



OPEN ACCESS

EDITED AND REVIEWED BY
Zaida Chinchilla-Rodríguez,
Spanish National Research Council
(CSIC), Spain

*CORRESPONDENCE

Mohamed Ben Aouicha
✉ mohamed.benaouicha@fss.usf.tn
Houcemeddine Turki
✉ turkiabdelwaheb@hotmail.fr
Mohamed Ali Hadj Taieb
✉ mohamedali.hajtaieb@fss.usf.tn

RECEIVED 10 August 2023

ACCEPTED 04 September 2023

PUBLISHED 15 September 2023

CITATION

Ben Aouicha M, Turki H and Hadj Taieb MA
(2023) Editorial: Linked open bibliographic data
for real-time research assessment.
Front. Res. Metr. Anal. 8:1275731.
doi: 10.3389/frma.2023.1275731

COPYRIGHT

© 2023 Ben Aouicha, Turki and Hadj Taieb. This
is an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Editorial: Linked open bibliographic data for real-time research assessment

Mohamed Ben Aouicha*, Houcemeddine Turki* and
Mohamed Ali Hadj Taieb*

Data Engineering and Semantics Research Unit, Faculty of Sciences of Sfax, University of Sfax, Sfax,
Tunisia

KEYWORDS

**bibliographic databases, knowledge graphs, real-time scientometrics, research
evaluation, open data, FAIR data, living scientometrics**

Editorial on the Research Topic

Linked open bibliographic data for real-time research assessment

So far, research evaluation has been very important as a means for deciding academic tenure, awarding research grants, tracking the evolution of scholarly institutions, and assessing doctoral students (King, 1987). However, this effort has been limited by the lack of findability and accessibility of bibliographic databases allowing such an assessment and the legal and financial burdens toward their reuse (Herther, 2009). By the Internet age, scholarly publications became issued online in an electronic format allowing the extraction of accurate bibliographic information from them (Borgman, 2008) as well as the tracking of their readership, download, sharing, and search patterns (Markscheffel, 2013). Online resources called bibliographic knowledge graphs have consequently appeared, providing free bibliographic data and usage statistics for scholarly publications (Markscheffel, 2013). These resources are structured in triples, making them manageable through APIs and query endpoints (Ji et al., 2021). They are kept up-to-date in near real-time through automated methods for enrichment and validation.

Currently, many of these resources are released under permissive licenses such as CC0, CC-BY 4.0, MIT, and GNU covering various aspects of research evaluation (Markscheffel, 2013) including citation data (Peroni and Shotton, 2020), patent information (Verluisse et al., 2020), research metadata (Stocker et al., 2022), bibliographic metadata (Hendricks et al., 2020), author information (Haak et al., 2012), and data about scholarly journals, and conferences (Ley, 2009). Multilingual and multidisciplinary open knowledge graphs provide large-scale information about a variety of topics, including bibliographic metadata thanks to user contributions and crowdsourcing within the framework of Linked Open Data (Turki et al., 2022). Due to their flexible data model, they can integrate and centralize knowledge from multiple open and linked bibliographic resources based on persistent identifiers (PID) to become a secondary resource for research data (Nielsen et al., 2017). They also include a large set of non-bibliographic information such as country and prize information that can be used to augment bibliographic data and study the effect of social factors on research efforts (Turki et al., 2022). Later, gathered information can be used to generate research evaluation dashboards that can be updated in real-time based on SPARQL queries (Nielsen et al., 2017) or API queries (Lezhnina et al.). This will allow the launching of a new generation of knowledge-driven living research evaluation (Markscheffel, 2013). Beyond online resources having permissive licenses, several bibliographic databases are available online but have an

All Rights Reserved license like Google Scholar, maintained by Google (Orduña-Malea et al., 2015), and PubMed, provided by the National Center for Biotechnology Information (Fiorini et al., 2017). These resources can be very useful to feed private research dashboards and real-time research evaluation reports for scholarly institutions.

Despite the value of open bibliographic resources, they can involve inconsistencies that should be solved for better accuracy. As an example, *OpenCitations* mistakenly includes 1,370 self-citations and 1,498 symmetric citations as of April 30, 2022.¹ As well, they can involve several biases that can provide a distorted mirror of the research efforts across the world (Martín-Martín et al., 2021). That is why these databases need to be enhanced from the perspective of data modeling, data collection, and data reuse. This goes in line with the current perspective of the European Union on reforming research assessment (CoARA, 2022). In this topical collection, we are honored to feature novel research works in the context of allowing the automatic generation of real-time research assessment reports based on open bibliographic resources. We are happy to host research efforts emphasizing the importance of open research data as a basis for transparent and responsible research assessment, assessing the data quality of open resources to be used in real-time research evaluation, and providing implementations of how online databases can be combined to feed dashboards for real-time scholarly assessment.

The four accepted papers in this Research Topic provide insight into the use of open bibliographic data to evaluate academic performance. Majeti et al. present an interface that harvests bibliographic and research funding data from online sources. The authors of this paper address systematic biases in collected data through nominal and normalized metrics and present the results of an evaluation survey taken by senior faculty. Porter and Hook explore the deployment of scientometric data into the hands of practitioners through cloud-based data infrastructures. The authors present an approach that connects Dimensions and World Bank data on Google BigQuery to study international collaboration between countries of different economic classifications. Schnieders et al. evaluate the readiness of research institutions for partially automated research reporting using open, public

research information collected via persistent identifiers (PIDs) for organizations (ROR), persons (ORCID), and research outputs (DOI). The authors use internally maintained lists of persons to investigate ORCID coverage in external open data sources and present recommendations for future actions. Lezhnina et al. propose a dashboard using scholarly knowledge graphs to visualize research contributions, combining computer science, graphic design, and human-technology interaction. The user survey showed the dashboard's appeal and potential to enhance scholarly communication through knowledge graph-powered dashboards in different domains.

The research papers featured here underscore the critical importance of open bibliographic data in transforming the landscape of research evaluation. These papers not only shed light on this pivotal role but also offer invaluable practical tools for both researchers and practitioners. By harnessing linked open data, these resources empower individuals within the academic community to navigate the intricacies of scholarly communication more effectively, ultimately leading to improved research assessment practices among scholars and institutions.

Author contributions

MB: Writing—original draft, Writing—review and editing. HT: Writing—original draft, Writing—review and editing. MH: Writing—original draft, Writing—review and editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

¹ A detailed list of deficient self-citations and symmetric citations at OpenCitations can be found at <https://github.com/csisc/OCDeficiency>.

References

- Borgman, C. L. (2008). Data, disciplines, and scholarly publishing. *Learn. Pub.* 21, 29–38. doi: 10.1087/095315108X254476
- CoARA (2022). *Agreement on Reforming Research Assessment*. Brussels, Belgium: Science Europe and European University Association.
- Fiorini, N., Lipman, D. J., and Lu, Z. (2017). Cutting edge: toward PubMed 2.0. *eLife* 6, e28801. doi: 10.7554/eLife.28801
- Haak, L. L., Fenner, M., Paglione, L., Pentz, E., and Ratner, H. (2012). ORCID: a system to uniquely identify researchers. *Learn. Pub.* 25, 259–264. doi: 10.1087/20120404
- Hendricks, G., Tkaczyk, D., Lin, J., and Feeney, P. (2020). Crossref: the sustainable source of community-owned scholarly metadata. *Quant. Sci. Stud.* 1, 414–427. doi: 10.1162/qss_a_00022
- Herther, N. K. (2009). Research evaluation and citation analysis: key issues and implications. *Elect. Lib.* 27, 361–375. doi: 10.1108/02640470910966835
- Ji, S., Pan, S., Cambria, E., Marttinen, P., and Philip, S. Y. (2021). A survey on knowledge graphs: representation, acquisition, and applications. *IEEE Transact. Neural Networks Learn. Sys.* 33, 494–514. doi: 10.1109/TNNLS.2021.3070843
- King, J. (1987). A review of bibliometric and other science indicators and their role in research evaluation. *J. Inform. Sci.* 13, 261–276. doi: 10.1177/016555158701300501

- Ley, M. (2009). DBLP: some lessons learned. *Proceed. VLDB Endow.* 2, 1493–1500. doi: 10.14778/1687553.1687577
- Markscheffel, B. (2013). New Metrics, a Chance for changing Scientometrics. A Preliminary Discussion of Recent Approaches. *Scientometrics: Status and Prospects for Development* (p. 37). Moscow, Russia: Institute for the Study of Science of RAS. Available online at: https://www.researchgate.net/publication/258926049_New_Metrics_a_Chance_for_Changing_Scientometrics_A_Preliminary_Discussion_of_Recent_Approaches (accessed August 10, 2023).
- Martin-Martín, A., Thelwall, M., Orduña-Malea, E., and Delgado López-Cózar, E. (2021). Google scholar, microsoft academic, scopus, dimensions, web of science, and opencitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics* 126, 871–906. doi: 10.1007/s11192-020-03690-4
- Nielsen, F. Å., Mietchen, D., and Willighagen, E. (2017). Scholia, scientometrics and wikidata. *European Semantic Web Conference*. Cham: Springer, 237–259. doi: 10.1007/978-3-319-70407-4_36
- Orduña-Malea, E., Ayllón, J. M., Martín-Martín, A., and Delgado López-Cózar, E. (2015). Methods for estimating the size of Google Scholar. *Scientometrics* 104, 931–949. doi: 10.1007/s11192-015-1614-6
- Peroni, S., and Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. *Quant. Sci. Stud.* 1, 428–444. doi: 10.1162/qss_a_00023
- Stocker, M., Heger, T., Schweidtmann, A., Cwiek-Kupczyńska, H., Penev, L., Dojchinovski, M., et al. (2022). SKG4EOSC-scholarly knowledge graphs for EOSC: establishing a backbone of knowledge graphs for FAIR scholarly information in EOSC. *Res. Ideas Out.* 8, e83789. doi: 10.3897/rio.8.e83789
- Turki, H., Hadj Taieb, M. A., Shafee, T., Lubiana, T., Jemielniak, D., Ben Aouicha, M., et al. (2022). Representing COVID-19 information in collaborative knowledge graphs: the case of Wikidata. *Sem. Web* 13, 233–264. doi: 10.3233/SW-210444
- Verluisé, C., Cristelli, G., Higham, K., and de Rassenfosse, G. (2020). *The Missing 15 Percent of Patent Citations*. Lausanne: EPFL.