



26<sup>th</sup> International Conference on Science and Technology Indicators  
"From Global Indicators to Local Applications"

#STI2022GRX

Full paper

## STI 2022 Conference Proceedings

*Proceedings of the 26<sup>th</sup> International Conference on Science and Technology Indicators*

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

## Proceeding Editors

Nicolas Robinson-Garcia  
Daniel Torres-Salinas  
Wenceslao Arroyo-Machado



**Citation:** Basson, I., Simard, M.A., Sugimoto, C.R., & Larivière, V. (2022). The relationship between open access publishing and referencing. In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Eds.), *26th International Conference on Science and Technology Indicators*, STI 2022 (sti22225). <https://doi.org/10.5281/zenodo.6957201>



**Copyright:** © 2022 the authors, © 2022 Faculty of Communication and Documentation, University of Granada, Spain. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#).

**Collection:** <https://zenodo.org/communities/sti2022grx/>

26<sup>th</sup> International Conference on Science and Technology Indicators | STI 2022

## “From Global Indicators to Local Applications”

7-9 September 2022 | Granada, Spain

#STI22GRX

### The relationship between open access publishing and referencing<sup>1</sup>

Isabel Basson\*, Marc-André Simard\*, Cassidy R. Sugimoto\*\*, and Vincent Larivière\*<sup>\*,\*\*,\*\*\*</sup>

\* [isabel.basson@umontreal.ca](mailto:isabel.basson@umontreal.ca); [marc-andre.simard.1@umontreal.ca](mailto:marc-andre.simard.1@umontreal.ca); [vincent.lariviere@umontreal.ca](mailto:vincent.lariviere@umontreal.ca);

École de bibliothéconomie et des sciences de l'information, Université de Montréal, Pavillon Lionel-Groulx  
3150, rue Jean-Brillant, Montréal, Québec, H3T 1N8 (Canada)

\*\* [sugimoto@gatech.edu](mailto:sugimoto@gatech.edu);

School of Public Policy, Georgia Institute of Technology, D.M. Smith Building, 685 Cherry Street, Atlanta,  
Georgia, 30332 - 0345 (USA)

DST-NRF Centre of Excellence in Scientometrics and STI Policy; and Centre for Research on Evaluation,  
Science and Technology, Stellenbosch University, Krotoa Building, 52 Ryneveld Street, Stellenbosch, Western  
Cape, 7600 (South Africa)

\*\*\*

Observatoire des sciences et des technologies, Université du Québec à Montréal, Pavillon Paul-Gérin-Lajoie,  
1205 rue Saint-Denis, Montréal, Québec, H2X 3R9 (Canada)

#### Introduction

The benefits of open access (OA) publishing have been touted by many researchers, funders, librarians, and governments (Archambault et al., 2014; Evans & Reimer, 2009; Piwowar et al., 2018; Suber, 2003). The advantages of OA are well established: it contributes to a higher visibility for research papers, it is associated with higher citation rates, and it benefits the general public, which is supporting basic research through taxes (Archambault et al., 2014; Piwowar et al., 2018; Suber, 2003).

Various studies have aimed to measure the growth of OA and the prevalence of OA across countries (Archambault et al., 2014; Piwowar et al., 2018; Robinson-Garcia et al., 2020; Simard et al., 2021; Van Leeuwen et al., 2019), though the majority of these have focused on the publication side and not the referencing behavior. The aim of the study is to measure whether researchers are using OA more than expected, with a focus on country of affiliation and discipline. Here usage refers both to both OA publication (i.e., the availability of their papers) as well as whether they incorporate OA papers into their work (i.e., cite OA papers). More specifically, this study investigates the following research questions:

- What proportion of papers published in 2019 and 2020 are available as OA?
- What proportion of references papers published in 2019 and 2020 are to OA papers?

<sup>1</sup> This work is based on research funded by the International Development Research Centre (IDRC) Science Granting Councils Initiative in Sub-Saharan Africa program under project 109272.

- What countries and regions render their papers OA more frequently than expected?
- What countries and regions reference OA papers more frequently often than expected?
- Do we see a pattern regarding OA usage for countries from different income groupings?

### Methodology

Dimensions was selected as the data source for this study, and documents were restricted to those listed as article publication types. To examine the publication and referencing of OA papers, two datasets were extracted. The first dataset consists of all Dimensions-indexed journal publications published in 2019 and 2020. This dataset contains the following data on the papers: publication year, country affiliation of first author, research field, and access status at the time of data extraction. The second dataset consists of data on the cited references to these publications limited to a 10-year citation window. This dataset contains data on the cited and citing paper, with data on the access status of the cited paper.

Unpaywall's five-category classification system (Piwowar et al., 2018) is used to distinguish the OA status of papers. When classifying OA in this paper, we refer to all papers that are not in the "closed" category. This potentially overcounts closed access papers, due to a lack of an explicit category for this in the metadata as well as the occurrence of missing metadata (Basson et al., 2022; Guerrero-Bote et al., 2021). These issues are mitigated by removing the papers and references with missing metadata. We assigned each paper only to one discipline, that is, the first to which it is associated in the database. To classify papers by region and by income categories we used the World Bank Country and Lending Groups (2022) and the country affiliation of the first author. A small number (27 out of 236) of the countries and territories present in the Dimensions data are not included in the World Bank categories and thus are not represented on the figures.

To measure whether authors from different regions have rendered their papers OA more often than expected we calculated the normalized open access publication score (NOAPub), which takes into account that the rates of OA vary drastically by discipline (Maddi, 2020). This was calculated by normalizing each country percentage of OA papers in a given specialty by the average percentage of OA papers in that specialty, and then weighing and aggregating the score. A similar approach was followed to determine the normalized open access referencing score (NOARef), referring to OA status of the referenced paper, and the field in which the referencing paper is published.

### Results

**¡Error! No se encuentra el origen de la referencia.** provides access types of the 5,278,411 papers indexed in Dimensions published in 2019 and 2020 as well as for papers published between 2010 and 2020. It also presents access types of the references (10-year citation window) cited in the 2019 and 2020 papers. **¡Error! No se encuentra el origen de la referencia.** demonstrates that 49.9% of papers published in 2019 and 2020 are currently available as OA and 51.3% of references from all papers published during those two years are to papers that are currently available as OA. These two percentages are more similar than the percentage of the papers published between 2010 and 2020 that are OA (i.e., 43.3%), suggesting that the OA percentage of the references of papers is not simply a reflection of the access status of the available papers. When we investigate by OA access type, we observe a similar pattern. The exception is gold OA, with a difference of 9.2 percentage points as opposed to 11.9

percentage points. The results suggest that references in recent papers are more open than one would expect, given OA publication practices in the last decade and that they are more open than the publications in which they appear. This demonstrates that the use of OA exceeds the production of it.

Table 1 Summary of open access publications and references by access type

access types	Publications (2019 & 2020)		Publications (2010 to 2020)		References (10y window for 2019 & 2020 papers)	
	count	%	count	%	count	%
Closed	2,643,173	50.1%	12,517,716	56.7%	40,947,160	48.7%
Gold	1,537,491	29.1%	4,394,123	19.9%	14,431,590	17.2%
Bronze	414,923	7.9%	2,104,687	9.5%	10,013,085	11.9%
Green-only	367,373	7.0%	2,107,786	9.5%	12,887,901	15.3%
Hybrid	315,451	6.0%	970,169	4.4%	5,730,445	6.8%
Total	5,278,411	100%	22,094,481	100%	84,010,181	100%

At the level of fields (see Table 2), we observe that, in most cases (17 out of 22 fields), a higher percentage of the references are to OA papers than the number of OA papers available for those years (2010 to 2020). The largest differences are observed for the fields of Philosophy and Religious Studies (15.3 percentage points) and Medical and Health Sciences (15 percentage points), suggesting that researchers draw heavily upon OA. This stands in contrast to the OA status of the citing papers themselves; only in a few fields (7 out of 22 fields) is the percentage of the references to OA papers higher than the percentage of OA papers published in 2019 and 2020, suggesting that OA publication practices are beginning to keep pace with referencing. The largest differences are observed for Built Environment and Design (27 percentage points) and Education (12 percentage points); that is, these fields have much higher practices of publishing OA than referencing OA. This illustrates differences in OA citing behavior and publishing behavior between the different fields. However, it should be noted that, even if we observed little difference on a global scale, we cannot assume the percentages to be the same for the references and the papers on the level of disciplines, as authors can cite from beyond their field of publication. These differences in percentage of OA publications and references to OA papers between fields and years confirms the need for normalization by field, year, and country. Therefore, the remaining comparisons are conducted using normalized measures.

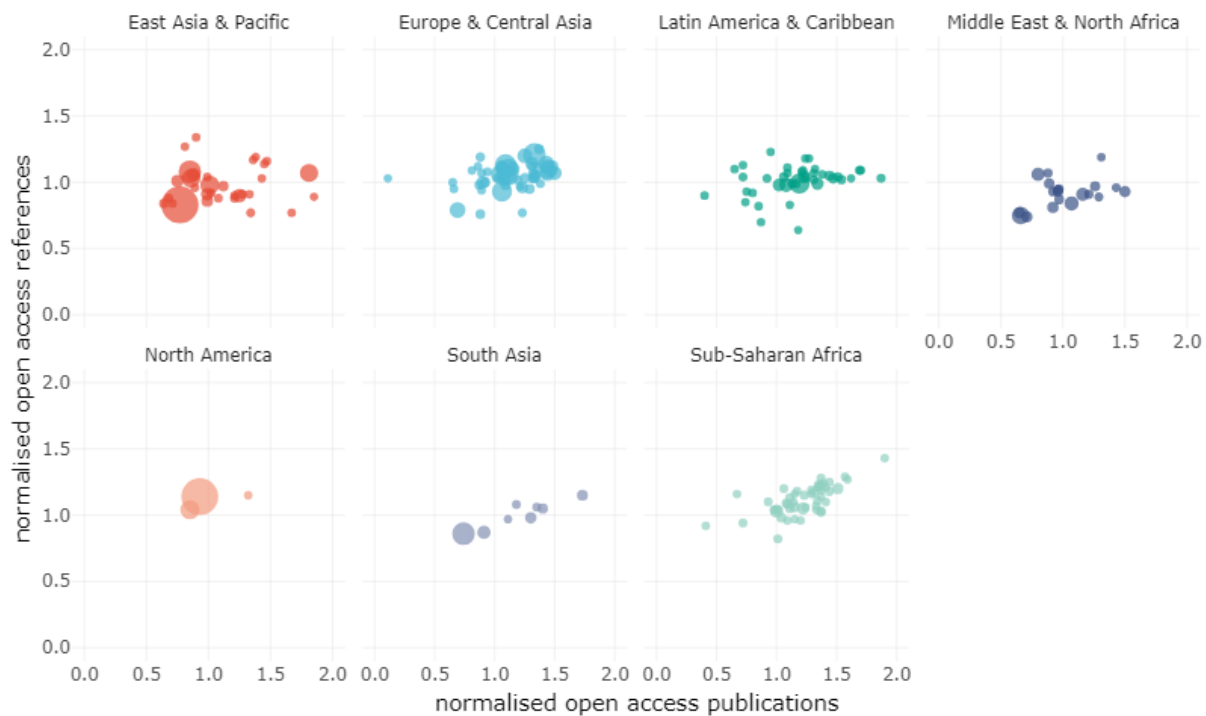
Table 2 Summary of open access publications and references by field

Research fields	Pubs. (2010 to 2020)	Refs. 10-year window (2019 & 2020 papers)	Pubs. (2019 & 2020)
Biological Sciences	56.9%	67.5%	60.0%
Physical Sciences	54.8%	62.0%	62.0%
Mathematical Sciences	50.5%	49.0%	51.8%
Built Environment and Design	50.4%	39.4%	66.4%
Medical and Health Sciences	48.4%	63.4%	55.0%
Psychology and Cognitive Sciences	46.4%	54.7%	51.9%
Economics	45.9%	48.3%	52.2%
Law and Legal Studies	45.9%	45.7%	57.5%
Agricultural and Veterinary Sciences	42.8%	48.5%	48.7%
Information and Computing Sciences	42.4%	41.9%	52.3%
Education	42.0%	38.9%	51.3%
Environmental Sciences	41.7%	44.9%	52.9%
Earth Sciences	40.0%	46.8%	45.4%
Technology	39.7%	43.9%	48.9%
Studies in Human Society	39.6%	43.1%	46.6%
Language, Communication and Culture	38.3%	41.9%	48.8%
Philosophy and Religious Studies	31.7%	47.0%	43.9%
Commerce, Management, Tourism and Services	31.2%	31.4%	37.9%
History and Archaeology	31.1%	43.8%	41.1%
Engineering	28.6%	29.3%	37.2%
Studies in Creative Arts and Writing	26.8%	39.5%	33.5%
Chemical Sciences	26.2%	31.8%	33.1%

### *Analysis by region*

Figure 1 shows that the normalized OA proportion of references (NOARef) displays fewer extremes than the normalization OA proportion of publication (NOAPub). We can also observe regional and country differences. Sub-Saharan Africa and most of Europe & Central Asia publish and reference OA papers (regardless of type) more than expected. Most of Latin America & Caribbean publish more in OA, but reference OA papers as expected based on the fields in which the authors publish. North America (the USA and Canada) shows a different behavior for the NOAPub than for the NOARef: North American authors publish less in OA than expected based on the fields that they are publishing in, but reference OA papers more than expected. Similar can be said about countries in the Middle East and North Africa. China both references and publishes less in OA considering the fields in which associated authors publish.

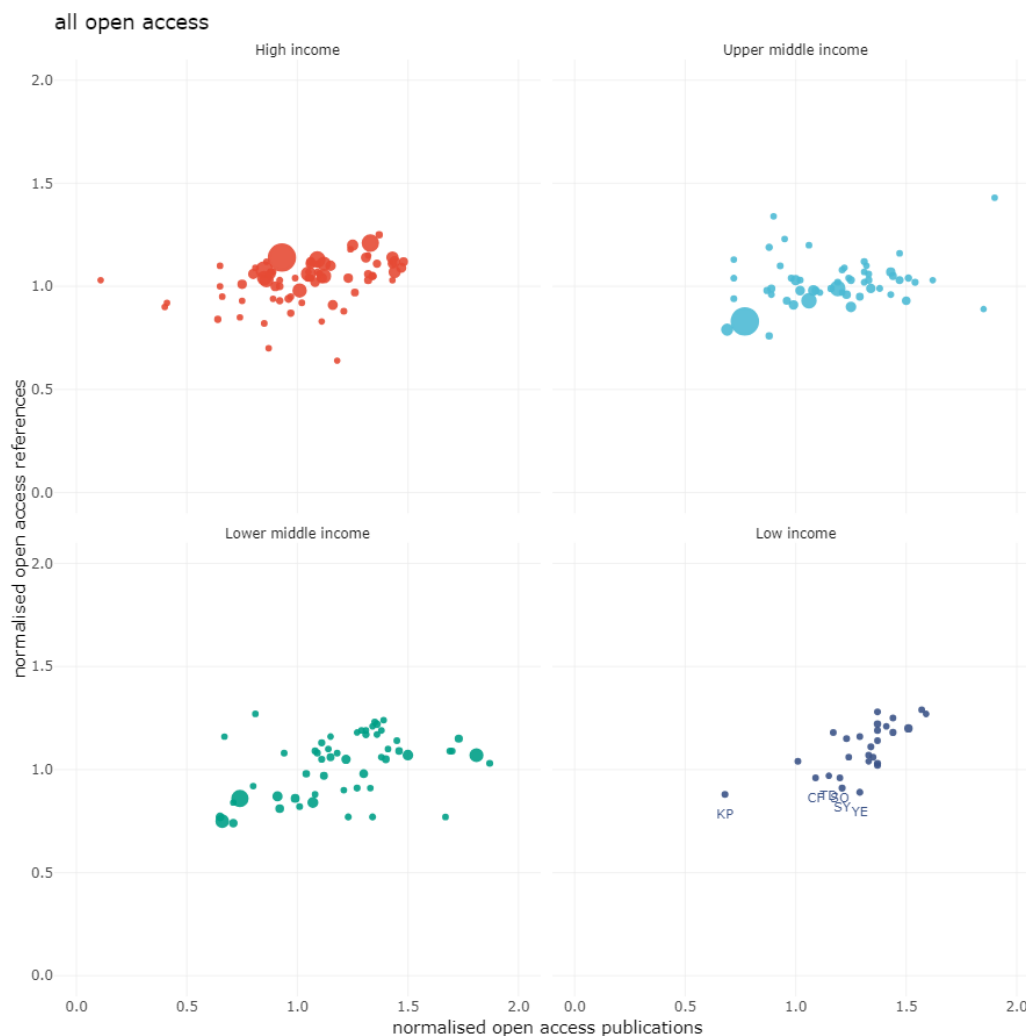
Figure 1: Scatter plots of NOAPub and NOARef of countries, faceted by World Bank regions, sized based on total publication output, for all access



#### *Analysis by income group*

Disaggregating countries by World Bank income group (with the total number of papers regardless of access type given in brackets) presents no uniform pattern, except for the low-income group (see Figure 2). Most of the countries in the low-income group follow a similar pattern, both publishing and referencing OA papers more often than expected. The exceptions are North Korea, (364), the Central African Republic (66), Chad (45), Somalia (66), Syria (687), and Yemen (555), of which all except North Korea publish OA papers more than expected. It should be noted however, that most low-income countries (23 of 27 low-income countries) are Sub-Saharan countries. A similar observation could be made for lower-middle income countries in Sub-Saharan Africa i.e., that most publish and reference OA more than expected (15 of 18), with the exception of the Comoros (8) (which publishes in OA less than expected, but reference it more than expected), Mauritania (371) (which publishes as expected but references less than expected), and Cameroon (2,445) (which publishes and references OA as expected). These observations potentially conflate patterns by region with income group or vice versa, with 41 of all Sub-Saharan countries (total 48 in data) falling into these two income groupings.

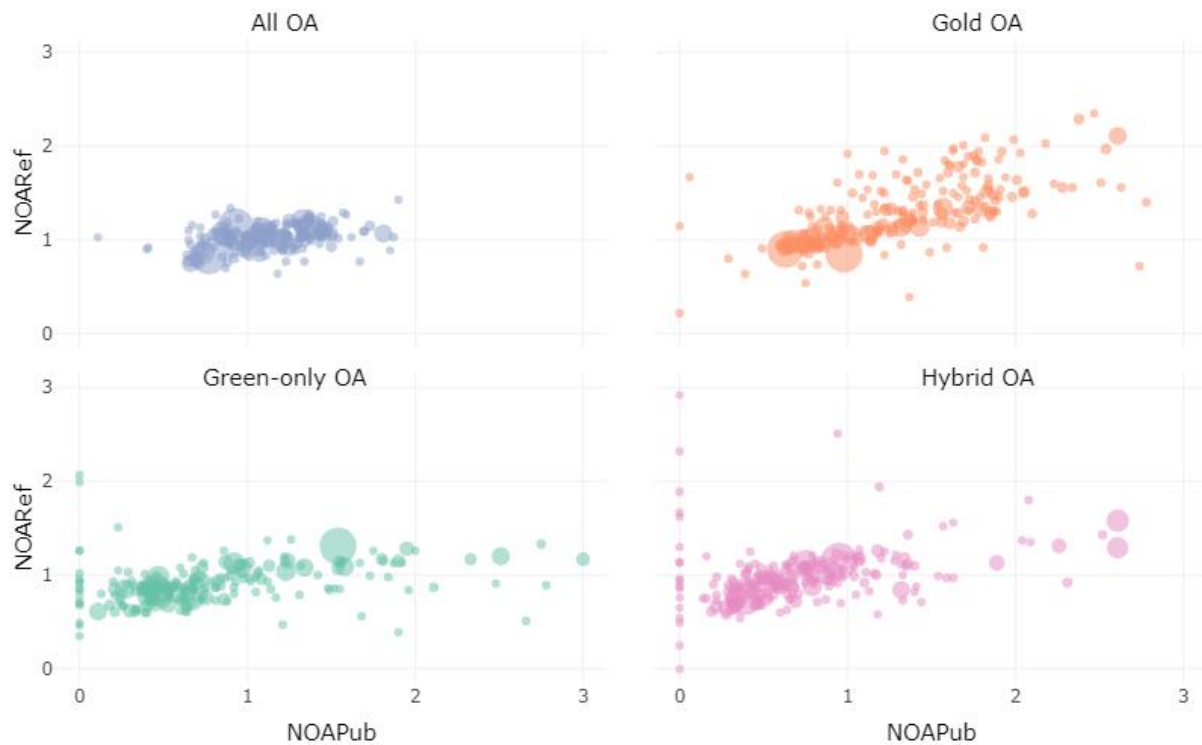
Figure 2: Scatter plots of NOAPub and NOARef of countries, faceted by World Bank income group, sized based on total publication output, for all open access



### *Analysis by access type*

We investigated the relationship between the NOAPub and NOARef of countries for each type of OA by calculating the Spearman's rank correlation between the scores. We observed a strong positive association between the scores for the gold OA ( $\rho=0.67$ ,  $p=0.00$ ), all OA ( $\rho=0.43$ ,  $p=0.00$ ), and green-only OA ( $\rho=0.42$ ,  $p=0.00$ ), and a moderate positive relationship for hybrid OA ( $\rho=0.33$ ,  $p=0.00$ ). Thus, at least for gold OA, publishing and referencing are highly correlated. The publication and referencing profiles of countries differ clearly between the different access types for most countries, as illustrated in Figure 4 by comparing countries' distribution for the green-only OA and gold OA. For all other income groupings besides high income countries, countries tend to publish and reference gold OA papers more than expected, whereas the opposite tends to be the case for green-only OA. Exceptions to this can also be observed, mainly with countries having a high NOARef, without having a high NOAPub for green-only OA.

Figure 3: Scatter plots of NOAPub and NOARef of countries, faceted by open access type, sized based on total publication output



Extreme values ( $\text{NOARef} > 3.0$ ) for the NOARef can be observed for green-only OA and hybrid OA. That is, countries that referencing above expected values. Many of these are for countries or territories with low publication output for 2019 and 2020, and which have been excluded from the analyses, as they are not categorised by the World Bank. Six countries have extreme values for the NOARef for green-only OA, two of which are categorized by the World Bank. These countries are the United Kingdom (249,771 total publications) and the Comoros (8 total publications). For hybrid OA, most of the countries with extreme values in OA publishing are in Europe & Central Asia: the Netherlands: 5.32, Sweden: 4.80, Austria: 4.32, Hungary: 4.04, Norway: 3.93, and Finland: 3.55.



Figure 4: Scatter plots of NOAPub and NOARef of countries, faceted by World Bank income group, sized based on total publication output, for gold OA and green-only OA, 2019 and 2020 (values >3 and = 0 are removed from figure)



## Discussion

This paper reinforces the necessity of normalizing OA adoption by discipline and country, as suggested and practiced in earlier works (Archambault et al., 2014; Maddi, 2020). Using such measures, we can identify countries for which OA has become characteristic of their publications, with OA papers informing their research as well as serving as an avenue to share their research with the world. There is, however, a danger to this normalization in that it creates both a methodological and conceptual expectation: that is, that disciplines with lower OA rates are expected to be and remain lower.

Aggregating countries by income categories also can create some distortion as there are large differences in GNI per capita within groupings. Using the GNI measure in OA studies has been criticized as GNI is not an accurate measure of internet access in a country; ICT development is arguably more related to OA access than GNI. Additionally, it has been noted that these measures both neglect the science communication practices and contexts in the countries and regions (Xia, 2012). Thus, even if we observe different behavior regarding types of OA used in lower middle- and low-income countries, other factors in these countries (e.g., the time of OA adoption, institutional mandates) could be a larger contributor to the observations made.

Despite these limitations, several key contributions are made by the present analysis. We show that, globally, scholars are slightly more likely to publish OA papers than reference OA papers. This may, of course, be explained by the trend towards increasing OA over the last decade. That is, there are more contemporary OA publications, which would suggest that the OA status of references would lag. However, we find that scholars draw from OA references more than would be expected, which reinforces the citation advance of OA papers. In short, papers are increasingly OA, and scholars reference OA papers to a higher degree than would be expected;

however, the proportion of OA referencing remains lower than contemporary OA publishing given the time it takes for citations to accumulate.

Marked differences are observed by disciplines and countries. For example, Law and Education are more likely to publish in OA than reference it. It is notable that these disciplines have slightly different scholarly communication practices and a strong orientation towards practice. This may suggest an intentionality about widening dissemination of their work to practitioner audiences that might not necessarily have access. Domains like medicine, health, and biological sciences draw more heavily from OA than they publish it. This could be problematic in that it suggests that the demand for openly available medical knowledge is higher than the supply. However, it could also be an artefact of the expansive OA availability of papers funded by the NIH, supported by the infrastructure of PubMed and enforced through policy mandates (Larivière and Sugimoto, 2018).

Publication behaviour, more so than referencing behaviour, is influenced by OA mandates and institutional policies. Just as institutions within countries can specialize in research areas, the science system within a country can be particularly supportive of OA practices. This includes initiatives to transform local journals into OA journals. Platforms like SciELO in Brazil may explain the higher rates of OA publishing, which is not matched by high levels of OA referencing. The publish-or-perish culture (Rawat & Meena, 2014) and the perception that OA journals have a faster turnaround time (Rowley et al., 2017) is also thought to drive the increase of publications in OA journals in academic systems that reward a high volume of publications (de Jager et al., 2017; Green, 2019; Vardi, 2012). This potentially accounts for the more extreme NOAPub values as opposed to NOARef values observed. The trend towards high OA publishing and referencing for green-only and gold OA in Hungary may reflect a concerted effort to increase participation in OA publishing. Several countries display close to expected values for green-only and gold OA which may reflect a cultural acceptance for this mode of communication. China represents the other end of the spectrum with lower-than-expected levels of both publication and OA referencing for each type of OA. Such pattern is likely due to an emphasis on the journal impact factor and other journal-level indicators, which drives scholars towards closed publication environments, as well as to the high degree of country-self citation (Larivière, Gong and Sugimoto, 2018). On the whole, this paper shows that OA publications exceed OA referencing and that the success of OA dissemination of a country is not limited to their OA mandates and policies, but are affected by the entire research evaluation system.

## References

- Archambault, É., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, L., & Roberge, G. (2014). *Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013*. [http://www.science-metrix.com/sites/default/files/science-metrix/publications/d\\_1.8\\_sm\\_ec\\_dg-rtd\\_proportion\\_oa\\_1996-2013\\_v11p.pdf](http://www.science-metrix.com/sites/default/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf)
- Basson, I., Simard, M.-A., Ouangré, Z. A., Sugimoto, C. R., & Larivière, V. (2022). The effect of data sources on the measurement of open access : A comparison of Dimensions and the Web of Science. *PLoS ONE*, *17*(3), e0265545. <https://doi.org/10.1371/journal.pone.0265545>
- de Jager, P., van der Spuy, P., & de Kock, F. (2017). Do not Feed the Predators. *South African Journal of Business Management*, *48*(3), 35–45. <https://doi.org/10.2139/ssrn.2858750>

- Evans, J. A., & Reimer, J. (2009). Open access and global participation in science. *Science*, 323(5917), 1025. <https://doi.org/10.1126/science.1154562>
- Green, T. (2019). Is open access affordable? Why current models do not work and why we need internet-era transformation of scholarly communications. *Learned Publishing*, 32(1), 13–25. <https://doi.org/10.1002/leap.1219>
- Guerrero-Bote, V. P., Chinchilla-Rodríguez, Z., Mendoza, A., & de Moya-Anegón, F. (2021). Comparative Analysis of the Bibliographic Data Sources Dimensions and Scopus: An Approach at the Country and Institutional Levels. *Frontiers in Research Metrics and Analytics*, 5. <https://doi.org/10.3389/frma.2020.593494>
- Larivière, V., Gong, K., Sugimoto, C.R. (2018). Citations strength begins at home. *Nature*, 564 (7735) S70-S71.
- Larivière, V., Sugimoto, C.R. (2018). Do authors comply with mandates for open access? *Nature*, 562, 483-486.
- Maddi, A. (2020). Measuring open access publications: a novel normalized open access indicator. *Scientometrics*, 124(1), 379–398. <https://doi.org/10.1007/s11192-020-03470-0>
- Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes. R package version 1.1-2*. CRAN. <https://cran.r-project.org/package=RColorBrewer>
- Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA : a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, 1–23. <https://doi.org/10.7717/peerj.4375>
- Rawat, S., & Meena, S. (2014). Publish or perish: Where are we heading? *Journal of Research in Medical Sciences*, 19(2), 87–89.
- Robinson-Garcia, N., Costas, R., & van Leeuwen, T. N. (2020). Open Access uptake by universities worldwide. *PeerJ*, 2020(7), 1–20. <https://doi.org/10.7717/peerj.9410>
- Rowley, J., Johnson, F., & Saffi, L. (2017). Academics' Behaviors and Attitudes Towards Open Access Publishing in Scholarly Journals. *Journal of the Association for Information Science and Technology*, 68(5), 1201–1211. <https://doi.org/10.1002/asi.23710>
- Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman and Hall/CRC. <https://plotly-r.com>
- Simard, M.-A., Ghiasi, G., Mongeon, P., & Larivière, V. (2021). Geographic Differences in the Uptake of Open Access. *Proceedings of the 18th International Conference on Scientometrics and Informetrics (ISSI 2021)*, 1033–1038. <https://issi2021.org/proceedings/>
- Suber, P. (2003). *The taxpayer argument for open access*. SPARC Open Access Newsletter. [https://dash.harvard.edu/bitstream/handle/1/4725013/suber\\_taxpayer.htm](https://dash.harvard.edu/bitstream/handle/1/4725013/suber_taxpayer.htm)

- Van Leeuwen, T., Costas, R., & Robinson-García, N. (2019). *Indicators of open access publishing in the CWTS leiden ranking 2019*. CWTS Blog.
- Vardi, M. Y. (2012). Predatory scholarly publishing. *Communications of the Association for Computing Machinery*, 55(7), 5–5. <https://doi.org/10.1145/2209249.2209250>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- World Bank. (2022). *World Bank Country and Lending Groups*. <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>
- Xia, J. (2012). Diffusionism and open access. *Journal of Documentation*, 68(1), 72–99. <https://doi.org/10.1108/00220411211200338>
- Xie, Y. (2014). knitr: A Comprehensive Tool for Reproducible Research in R. In V. Stodden, F. Leisch, & R. D. Peng (Eds.), *Implementing Reproducible Computational Research*. Chapman and Hall/CRC.
- Xie, Y. (2015). *Dynamic Documents with R and knitr* (2nd Editio). Chapman and Hall/CRC.