Dorsch, Isabelle; Lemke, Steffen; Peters, Isabella

Conference Paper — Published Version

Analysis of "Open Access Publishing Characteristics" for Covid-19 and cancer publications in Web of Science

Suggested Citation: Dorsch, Isabelle; Lemke, Steffen; Peters, Isabella (2023) : Analysis of "Open Access Publishing Characteristics" for Covid-19 and cancer publications in Web of Science, In: Semar, Wolfgang (Ed.): Nachhaltige Information – Information für Nachhaltigkeit. Proceedings des 17. Internationalen Symposium für Informationswissenschaft (ISI 2023), ISBN 978-3-86488-200-5, Verlag Werner Hülsbusch, Glückstadt, pp. 387-392, https://zenodo.org/doi/10.5281/zenodo.10009337

This Version is available at:
http://hdl.handle.net/11108/610

Analysis of "Open Access Publishing Characteristics" for COVID-19 and Cancer Publications in Web of Science

Isabelle Dorsch[1], Steffen Lemke[2], Isabella Peters[1]

1 ZBW Leibniz Information Centre for Economics, Kiel, Germany

{i.dorsch, i.peters}@zbw-online.eu

2 CAU Kiel University, Germany

sle@informatik.uni-kiel.de

Abstract

The COVID-19 pandemic affected the discourse on the openness of science. Although many publishers allowed free access to research publications, this change of access might only be temporary since publications often lack licenses or are only limited to the topic of COVID-19. This poster compares the open access status and bibliometric characteristics of research articles and reviews concerned with "cancer" and "COVID-19" published within the three-year period from 2020 to 2022 and obtained from Web of Science. To determine the two sets of publications a search strategy from PubMed for cancer publications and a COVID-19 document dataset from Dimensions were selected. The final amount of publications indexed in Web of Science results in 87,821 cancer research-related and 461,205 COVID-19-related publications with an overlap of 4,670 publications existing in both sets. For the open access categories Gold and Hybrid, overall comparable shares for both, cancer and COVID-19 publications, are observed. For the COVID-19 publications higher shares of Green and Bronze open access and lower shares of publications not classified into one of the OA categories are further observed. The study contributes to the analysis of OA statuses of publications made available during the COVID-19 pandemic.

Keywords: cancer research, COVID-19 research, open access, open science, scientometrics, accessibility

1 Introduction

With the beginning of the COVID-19 pandemic, many publishers allowed free but temporary access to their publications. Preprints and open data became key aspects of the pandemic too – all aiming at accelerating research on the COVID-19 virus and vaccines (Blasetti et al., 2020). Thus, the COVID-19 pandemic had an enormous impact on the meaningfulness of openness in science. In general, the open science movement is constantly developing with the goal of reducing barriers in all areas of science (Tochtermann & Höfler, 2022) by "opening up the research process by making all of its outcomes, and the way in which these outcomes were achieved, publicly available on the World Wide Web" (Kraker et al., 2011, p. 645). One part of this is open access (OA), to make research publications openly accessible without any financial, technical, and legal barriers. There are, however, different OA variants and definitions (Herb & Pampel, 2022). In this study, we work with the following OA definitions used in Web of Science, which are based on data from OurResearch (Unpaywall)[1]: Gold (published in an all-OA journal), Hybrid (open license, with APC), Bronze (no or unclear usage license; free to read, can be withdrawn), and Green (submitted/accepted/published to/in a journal and archived in an OA repository). Thereby, not having a proper usage license, as with Bronze OA, is problematic since there is no guarantee of how long this publication is accessible and to what extent.

Already a few weeks after the outbreak of COVID-19, researchers started to capture the pandemics' impact on the publication system scientometrically (e.g., Aviv-Reuven & Rosenfeld, 2021; Belli et al., 2020). For medical and biomedical COVID-19 publications (mostly journal articles) from PubMed (n = 95,605), published between January 1, 2020 and March 1, 2021, a decreasing trend for Green OA publications during the pandemic emerged, whereas Bronze, followed by Gold publishing models "became more prominent, with a significant increase of the Bronze model from the second quarter of 2020 onward" (San Torcuato et al., 2022, p. 5). Trying to forecast the expected growth of COVID-19 literature, Nane et al. (2023) did not find a shift towards OA in general or specifically for Green OA during the pandemic.

While the pandemic puts openness into the spotlight, it is important to empirically investigate how this affects the publication landscape at large.

1 https://webofscience.help.clarivate.com/en-us/Content/open-access.html

Following the related literature, we compared the shares of different OA classes among publications topically related to COVID-19 or to cancer research (as an example of a large biomedical domain with ongoing relevance and without a direct connection to the topic of COVID-19) in terms of differences in the OA status and bibliometric characteristics of COVID-19 and cancer publications published between 2020 and 2022.

2 Methods

For our analysis of COVID-19-related publications, the COVID-19 Document Dataset of Global Research, openly published by Dimensions via Google BigQuery[2], was used. From 2020 to 2022, 1,613,720 unique DOIs were registered. We used this time frame since the COVID-19 outbreak started at the end of 2019 and to observe full years.

To be able to compare this set of COVID-19-related publications with literature that is most likely concerned with an exemplarily biomedical topic different to COVID-19, we created a second dataset, which shall be centered around publications from the domain of cancer research. To do so, we made use of Dimensions' concepts, which are descriptive noun phrases derived from documents' abstracts via machine learning techniques,[3] and an adaptation of the search strategy applied by PubMed to create its filter for cancer-related research.[4] Searching for the MeSH terms included in this strategy over the concepts of documents from Dimensions published between 2020 and 2022 retrieved 213,602 unique DOIs, supposedly referring to documents from the domain of cancer research.

With these two sets of DOIs we queried WoS for document types 'article' or 'review', again filtering for publication years 2020 to 2022, and retrieved their metadata and OA status (which WoS obtains from Unpaywall). This resulted in two datasets of 461,205 COVID-19-related and 87,821 cancer

2 https://console.cloud.google.com/marketplace/product/digitalscience-public/covid-19-dataset-dimensions

3 https://api-lab.dimensions.ai/cookbooks/1-getting-started/7-Working-with-concepts.html

4 https://www.nlm.nih.gov/bsd/pubmed_subsets/cancer_strategy.html

research-related publications indexed in WoS, with an overlap of 4,670 publications (so 5.32% of the cancer research dataset) that exist in both sets. Subsequently, we compared the OA statuses between the two datasets. The data analyzed in this study was reflective of April 2023.

3 Early Results of the Work in Progress

Figure 1 shows the relative shares among the OA statuses among articles and reviews from the two analyzed datasets. While overall comparable shares are observed for the types Gold and Hybrid, for COVID-19 publications we see increased shares of Green and Bronze OA as well as decreased shares of 'No known open access' publications.

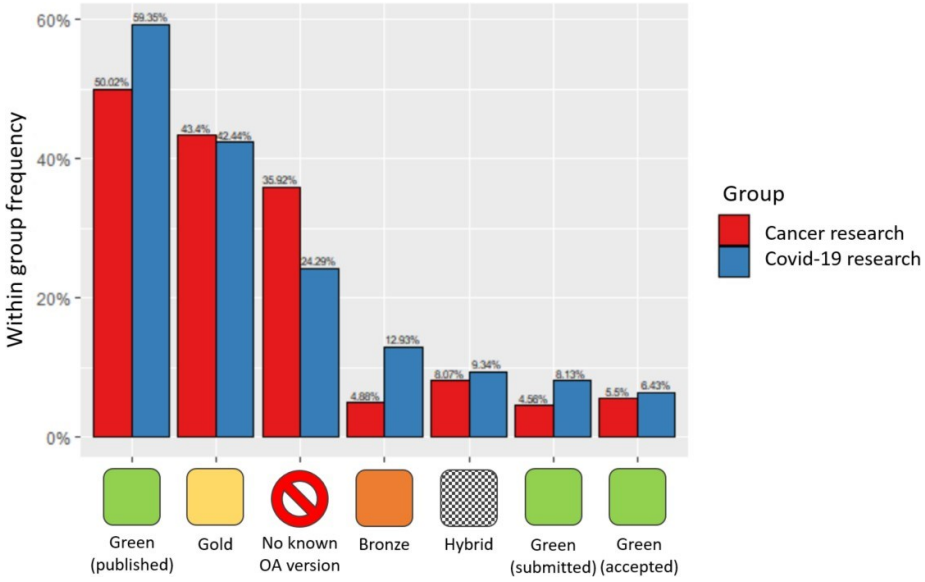

Fig. 1 OA statuses' relative shares among publications of the two article groups ($n = 549{,}026$). OA colors according to https://webofscience.help.clarivate.com/en-us/Content/open-access.html

4 Discussion and Outlook

By contrasting COVID-19 publication with those on cancer, this research sheds light on topic-specific, and often temporary, open access-practices of publishers. Contrary to what has been highlighted in previous literature, the Bronze category is not the most assigned category for both groups. As our data reflects a status of April 2023 we speculate that the OA status already re-shifted for parts of COVID-19 publications, resulting in a reduced frequency of Bronze OA. Nevertheless, the selection of database, document types, and disciplines also impacts the results. We can already see differences for the cancer and COVID-19 groups in terms of Bronze, Green and No known OA (which also includes closed access) from these early results. Based on this we further want to analyze the OA status, publication years, and their journal and publisher information to deepen our knowledge about the observed OA differences.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 495515545. We thank the German Kompetenznetzwerk Bibliometrie for providing our project with access to their infrastructure of bibliometric metadata based on WoS.

References

Aviv-Reuven, S., & Rosenfeld, A. (2021). Publication patterns' changes due to the COVID-19 pandemic: A longitudinal and short-term scientometric analysis. *Scientometrics, 126*(8), 6761–6784. https://doi.org/10.1007/s11192-021-04059-x

Belli, S., Mugnaini, R., Baltà, J., & Abadal, E. (2020). Coronavirus mapping in scientific publications: When science advances rapidly and collectively, is access to this knowledge open to society? *Scientometrics, 124*(3), 2661–2685. https://doi.org/10.1007/s11192-020-03590-7

Blasetti, A., Droß, P. J., Fräßdorf, M., & Naujoks, J. (2020). Offenheit im globalen Lockdown: Ein Zukunftsmodell für die Wissenschaft? [Research Report]. Wissenschaftszentrum Berlin für Sozialforschung, WZB. https://www.econstor.eu/handle/10419/223164

Herb, U., & Pampel, H. (2022). Open Access. In R. Kuhlen, D. Lewandowski, W. Semar, & C. Womser-Hacker (Eds.), *Grundlagen der Informationswissenschaft* (pp. 715–726). Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110769043-061

Kraker, P., Leony, D., Reinhardt, W., & Günter, B. (2011). The case for an open science in technology enhanced learning. *International Journal of Technology Enhanced Learning, 3*(6), 643–654. https://doi.org/10.1504/IJTEL.2011.045454

Nane, G. F., Robinson-Garcia, N., van Schalkwyk, F., & Torres-Salinas, D. (2023). COVID-19 and the scientific publishing system: Growth, open access and scientific fields. *Scientometrics, 128*, 345–362. https://doi.org/10.1007/s11192-022-04536-x

San Torcuato, M., Bautista-Puig, N., Arrizabalaga, O., & Méndez, E. (2022). Tracking Openness and Topic Evolution of COVID-19 Publications January 2020–March 2021: Comprehensive Bibliometric and Topic Modeling Analysis. *Journal of Medical Internet Research, 24*(10), e40011. https://doi.org/10.2196/40011

Tochtermann, K., & Höfler, A. M. (2022). Open Science. In R. Kuhlen, D. Lewandowski, W. Semar, & C. Womser-Hacker (Eds.), *Grundlagen der Informationswissenschaft* (pp. 703–714). Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110769043-060

In: W. Semar (Hrsg.): Nachhaltige Information – Information für Nachhaltigkeit. Tagungsband des 17. Internationalen Symposiums für Informationswissenschaft (ISI 2023), Chur, Schweiz, 7.–9. November 2023. Glückstadt: Verlag Werner Hülsbusch, S. 387–392. DOI: https://dx.doi.org/10.5281/zenodo.10009338