“O Author, Where Art Thou?” An Analysis of Affiliation Indexing in Canadian Journals and Bibliometric Research Potential

Simon van Bellen

simon.van.bellen@erudit.org
Consortium Érudit, Université de Montréal, Pavillon 3744 Jean-Brillant, bureau 6500, Montréal, Québec, H3C 3J7 (Canada)

Abstract

Bibliometrics in the social sciences, humanities and arts (SSHA) are hampered by the limited presence of scholarly journals in analysis tools traditionally used. We analyzed the level of indexing of Canadian journal author affiliations in Dimensions.ai and OpenAlex to assess effects on bibliometric research. Annually, around 3,500 articles signed by Canadian researchers and published in Canadian journals remained irretrievable. Incomplete indexing particularly affects journals associated with not-for-profit publishers and those publishing in French. A fair representation of national SSHA research could enhance our understanding of publishing trends and contribute to the sustainability of the Canadian journals.

1. Introduction

Canadian scholarly journals in the social sciences, humanities and arts (SSHA) play a particular role in the dissemination of knowledge, because of the specificity of the research objects and the relative importance of French as opposed to English (Larivière, 2018). Canada is home to approximately 800 scientific journals, including 600 in SSHA, most of which are independent of large commercial structures and almost half offer their content in open access (Larivière et al., 2021). However, the indexing of Canadian journals, especially those in SSHA, is fragmentary in traditional bibliometric tools such as Web of Science (WoS) and Scopus (Tennant, 2020). For example, of the 207 active journals on the Érudit platform, the major disseminator of Canadian SSHA journals, only 56 were indexed in WoS as of January 2023. If these tools are used for bibliometrics in SSHA, they will create a distorted image of the SSHA publishing landscape.

The continued development of digital technology, persistent identifiers and increasingly powerful algorithms have great potential to improve the discoverability of publications. Dimensions.ai (Digital Science, 2018) and OpenAlex (Priem et al., 2022) have been created as alternatives to WoS or Scopus. Whereas Dimensions.ai has a for-profit structure, it is freely available for research purposes. OpenAlex was launched in 2022 as a fully open, not-for-profit initiative, mainly based on data from Microsoft Academic and Crossref. Using both tools, we assessed the level of indexing of author affiliations identified by persistent identifiers. Affiliations, or author addresses, are frequently used in bibliometrics, for example in studying internationalization patterns (Zhang et al., 2021) or in evaluating research patterns at the level of institutions (Zancanaro et al., 2015). Indexing author addresses using persistent identifiers in bibliometric tools may be critical since, unlike many European universities, few Canadian

institutions maintain publication directories of their researchers, also referred to as Current Research Information Systems (Carr-Wiggin et al., 2019). Incomplete indexing of affiliations highly complicates the acquisition of a complete set of articles and inevitably leads to a biased picture of where researchers are published.

The objective of this research was to quantify the indexing patterns of author affiliations with persistent identifiers for articles published by Canadian journals, covering the 2012–2021 period. Differences in coverage between publishers, and variations according to the language of the article, the year of publication and open access status were explored. We also aimed at providing an estimation of the number of articles that could be irretrievable and potential effects this may have on research evaluation using bibliometrics.

2. Methodology

To assess the level of indexing of Canadian journals, we obtained data from Dimensions.ai, accessed through BigQuery, and OpenAlex, accessed through the platform developed by the Curtin Open Knowledge Initiative (COKI). Queries were aimed at retrieving all articles published by Canadian scholarly journals for the period 2012–2021. The journals targeted were based on Larivière et al. (2021); ISSN and EISSN of each journal were added to the query. All article data were extracted in January 2023. Article languages were not available in Dimensions.ai or OpenAlex but were added by matching DOIs with the COKI Language Dataset (Diprose & Neylon, 2022).

Author affiliations can take multiple forms, which needed to be standardized by assigning unique, persistent identifiers. Specifically created for the identification of research organizations, we targeted the presence of the persistent identifiers ROR (Research Organization Registry; https://ror.org) and GRID (Global Research Identifier Database; https://www.grid.ac) in Dimensions.ai and OpenAlex. Although these identifiers were developed independently, they are interchangeable according to a conversion table. At the article level, affiliations were considered as “present” if at least one ROR or GRID identifier could be retrieved.

3. Results

We first analyzed the importance of DOIs for indexing in Dimensions.ai and OpenAlex. Having access to the complete record of DOIs of articles disseminated by Érudit, we were able to verify the level of indexing by both tools. Almost all articles on Érudit have a DOI and among these, 97% are indexed in Dimensions.ai and 95% in OpenAlex. It appears the presence of a DOI nearly guarantees indexing in Dimensions.ai and OpenAlex.

Querying all Canadian journals, Dimensions.ai returns 204,199 articles, a higher number than OpenAlex, which shows 192,204 articles (Figure 1). However, the indexing of affiliations in OpenAlex exceeds that of Dimensions.ai, leading to a higher number of articles retrievable by author affiliation for OpenAlex, at 121,885 (63% of the total), than for Dimensions.ai, at 107,674 (53%).

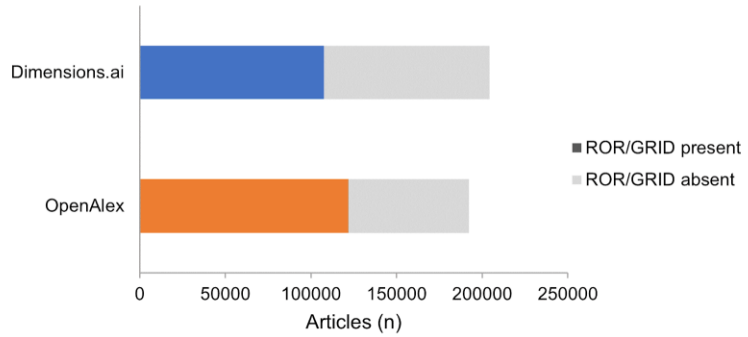

Figure 1: Numbers of Canadian journal articles (2012–2021) that have at least one, or no ROR/GRID identifier, according to Dimensions.ai and OpenAlex.

The presence of affiliations with persistent identifiers in Dimensions.ai and OpenAlex varies widely between publishers. Érudit has very low coverage in Dimensions.ai, at 17%, while it reaches 59% in OpenAlex (Figure 2). In OpenAlex, U. of Alberta Libraries journals show the lowest proportion of ROR/GRID, at 26%. At the other end of the spectrum, JMIR publications, publishing the Journal of Medical Internet Research, has affiliations with persistent identifiers for 98% of the articles indexed in Dimensions.ai and 91% in OpenAlex. Canadian Science Publishing and Springer Nature are also indexed at over 90%. These discrepancies in indexing affect recall values and thus have major implications for bibliometric research: when searching for articles with a specific author affiliation using Dimensions.ai, an article from JMIR publications is five to six times more likely to be traced than an article on Érudit.

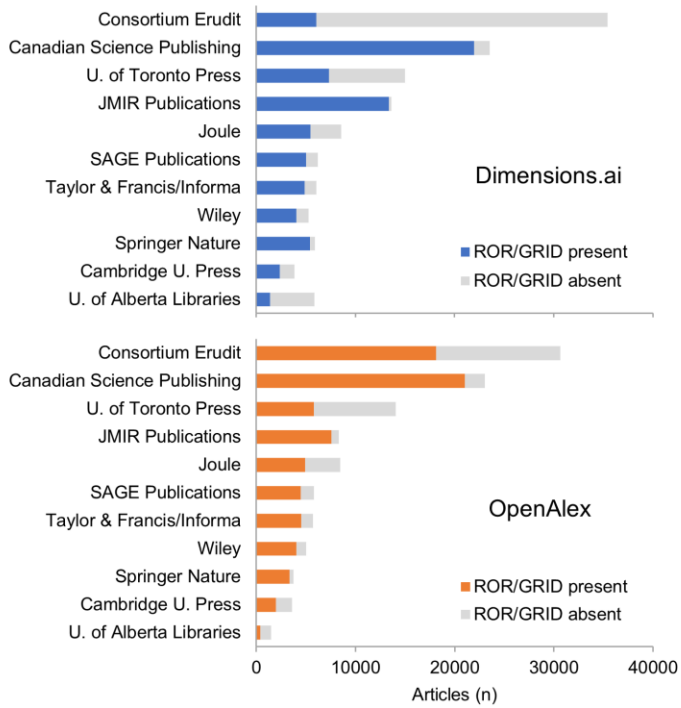

Figure 2: Numbers of Canadian journal articles (2012–2021) that have at least one, or no ROR/GRID identifier, according to Dimensions.ai and OpenAlex, for the main Canadian publishers.

The language of publication is also associated with the indexing of affiliations with persistent identifiers. In Dimensions.ai, 62% of articles in English have an ROR/GRID identifier compared to 26% of articles in French (Figure 3). OpenAlex also shows a difference in favour of English, but it is much smaller, at 67% for English versus 55% for French. OpenAlex has a much higher level of affiliation indexing for French content than Dimensions.ai.

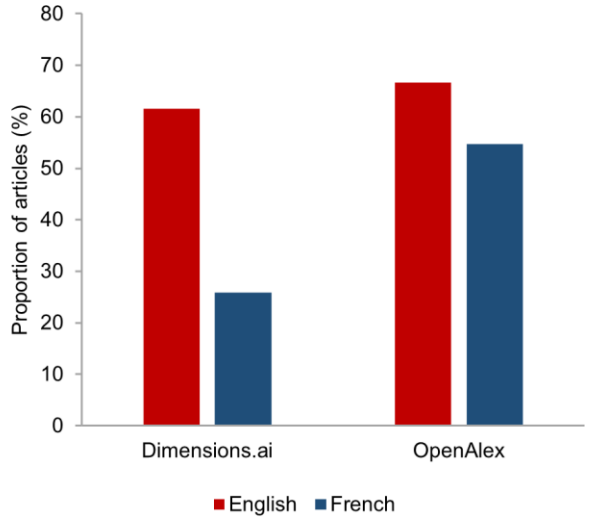

Figure 3: Proportion of Canadian journal articles (2012–2021) that have at least one ROR/GRID identifier, according to Dimensions.ai and OpenAlex, by language of publication.

The year of publication partly explains the level of indexing, but only for OpenAlex, where ROR/GRID identifiers are rarer for older articles (Figure 4).

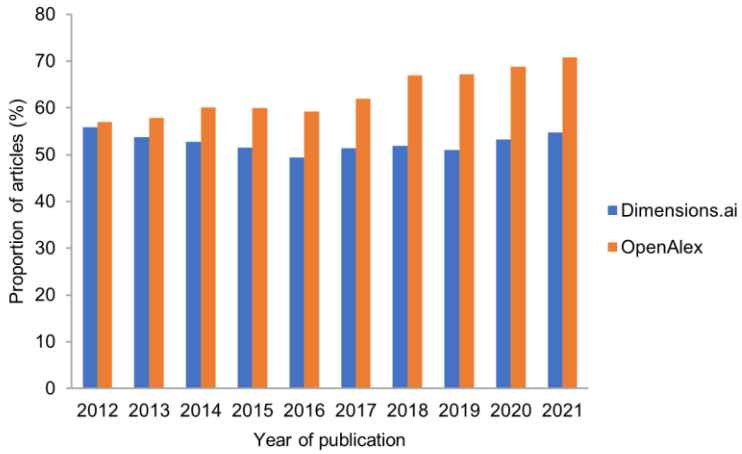

Figure 4: Proportion of Canadian journal articles (2012–2021) that have at least one ROR/GRID identifier, according to Dimensions.ai and OpenAlex, by year of publication.

Finally, Dimensions.ai indexes author's addresses for open access content less often than for articles requiring a subscription. Whereas 62% of "closed" articles have at least one indexed affiliation, we found affiliations for only 47% of open access articles. In OpenAlex, 63% of

articles requiring a subscription have an associated ROR/GRID, compared to 64% of open access articles.

How reliable are these tools for tracking and evaluating the articles published in national journals by Canadian researchers? Since OpenAlex generates the highest general recall of the two tools, we used it for subsequent analyses.

Of the 121,885 articles from Canadian journals with a ROR/GRID identifier in OpenAlex, 59,738 have at least one author affiliated to a Canadian institution, which represents 49%. Assuming that the proportion of Canadian authors is also 49% for articles lacking a ROR/GRID, the absence of affiliations for 70,319 articles implies that around 35,000 articles from Canadian journals signed by Canadian authors cannot be retrieved in OpenAlex for the period 2012–2021. Annually, this would represent at least 3,500 publications.

To put these numbers in perspective, we assessed the presence of affiliations for articles in all journals—both Canadian and foreign—from two of the world's largest publishers, Elsevier and Taylor & Francis. Elsevier is the largest publisher in the world yet publishes few Canadian journals. Taylor & Francis, a division of Informa and indexed as such in OpenAlex, specializes in SSHA. For both publishers, at least 80% of articles have a ROR/GRID, in both tools, with Dimensions.ai showing slightly higher proportions than OpenAlex. The analyses also show that the Taylor & Francis/Informa journals based abroad are better indexed than those from Canada: 79% (OpenAlex) and 81% (Dimensions.ai) of the Canadian journals have at least one author affiliation indexed, whereas the proportions are 81% and 85%, respectively, when foreign journals are added.

4. Discussion

The shortcoming in indexing is not limited to the absence of ROR/GRID identifiers in the tools. Some articles cannot be linked to an affiliation because many institutions simply do not have ROR/GRID identifiers. In addition, independent authors typically cannot be linked to a country. For example, in journals published on Érudit, 6% of authors cannot be associated with a ROR/GRID for these reasons, making them the fourth largest "affiliation" (of non-affiliates) among those publishing on the platform.

This study did not target indexing quality. Moreover, as we only considered the presence of at least one identifier per article, we cannot specify the degree of coverage at the article level. Further analysis would be required to validate these data. Since the causes of these variations in indexing levels are not transparent nor easy to evaluate, it is important that bibliometric research interprets query results with caution.

5. Conclusion

The indexing of author affiliations with persistent identifiers in Canadian journals shows major shortcomings, with half to a third of the articles published from 2012 to 2021 not showing any unique affiliation identifier. Nevertheless, indexing quality varies greatly depending on the publisher, the language of publication and the tool used. In general, French-language articles and those of not-for-profit publishers show the greatest deficiencies in the indexing of ROR/GRID identifiers and they are particularly disadvantaged when Dimensions.ai is used, in comparison

with OpenAlex. This incomplete indexing leads to an important underestimation of the number of Canadian articles published in Canadian scholarly journals. Annually, Canadian journals publish approximately 3,500 articles signed by Canadian researchers that remain irretrievable, because deficient indexing makes it challenging to identify the address of the author, according to the most effective tool used, OpenAlex.

An underestimation of the number of articles published in Canadian journals leads to an underestimation of the importance of these journals, for Canadian researchers as well as for those from elsewhere. The invisibility of their publications could compromise government and institutional support for these journals, which may negatively affect their sustainability in the long term.

6. Acknowledgments

Many thanks to Cameron Neylon for access to and assistance with the COKI tool.

7. References

Carr-Wiggin, M., Rothfus, M., Barrett, A., & Bourne-Tyson, D. (2019). Implementing a Current Research Information System (CRIS) in Canada. *Proceedings of the IATUL Conferences, Paper 3*. https://docs.lib.purdue.edu/iatul/2019/value/3

Digital Science. (2018). *Dimensions*. https://app.dimensions.ai

Diprose, J. P., & Neylon, C. (2022). *COKI Language Dataset* [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.6636625

Larivière, V. (2018). Le français, langue seconde ? De l'évolution des lieux et langues de publication des chercheurs au Québec, en France et en Allemagne. *Recherches sociographiques*, *59*(3), 339-363. Érudit. https://doi.org/10.7202/1058718ar

Larivière, V., Delmas, È., Beth, S., Paquin, É., & van Bellen, S. (2021). Canadian scholarly journals in the humanities and social sciences. A quantitative and qualitative portrait. https://www.erudit.org/public/documents/Canadian_scholarlyjournals_HSS_2021.pdf

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex : A fully-open index of scholarly works, authors, venues, institutions, and concepts*. https://arxiv.org/abs/2205.01833

Tennant, J. P. (2020). Web of Science and Scopus are not global databases of knowledge. *European Science Editing*, *46*, e51987. https://doi.org/10.3897/ese.2020.e51987

Zancanaro, A., Todesco, J. L., & Ramos, F. (2015). A Bibliometric Mapping of Open Educational Resources. *International Review of Research in Open and Distributed Learning*, *16*(1), 1-23. Érudit. https://doi.org/10.19173/irrodl.v16i1.1960

Zhang, L., Shang, Y., Huang, Y., & Sivertsen, G. (2021). Toward internationalization : A bibliometric analysis of the social sciences in Mainland China from 1979 to 2018. *Quantitative Science Studies*, *2*(1), 376-408. https://doi.org/10.1162/qss_a_00102