MORE READERS IN MORE PLACES: THE BENEFITS OF OPEN ACCESS FOR SCHOLARLY BOOKS

Alkim Ozaygena

Lucy Montgomerya, Cameron Neylona, Chun-Kai (Karl) Huanga, Ros Pyneb, Christina Emeryb and Mithu Lucraftb

Centre for Culture and Technology, School of Media, Creative Arts and Social Inquiry, Curtin University, Bentley 6102, Western Australia bSpringer Nature, 4 Crinan Street, London N1 9XW, UK

Abstract: *Open access to scholarly contents has grown substantially in recent years. This includes the number of books published open access online. However, there is limited study on how usage patterns (via downloads, citations and web visibility) of these books may differ from their closed counterparts. Such information is not only important for book publishers, but also for researchers in disciplines where books are the norm. This article reports on findings from comparing samples of books published by Springer Nature to shed light on differences in usage patterns across open access and closed books. The study includes a selection of 281 open access books and a sample of 3,653 closed books (drawn from 21,059 closed books using stratified random sampling). The books are stratified by combinations of book type, discipline and year of publication to enable likewise comparisons within each stratum and to maximise statistical power of the sample. The results show higher geographic diversity of usage, higher numbers of downloads and more citations for open access books across all strata. Importantly, open access books have increased access and usage for traditionally under-served populations.*

Keywords: *open access, books, Springer Nature, usage pattern, download, citation.*

INTRODUCTION

Open access (OA) to scholarly outputs has taken the central stage in recent years, with numerous international, regional and local initiatives leading the way in advancing rapid changes to the publishing landscape. Yet, despite the high volume of research available on journal articles (and academic outputs in general), relatively little has focused on OA books. In particular, there is limited information on the level of online usage, their geographic distribution

and, importantly, how usage may be influenced by publishing books in OA forms.

There are numerous potential proxies for measuring the usage of scholarly work. These include citations, downloads, website visits, social media mentions and their various forms. Through a randomised controlled trial, Davis, Simon & Connolly (2008) is able to show that OA articles have higher numbers of downloads and more unique web page visitors than non-OA articles1. Wang et al. (2015) further finds that the increased level of downloads for OA articles is sustained over time. This is found in addition to OA articles attracting more social media attention2. However, research also found the OA advantage of altmetric activities to have significant differences across disciplines3. The citation advantage of OA publishing remains a hotly debated issue. Although, a recent literature review shows there is relatively more research in support of the OA advantage, with the caveat that there may be a large variability across disciplines4.

Most of the above findings have a strong focus on journal articles. Yet, it remains unclear whether these results can be generalised to books. In particular, there are significant differences between journal articles and books in terms of how they are hosted, shared and used online, and how they can be identified and tracked5. These make the integration of usage data for books a challenging task. Counting Online Usage of Networked Electronic Resources (COUNTER) is an international effort to overcome some of these problems. It is a code of practice for compiling online usage statistics of electronic resources6. Benchmarking book usage levels is another important aspect to consider. Books with different attributes (such as different languages and research fields) can have vastly different target audiences. Hence, the ability to compare books with similar attributes is essential for deep understandings of book usage.

There is a limited amount of previous work comparing downloads of OA and non-OA books with the goal of understanding the impacts of OA on the geographies of usage. The work of Snjider (2013) showed increased usage for OA books as well as some evidence of an increase in sales7. Using a sample of 180 books Snijder (2013) showed that OA led to increased proportions of usage in developing countries as well as demonstrating a "digital divide" in discovery and use.

This article, which extends the findings of Snijder (2013), provides a timely update to evidence-based arguments for the benefits of OA to scholarly books. Our analysis of a larger sample allows us to investigate these effects,

particularly the geographic effects, in much greater detail. Using books available from a common source (i.e., Springer Nature) also alleviate some of the challenges discussed above. Having download data by month and various disciplines for all books allows us to confirm that downloads are higher for OA books across their whole history and across all disciplines. We also update analysis on the effects of OA across downloads, citations, and web visibility for a single large sample, following on the work undertaken by Springer Nature in 20178.

MAIN FINDINGS

This article reports on the analysis of usage (with downloads, citations and web visibility as proxies) and related indicators for a sample of books that is stratified by mixtures of book type, discipline and year of publication. In particular, the analysis considers the geographic usage of OA and non-OA books, examining whether OA facilitates the takeup of books by countries or regions that are traditionally underrepresented in the production and use of scholarly content.

To the best of our knowledge, this is the largest independent analysis ever conducted on the usage of OA and non-OA books. The sample size and sampling procedure allow us to be significantly more confident that there are substantial effects connecting OA status with downloads and citations for this set of books.

The main findings of our analysis are:

•OA books as a group show a higher geographic diversity of usage and reach more countries, i.e., they have a greater proportion of usage in a wider range of countries.

•OA books have increased access and usage for under-served populations and low or middle income countries, including a high number of countries from Africa.

•OA books as a group have ten times more downloads than non-OA books and more than double the number of citations.

•There is higher (at least 2.7-fold) usage (via downloads) of OA books across every stratum in our sample. That is for every type of book, every discipline, and each of the three years of publication in the sample, OA books show more usage than their non-OA comparison groups. This holds for every month after publication and for alternate categories such as imprints.

•Books that contain the name of a country or region in their title generally show enhanced usage in that country or region. This effect is clearest for Latin America and Africa and is greater for OA titles.

•Anonymous downloads are generally around double that of logged downloads. This means reporting that relies on institutional identification will be substantially undercounting the usage of OA books.

These findings are important for stakeholders by providing a robust understanding of the benefits of publishing books in OA forms. They give support to evidence-based publishing and marketing strategies for publishers. They also equip authors with enhanced knowledge for making decisions about publishing venues, formats and titles, etc. It is our hope that these findings will facilitate the advancement towards a greater diversity of readership and accessibility.

DATA AND METHODOLOGY

Springer Nature provided a set of 281 English-language OA titles published by its various imprints (e.g. Palgrave Macmillan, Springer, Birkhäuser) in 2015, 2016 and 2017. The titles were divided into three book types ("monographs", "contributed volumes" and "briefs"9); as well as five discipline clusters: "humanities", "social sciences", "business and economics", "medical, biomedical and life sciences" and "physical sciences, engineering, mathematics and computer science". Springer Nature also provided access to metadata relating to an additional 21,059 non-OA titles for the purposes of the study. Of the 21,059 non-OA books, a comparison set of 3,653 non-OA books was selected for closer analysis. The non-OA books were selected using a stratified random sampling procedure (stratified across combinations of book type, discipline cluster and year of publication) aimed at maximising statistical power of the sample and maintaining a consistent ratio of OA to non-OA books in each stratum.

There are three primary metrics that are of interest to this study; namely downloads, citations and web visibility. The first two of these are supplied by Springer Nature. The Springer Nature downloads data includes country information for logged access (known institutional subscriber to Springer Nature). This is supplemented with the use of the IP2Location database10 to determine country locations of anonymous downloads. Web visibility is determined through analysis performed by a webometrics11 tool. In particular, we analyse URLs mentioning each book to extract information such as the

number of unique domain names12 that references the book and country of the domain name.

We compare the average number of downloads, citations and unique domains, as well as the average downloads over time, between OA and non-OA books across different book types and discipline clusters. The geographic distributions of downloads across countries are visualised and are compared using the Gini coefficient. Further details of the data and methodology are provided in the Appendix.