On the Peripheries of Scholarly Infrastructure: A look at the Journals Using Open Journal Systems[1]

Juan Pablo Alperin[*], Kevin Stranack[**] and Alex Garnett[**]

[*]juan@alperin.ca
Canadian Institute for Studies in Publishing, Simon Fraser University, 515 West Hastings Street, Vancouver, 6B 5K3 (Canada)
Public Knowledge Project, Simon Fraser University, 8888 University Drive, Burnaby, V5A 1S6 (Canada)

[**] kstranac@sfu.ca; garnett@sfu.ca
Public Knowledge Project, Simon Fraser University, 8888 University Drive, Burnaby, V5A 1S6 (Canada)

Background

The Public Knowledge Project (PKP) is a research and development initiative of Simon Fraser University and Stanford University, with a focus on understanding and building enhanced modes of scholarly communication that facilitate open access, high quality publishing, and local capacity building and participation.

One of the most significant contributions from PKP has been the development of its free, open source Open Journal Systems (OJS) software.[i] OJS is a professional journal publishing platform that is easy to download, install, and operate with minimal server requirements. OJS allows for online submissions, peer review, copyediting and layout, and publishing. It also provides connections to indexing (e.g., PubMed, DOAJ), digital identifiers (e.g., CrossRef, ORCID), and preservation services (e.g., LOCKSS, Portico).

Because OJS is free and is designed to maximize efficiencies for publishing activities, publishers have been able to launch journals that would otherwise have been financially untenable. Although some existing publishers have made use of the system, the majority of OJS users are new to publishing and are based in academia, either through library publishing programs or by individual "scholar-publishers," determined to build communities of interest around their research areas (Edgar & Willinsky, 2010). Most are small-scale, often operating with in-kind contributions and minimal budgets (Edgar & Willinsky, 2010). This type of use suggests an increased participation in scholarship, both in terms of who publishes and who read.

In a sense, OJS can be already be said to form part of the scholarly infrastructure. It provides a layer over which scholarly activity takes place, and it brings the use of standards and best practices to those activities. It acts as one layer of infrastructure especially for those journals who would are not formally part of the scholarly infrastructure

Methods

---

[1] In the interest of full disclosure, it should be noted that the authors of this work are all affiliated with the Public Knowledge Project, the creators of Open Journal Systems.

We have sought to use automated methods for identifying OJS journals, and subsequently use a combination of web scraping[ii] and the journal's OAI PMH endpoint[iii] to collect publicly available information about the journal, including article metadata. Because this process is entirely automated, it can be continuously run, yielding a dataset of journals and articles data that will eventually include longitudinal data. These data are then processed to identify the number of journals, articles, and their geographic location.

In a latter phase, yet to be completed (and thus not presented here) these data will be complemented with an online survey, loosely based on a similar study of OJS journals done by Edgar & Willinsky (2010), and targeted at the email addresses collected through the first phase data collection.

The first challenge in studying these journals is to learn of their existence. Because the OJS software is open source, each journal or publisher can install the software on their own server, without ever needing to register with PKP or elsewhere. Only once a journal's web address (URL) has been identified is it possible to learn how many articles it publishes in any given year, its geographic location, and its identifying information and other metadata. To collect this data, we have devised a necessarily complex method of extracting and processing the information (Figure 1).

Figure 1. Steps to collect and process OJS journal information

1. Process the PKP website logs for referrer URLs that 'look like' OJS journals
2. Attempt to contact the OAI PMH URL corresponding to the journal URL (following known OJS URL patterns) to verify if it is an OJS journal
3. Save the repository identifier, I.P. address, OJS version number
4. Identify all the journals for this installation using the OAI verb "ListSets"
5. Save the journal name, and journal contact email address from the OAI response for later use
6. Add known OJS OAI URLs to an instance of the PKP Harvester
7. Look up the journal's country
8. Collect the article metadata for every journal using OAI PMH
9. Process the article data to identify number of articles published per year, the country of origin of the journal, etc.

Note: Code for step 1 can be found at: https://github.com/pkp/ojsstats/blob/master/pkp-log-parser.php; Code for steps 2-5 can be found at: https://github.com/pkp/ojsstats/blob/master/checkOJS.py; Code for step 6 can be found at: https://github.com/pkp/ojsstats/blob/master/harvesterInsert.py; Code for step 7 can be found at: https://github.com/pkp/ojsstats/blob/master/countryLookup.py; Step 8 is handled by the PKP Harveser, which can be found at: https://github.com/pkp/harvester; Code for step 9 can be found at: https://github.com/pkp/ojsstats/blob/master/statscrunch.py. The process is coordinated by a "cron" job, which can be found at: https://github.com/pkp/ojsstats/blob/master/cron.sh. Detailed description of the each step can be found in the README: http://github.com/pkp/ojsstats.

There are several known ways in which the above system can fail to correctly identify and query a valid OJS journal. It is possible, for example, that a journal removes the links to the PKP website from their installation altogether, rendering it invisible to our logs; similarly, it is possible that journals have modified the code to change the way URLs are structured, making them difficult to correctly identify; it is also possible that journals have disabled their OAI endpoint, thereby invalidating the way in which we verify an OJS installation and subsequently collect data; or, a journal may simply be unavailable online at the time when we

attempt to crawl it, leading our system to believe that the site no longer exists. Any of these circumstances, and possibly others, would result in us undercounting the number of OJS journals.

While we realize there are simpler ways of collecting this data about OJS journals (i.e., the data could be pushed from the OJS installation to a centralized system), it has always been PKP's approach to give control of the data to the journals, and to ask for nothing in return for using the software. In recent years, however, PKP has recognized the need to better understand its journal community, and, as a result, has included a "phone home" feature that provides PKP with a minimal set of publicly available data (with an opt-out option). This feature has only been available since 2015, so it was still necessary to develop the methods outlined above to produce the results that follow.

Results

We identified 6,271 installations of OJS with some content, spread across 136 countries on 6 continents. These installations collectively host 9,828 "journals" that meet our arbitrary threshold of at least 18 articles published in the previous two years.[iv] At the time of writing, there were 7,491 journals that met this threshold for 2015, and 9,315 that met it for 2014 (some journals appear to add content with a delay in publication resulting in the appearance of a drop-off in numbers in 2015). Of these, the top 3 countries, by number of journals published using OJS with recent content, are 1,426 in Brazil, 1,075 in Indonesia, and 912 in the US (Figure 2). No other country had more than 500 journals that met the criteria in 2015. Interestingly, Latin America (led by Brazil) publishes approximately one third of all OJS journals.

Naturally, the corresponding country income level accompanies the geographic distribution. Approximately 42% of the journals are published by high-income countries, 38% by upper-middle income countries, and the final 20% by low or low-middle income countries (Table 2).[v]

These journals have published 2,565,300 articles since 1990. As might be expected, more articles are published using the software in more recent years, with over 300,000 articles published in 2013 and 2014, and over 250,000 in 2012 and 2015 (Figure 3). This is due primarily to an increase in the number of journals that have gone (and stayed) online over time (Figure 4). Given that the software was originally released in 2001, it seems that journals come online, add some amount of back-content (archives), and then continue to publish. In recent years (since 2010), OJS journals that meet the 18-article threshold have published 29.8 articles (up from 25.6 in the 15 years prior) (Figure 5).

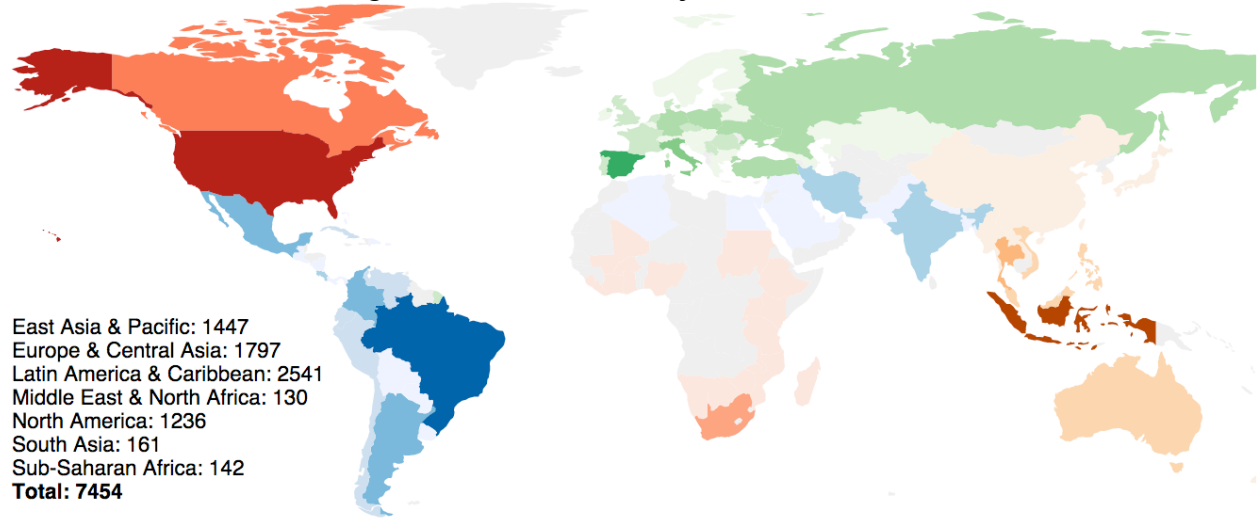Figure 2. Location of OJS journals in 2015

East Asia & Pacific: 1447
Europe & Central Asia: 1797
Latin America & Caribbean: 2541
Middle East & North Africa: 130
North America: 1236
South Asia: 161
Sub-Saharan Africa: 142
Total: 7454

Table 2. Country Income-Level of OJS Journals for 2015

Income Category	Number and percent of journals (2015)	Number and percent of articles (2015)	Average Number of Articles per journal (2015)
Low	53 (0.7%)	1,225 (0.4%)	23.1
Lower-middle income	1,525 (19.7%)	68,571 (23.8%)	46.6
Upper middle income	2,832 (37.8%)	94,945 (33.0%)	33.5
High income	3,129 (41.8%)	123,211 (42.8%)	39.4

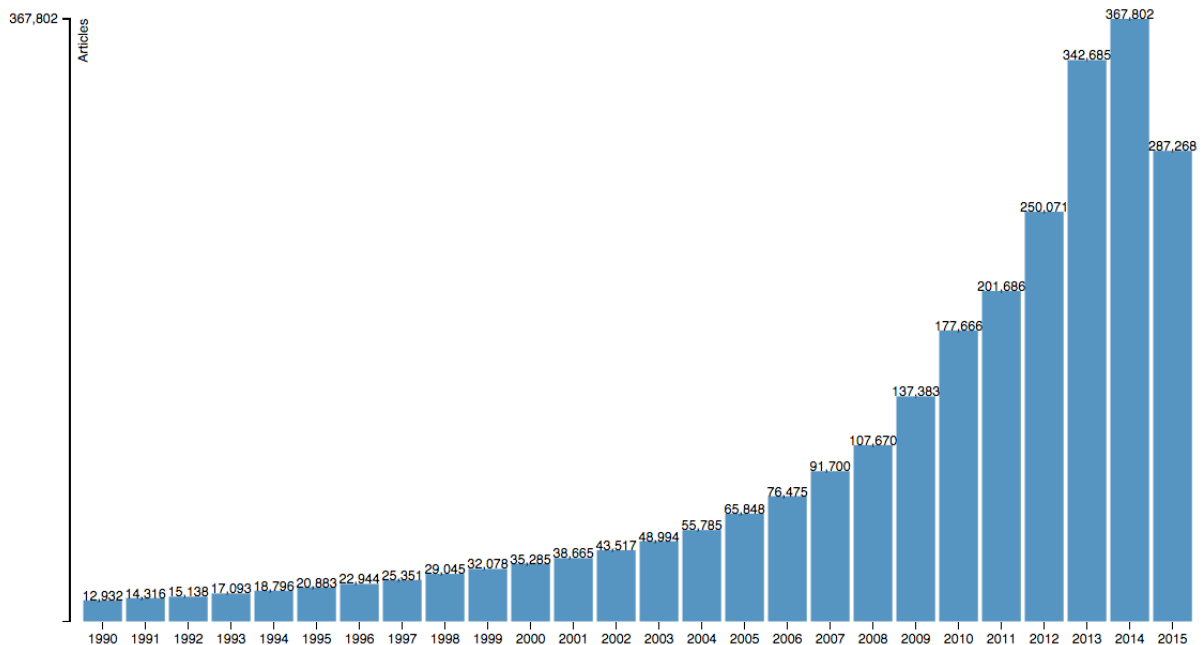Figure 3. Number of Articles Published in OJS Journals By Year

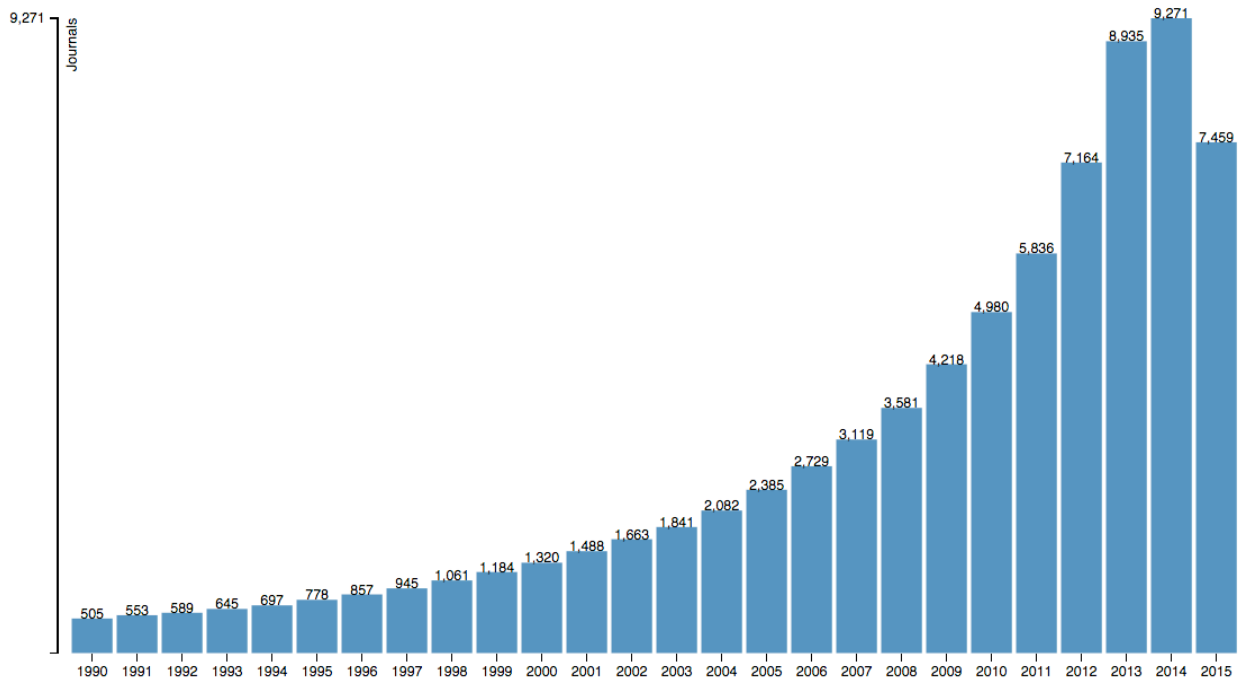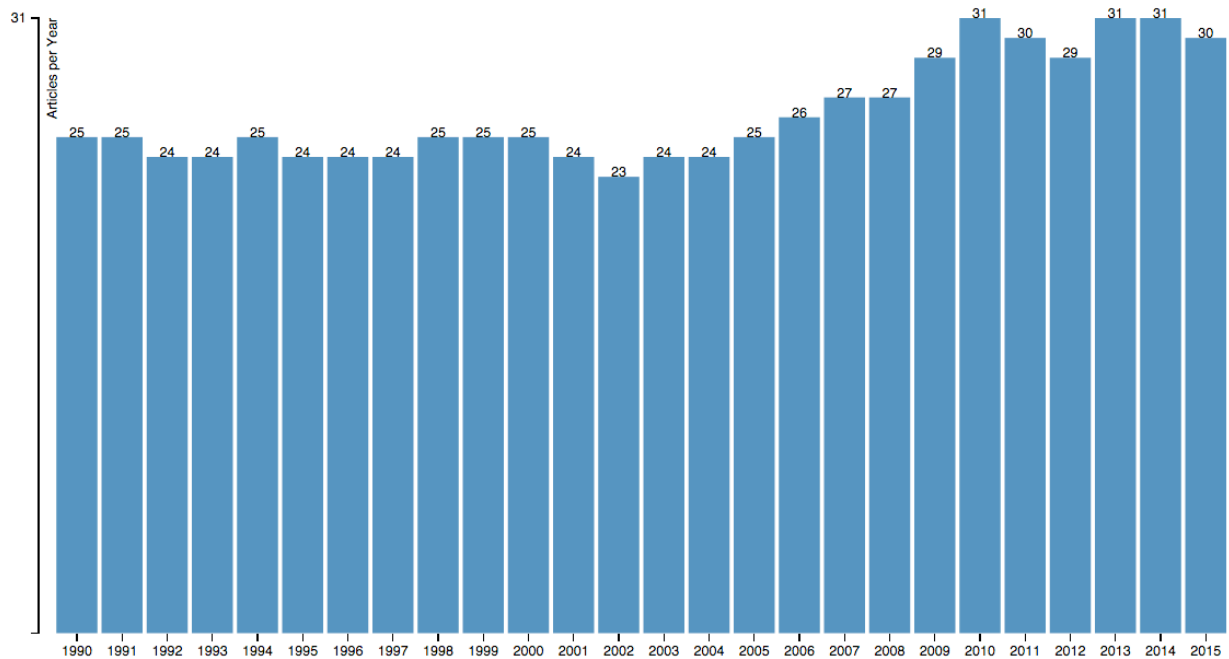Figure 4. Number of Journals Using OJS by Year

Figure 5. Average Number of Articles Published per Year in OJS Journals

Discussion and Future Work

Because of the imperfect criteria and roundabout method for collecting the data, these numbers should be treated as estimates. However, although we are surely counting some instances that are not really journals as have been traditionally understood, we are also certain that we have missed some instances altogether, and are reasonably comfortable that they provide a sufficiently accurate estimate to provide a better understanding of the type of journals that are in need of access to the scholarly infrastructure. Future studies and improved methods, including the better automated data collection, will allow us to assess these estimates.

The geographic extent and sheer number of journals and articles—most of which are not likely to be found in international citation databases—should serve to open our eyes to the need to think more broadly about who is actively trying to communicate scholarship.

The extent to which these journals are currently outside the scholarly infrastructure cannot be overstated. Not only are these journals excluded from major citation databases (there is no need to test their inclusion, as the sheer volume of journals from outside North America and Europe dwarf the total number of journals from these regions in Web of Science and Scopus), they are also lacking some of the basic elements of the scholarly infrastructure, such as DOIs. As of November 2015, there were only 837 CrossRef members that used OJS (CrossRef, personal communication). These members collectively minted 701,622 DOIs (out of over 2.5M articles).[vi] Even under the generous assumption that each DOI minted corresponds to an article (i.e., not supplementary files, data, etc.), this amounts to less than a third of the total articles published with OJS journals.

The geographic and economic distribution of access to elements of the scholarly infrastructure, such as DOIs, remains to be studied.[vii] It is nonetheless striking (although likely coincidental) that the number of DOIs minted by CrossRef members using OJS in 2015 (127,026) coincides with the number of articles published by OJS journals from high-income countries that year (123,211).[viii] While PKP works towards improving the adoption of DOIs for these journals and CrossRef works to revamp its small-publisher tools, the reality is such that the scholarly metrics and indicators that can be calculated today do not consider these journals on the peripheries of the scholarly communication infrastructure.

As can be seen from the complexity of the data collection described above, deriving metrics from these small, decentralized publishers is a challenge. Attempting to extend our efforts beyond the journals published with OJS would make the task nearly impossible. In this way, OJS is itself a piece of the infrastructure, one that allows us to account for over 2.5M articles and gives us the opportunity to learn more about them.

If these journals were part of the larger metrics infrastructure and systems, it would fundamentally change the peripheral nature of the scholars and scholarly work contained therein. As these figures highlight, there are literally millions of articles published from this periphery. The over 20% of OJS journals in low and low-middle income countries (corresponding to nearly 25% of all OJS-published articles) need to be included if the indicators and metrics are to be truly representative of the scholarship that is out there.

Being inclusive in the indicators and metrics infrastructure is necessary if we wish to use these indicators to understand how scholarship is carried out around the world. Being inclusive, however, is not as simple as putting out an open invitation to participate. It is also necessary to lower the barriers to access by making sure the infrastructure is suitable and adapted to the needs of everyone. To do that effectively, we need to know who we are trying to include, so that they can be consulted and can participate from the beginning as the infrastructure and metrics are designed and built. The research described here is a first look at many of the thousands of journals who are currently not being included.

This work is very much ongoing. In the coming months, we will conduct a survey of these journals to learn about the practices, challenges, funding, and other information that will be directly relevant to those seeking to build truly inclusive scholarly infrastructure. While this first phrase has given a general sense of the scale and extent of journals, it is only the beginning of a longer process of trying to understand the global research landscape.

References

Alperin, J. P. (2013, October 11). Are ALMs/altmetrics propagating global inequality? Retrieved from https://speakerdeck.com/jalperin/altmetrics-propagating-global-inequality

Alperin, J.P. (2014a). Altmetrics could enable scholarship from developing countries to receive due recognition. *LSE Impact Blog.* Retrieved from:
http://blogs.lse.ac.uk/impactofsocialsciences/2014/03/10/altmetrics-for-developing-regions/

Alperin, J. P. (2014b). South America: Citation databases omit local journals. *Nature*, *511*(7508), 155–155. http://doi.org/10.1038/511155c

Bilder, G., Lin, J., & Neylon, C. (2015). Principles for Open Scholarly Infrastructures-v1. http://doi.org/10.6084/m9.figshare.1314859.v1

Cetto, A. M., & Alonso-Gamboa, J. O. (1998). Scientific periodicals in Latin America and the Caribbean: a global perspective. *Interciencia*, *23*(525), 84–93. Retrieved from http://www.interciencia.org/v23_02/cetto.pdf

Collazo-Reyes, F., Luna-Morales, M. E., Russell, J. M., & Pérez-Angón, M. (2008). Publication and citation patterns of Latin American & Caribbean journals in the SCI and SSCI from 1995 to 2004. *Scientometrics*, *75*(1), 145–161. http://doi.org/10.1007/s11192-007-1841-6

Edgar, B. D., & Willinsky, J. (2010). A Survey of the Scholarly Journals Using Open Journal Systems. *Scholarly and Research Communication*, *1*(2), 1–22.

Haustein, S., Sugimoto, C., & Larivière, V. (2015). Guest editorial: social media in scholarly communication. *Aslib Journal of Information Management*, *67*(3). http://doi.org/10.1108/AJIM-03-2015-0047

Russell, J. M., Madera Jaramillo, M. J., & Shirley, A. (2009). El análisis de redes en el estudio de la colaboración científica. *Redes: Revista Hispana Para El Análisis de Redes*

Sociales, *17*(2), pp. 39–47. Retrieved from
http://dialnet.unirioja.es/servlet/articulo?codigo=3165613&orden=243060&info=link

Terra Figari, L. I. (2008). Diseminación del Conocimiento Académico en América Latina. In *Antropología Social Y Cultural En Uruguay 200*7. Montevideo, Uruguay: UNESCO. Retrieved from
http://www.unesco.org.uy/shs/fileadmin/templates/shs/archivos/anuario2007/articulo_15.pdf

---

[i] http://pkp.sfu.ca/ojs

[ii] Using a script to make web requests, like a browser would, to access content from the journal pages.

[iii] OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) is a standard used by many journal and repository systems to expose metadata in a machine-readable format.

[iv] This 18 articles in the previous two years is based on Canada's Social Science Research Council's Aid to Scholarly Journals' guideline that stipulates 12 peer reviewed articles in the previous two years (see http://www.sshrc-crsh.gc.ca/funding-financement/programs-programmes/scholarly_journals-revues_savantes-eng.aspx#a5 for details). We assume, therefore, that two thirds of articles are peer reviewed to arrive at an 18 article threshold). This minimum was set in an effort to filter out test installations, or installations that have just gone online and have not fully gotten underway. Since we are not verifying any other criteria, the word "journal" is therefore used loosely.

[v] The income levels are drawn from the World Bank classifications, with both high-income categories (OECD and non-OECD combined) and the low and lower-middle income categories combined.

[vi] Note that a member could correspond to more than one journal.

[vii] Ironically, the most suitable way of cross-checking the OJS database and others would be to use DOIs.

[viii] Even if this were not a coincidence, and the majority DOIs were minted by journals from high-income countries, it should be noted that DOIs minted in 2015 do not necessarily correspond to content published in 2015, so it is unlikely there would be a one-to-one correspondence.