ZBW *Publikationsarchiv*

Publikationen von Beschäftigten der ZBW – Leibniz-Informationszentrum Wirtschaft *Publications by ZBW – Leibniz Information Centre for Economics staff members*

Fraser, Nicholas; Momeni, Fakhri; Mayr, Philipp; Peters, Isabella

Conference Paper — Published Version Altmetrics and Open Access: Exploring Drivers and Effects

Suggested Citation: Fraser, Nicholas; Momeni, Fakhri; Mayr, Philipp; Peters, Isabella (2019) : Altmetrics and Open Access: Exploring Drivers and Effects, In: Proceedings of the Workshop altmetrics 19 at the 6th Altmetrics Conference (6:AM), The University of North Carolina at Chapel Hill, Chapel Hill, http://altmetrics.org/wp-content/uploads/2019/10/Fraser_altmetrics19_paper_3.pdf

This Version is available at: http://hdl.handle.net/11108/428

Kontakt/Contact ZBW – Leibniz-Informationszentrum Wirtschaft/Leibniz Information Centre for Economics Düsternbrooker Weg 120 24105 Kiel (Germany) E-Mail: info@zbw.eu https://www.zbw.eu/de/ueber-uns/profil-der-zbw/veroeffentlichungen-zbw

Standard-Nutzungsbedingungen:

Dieses Dokument darf zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen dieses Dokument nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Sofern für das Dokument eine Open-Content-Lizenz verwendet wurde, so gelten abweichend von diesen Nutzungsbedingungen die in der Lizenz gewährten Nutzungsrechte.

Terms of use:

This document may be saved and copied for your personal and scholarly purposes. You are not to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. If the document is made available under a Creative Commons Licence you may exercise further usage rights as specified in the licence.





Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Altmetrics and Open Access: Exploring Drivers and Effects

Nicholas Fraser¹, Fakhri Momeni², Philipp Mayr², Isabella Peters^{1,3}

¹ZBW – Leibniz Information Centre for Economics, Kiel, Germany
²GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany
³Kiel University, Kiel, Germany

Correspondence: n.fraser@zbw.eu

1. Introduction

The movement to provide open access (OA) to research is posited to have a number of net benefits for science and society. In the scientific sense, OA is linked to faster and wider dissemination and usage of research results, which is manifested in increased citations of articles that are openly available versus those that are not (c.f. Swan, 2010; Archambault et al., 2016; Piwowar et al., 2018). In the societal sense, it has been suggested that OA can increase engagement with scientific research leading to positive impacts for industry, the economy and the public (ElSabry, 2017). One way to understand the impact of science in non-research contexts is by studying indicators of engagement with scientific research on online platforms, collectively termed 'altmetrics'.

To date there have been relatively few large-scale analyses aiming to quantify the relationship between OA and altmetrics, taking into account important factors associated with OA publishing practices, such as discipline, authorship properties (e.g. country of authorship, collaboration networks), publication properties (e.g. Journal Impact Factor; JIF) and document properties (e.g. paper length). The importance of these factors is underscored in a survey of more than 1800 UK-based researchers by Zhu et al. (2017), which found significant differences in OA publishing practices between subject areas, genders, age and seniority of academics; those that reported the greatest experience with Gold OA publishing were, respectively, in the Medical and Life Sciences, Male, 45-54 and Professors.

In measuring these effects, the choice of data source plays an important role; no single bibliometric data source is ubiquitous and all have caveats with respect to their coverage, e.g. in terms of subject areas or publication types. This contribution is a case study of the relationship between altmetrics and OA, using multiple data sources. We provide a description of a large dataset, and initial exploratory results of variability in counts of multiple altmetric indicators considering two major influencing factors: subject classification and country of authorship.

2. Methods

Article Sample

We extracted a random sample of 1,000,000 articles from the Unpaywall service (https://unpaywall.org), limited to years 2013-2017, and 'journal-article' types. Unpaywall is a service developed and maintained by Our Research (Formerly ImpactStory; https://our-

research.org) which provides links to OA versions of an article derived from crawling and aggregation of numerous data sources including PMC, BASE, DOAJ and institutional repositories and publishers (Piwowar et al., 2018). The Unpaywall service contains records for all Crossref DOIs, which can be queried via a free public API. Unpaywall also provide regular database 'snapshots' or 'dumps' which are more suitable for large-scale analysis. In this study we used the Unpaywall database snapshot from April 2018, which was subsequently parsed to an SQL table for fast querying.

Articles were classified into five access types, following the schema in Figure 1, in line with the classification schema used by Piwowar et al. (2018). Closed access articles were those found to not have any openly (legally) available version; Gold as those found in a journal indexed by DOAJ; Green as those found in an OA repository; Hybrid as those published under an OA license in a journal that is not listed within the DOAJ (typically subscription journals that offer a paid OA option), and Bronze as those found to be OA on the publishers' page but with no clearly identifiable license information.



Figure 1: Workflow for classification of article access types using Unpaywall, following Piwowar et al. (2018).

Bibliometric databases

For extracting article and author features, we matched our Unpaywall sample to bibliometric databases, in this case the Web of Science (WoS) and Scopus (leveraging the data infrastructure of **Bibliometrics**: the German Competence Centre for http://www.forschungsinfo.de/Bibliometrie/en/index.php). Articles in both bibliometric databases were limited to publication years 2013-2017, to 'journal' publication types and 'article' or 'review' document types. Matching was performed via direct, case-insensitive matching of DOIs. A total of 461,238 articles from our initial Unpaywall sample were matched to a WoS record (46 %), whilst 537,227 articles were matched to a Scopus record (54 %). It should be noted that neither of these values is representative of the true coverage of Unpaywall in WoS or Scopus, due to the limitations on publication and document types, and the relatively simple matching process which relies solely on direct DOI correspondence.

WoS and Scopus both provide their own subject classification systems, although documentation of the construction of the classification systems is relatively lacking (Wang and Waltman, 2016). The WoS classification system consists of approximately 250 categories

covering the sciences, social sciences, arts and humanities. The Scopus classification system (known as the 'All Science Journal Classification' (ASJC) system) consists of two levels – an upper level with 27 broader categories, and a more granular lower level of 304 categories. Both of these classification systems function on the journal rather than article level, meaning that articles in highly multidisciplinary journals (e.g. Nature, PNAS, Science, PLOS One) may not fit well within categories they are assigned to, and that a single journal may be classified as belonging to more than one category.

To allow field-level comparison between our WoS and Scopus samples, we reclassified journals in both samples into the classification system of the Revised Fields of Science and Technology Classification of the Frascati Manual 2002 (http://www.oecd.org/science/inno/38235147.pdf) (herein referred to as the OECD classification system). The OECD classification system is a two-level system, with 6 upperlevels (Natural Sciences, Engineering and Technology, Medical and Health Sciences, Agricultural Sciences, Social Sciences, and Humanities) and 42 lower-levels. Both Clarivate Analytics and Scopus provide schemas to map their own classification systems to the OECD classification system (note that 7,539 items in the Scopus 'multidisciplinary' category (~1 % of total items) were excluded as this category is not mapped to any field or discipline in the OECD classification). In our analysis we utilise only the upper-level categories of the OECD classification system for broader field-level comparisons.

For analysis of authorship countries, we additionally extracted from WoS and Scopus the country of the affiliation of the first author. Where an author had multiple affiliations in multiple countries, both countries were extracted (no fractional counting was applied).

Altmetrics data

Altmetrics data were retrieved by querying individual article DOIs against the Altmetric.com public API. Altmetric.com provide total counts of various indicators of social media and online sharing. In this study we limited results to 6 main indicators: the number of tweets, mentions in blog posts, mentions in mainstream media news articles, mentions in Wikipedia, mentions in policy documents, and the number of Mendeley reads. As distributions of altmetric counts are usually highly skewed, counts of individual articles were log-transformed (log+1) prior to summarizing and reporting.

3. Preliminary results

Unpaywall sample

A breakdown of our random Unpaywall sample in terms of publications per year and access types is shown in Figure 2. The sample size increases year-on-year between 2013 and 2016, in line with general trends of annually increasing scientific publication output. The slight decrease in 2017 may be a result of the delay between publication and indexing of articles in Crossref, upon which Unpaywall is dependent. In terms of access types, Gold OA grew as a proportion of total output from ~10 % in 2013, to ~13 % in 2017. Green OA remained relatively static in terms of proportion of output (~7 %), with the exception of 2017 where

Green OA dropped to \sim 5 %, likely due to the delay between article publication and deposition by the authors in an OA repository, known as 'backfilling' (Archambault et al., 2014). Bronze OA is also reduced in 2017 compared to previous years – Piwowar et al. (2018) found that Bronze OA often represents delayed OA articles (articles published in a subscription journal, but made free to read after an embargo period), which may explain this decrease in recent years as articles remain under an embargo.



Figure 2: Publications per year and access type for the random sample of 1,000,000 Unpaywall records used in this study.

Altmetrics and OA: Crude Analysis

Figure 3 shows mean (log-transformed) counts of different altmetric indicators as a function of access type for our Unpaywall sample. Similar patterns are observed across all indicators – generally highest counts are observed in Bronze OA and then Green OA, whilst Closed Access publications have the lowest counts for all indicators. Although the general patterns are similar, the y-scale differs between each indicator and thus the absolute magnitude of counts differs greatly between them; Mendeley reads and tweets generally have the highest mean counts per article, whilst policy and Wikipedia mentions have the lowest counts.



Figure 3: Mean, log-transformed counts of six different altmetric indicators as a function of access type.

Subject Categories

Altmetric indicators in different subject areas and access types are showing in Figure 4, using our Unpaywall sample matched to WoS (Figure 4A) and Scopus (Figure 4B). Differences are observed in results from samples using the two databases, for example, using the WoS sample it appears that articles in Medical and Health Sciences, Natural Sciences and Social Sciences receive overall more tweets and Mendeley reads than Humanities, Engineering and Technology, and Agricultural Sciences. Using the Scopus sample, the picture is less clear. Differences between the two may result partly from differences in subject coverage of the databases themselves, and partly from differences in subject classification schemas and their mappings to the OECD schema.

Major differences are also observed in behaviour of altmetric indicators between access types for different subject areas. For example, Closed Access articles are generally the least tweeted amongst all subject areas (for both WoS and Scopus samples), with the exception of the Humanities, where Gold OA articles are the least tweeted. The same effect is observed in Mendeley reads, which appears to closely match tweets. Such results can help us to understand how articles published under different access types are engaged with in different research communities.



Figure 4: Mean, log-transformed counts of six altmetric indicators for six subject areas, as a function of access type, using our Unpaywall sample data matched to WoS (A) and Scopus (B). Subject areas correspond to the upper-level classifications of the Revised Fields of Science and Technology Classification of the Frascati Manual 2002.

Author Countries

Altmetric indicators as a function of access type for the 20 most productive countries (i.e. those with the highest publication output) in our Unpaywall sample matched to WoS and Scopus are shown in Figure 5A and Figure 5B, respectively.



OA type O Closed O Bronze O Hybrid O Gold O Green

Figure 5: Mean, log-transformed counts of six altmetric indicators as a function of access type for the 20 most productive countries within our Unpaywall sample, matched to WoS (A) and Scopus (B). Countries refer to the country of the affiliation of the first author.

Clear differences are observed between countries and their interactions with different types of access. For example, countries with the highest overall Twitter counts (e.g. UK, Australia, USA, Switzerland, Netherlands, Canada, Germany, France) tend to have higher Twitter counts for Gold OA publications, which is surprising given that overall, Twitter counts for Green and Bronze OA tend to be higher (Figure 3). A similar pattern is noted for Mendeley reads. Conversely, for the same countries Gold OA generally does not perform as well in blogs and news articles. Closed access articles tend to have lower counts than other access types for nearly all indicators and countries. For Brazil, Russia, Poland and Iran, Green OA tends to receive more Twitter attention than other access types, but this is not reflected in counts of other indicators.

4. Outlook

This contribution presents initial exploratory results into factors affecting the relationship between OA and altmetrics. The results show clear differences in engagement with scientific research on various online platforms dependent on how the research can be accessed. However, many factors, for example subject areas and country of authorship, can strongly influence these relationships.

Several previous smaller-scale analyses have found an 'altmetric advantage' of OA over non-OA articles; Wang et al. (2015) used a sample of 1761 articles from *Nature Communications*, Alhoori et al. used a sample of 42,582 from 23 top-ranked journals from Google scholar, and Robinson-Garcia et al. (2018) used a sample of 40,044 articles from *Physical Review B* and *Physical Review X*. Our findings here suggest that these previous results cannot be easily generalized, for example, the results from Wang et al. (2015) are based upon a multidisciplinary journal, whilst those from Robinson-Garcia et al. (2018) are based upon specialized Physics journals, and thus the two studies represent different communities with different sharing practices.

A number of open questions remain which will form the basis for discussion at the altmetrics19 workshop and direct future work. For example, what additional factors might be the most important influencers of the relationship between OA and altmetrics (e.g. publication venue, collaboration networks)? Is our OA classification schema appropriate? Are the data sources using (e.g. Unpaywall, WoS and Scopus) appropriate, or should we consider other data sources (e.g. Dimensions) too? What effect does the coverage of each data source have on our results?

5. Acknowledgements

This work is supported by BMBF project OASE, grant number 01PU17005A.

6. References

Alhoori, H., Ray Choudhury, S., Kanan, T., Fox, E., Furuta, R., & Giles, C. L. (2015). On the relationship between open access and altmetrics. iConference 2015 Proceedings.

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.

Archambault, É., Côté, G., Struck, B., & Voorons, M. (2016). Research impact of paywalled versus open access papers. Copyright, Fair Use, Scholarly Communication, Etc. Retrieved from https://digitalcommons.unl.edu/scholcom/29/

ElSabry, E. (2017). Claims About Benefits of Open Access to Society (Beyond Academia). In Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices. IOS Press.

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., ... 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

Robinson-Garcia N., Arroyo-Machado W., Moed H.F., Torres-Salinas D. (2018). Do altmetrics promote Open Access? An exploratory analysis on altmetric differences between types of access in the field of Physics. STI 2018 Conference Proceedings. http://hdl.handle.net/1887/65221

Swan, A. (2010). The Open Access citation advantage: Studies and results to date [Monograph]. Retrieved from https://eprints.soton.ac.uk/268516/

Wang, X., Liu, C., Mao, W., & Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. Scientometrics, 103(2), 555–564. https://doi.org/10.1007/s11192-015-1547-0

Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. Journal of Informetrics, 10(2), 347–364. https://doi.org/10.1016/j.joi.2016.02.003

Zhu, Y. (2017). Who support open access publishing? Gender, discipline, seniority and other factors associated with academics' OA practice. Scientometrics, 111(2), 557–579. https://doi.org/10.1007/s11192-017-2316-z