “closed” – journals usually list them on their websites – neither are they “open”



Such stories about scientific gatekeepers, however, often remain anecdotal, or the evidence remains limited to single-case studies, to specific sub-disciplines, to a narrow range of journals. The aggregate extent of such patterns across the wider scientific system remains unknown. Ideally, one could uncover such potentially unethical activities with large-scale data about editorial boards in a highly structured format. Names and ORCID and affiliations could then be connected *en masse* to broad publication patterns to detect anomalies. However, such “**editormetric**” investigations can hardly be conducted. While data about editors are not “closed” – journals usually list them on their websites – neither are they “open” in the sense that approximates the **FAIR** principles of open data: they are not trivially findable (F), accessible (A), interoperable (I) and re-useable (R) on a grand scale. Instead, they are scattered across tens of thousands of journal websites in different formats so that one would have to collect the data manually – a dauntingly laborious, time-consuming task.

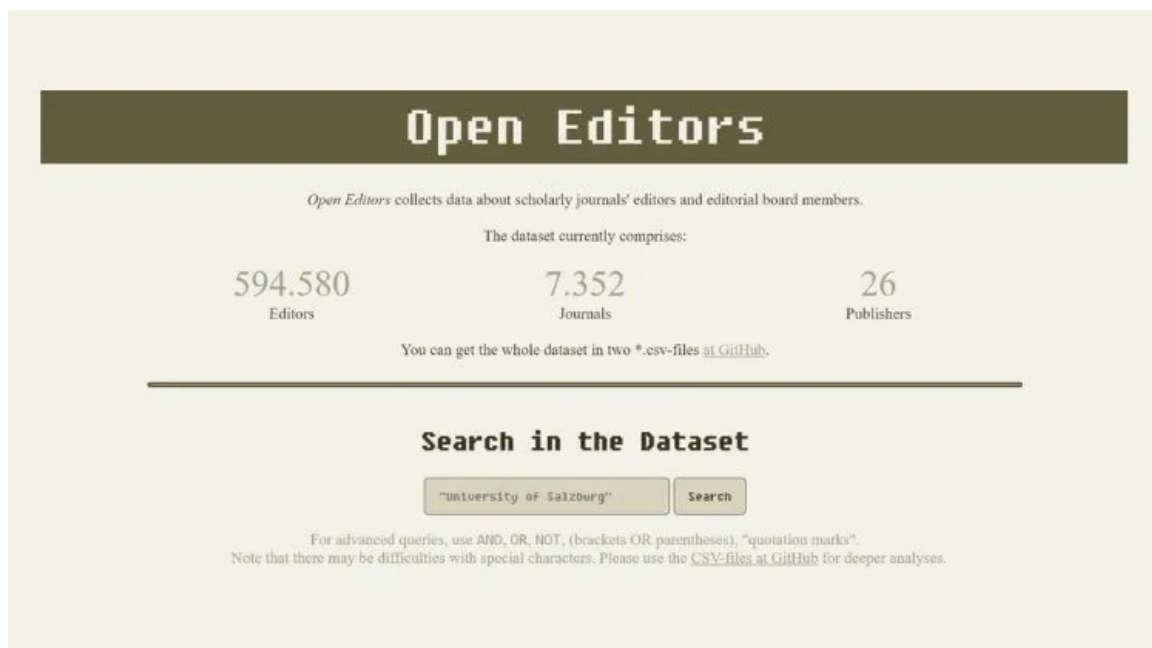
Open Editors: A second-best solution

A second-best solution would be to try to webscrape data about editors from the websites of the journals. This is what we did with the project “Open Editors” (funded by Wikimedia Deutschland’s Open Science Programme), about which we recently published a **data paper**. Scripts were programmed that accessed the websites of more than 7.000 journals across 26 publishers so as to gather data about more than half a million editorial board members.

A dedicated website was then set up so that anyone could search in the database, such as by typing in an affiliation. **A search for “London School of Economics” lists 455 editorial board memberships**, for example, from “Chief Editors” and “Honorary Editors” to “Book Review Editors” and “Associate Editors”.

This example already shows that the dataset can not only be used for finding unethical conduct, but also for many other, positive uses. Since our preprint was put online two years ago ([see the coverage in Nature Index](#)), academic publishers and university librarians have used "Open Editors" to find peer-reviewers, to organize a meetup of local editors, or simply to get an overview of a given institute's community engagement beyond mere paper outputs.

The broad numbers of a descriptive statistics convey interesting findings. We have already hinted at the various labels of editorial roles – the total dataset contains a whopping 4.024 different labels for editorial board roles! We also looked at the geographical distribution and found that some publishers exhibit overly high shares of Anglo-American editors. This includes eLife (64.5%), SAGE (70.7%), Cambridge University Press (72.7%), and APA (90.3%), raising questions about global diversity. (Note, however, that the frequency with which countries are mentioned in the affiliations of editors correlate positively with the countries' worldwide share of scientific output.) In general, the median journal lists 34 editors – albeit with extreme outliers like *Frontiers in Psychology* that had almost 14.000 editorial board members at the time of data-collection – representing affiliations in 11 countries. A standard deviation of 467 editors indicates, however, that scientific journals are extremely heterogeneous when it comes to the composition of their editorial boards.



Open Editors

Open Editors collects data about scholarly journals' editors and editorial board members.

The dataset currently comprises:

594.580 Editors	7.352 Journals	26 Publishers
--------------------	-------------------	------------------

You can get the whole dataset in two *.csv-files [at GitHub](#).

Search in the Dataset

For advanced queries, use AND, OR, NOT, (brackets OR parentheses), "quotation marks".
Note that there may be difficulties with special characters. Please use the [CSV-files at GitHub](#) for deeper analyses.

Unfortunately, the data remain incomplete. While we do assess that the editors covered in the dataset may have processed more than 20% of the total scholarly output in 2021, the total number of journals *not* covered by our dataset must be (if it can be known) immensely high. The reason behind this omission is that many websites of scholarly publishers do not enforce a uniform structure in listing editors,

thus rendering it difficult to webscrape the data with automated scripts. The difficulty arises through trivial issues like punctuation – is the affiliation of an editor listed after a comma or rather after a dash? Is it written in italic and if so, does it use the HTML tag “” or the HTML tag “<i>”? Each format requires a different script – and if thousands of journals follow different data displays, then thousands of scripts would be required, which would hardly be advantageous to a manual data-collection.

Some of the huge publishers like Taylor & Francis, Springer or Wiley may publish more than 1.000 journals each, but they do not offer a uniform enumeration of their editors. Information about these journals thus remain missing in the Open Editors dataset. There are, thus, certainly way more than just 455 researchers from the LSE across the scientific journal landscape and their editorial boards – they just could not be scraped by Open Editors because of the prevalence of unclean data structures.

At least some publishers do have a homogenous way of displaying data about editors. Cambridge University Press, SAGE and Elsevier are a few examples among the big publishers. What is more, even notorious predatory publishers have a surprisingly friendly data structure, which allowed us to scrape data about editors listed in a few hundreds of bogus journals – which, in turn, points to another use case of the dataset, namely to detect whether some researchers of one's institute fell prey to a questionable journal (and to alert them about the risks of being associated with them).

Towards a FAIR solution

What is even more promising is that there is now a heightened awareness about the need for high-quality data about the overall journal infrastructure (cf. the **Journal Observatory** initiative). Admittedly, the webscraping solution offered by Open Editors will not be sustainable over the longer term – publishers' websites change their design and URL patterns regularly so that the scripts need to be re-programmed as well. And, ultimately, Open Editors remains an amateur project that cannot guarantee a thorough data-curation lasting for years and decades.

Rather than relying on individual-led projects like Open Editors, a community-driven effort to render the data display about editors uniform across all journals and publishers would be preferable. The best solution may be a central registry where authoritative information about editorial board memberships can be stored according to FAIR principles. **CrossRef has already started thinking about it** – and with its remarkable developments surrounding **open citations** and **open abstracts**, it is not implausible to believe that CrossRef may indeed achieve an opening up of large-scale

data about scientific journal editors one day. Then, and only then, can we finally test our suspicions about the extent of 'gatekeeping' in our least/favourite journals systematically.

Readers can find out more about the Open Editors project and explore the dataset here: <https://openeditors.ooir.org/>

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: Screenshot of Open Editors website, reproduced with permission of the author.

About the author



Andreas Nishikawa-Pacher

Andreas Nishikawa-Pacher conducts scientometric analyses at the TU Wien Bibliothek. In addition, he is a DOC-Fellow of the Austrian Academy of Sciences at the University of Vienna (Department of Legal and Constitutional History) in a joint programme with the Vienna School of International Studies.

Posted In: Academic publishing | Metrics

2 Comments



Related Posts