

Post-pandemic changes in the adoption of OA models – a case study on Covid-19 and cancer research

Lemke, Steffen
Dorsch, Isabelle
Peters, Isabella

CAU Kiel University, Germany | sle@informatik.uni-kiel.de
ZBW Leibniz Information Centre for Economics, Germany | i.dorsch@zbw-online.eu
ZBW Leibniz Information Centre for Economics, Germany | i.peters@zbw-online.eu

ABSTRACT

During the Covid-19 pandemic many publishers granted free access to research publications, most often in the form of temporary Bronze open access (OA). It remains unclear how stable such shifts in adoptions of certain OA models during the pandemic are, and whether they apply to Covid-19-related and -unrelated fields to varying extents. We compare the OA statuses and these statuses' consistency across two Web of Science snapshots from April 2022 and April 2023 for sets of Covid-19 and cancer-related research articles that were published in 2020 and 2021. First results indicate noteworthy differences between both groups' developments regarding frequent OA models.

KEYWORDS

scientometrics, open access, open science, covid-19, cancer research

INTRODUCTION

Having open access (OA) to research publications is an increasing claim in the landscape of academia. Such access became even more relevant during the Covid-19 pandemic. To accelerate research on the virus and its impact, several scholarly publishers provided temporary free access to their Covid-19-related publications (Blasetti et al., 2020). For example, 91.4% of coronavirus publications from the first twelve weeks of 2020 (n = 233) are published as OA. Within this subset, 53.6% are published as Bronze OA (Belli et al., 2020), i.e., as articles that are free to read on publisher websites, but lack an explicit open license (Piwowar et al., 2018). Similarly, for a sample of 5,612 coronavirus-related articles published on PubMed during the first quarter of 2020, Arrizabalaga et al. (2020) found 97.4% to be published as OA, with 68.3% of them being classified as Bronze. Likewise, San Torcuato et al. (2022) found that from a larger subset of biomedical Covid-19-related PubMed journal articles (n = 94,015) published between January 2020 and March 2021, 94.08% are published in an OA format, the majority of them again as Bronze OA (44.8%), followed by Gold OA (31.9%), Green OA (14.1%), and Hybrid OA (9.3%).

Although the overall high shares of OA articles reported by aforementioned studies can be considered optimistic signs for the global transformation towards openly accessible research, the observed high proportions of Bronze OA might tarnish this view. Bronze OA publications usually lack licenses granting any other rights than their temporary free readability, which can be removed unilaterally by their respective publishers. Furthermore, timeframes and criteria for which publishers grant or revoke this access are intransparent. Such absence of fully OA compliant public licensing restricts research's reusability as well as visibility, and thus ultimately slows scientific searches for solutions to pressing societal problems (San Torcuato et al., 2022). Also, the publisher-choice based OA model Bronze affects consistency and usefulness of OA data. While author-choice based OA models (i.e., closed access, Gold, Green, Hybrid OA) mostly remain stable over time, Bronze OA is volatile, leading to increased temporal fluctuations in analyses of publications' OA statuses (Sanford, 2022). Moreover, in light of known citation advantages associated with articles' open access (see for instance Fraser et al., 2020), Bronze OA can be considered an interference with the scientific reputation system outside the control of the authors themselves.

We aim to contribute to the research on the effects of the Covid-19 pandemic on the OA transformation, considering both Covid-19-related as well as -unrelated fields of research. We start by comparing the consistency of OA models across two years (2022 and 2023) among recently published articles of two groups: one group of articles topically related to Covid-19, and one group of articles related to cancer research (as an example of a large biomedical domain without direct connection to Covid-19). We analyze articles published in 2020 and 2021, while we compare their OA statuses and their consistency in Web of Science (WoS) snapshots from April 2022 and April 2023.

METHODS

We start with the Covid-19 Document Dataset of Global Research openly published by Dimensions via Google BigQuery.¹ Querying this dataset for all publications published in our analysis timeframe of 2020 to 2021 led to the retrieval of 911,392 unique DOIs. We also create a second dataset centered around publications from the domain of cancer research. To do so, we make use of Dimensions' concepts, which are descriptive noun phrases derived from documents' abstracts via machine learning techniques,² and an adaptation of the search strategy applied by PubMed

¹ <https://console.cloud.google.com/marketplace/product/digitalscience-public/covid-19-dataset-dimensions>

² <https://api-lab.dimensions.ai/cookbooks/1-getting-started/7-Working-with-concepts.html>

to create its filter for cancer-related research.³ Searching for the MeSH terms included in this strategy over the concepts of documents from Dimensions published between 2020 and 2021 led to the retrieval of 142,008 unique DOIs, supposedly referring to documents from the domain of cancer research.

With these two sets of DOIs we subsequently query WoS for publications of the types *article* or *review*, again filtering for publication years 2020 to 2021, and retrieve their metadata and OA status (which WoS obtains from Unpaywall). We perform both of these queries twice, once on a WoS snapshot from April 2022 and once on a snapshot from April 2023 (see Table 1). In comparisons between the years 2022 and 2023, we only consider publications present in both snapshots of WoS, leading to n=191,117 for Covid-19 and n=48,614 for cancer research articles. We compare the OA statuses between the retrieved datasets to explore possible differences related to the adoption and prevalence of OA models within biomedical research since the beginning of the Covid-19 pandemic.

Table 1: Numbers of retrieved articles within the four analysis datasets.

	Covid-19/WoS 2022	Covid-19/WoS 2023	Cancer/WoS 2022	Cancer/WoS 2023
# of unique DOIs	201,908	191,528	49,691	48,639

PRELIMINARY RESULTS

For both cancer and Covid-19 research, Green OA and Gold OA dominate in both WoS snapshots (Figure 1). Between the years, the shares of publications without known OA versions decrease in both groups (by 4.32% and 2.77% respectively). For the group of Covid-19 publications, we moreover observe a decrease for Bronze OA (by 6.46%) and increases in the categories Green OA (6.99%) and Gold OA (5.54%) between the years. For cancer research, Green OA denotes a 5.54% increase, while the other categories remain relatively stable. Particularly striking overall differences between the two article groups can be seen regarding the shares of Bronze OA, Green OA, and publications without known OA version.

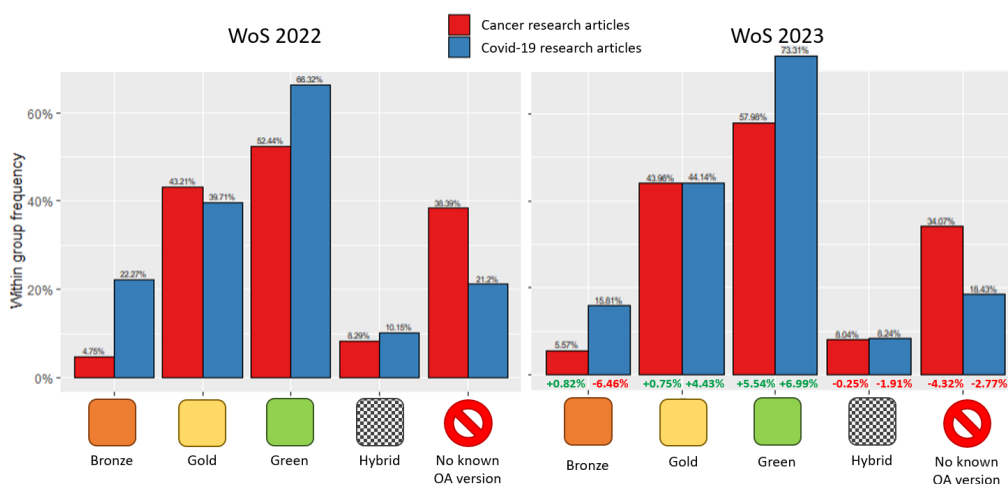


Figure 1: OA statuses' relative shares among publications of the two article groups within the two snapshots of WoS data. Green and red percentage values (on the right) indicate the increase or decrease of an OA category between snapshots. OA colors according to <https://webofscience.help.clarivate.com/en-us/Content/open-access.html>.

DISCUSSION

The noteworthy increase of Green OA seen in our results might be explained by the “backfilling” effect described by Archambault et al. (2014), i.e., researchers uploading OA versions of their already published articles to repositories. Of particular interest is the drop of Bronze OA publications among Covid-19-related articles, indicating that publishers' interference is particularly apparent regarding the hot topic of Covid-19. Our next step will be to examine to which publication models such articles typically transfer, as well as an analysis of specific publishers' prevalence within subgroups. A limitation of our study lies in its preliminary reliance on WoS as a data source, which only covers OA literature to a limited extent. The inclusion of additional data sources to mitigate this limitation is part of our future work.

ACKNOWLEDGMENTS

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), grant number 49551545. We thank the German Kompetenznetzwerk Bibliometrie for providing us with access to their data infrastructure.

³ https://www.nlm.nih.gov/bsd/pubmed_subsets/cancer_strategy.html

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. European Commission.
- Arrizabalaga, O., Otaegui, D., Vergara, I., Arrizabalaga, J., & Méndez, E. (2020). Open Access of COVID-19-related publications in the first quarter of 2020: A preliminary study based in PubMed. *FI000Research*, 9, 649. <https://doi.org/10.12688/fi000research.24136.2>
- 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?* (Corona und die gesellschaftlichen Folgen: Schlaglichter aus der WZB-Forschung) [Research Report]. Wissenschaftszentrum Berlin für Sozialforschung, WZB. <https://www.econstor.eu/handle/10419/223164>
- Fraser, N., Momeni, F., Mayr, P., & Peters, I. (2020). The relationship between bioRxiv preprints, citations and altmetrics. *Quantitative Science Studies*, 1(2), 618–638. https://doi.org/10.1162/qss_a_00043
- 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*, 6, e4375. <https://doi.org/10.7717/peerj.4375>
- 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>
- Sanford, H. (2022). *The state of unpaywall: Analyzing the consistency of open access data*. 26th International Conference on Science, Technology and Innovation Indicators (STI 2022), Granada, Spain, 7-9 September 2022.