Open Editors: A dataset of scholarly journals' editorial board positions

Andreas Nishikawa-Pacher (1,2,3,*, Tamara Heck⁴, Kerstin Schoch^{5,6}

¹Department of Constitutional and Legal History, University of Vienna, 1010 Vienna, Austria

²TU Wien Bibliothek, Resselgasse 4, 1040 Vienna, Austria

³Department of International Relations, Vienna School of International StudiesAustria

⁴DIPF | Leibniz Institute for Research and Information in Education, 60323 Frankfurt, Germany

⁵Pop-Up Institute, HKS—University of Applied Sciences and Arts Germany

⁶Department Psychology and Psychotherapy, Witten/Herdecke University, 58455 Witten, Germany

*Corresponding author. Email: andreas.pacher@tuwien.ac.at

Abstract

Editormetrics analyses the role of editors of academic journals and their impact on the scientific publication system. Such analyses would best rely on open, structured, and machine-readable data about editors and editorial boards, which still remains rare. To address this shortcoming, the project *Open Editors* collects data about academic journal editors on a large scale and structures them into a single dataset. It does so by scraping the websites of 7,352 journals from 26 publishers (including predatory ones), thereby structuring publicly available information (names, affiliations, editorial roles, ORCID etc.) about 594,580 researchers. The dataset shows that journals and publishers are immensely heterogeneous in terms of editorial board sizes, regional diversity, and editorial role labels. All codes and data are made available at *Zenodo*, while the result is browsable at a dedicated website (https://openeditors.ooir.org). This dataset carries implications for both practical purposes of research evaluation and for meta-scientific investigations into the landscape of scholarly publications, and allows for critical inquiries regarding the representation of diversity and inclusivity across academia.

Keywords: Editors; scientometrics; dataset; journals; publications; metadata

1. Introduction

Editors of scientific journals play an important role in the publication system: They legitimize selected research (Card and DellaVigna 2020), oversee the academic peer review procedure (Horbach and Halffman 2020), and influence the careers of scholars by (not) allowing them to publish in their journals (Zhang et al. 2021b). Editorial decisions may sustain intellectual and societal biases (Petersen 2017); for instance, their judgements might be driven by homophily, or personal affections that favour colleagues they work with, who come from the same region, or who promote research approaches similar to one's own studies (Yoon 2013). But if 'scientific rewards are based [merely] on publications' (Crane 1967), then any favouritism granted by editors to individual researchers or specific groups might seem unethical (Laband and Piette 1994).

It is in this context that researchers have investigated the impact and consequences of editorial work by analysing the composition of academic journals' editorial boards (Mazov and Gureev 2016), including in terms of gender representation (Feeney, Carson and Dickinson 2019; Stegmaier, Palmer and van Assendelft 2011; Topaz and Sen 2016), international diversity (Espin et al. 2017; Okagbue et al. 2018; Gutiérrez and López-Nieva 2001), economic inequalities (Horton 2003; Xu et al. 2019), social networks of researchers (Brogaard, Engelberg and Parsons 2014; Erfanmanesh and Morovati 2018; Baccini et al. 2020; Goyanes and de-Marcos 2020), the reputation of their institutional affiliations (Hodgson and Rothman 1999; Petersen, Hattke and Vogel 2017), and issues of publication ethics (Bornmann and Daniel 2009; Wong and Callaham 2012; Bishop 2020). They have also turned their eyes on predatory journals that wrongfully list reputable scientists as editors so as to heighten the journals' prestige (Ruiter-Lopez, Lopez-Leon and Forero 2019).

Research evaluators might also be interested in such 'editormetric' studies (Mendonça, Pereira and Ferreira 2018). They may wish to gain an overview of the representation of scientists across scholarly outlets. They may want to identify the extent to which they are (erroneously) named as editors of predatory journals. They may want to assess the mentoring and community services that come with editorial workloads. They may want to consider the different roles, like chief editorships or peer review contributions, and relate them to Open Science practices: an editor of a journal that just transited to a 'Diamond Open Access'-model or that experiments with 'Open Peer Review' could obtain special recognition. Data about journal editorships can thus add important nuances to the scientific system—especially since '[t]he choice by funders and institutions of what to measure for assessment directly influences research culture and behaviours' (European Commission 2021: 4). And instead of rewarding only 'a certain type of research outputs'-namely paper publications-research evaluation could draw from editormetrics to also acknowledge vital contributions in 'research as a process' (European Research Area and Innovation Committee 2021: 12).

Analysing the 'gatekeepers of knowledge' (McGinty 1999), however, comes with one major problem: the lack of structured data availability on scholarly journals' editors and

 $[\]ensuremath{\mathbb{O}}$ The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

editorial boards. While basic information about editors are available in the online presences of scholarly journals, the data are not uniformly structured and therefore not machinereadable-a major issue in the broader movement towards Open Data and Open Science as it relies on a quick processing of huge amounts of meta-scientific data (Aryani et al. 2020; Hendricks et al. 2020). To give an illustrative example, one publisher may list the editors by mentioning the given name first and the last name thereafter, followed by a comma and the editors' institutional affiliation, as in Jan Melissen, Leiden University, while another may first list a title (like Prof. or Dr), then the last name with an abbreviated given name, followed by the institutional affiliation in parentheses rather than after a comma, as in Dr Melissen, J. (Leiden University). The heterogeneous way of displaying information not only pertains to inter-publisher, but in many cases even to intra-publisher differences. To name but one example, the publisher Taylor & Francis shows editorial board details in different formats despite the journals belonging to one and the same publisher (Figure 1). In the case of Springer Nature, even the URLs linking to the information about the editors are heterogeneous, which makes automated data generation rather burdensome (see Table 1).

As a consequence of the non-uniform data structure, every study that analysed the impact of journal editors and editorial boards relied on laborious processes of *manual* data collection (Topaz and Sen 2016: 4; Petersen 2017: 258; Baccini et al. 2020: 279). Not only is this a highly time-consuming and monotonous task (Baccini and Barabesi 2010: 369–70; Xie, Wu and Li 2019; 1338) that only allows for a small sample size; but what is more, this approach offers 'just a snapshot in [a] specific moment' (Mendonça, Pereira and Ferreira 2018: 1518). The data collected, even if prepared properly and shared openly, get outdated quickly and thus are not appropriate for further re-use in research. Follow-up studies would need to manually generate the data again, hampering a reproduction of previous studies on editorial boards.

To mitigate some of those shortcomings, this article presents an automated data-collection project called *Open Editors*. It utilizes webscraping to collect and provide a structured dataset of scholarly journals' editors and editorial board members. The data are freely available for any use under a CC0-license so as to enable large-scale analyses. The data collection will be performed on a regular basis through a scripted programme whose codes are available in a public repository. In the latest release, *Open Editors* scraped the websites of 7,352 journals from 26 publishers (including five predatory publishers), comprising basic information about 594,580 editorial board positions.

The present effort goes hand in hand with other metascientific initiatives that systematically document key practices of journals and editorial guidelines. They often do so as grassroots projects crowdsourcing related data, such as with regards to peer review models (cf. Horbach and Halffman 2018), preprint guidelines (Nosek et al. 2015; Klebel et al. 2020), or Open Access policies (Curry 2017; Marchitelli et al. 2017). *Open Editors* serves as an additional step in this broader effort to increase the transparency of the structure of scientific publishing through bottom-up data collections—at least as long as a centralized, top-down data-collection approach remains absent (Michaud 2022).

The following section outlines the methodical approach to generate the dataset in detail, followed by an explanation of how to access the data. The results section thereafter assesses the representativeness of the sample and presents some summary statistics regarding the overall dataset. It also outlines publisher-level data about editorial board sizes, highlights the immense pluralism regarding editorial role labels, and shows publisher-level statistics about the geographical distribution of the editors' affiliations with a special look at predatory publishers and at the Anglo-American shares represented in the editorial boards; all in all, the descriptive data indicate that the journal landscape is highly heterogeneous. The article then discusses how Open Editors could aid in various editormetric analyses, including in investigations of diversity and inclusivity aspects as well as possible use cases for research evaluations and scientometric assessments. That section also points to various limitations and further desiderata, such as the challenging task of linking the dataset with persistent identifiers (PIDs), or the ethical problems that arise with using automated 'diversity guesser' tools. The article finally concludes with a call to push forward the broader Open Science agenda of opening up more and more data about the scholarly publication process so as to enable a culture of inclusivity, transparency, and knowledge equity across the system of science.

2. Methods

2.1 Data sample

This method section outlines the approach taken for the release of *Open Editors* in early 2022. As the project intends to iterate the data collection annually, it is possible that future releases will add minor adjustments to enhance the approach and to enlarge the sample. New publishers were indeed



Figure 1. Three journals from the same publisher (Taylor & Francis) displaying information about their editors in three different ways, as of December 2020. Left: *Central Asian Survey*. Centre: *The Cartographic Journal*. Right: *The Information Society*.

Table 1. Heterogeneously structured links leading to information about editors at four Springer Nature journals

Journal	URL	Comment
British Dental Journal	nature.com/bdj/about#editors	URL contains a hash (#)
Human Genome Variation	nature.com/hgv/about/editor	'editor' in singular
Bone Marrow Transplantation	nature.com/bmt/editors	URL does not contain the 'about'-tag
Nature Cell Biology	nature.com/ncb/about/editors	'editors' in plural

constantly added to the dataset. All changes will be documented transparently through the version control system on *Zenodo* where all codes and data are available under a CC0 license.

As of mid-2022, *Open Editors* collected data about editors and editorial board members from journals belonging to 26 major publishers. The major criteria for the choice of publishers was that they provide data about journal editors in a homogeneous structure, thus allowing machines to read them without major difficulties.

The first 21 of these publishers were American Psychological Association (APA), American Society of Chemical Engineers (ASCE), BioMedCentral, Brill, Cambridge University Press (CUP), eLife, Elsevier, Emerald, Frontiers, Hindawi, Inderscience, John Benjamins, Karger, Multidisciplinary Digital Publishing Institute (MDPI), PeerJ, Pleiades, Public Library of Science (PLoS), Royal Society of Chemistry (RSC), SAGE, and Springer Nature. In addition, five publishers known to be predatory were likewise included in the sample, namely Allied Academies, iMedPub, Longdom, Scientific Research Publishing (SCIRP), and SciTechnol; all of them but SCIRP are imprints of the larger predatory publisher OMICS, and both SCIRP and OMICS are widely known to host 'fake journals' (Simón 2016; Masic 2017).

The goal of this work is not primarily the inferential analysis of the data themselves but the development of a methodical tool that will allow scientists to gain samples for their research without the tedious process of manual data collection. The current subset is only the beginning and offers a convenient sample of an ultimately unknown population of publishers. We will nevertheless give a rough estimate of the dataset's representativeness by comparing it with the publication data as documented by *OpenAlex*, a new database covering metascientific information on a grand scale (Priem, Piwowar and Orr 2022); and the discussion section will argue that for exploratory purposes, it is better to collect any available data based on a convenient sample than to not undertake any collection efforts at all.

2.2 Data collection: webscraping

Two data subsets were collected with regards to the 26 publishers: the first subset contains each of the publishers' journals as well as the URLs to the journals' editorial board information. The second subset contains seven variables pertaining to every editorial board member. The variables are:

- the name (without disentangling titles, given names, middle names, and surnames),
- the institutional affiliation (if available, the research institute, often the university, at which the editor is employed),
- the ORCID ID (Open Researcher and Contributor ID), if available in a uniformly structured format (this only applies to journals from *Hindawi*, *John Benjamins*, *PLoS*, and *RSC*),

- the editorial position (the role the editor occupies at the journal, such as 'Editor-in-Chief', 'Associate Editor', or 'Book Review Editor'),
- the journal name,
- the ISSN (if available on the same webpage as the information about editors),
- the publisher,
- the URL (containing a link to the scraped website),
- and the date of scraping (in the format YYYY-MM-DD).

Open Editors mainly utilized the R library rvest (Wickham 2022) to fetch the data from the journal websites. For each publisher, a different code had to be used to adapt to their respective webpage design and structure. The webscraping process took place in early 2022.

2.3 Data presentation and accessibility

After the webscraping process, the data were structured into machine-readable files based on comma-separated values (CSV). As a basic quality check, duplicate data that carried the identical values in the editor-, journal-, and publisherfields were eliminated, and incomplete data where both the editor and the affiliation remained unavailable during the crawling were removed as well. In addition, HTML tags, or codes that aim at layouting a given website such as by making fonts italic through the tag $\langle i \rangle$ sometimes accidentally slipped into the dataset; they were likewise deleted. All the remaining data were merged into a CSV-file and made available at *GitHub* and *Zenodo* under a Creative Commons license (CC0) so as to enable anyone to download the whole data dump without any restrictions.

In addition to the CSV-data dump, *Open Editors* also created a website that enables access to the data via a form with which end users can search for names, journals, or institutional affiliations. For that purpose, the collected dataset was put into a MySQL database. The website is available at https://openedi tors.ooir.org. As an example, users can type in 'University of Salzburg' into the search bar to find a list of all editors in the dataset who have 'University of Salzburg' in their names, affiliations, or journal name. The search function also supports basic Boolean operators such as AND, OR, NOT, brackets, and quotation marks. For instance, the query 'Salzburg' AND (Biology OR Geoinformatics)' would look for all data that contain 'Salzburg' and either 'Biology' or 'Geoinformatics'.

Note that a variant of the code at *GitHub*—added *ex post* by Bianca Kramer from the Utrecht University Library allows one to loop the affiliation data through the so-called *Research Organization Registry*, or ROR (Lammey 2020). This integration of *Open Editors* with ROR aids in cleaning the data, for a single affiliation can be written in numerous ways. The TU Wien, for instance, could be spelt out as 'Technische Universität Wien', or translated to 'Vienna University of Technology' or mixed as 'TU Vienna, Österreich'. Without ROR, searching for all editors that belong to the TU Wien may necessitate dozens of search queries taking into account all possible variants of spelling the university name, with the risk of omitting multiple unknowns. ROR allows one to homogenize multiple spellings into a single identifier.

Having outlined the methodical approach, the next section will present some basic information about the data underlying *Open Editors*' latest round of webscraping.

3. Results

3.1 Representativeness of the current sample

It is estimated that there exist circa 16,780 scientific publishers globally (according to data based on OpenAlex, cf. Pollock 2022), which is barely possible to be represented in one dataset. However, the current sample includes, inter alia, 15 of the 44-or almost one-third of the-large-scale publishers that have more than 100 journals. While this group represents only 0.26% of all academic publishers, they were responsible for almost two-thirds (65%) of the total scientific output in 2021 (Pollock 2022). In an admittedly rough estimation, the editors covered by Open Editors can be said to have been involved in at least one-third of that share, i.e. in circa 21.67% of the scholarly publication output (for the sake of cautiousness and simplicity, we leave out the 11 additional small- to mid-sized publishers covered by Open Editors). Since OpenAlex itself might inhere omissions of an unknown scale, one may remain on the safe side and estimate that the editors covered by our sample may have processed approximately a fifth of the total scientific output in 2021.

Given the immense heterogeneity of the publishers' and journals' publication patterns, this section will continue with descriptive (rather than inferential) statistics.

3.2 Overall number of editors

Open Editors found 594,580 editorial membership positions from 7,352 journals pertaining to 26 publishers, with an average of 81 editors and a median of 34 editors per journal (see Table 2).

The standard deviation of 467 editors indicates that scientific journals are highly heterogeneous in terms of their editorial board sizes; Figure 2 offers a glimpse into the strongly long-tailed distribution, and Figure 3 shows a closer look at all journals with no more than 100 editors (i.e. 6,860 journals, or ca. 93% of the dataset). It reveals that the modal journal has 11 editorial board members, followed by journals with 22 members, 29 members, 12 members, and 24 members. Most journals exhibit an editorial board size ranging between 10 and 45 editorial board members.

There are extreme outliers with regard to the number of editorial board members. Most of these outliers belong to the publisher *Frontiers*, where the journal with the highest number of editorial board members is *Frontiers in Psychology* (13,967 individuals). Other than *Frontiers* outlets, megajournals such as *PLoS ONE* (9,001 researchers on their board),

Table 2. Summary statistics showing number of editors per journal (n = 7,352).

Mean	Median	Mode	SD	Minimum	Q1	Q3	Maximum
80.9	34	11	467	1	20	53.2	13,967

PeerJ (1,673 editorial board members), and *eLife* (870 persons) can likewise be found within the far end of the distribution.

3.3 Publisher-level summary statistics

Grouped by publisher, basic descriptive data are visible in Table 3. The number of journals and editors varies quite starkly between the publishers, though most of them have a median count of editors per journal of around 25. There are some conspicuous outliers (mostly unsurprising due to their nature as megajournals), such as *PeerJ* (Mdn = 1,673), *eLife* (Mdn = 870), *Frontiers* (Mdn = 805.5), and *PLoS* (Mdn = 199.5). Not showing these outliers, Figure 4 visualizes the distribution of the numeric composition of editorial boards by publisher.

3.4 Editorial roles and diversity

Not all editorial board members are of equal weight. Instead, some roles typically convey greater responsibility and a more intense workload than others. The dataset contains various labels and spellings with regard to these editorial positions. It found 4,024 distinct names for the respective roles of the editors, the most common ones being 'Review Editor', 'Editorial Board', and 'Guest Associate Editor'. Some less frequent attributes are 'Post-doctoral Board', 'Media Review Editors', 'International Consultant Editors', and 'Editorial Board Assistant'. With thousands of different titles, it seems difficult to generate a manageable typology of the various journal positions that researchers occupy with their editorial tasks.

Some publishers, however, seem to have taxonomized the labels of editorial tasks across their journal portfolio. The publisher *Frontiers* offers an example, as it only uses nine distinctly labelled roles for the almost 280,000 editorial board members across 92 journals (Table 4).

3.5 Geographical distribution

The researchers' affiliations often disclose the countries that harbour their respective institution. Using the 'world cities'-dataset from the R library *maps* (Becker et al. 2021), one can count the frequency with which each country is mentioned. In total, a superficial analysis finds 194 countries and territories listed across ca. 94.3% of all editorial board positions, with the remaining 5.7% not disclosing any country names.

At the continent-level, Europe occupies over a third (35.7%) of the editorial positions; North America more than a fourth (27.5%); and Asia more than a fifth (20.4%). Latin America and Oceania have a share of roughly 4% each, while African affiliations are visible in only 1.5% of the editorial board positions in the present dataset. Juxtaposing them with the global population as taken from the World Bank would suggest large discrepancies between editorial representation and global population share. Figure 5 visualizes that lack of correlation, with the left bars showing the editorial representation, and the right ones showing the world population share.

Unsurprisingly, therefore, even on the country-level rather than continent-level, there is no statistically significant correlation between representation in the editorial boards of scientific journals and population size, according to a Kendall's tau test ($r_{\tau} = 0.4, p > 0.05$).

Speaking of which, treading from the continent-level to country-level data leads to the aggregation in Figure 6. It



Figure 2. Histogram showing the distribution of the number of editors per journal, based on 99% of journals in the dataset (the top 1% of journals, 74 in total, is not shown because their extremely large numbers would skew the graph too strongly).



Figure 3. Zoom-in into the head of the histogram in Figure 2, showing the distribution of the number of editors per journal.

offers a view of the 30 most frequently mentioned countries (see the bars on the left-hand side). The USA leads the ranking, occupying more than a fourth of all editorial board positions in the dataset. China comes second with slightly less than 10%, followed by the UK in the third position. Italy, Germany, Australia, Canada, Spain, France, and India round up the top 10 of the most frequently mentioned affiliationcountries.

To put these numbers into a perspective beyond world population counts, the right-hand bars in Figure 6 demonstrate the respective countries' total scientific outputs (according to the numbers taken from Scimago Journal & Country Rank 2022). Not only is there a graphical symmetry visible; there is also a strong positive correlation between a country's total scientific output and their numerical representation in the editorial boards covered by *Open Editors* ($r_{\tau} = 0.7, p < 0.01$).

One could also look at the frequency of countries listed in the affiliations of those researchers who (allegedly) serve in the boards of predatory journals. The present article can only offer a superficial glimpse to illustrate the dataset's utility, but the list of the 10 countries most present at the five predatory publishers covered by the dataset (Allied Academies, iMedPub, Longdom, SCIRP, and SciTechnol) indicates that some countries such as Turkey and Egypt are overrepresented at predatory journals when compared to their presence at all the other publishers (see Figure 7, with the left bars indicating the share of editorships at non-predatory, and the right bars indicating the share of editorships at predatory outlets). But with the USA leading that enumeration, and with countries like Italy, Canada, UK, Japan, and Spain included in the top 10 list as well, the graph indicates that the issue of predatory publishers is not a problem merely of the 'Global South', but

Table 3. Basic information about the number of journals and editors by publisher.

Publisher	Journals	Editors	Mean ^a	Median ^a	Mode ^a	SD ^a	Min ^a	Q1 ^a	Q3 ^a	Max ^a
Allied Academies	27	460	17.0	12.0	19	14.6	4	8.5	19.0	70
APA	53	3,412	64.4	56.0	39	31.8	19	40.0	79.0	160
ASCE	16	754	47.1	50.5	63	16.8	16	36.5	61.5	67
BioMedCentral	216	13,308	61.6	50.0	28	56.6	1	36.8	69.2	614
Brill	314	7,589	24.2	22.0	20	14.0	1	14.2	32.0	87
Cambridge UP	383	12,485	32.6	29.0	24	21.3	1	19.0	42.0	173
eLife	1	870	870.0	870.0	870	NA	870	870.0	870.0	870
Elsevier	2,347	112,947	48.1	41.0	36	39.5	1	25.0	61.0	726
Emerald	220	11,565	52.6	40.5	29	37.2	12	30.0	61.0	252
Frontiers	134	278,842	2,080.9	805.5	741	2,688.2	1	446.5	2,625.0	13,967
Hindawi	240	14,939	62.2	25.0	16	149.5	1	16.0	51.8	1,719
IGI Global	228	10,200	44.7	42.0	37	21.6	3	31.0	53.2	189
iMedPub	160	3,623	22.6	20.0	15	15.6	4	14.0	25.0	130
Inderscience	468	16,535	35.3	32.0	21	19.3	1	22.0	43.0	190
John Benjamins	90	2,549	28.3	27.0	24	10.9	5	22.0	33.8	61
Karger	99	3,502	35.4	33.0	44	23.3	1	17.5	48.0	129
Longdom	171	6,838	4.00	33.0	21	31.9	8	23.0	47.0	347
MDPI	376	8,724	23.2	14.0	11	22.6	10	12.0	24.0	215
PeerJ	1	1,673	1,673.0	1,673.0	1,673	NA	1,673	1,673.0	1,673.0	1,673
Pleiades	115	3,182	27.7	26.0	20	9.5	10	21.0	32.0	58
PLOS	12	11,059	921.6	199.5	39	2,547.6	39	101.5	244.8	9,001
RSC	42	2,932	69.8	65.0	54	32.3	12	50.2	86.2	193
SAGE	1,195	57,859	48.4	38.0	37	38.5	1	26.0	60.0	481
SCIRP	247	6,145	24.9	23.0	23	9.6	5	18.0	30.5	58
SciTechnol	81	1,280	15.8	15.0	15	7	1	12.0	20.0	39
Springer Nature	116	1,308	11.3	6.0	2	13.4	1	3.0	12.2	68

a Journals per Editor.



Figure 4. Boxplots showing the distribution of the number of editors per journal grouped by publishers (outliers are blended out; and the publishers *eLife*, *Frontiers*, *PeerJ*, and *PLoS* are likewise missing because their inclusion would distort the visualization too strongly).

Table 4. Editorial roles and their frequency at the journals of the publisher

 Frontiers, offering an example of standardized editorial role labels.

Role	Frequency	Share (in %)
Review Editor	162,627	58.322
Guest Associate Editor	77,608	27.832
Associate Editor	36,909	13.237
Specialty Chief Editor	1,455	0.522
Field Chief Editor	125	0.045
Reviewer	73	0.026
Assistant Specialty Chief Editor	23	0.008
Editor-in-Chief	8	0.003
Chief Editor	5	0.002
Managing Editor	4	0.001
Assistant Field Chief Editor	3	0.001
Assistant Chief Editor	2	0.001

also an acute concern for research culture in Europe and North America as well.

Indeed, the geographical representation in the dataset's predatory and the non-predatory publishers correlates strongly, with Kendall's tau $r_{\tau} = 0.95$, p < 0.01.

Looking at the geographical diversity at the journal-level, one can again see how heterogeneous the scholarly publication system is. Grouped by publisher, *Frontiers'* median journal exhibits the largest number of distinct countries (almost 60), followed by the megajournals *PeerJ* (46), *eLife* (41), and the median *PLoS* journal (24.5), cf. Figure 8. Across all journals and publishers, the median number of distinct countries represented in an editorial board is 11.

To give a closer illustration, Table 5 enumerates the top five countries listed in every journal grouped by publisher (discounting missing values). The USA dominates most of the publishers' editorial boards, or 22 out of 26, and it is always among the top five represented countries. China is listed in 20 of these enumerations, the UK in 18, and Italy in 14.

3.6 Share of Anglo-American affiliations

With Australia and Canada likewise present in many of these top five-lists, one may explore the extent to which the scientific publication system exhibits an Anglo-American dominance (based on the sum of editorial board affiliations located in the USA, UK, Canada, Australia, and New Zealand). In Figure 9, one can see that there are some publishers where at least half the editorial teams are based at Anglo-American institutes. That list is led by the journals of the APA with a median Anglo-American share of 90.3%, followed by *Cambridge University Press* with 72.7%, *SAGE* with 70.7%, *eLife* with 64.5%, the predatory publisher *SciTechnol* with 59.6%, *PLoS* with 58,85%, *Emerald* with 51.75%, and *Springer Nature* with 50% (the percentages discount those affiliations where no country name could be extracted automatically).

All in all, the summary statistics presented in this section point to a highly heterogeneous landscape of journals and their editorial boards, of publishers and their regional diversity, of editorial roles and labels. The descriptive data merely offer some superficial glances, but they thereby illustrate the analytical utility of the dataset for further research and practice.

4. Implications on dataset usage and measuring diversity

4.1 Four use cases to further utilize the dataset

Open Editors scraped the websites of thousands of academic journals to collect and structure data about half a million editorial positions, making them available under a CC0 license for anyone to use and analyse, with a pledge to keep the data up to date by iterating the scraping on a regular basis. With this large-scale dataset, the conduct of editormetric analyses should become more efficient, more effortless, and more reproducible. Possible investigations that could draw from this dataset include analyses regarding the gender composition of editorial boards; the regional diversity among editors; the possibility to detect the frequency to which editors publish in their own journals; the representation of early career researchers in the landscape of scholarly publications; the degree to which interdisciplinary publishing is realized; and other issues pertinent to research evaluation and research assessment.

Four (semi-)fictitious examples may illustrate the utility of the dataset for both researchers and for practitioners (such as university librarians and research evaluators), examples that are partly in line with recent calls to '[r]ecognize the diversity of research activities and practices, with a diversity of outputs, and [... to] valorize the diversity in research roles and careers' (European Commission 2021: 12).

First, say that the Ministry of Science in the Czech Republic initiates an awareness-raising campaign regarding predatory publishers. Before doing so, however, they wish to explore the presence of Czech-affiliated scholars across predatory journals. Using Open Editors, the Ministry could filter the dataset for the five predatory publishers contained therein and query it in conjunction with the country name. Using the web version of Open Editors, a possible query could perhaps be ('Allied Academies' OR iMedPub OR Longdom OR SCIRP OR SciTechnol) AND (Czech OR Czechia OR Brno OR Praque OR Praq OR Praha OR Ostrava OR Olomouc), which leads to 40 supposedly Czech-affiliated researchers. In a further step, the Ministry could verify the identity of these 40 persons listed at the predatory journals and warn them about the nature of the journals that boast their names (Downes 2020).

A second fictitious use case may relate to a small university library's need to assess the presence of their institute's scholars across a certain publisher's editorial boards (e.g. Karger) because negotiations about a renewal of an institutional subscription of their journals are imminent. Imagine that the university's scholars hardly publish in Karger journals, which suggests that perhaps a renewal of the subscription might be dispensable, or only expedient if the publisher's proposed price was lower than before. Using data from Open Editors, the university's research evaluators could have an additional source at hand, either to corroborate the library's argument for a lower price, or to offer a counterargument if it turns out that the institution's scholars actually occupy active editorial positions at Karger journals. Whatever the outcome of this editormetric analysis, tapping the dataset will always offer a richer view of an institute's linkages to a given publisher, thus providing crucial data to support negotiations between libraries and publishers with further evidence.



Figure 5. The most frequently mentioned continents in the editors' affiliations (left bars, black) juxtaposed with the share of world population (right bars, grey).

A third example of how the dataset could be (and has already been) practically used would involve research administrators organizing a workshop on scholarly publication ethics. Using Open Editors, they could find out which local researchers serve in journals' editorial boards so as to invite them to that workshop and target them for possible discussions. The participants could be further categorized into various groups depending on whether they serve as chief editors or section editors, as managing editors, or as editorial board members who are merely tasked with reviewing submissions. Other disaggregations could be based on whether the journals they serve on are particularly innovative with regards to Open Science practices (e.g. by publishing Registered Reports or by supporting Open Peer Reviews). Such approaches add greater nuances to the evaluation of academic community services.

Fourth, scientometric research could tap data from Open Editors for various basic scientific purposes. For instance, one study already used this dataset to discover, inter alia, that journal-level open access policies do not necessarily correlate with greater diversity of editorial leadership (Altman and Cohen 2021). Editormetric research has so far explored many other issues, albeit with a smaller sample size than Open Editors could deliver-such as the extent to which journals exhibit interdisciplinary editorial boards (Wu et al. 2020; Zhang et al. 2021a). As many emerging research fields are of an inter- and trans-disciplinary nature, the overrepresentation of a specific discipline's researchers in an editorial board may turn out to be problematic for a journal. Research published in that respective outlet might then not represent the discipline's broader research achievements properly. Relatedly, the dataset could be used to detect the geographical diversity of editorial board compositions (Harzing and Metz 2013; Mazov and Gureev 2016). There is evidence that case studies and research with cohorts from only one (socalled 'non-Western') country often have difficulties to find acceptance from traditional journals (Wilson and Knutsen 2022), as editorial boards fear such kind of research may not be representative for their readership. If a journal wishes to publish research with regional diversity, then an editorial board with members from different countries might help in attracting submissions with a greater geographical scope. Open Editors may aid in rendering such regional biases visible. A final example of a basic scientometric analysis drawing from the present dataset would be to analyse a given journal's innovativeness in correlation with the diversity of the 'academic age' of the editorial board members (Nicholas et al. 2020). As innovative research or emerging topics might find difficulty in getting published in long-standing high impact journals in specific fields (Petersen 2017), this can affect early career researchers with their emerging approaches more often than senior ones. More diverse boards that include early career researchers might smoothen the danger of a traditional topical and methodological bias. Moreover, including more early career researchers as editorial board members (Pagedar et al. 2022), and thus indirectly as reviewers, might help to manage the ever increasing amount of manuscripts submitted to scientific journals, and to shorten the review process (Nguyen et al. 2015). While the present dataset alone does not allow such analyses due to the lack of metadata about the 'academic age' of each person in the editorial board, one could link the dataset to ORCID (Haak et al. 2012), or the Open Researcher and Contributor ID, in order to do conduct such investigations, such as by fetching data about the year of graduation as saved in ORCID's metadata.

4.2 Dataset limitations

Speaking of which, the status quo of the dataset is certainly not yet perfect as such. Instead, Open Editors would enable more elaborate examinations if it was linked with other datasets, especially via PIDs. ORCID would allow one to disambiguate researchers who carry the same names at least in their Latin transcriptions (e.g. 'Anna Smith', 'Wang Li'), and it would also open a path towards the potentially rich metadata associated with ORCID profiles, such as on individual scholars' peer review activities. An analysis that draws from both Open Editors and ORCID could examine the referees' increased workload that accompanies an editorial position. Or, if ORCID profiles show former institutional affiliations of editors, one could detect the extent to which editors allow researchers from their former organizations to publish in their journals. The addition of PIDs would thus greatly facilitate the task of comprehensive searches and complex analyses based on trustworthy data.

Another limiting issue of the dataset is the underlying data sample. The current approach is based on a necessarily incomplete coverage. While the total number of academic publishers and scholarly journals can only be estimated (Johnson,



Figure 6. Most frequently mentioned countries in the affiliations of the Open Editors dataset, juxtaposed with the respective country's share of total scientific outputs in 2021.

Watkinson and Mabe 2018; Nishikawa-Pacher forthcoming), we showed that the editors covered by the dataset may be said to have processed a fifth of the total scientific output in 2021. One should thus keep in mind that the sample remains fragmentary because the journals' heterogeneous way of displaying data about editors inhibits a simple webscraping approach. Major examples outside of *Open Editors*' sample include *Oxford University Press, Taylor & Francis*, and *Wiley*. Nevertheless, even though it may be impossible to get any statistically decent sample given the unknown population, it seems better to collect any available data rather than to discourage related efforts.

The best solution against this issue of the dataset's sample size would be to have a centralized infrastructure that would demand the opening up of uniformly structured editorial data in a top-down manner across every single publisher. Barring such an implementation, another mitigation could be to use webscraping with a default structures of editorial data display by accepting the risk of omitting a large number of journals whose website structure deviate from the default's one (as was done by Philipp Zumstein and Jan Kamlah with regards to Springer and Wiley; cf. Kamlah and Zumstein 2019) integrate a machine learning approach that would be capable of reading the data about editors regardless of the specific HTML and CSS arrangements (Safder et al. 2020).

As long as this is not achieved, *Open Editors* provides a sufficient solution: the data are incomplete, the sample's underlying population is unknown, and the dataset may be biased towards well-structured publishers. However, once these limitations are accounted for, the dataset is useful in generating some insights for research assessments (e.g. on the presence of an institute's scholars in the editorial boards of serious



Figure 7. Most frequently mentioned countries in the affiliations of the editors at the five predatory publishers Allied Academies, iMedPub, Longdom, SCIRP, and SciTechnol, compared with their representation in non-predatory outlets.

and predatory journals), in coming up with hypotheses (e.g. on the gendered patterns in chemistry journals versus sociology journals), and in pointing to further avenues for research and practice (e.g. for taxonomizing editorial board roles, or for harmonizing the data display of editorial boards).

4.3 Critical thoughts on measuring diversity

A final aspect we wish to discuss pertains to the potential investigations of social diversity issues that might or might not be possible with the dataset thus accrued. Many editormetric analyses aim to detect biases and inequalities that are sustained by editorial power within the scientific publication system. In doing so, however, they are dependent on the available metadata which, as of now, are largely confined to the editors' names, affiliated institutions, and roles. Such data, especially if they find linkage to PIDs, can be harnessed to illuminate supposedly objective scientometric patterns such as the ones mentioned above—regional diversity, interdisciplinarity, or the presence of early career researchers based on a given person's 'academic age'. But to obtain an even broader, overall picture on the state of a less 'objective' and more social constructivist conceptualization of diversity among editorial boards, more information would be needed—but *what* information?

Some readers might ask why diversity and not merit is the crucial construct in this work. The answer is simple: In an ideal world, the two should be equivalent, but in fact they are not. If we have a look at today's academia, it is still dominated by a *white*, cis-male gaze. Many senior researchers and editors owe their positions not only to their hard work, but also to their privileges, while people from (multi-)marginalized groups have a harder time getting there. People affected by classism, racism, ableism, and other forms of discrimination are less likely to gain access to higher education, become a researcher, and reach tenure or editor positions—regardless of their competence. This is not because they are less skilled, but because on top of their work, they have to tackle individual, institutional, and structural discrimination. Due to



Figure 8. Boxplot showing the distribution of the number of affiliation-countries represented in the editorial boards of journals, grouped by publisher.

homophily and other effects, this bias continues, e.g. editors are more likely to choose research to be published which fits their perspectives and to invite co-editors that think alike. *Open Editors* as a measurement for diversity might become a useful indicator for (lacking) diversity of editorial boards and a reminder for change, for example if combined with quotas.

However, diversity refers to a multitude of variables that intersect and go beyond region of origin. This is where dilemmas arise, as this category of diversity goes hand in hand with privileges of those who are considered to belong to a societal 'default', which currently means a systemic bias in favour of being cis, male, heterosexual, white, middle or upper class, able-bodied, socialized as Christian, and Euroamerican. The deviation from these norms correlates with exclusion mechanisms, and it is here where editormetric analyses could find a fundamentally important societal use. One might believe that these aspects may be easily analysable once we collect sufficient data, but a datafication of diversity may find its limitation when encountering issues of intersectionality (Cho, Crenshaw and McCall 2013; Crenshaw 2015). The framework of intersectionality states that a person's social and political identities create different modes of discrimination and privileges. Individuals or groups are marked as 'others'

and thus experience discrimination based on actual or perceived characteristics. Those identities might refer to aspects of gender, racialization, class, sexual orientation and identification, religion, ability, and other assigned characteristics. The framework describes the overlap and concurrency of more than one of these categories of discrimination, which results in more than one axis and thus lead to multiple discriminations. If several of these attributes apply to a person, they can be multiply burdened. For example, a white woman is a target of sexism, but will never experience racism, in contrast to a disabled woman of colour who experiences sexism, racism, and ableism simultaneously. In addition, any measurement of this category of diversity is hampered by the fact that many of the categories involved are in large parts based on subjective, historical, and socially constructed characteristics linked to individual, institutional, and structural discrimination. All these issues and their intersections render it difficult to measure them objectively, including in metascientific and editormetric investigations, without oneself promoting 'forced outings' (Skinner-Thompson 2015), further otherings, or similar discrimination processes.

To address this issue, one could use tools designed to automatically assign specific aspects related to diversity given a

Table 5.	The five most-frequently	mentioned countries ir	n the affiliations of the j	journals' editorial boards,	grouped by publisher.
----------	--------------------------	------------------------	-----------------------------	-----------------------------	-----------------------

Publisher	#1	#2	#3	#4	#5
Allied Academies	India	USA	Italy	China	Egypt
	71 (15.9%)	69 (15.5%)	44 (9.9%)	21 (4.7%)	20 (4.5%)
APA	USA	Canada	UK	Australia	Germany
	2,450 (73.3%)	228 (6.8%)	121 (3.6%)	84 (2.5%)	78 (2.3%)
ASCE	Australia	China	Hong Kong	Taiwan	USA
	11 (11.7%)	11 (11.7%)	10 (10.6%)	6 (6.4%)	6 (6.4%)
BioMedCentral	USA	UK	China	Australia	Italy
	3,530 (25.4%)	1,135 (8.2%)	1,029 (7.4%)	743 (5.4%)	727 (5.2%)
Brill	USA	UK	China	Germany	Hong Kong
	572 (22.3%)	206 (8%)	168 (6.5%)	150 (5.8%)	126 (4.9%)
Cambridge UP	USA	UK	Canada	Australia	Germany
	4,536 (37.5%)	2,182 (18%)	678 (5.6%)	677 (5.6%)	375 (3.1%)
eLife	USA	UK	Germany	China	Canada
	434 (51.2%)	61 (7.2%)	41 (4.8%)	33 (3.9%)	29 (3.4%)
Elsevier	USA	China	UK	France	Australia
	35,191 (30.2%)	9,658 (8.3%)	8,676 (7.4%)	4,770 (4.1%)	4,652 (4%)
Emerald	USA	UK	Australia	China	Canada
	2,367 (19.8%)	2,236 (18.7%)	901 (7.5%)	714 (6%)	344 (2.9%)
Frontiers	USA	China	Italy	UK	Germany
	66,554 (21.9%)	34,276 (11.3%)	24,390 (8%)	17,424 (5.7%)	13,223 (4.4%)
Hindawi	Italy	USA	China	Spain	Japan
	2,775 (17.2%)	2,415 (15%)	1,652 (10.2%)	803 (5%)	689 (4.3%)
IGI Global	USA	India	UK	China	Italy
	2,364 (21.4%)	1,179 (10.7%)	658 (6%)	527 (4.8%)	385 (3.5%)
iMedPub	USA	India	China	Italy	Turkey
	725 (21.1%)	453 (13.2%)	211 (6.2%)	186 (5.4%)	168 (4.9%)
Inderscience	USA	UK	China	India	Italy
	4,029 (22.3%)	1,514 (8.4%)	1,295 (7.2%)	832 (4.6%)	782 (4.3%)
John Benjamins	USA 140 (17.2%)	UK 92 (11.3%)	Hong Kong 56 (6.9%)	China 38 (4.7%)	France 36 (4.4%)
Karger	USA	Germany	Japan	Italy	UK
	775 (22.2%)	479 (13.7%)	234 (6.7%)	230 (6.6%)	175 (5%)
Longdom	USA	India	China	Italy	Japan
	2,864 (42%)	391 (5.7%)	357 (5.2%)	314 (4.6%)	216 (3.2%)
MDPI	USA	Italy	Spain	UK	France
	1,844 (20.1%)	1,641 (17.9%)	731 (8%)	616 (6.7%)	444 (4.8%)
PeerJ	USA	Australia	Hong Kong	China	Italy
	25 (15.8%)	12 (7.6%)	10 (6.3%)	9 (5.7%)	9 (5.7%)
Pleiades	Russia	USA	Germany	Belarus	France
	2,310 (79.1%)	130 (4.5%)	60 (2.1%)	45 (1.5%)	44 (1.5%)
PLOS	USA	UK	Australia	India	China
	950 (33.1%)	228 (7.9%)	132 (4.6%)	112 (3.9%)	110 (3.8%)
RSC	USA	China	UK	Germany	Japan
	634 (24.2%)	408 (15.6%)	292 (11.1%)	150 (5.7%)	113 (4.3%)
SAGE	USA	UK	Australia	Canada	India
	25,590 (44.8%)	6,972 (12.2%)	2,746 (4.8%)	2,481 (4.3%)	1,615 (2.8%)
SCIRP	USA	China	India	Italy	Canada
	1,908 (29%)	768 (11.7%)	331 (5%)	314 (4.8%)	227 (3.5%)
SciTechnol	USA	UK	Canada	Italy	China
	532 (47.1%)	64 (5.7%)	50 (4.4%)	46 (4.1%)	40 (3.5%)
Springer Nature	USA	UK	Germany	Italy	China
	90 (17.3%)	88 (17%)	39 (7.5%)	34 (6.6%)	30 (5.8%)



Figure 9. The distribution of the Anglo-American shares in the affiliation-countries represented in the editorial boards of journals, grouped by publisher.

certain input. Gender guesser tools may serve as prime examples (Sarawgi, Gajulapalli and Choi 2011; Lin and Serebrenik 2016). While their practicability cannot be denied, they also carry fundamental deficiencies. The guesser tool, for instance, usually parses the persons' gender in terms of a binary system, and studies report on inequalities between female and male researchers within the scholarly publication process (Squazzoni et al. 2021). But this approach excludes a whole range of non-binary and trans identities; also a person whose name is coded as male might not be cis, but trans. In other words, the coded and the actual self-identified gender can differ widely. The same applies for aspects of so-called race, which is not a biological but a historical and racist social construct. This becomes clear when one and the same person can be coded as white, Person of Colour, or Black depending on the region's population they are referred to. To go even further, guessing a person's physical ability or their BIPoC (Black/Indigenous/People of Colour) identity merely based on their names in the editorial board information is not possible at all.

All in all, automated and accurate estimations of diversity are not only technically difficult, but also ethically questionable. While a more promising venue might be to resort to self-reported criteria, there are ethical concerns here as well. A demand for detailed information on diversity is a doubleedged sword: On the one hand, particularized metadata can provide an actual state of equality, detect biases, and thus serve as an inducement to make sustainable changes towards it. On the other hand, stipulating information on sensitive, private criteria such as gender, sexual preference and identification, racialization, ethnicity, or physical abilities can result in a forced outing, such as when that outing is required before one occupies an editorial position at a scholarly journal. Contrary to the original goal, this would lead to even more discrimination. The value of an 'outing privacy' (Skinner-Thompson 2015) necessitates one to find other ways to measure diversity.

In summary, there is no easy solution of measuring this social-constructivist category of diversity based on metadata. It is even paradoxical that in a world in which discrimination persists, it can be even more stigmatizing to ask people to report actual or perceived characteristics on the basis of which they are marginalized. In addition, while diversity is an important issue, achieving it does not necessarily equal inclusivity; representation does not automatically lead to structural equality (Rose 1988). The mere addition of diverse

researchers to editorial boards does not mean that scholarly publishing has become more equal. To achieve actual equality, the whole system needs a sustainable change in its structures.

This is certainly a key limitation inherent to *Open Editors*, or to any scientometric endeavour. This kind of societal diversity, while fundamental, persists, regardless of the scientific system's growing richness in metadata. This is where future reflections on meta-scientific datafication can further tread on in order to find an equitable, long-term solution.

5. Conclusion

There are multiple possibilities with which to proceed on the basis of the dataset thus collected, especially if it becomes linked with PIDs and other open datasets. To attain even more powerful studies and greater utility for research evaluation, it would be necessary to go beyond the bottom-up data collection approach used here and in related grassroots and crowdsourcing initiatives. Instead, publishers themselves should become conscious of the way they display their journals' editors and editorial boards. A uniform structure would enable scientometric investigations to an even more effective scale, enabling one to illuminate how knowledge (in-)equity fares in the landscape of scholarly publications with novel data. Many publishers have recently made their metadata on references available (thanks to the Initiative for Open Citations, or I4OC; Heibi, Peroni and Shotton 2019), and have even opened up metadata on full abstract texts (thanks to the Initiative for Open Abstracts, or I4OA; Tay, Kramer and Waltman 2020). Perhaps an Initiative for Open Editors, or I4OE, might also be called for. While only providing a small step, it may be an indispensable one towards the greater goal of achieving full and inclusive knowledge equity within the broader agenda of Open Science.

Author's contribution

A. N.-P.: Conceptualization, data curation, methodology, software, writing—original draft; T. H.: conceptualization, writing—review and editing; and K. S.: conceptualization, supervision, writing—review and editing.

Funding

This work was supported by Wikimedia Deutschland e. V. within the Open Science Fellows Programme. Open Access funding provided by University of Vienna. In addition, A. N.-P. is a recipient of a DOC Fellowship of the Austrian Academy of Sciences.

Acknowledgements

The authors thank Philipp Zumstein and Parthasarathi Mukhopadhyay for helpful comments, and Bianca Kramer for linking the dataset with ROR.

Conflict of interest statement. None declared.

Data availability statement

The data and the underlying codes are available at *GitHub* and at *Zenodo* via https://doi.org/10.5281/zenodo.4619374.

References

- Altman, M. and Cohen, P. N. (2021) 'Openness and Diversity in Journal Editorial Boards', SocArXiv, DOI: 10.31235/osf.io/4nq97.
- Aryani, A., Fenner, M., Manghi, P., Mannocci, A., and Stocker, M. (2020) 'Open Science Graphs Must Interoperate!', in L. Bellatreche, M. Bieliková, O. Boussaïd, B. Catania, J. Darmont, E. Demidova, F. Duchateau, H. Mark, T. Merčun, B. Novikov, C. Papatheodorou, T. Risse, O. Romero, L. Sautot, G. Talens, R. Wrembel, and M. Žumer (eds) ADBIS, TPDL and EDA 2020 Common Workshops and Doctoral Consortium, pp. 195–206. Cham: Springer International Publishing. DOI: 10.1007/978-3-030-55814-7_16.
- Baccini, A., and Barabesi, L. (2010) 'Interlocking Editorship. A Network Analysis of the Links between Economic Journals', *Scientometrics*, 82: 365–89.
- Baccini, A., Barabesi, L., Khelfaoui, M., and Gingras, Y. (2020) 'Intellectual and Social Similarity among Scholarly Journals: An Exploratory Comparison of the Networks of Editors, Authors and co-Citations', *Quantitative Science Studies*, 1: 277–89.
- Becker, R. A., Wilks, A. R., Brownrigg, R., Minka, T. P., and Deckmyn, A. (2021) 'maps: Draw Geographical Maps', CRAN (R Package, version 3.4.0), https://cran.r-project.org/package=maps.
- Bishop, D. (2020). 'PEPIOPs—Prolific Editors Who Publish in Their Own Publications'. *BishopBlog.* http://deevybee.blogspot.com/2020/08/pepiops-prolific-editors-who-publish-in.html accessed 25 Feb 2021.
- Bornmann, L., and Daniel, H.-D. (2009) 'Reviewer and Editor Biases in Journal Peer Review: An Investigation of Manuscript Refereeing at Angewandte Chemie International Edition', *Research Evaluation*, 18: 262–72.
- Brogaard, J., Engelberg, J., and Parsons, C. A. (2014) 'Networks and Productivity: Causal Evidence from Editor Rotations', *Journal of Financial Economics*, 111: 251–70.
- Card, D., and DellaVigna, S. (2020) 'What Do Editors Maximize? Evidence from Four Economics Journals', *The Review of Economics* and Statistics, 102: 195–217.
- Cho, S., Crenshaw, K. W., and McCall, L. (2013) 'Toward a Field of Intersectionality Studies: Theory, Applications, and Praxis', Signs: Journal of Women in Culture and Society, 38: 785–810.
- Crane, D. (1967) 'The Gatekeepers of Science: Some Factors Affecting the Selection of Articles for Scientific Journals', *The American Sociologist*, 2: 195–201.
- Crenshaw, K. (2015) 'Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics', *University of Chicago Legal Forum*, 1989: 8.
- Curry, C. L. (2017) 'SHERPA Services and SHERPA/RoMEO', Journal of Electronic Resources in Medical Libraries, 14: 135–8.
- Downes, M. (2020) 'Thousands of Australian Academics on the Editorial Boards of Journals Run by Predatory Publishers', *Learned Publishing*, 33: 287–95.
- Erfanmanesh, M., and Morovati, M. (2018) 'Interlocking Editorships in Scientific Journals', *Science and Engineering Ethics*, 24: 1665–7.
- Espin, J., Palmas, S., Carrasco-Rueda, F., Riemer, K., Allen, P. E., Berkebile, N., Hecht, K. A., Kastner-Wilcox, K., Núñez-Regueiro, M. M., Prince, C., Rios, C., Ross, E., Sangha, B., Tyler, T., Ungvari-Martin, J., Villegas, M., Cataldo, T. T., and Bruna, E. M. (2017) 'A Persistent Lack of International Representation on Editorial Boards in Environmental Biology', *PLOS Biology*, 15: e2002760.
- European Commission. (2021) *Towards a reform of the research assessment system*. Luxembourg: Publications Office of the European Union. https://data.europa.eu/doi/10.2777/707440.
- European Research Area and Innovation Committee. (2021) *Research evaluation in a context of Open Science and gender equality*. Luxembourg: Publications Office of the European Union.
- Feeney, M. K., Carson, L., and Dickinson, H. (2019) 'Power in Editorial Positions: A Feminist Critique of Public Administration: Power in Editorial Positions: A Feminist Critique of Public Administration', *Public Administration Review*, 79: 46–55.

- Goyanes, M., and de-Marcos, L. (2020) 'Academic Influence and Invisible Colleges through Editorial Board Interlocking in Communication Sciences: A Social Network Analysis of Leading Journals', *Scientometrics*, 123: 791–811.
- Gutiérrez, J., and López-Nieva, P. (2001) 'Are International Journals of Human Geography Really International?' *Progress in Human* Geography, 25: 53–69.
- Haak, L. L., Fenner, M., Paglione, L., Pentz, E., and Ratner, H. (2012) 'ORCID: A System to Uniquely Identify Researchers', *Learned Publishing*, 25: 259–64.
- Harzing, A.-W., and Metz, I. (2013) 'Practicing What We Preach', Management International Review, 53: 169–87.
- Heibi, I., Peroni, S., and Shotton, D. (2019) 'Software Review: COCI, the OpenCitations Index of Crossref Open DOI-to-DOI Citations', *Scientometrics*, 121: 1213–28.
- Hendricks, G., Tkaczyk, D., Lin, J., and Feeney, P. (2020) 'Crossref: The Sustainable Source of Community-Owned Scholarly Metadata', *Quantitative Science Studies*, 1: 414–27.
- Hodgson, G. M., and Rothman, H. (1999) 'The Editors and Authors of Economics Journals: A Case of Institutional Oligopoly?' The Economic Journal, 109: 165–86.
- Horbach, S., and Halffman, W. (2018) 'The Changing Forms and Expectations of Peer Review', *Research Integrity and Peer Review*, 3: 8.
- Horbach, S. P. J. M., and Halffman, W. (2020) 'Journal Peer Review and Editorial Evaluation: Cautious Innovator or Sleepy Giant?', *Minerva*, 58: 139–61.
- Horton, R. (2003) 'Medical Journals: Evidence of Bias against the Diseases of Poverty', *The Lancet*, 361: 712–3.
- Johnson, R., Watkinson, A., and Mabe, M. (2018). 'The STM report: An overview of scientific and scholarly publishing'. The Hague: International Association of Scientific, Technical and Medical Publishers.
- Kamlah, J., and Zumstein, P. (2019) 'Scrape-editorial-board', GitHub, https://github.com/JKamlah/scrape-editorial-board/releases/tag/2019-10.
- Klebel, T., Reichmann, S., Polka, J., McDowell, G., Penfold, N., Hindle, S., and Ross-Hellauer, T. (2020). 'Peer review and preprint policies are unclear at most major journals', *bioRxiv*, DOI: 10.1101/2020.01.24.918995.
- Laband, D. N., and Piette, M. J. (1994) 'Favoritism versus Search for Good Papers: Empirical Evidence regarding the Behavior of Journal Editors', *Journal of Political Economy*, 102: 194–203.
- Lammey, R. (2020) 'Solutions for Identification Problems: A Look at the Research Organization Registry', *Science Editing*, 7: 65–9.
- Lin, B., and Serebrenik, A. (2016). 'Recognizing gender of stack overflow users'. Proceedings of the 13th International Conference on Mining Software Repositories, MSR '16, pp. 425–9. New York, NY: Association for Computing Machinery.
- Marchitelli, A., Galimberti, P., Bollini, A., and Mitchell, D. (2017) 'Improvement of Editorial Quality of Journals Indexed in DOAJ: A Data Analysis', *Italian Journal of Library, Archives, and Information Science*, 8: 1–21.
- Masic, I. (2017) 'Predatory Publishing—Experience with OMICS International', *Medical Archives*, 71: 304–7.
- Mazov, N. A., and Gureev, V. N. (2016) 'The Editorial Boards of Scientific Journals as a Subject of Scientometric Research: A Literature Review', *Scientific and Technical Information Processing*, 43: 144–53.
- McGinty, S. (1999). Gatekeepers of Knowledge: Journal Editors in the Sciences and the Social Sciences. Westport, CT: Greenwood.
- Mendonça, S., Pereira, J., and Ferreira, M. E. (2018) 'Gatekeeping African Studies: What Does "Editormetrics" Indicate about Journal Governance?', *Scientometrics*, 117: 1513–34.
- Michaud, F. (2022). 'A Registry of Editorial Boards: A new trust signal for scholarly communications?'. *Crossref.* https://www.crossref. org/blog/a-registry-of-editorial-boards-a-new-trust-signal-for-schol arly-communications/> accessed 11 May 2022.
- Nguyen, V. M., Haddaway, N. R., Gutowsky, L. F. G., Wilson, A. D. M., Gallagher, A. J., Donaldson, M. R., Hammerschlag, N., and

Cooke, S. J. (2015) 'How Long is Too Long in Contemporary Peer Review? Perspectives from Authors Publishing in Conservation Biology Journals', *PLoS One*, 10: e0132557.

- Nicholas, D., Herman, E., Jamali, H. R., Abrizah, A., Boukacem-Zeghmouri, C., Xu, J., Rodríguez-Bravo, B., Watkinson, A., Polezhaeva, T., and Świgon, M. (2020) 'Millennial Researchers in a Metric-Driven Scholarly World: An International Study', *Research Evaluation*, 29: 263–74.
- Nishikawa-Pacher, A. (forthcoming). 'Who are the 100 largest scientific publishers by journal count? A webscraping approach'. *Journal of Documentation.*
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T. A., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E. L., Simonsohn, U., Soderberg, C., Spellman, B. A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E. J., Wilson, R., and Yarkoni, T. (2015) 'Promoting an Open Research Culture', *Science*, 348: 1422–5.
- Okagbue, H. I., Atayero, A. A., Adamu, M. O., Bishop, S. A., Oguntunde, P. E., and Opanuga, A. A. (2018) 'Exploration of Editorial Board Composition, Citescore and Percentiles of Hindawi Journals Indexed in Scopus', *Data in Brief*, 19: 743–52.
- Pagedar, A., Sanghi, S., and Gangwar, S. (2022) 'Nurturing an "ethic of collaboration": Dispatches from a case study on Jindal Global Law Review', *Learned Publishing*, 35: 367–75.
- Petersen, J. (2017) 'How Innovative Are Editors?: Evidence across Journals and Disciplines', *Research Evaluation*, 26: 256–68.
- Petersen, J., Hattke, F., and Vogel, R. (2017) 'Editorial Governance and Journal Impact: A Study of Management and Business Journals', *Scientometrics*, 112: 1593–614.
- Pollock, D. (2022). 'News & Views: Publishers and Market Consolidation – Part 1 of 2'. *Delta Think*. https://deltathink.com/ news-views-publishers-and-market-consolidation-part-1-of-2/> accessed 6 July 2022.
- Priem, J., Piwowar, H., and Orr, R. (2022). 'OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts'. *arXiv*. DOI: 10.48550/arXiv.2205.01833.
- Rose, J. (1988) 'Margaret Thatcher and Ruth Ellis', *New Formations*, 6: 3–29.
- Ruiter-Lopez, L., Lopez-Leon, S., and Forero, D. A. (2019) 'Predatory Journals: Do Not Judge Journals by Their Editorial Board Members', *Medical Teacher*, 41: 691–6.
- Safder, I., Hassan, S.-U., Visvizi, A., Noraset, T., Nawaz, R., and Tuarob, S. (2020) 'Deep Learning-Based Extraction of Algorithmic Metadata in Full-Text Scholarly Documents', *Information Processing & Management*, 57: 102269.
- Sarawgi, R., Gajulapalli, K., and Choi, Y. (2011) 'Gender Attribution: Tracing Stylometric Evidence beyond Topic and Genre' in Proceedings of the Fifteenth Conference on Computational Natural Language Learning, CoNLL '11, pp. 78–86. Stroudsburg, PA: Association for Computational Linguistics.
- Scimago Journal & Country Rank. (2022) 'International Science Ranking'. <<u>https://www.scimagojr.com/countryrank.php</u>> accessed 6 July 2022.
- Simón, A. (2016) 'Pitfalls of Predatory Journals: A Personal Account', Comprehensive Psychology, 5: 216522281663169.
- Skinner-Thompson, S. (2015) 'Outing Privacy', Northwestern University Law Review, 110: 159–222.
- Squazzoni, F., Bravo, G., Farjam, M., Marusic, A., Mehmani, B., Willis, M., Birukou, A., Dondio, P., and Grimaldo, F. (2021) 'Peer Review and Gender Bias: A Study on 145 Scholarly Journals', *Science Advances*, 7: eabd0299.
- Stegmaier, M., Palmer, B., and van Assendelft, L. (2011) 'Getting on the Board: The Presence of Women in Political Science Journal Editorial Positions', PS: Political Science & Politics, 44: 799-804.

- Tay, A., Kramer, B., and Waltman, L. (2020) 'Why Openly Available Abstracts Are Important—Overview of the Current State of Affairs'. *Medium.* https://medium.com/a-academic-librarians-thoughts-on-open-access/why-openly-available-abstracts-are-important-overview-of-the-current-state-of-affairs-bb7bde1ed751> accessed 25 Feb 2021.
- Topaz, C. M., and Sen, S. (2016) 'Gender Representation on Journal Editorial Boards in the Mathematical Sciences', *PLoS One*, 11: e0161357.
- Wickham, H. (2022) 'rvest: Easily Harvest (Scrape) Web Pages', CRAN (R Package, version 1.0.3), https://cran.r-project.org/ package=rvest.
- Wilson, M. C., and Knutsen, C. H. (2022) 'Geographical Coverage in Political Science Research', *Perspectives on Politics*, 20: 1024–39.
- Wong, V. S. S., and Callaham, M. L. (2012) 'Medical Journal Editors Lacked Familiarity with Scientific Publication Issues despite Training and Regular Exposure', *Journal of Clinical Epidemiology*, 65: 247–52.

- Wu, D., Lu, X., Li, J., and Li, J. (2020) 'Does the Institutional Diversity of Editorial Boards Increase Journal Quality? The Case Economics Field', *Scientometrics*, 124: 1579–97.
- Xie, Y., Wu, Q., and Li, X. (2019) 'Editorial Team Scholarly Index (ETSI): An Alternative Indicator for Evaluating Academic Journal Reputation', *Scientometrics*, 120: 1333–49.
- Xu, B., Meng, H., Qin, S., Liu, Y., Li, Z., Cao, J., Lin, Y., Zhang, Y., and Wang, Z. (2019) 'How International Are the Editorial Boards of Leading Spine Journals? A STROBE-Compliant Study', *Medicine*, 98: e14304.
- Yoon, A. H. (2013) 'Editorial Bias in Legal Academia', Journal of Legal Analysis, 5: 309–38.
- Zhang, L., Sun, B., Jiang, L., and Huang, Y. (2021a) 'On the Relationship between Interdisciplinarity and Impact: Distinct Effects on Academic and Broader Impact', *Research Evaluation*, 30: 256–68.
- Zhang, T., Shi, J., and Situ, L. (2021b) 'The Correlation between Author-Editorial Cooperation and the Author's Publications in Journals', *Journal of Informetrics*, 15: 101123.