

More readers in more places: the benefits of open access for scholarly books

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 maximize 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 underserved populations.

Keywords

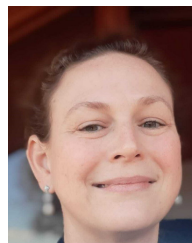
open access; books; Springer Nature; usage pattern; download; citation



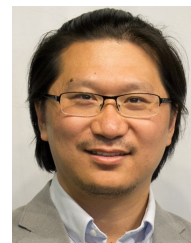
CAMERON NEYLON
Professor of Research Communications
Centre for Culture & Technology
Curtin University



ALKIM OZAYGEN
Research Fellow
Centre for Culture & Technology
Curtin University



LUCY MONTGOMERY
Professor of Knowledge Innovation
Centre for Culture & Technology
Curtin University



CHUN-KAI (KARL) HUANG
Research Fellow
Centre for Culture & Technology
Curtin University



ROS PYNE
Director
OA Books and Book Policies
Springer Nature



MITHU LUCRAFT
Director Content
Marketing Strategy
Springer Nature



CHRISTINA EMERY
Senior Marketing Manager
Thought Leadership Programmes
Springer Nature

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 randomized controlled trial, Davis, Simon and Connolly showed that OA articles have higher numbers of downloads and more unique web page visitors than non-OA articles.¹ Wang et al. 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 attention.² However, Holmberg et al. also found the OA advantage of altmetric activities to have significant differences across disciplines.³ The citation advantage of OA publishing remains a hotly debated issue. A recent literature review showed there is relatively more research in support of the OA advantage, with the caveat that there may be a large variability across disciplines,⁴ and arguments continue as to the reliability of different approaches to this question.

Most of the above findings have a strong focus on journal articles. It remains unclear whether these results can be generalized 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 tracked.⁵ 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 resources. (Note 1) 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 a deep understanding of book usage.

'the ability to compare books with similar attributes is essential for a deep understanding 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 Snijder showed increased usage for OA books as well as some evidence of an increase in sales.⁶ Using a sample of 180 books, Snijder 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, provides an 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 alleviates 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 2017.⁷

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

3 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 take-up of books by countries or regions that are traditionally under-represented 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 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 underserved 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 increased 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 as they provide 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' [Note 2]); 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 maximizing the statistical power of the sample and maintaining a consistent ratio of OA to non-OA books in each stratum.

'examining whether OA facilitates the take-up of books by countries or regions that are traditionally under-represented'

'These findings are important for stakeholders as they provide a robust understanding of the benefits of publishing books in OA forms'

4 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 include country information for logged access (known institutional subscriber to Springer Nature). This is supplemented with the use of the IP2Location database (Note 3) to determine country locations of anonymous downloads. Web visibility is determined through analysis performed by a webometrics (Note 4) tool. In particular, we analyse URLs mentioning each book to extract information such as the number of unique domain names (Note 5) that refer to the book and the countries of those domain names.

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 visualized and the diversity of geographic usage for each book is investigated. Further details of the data and methodology are provided in the Supplementary Materials.

The article focuses on four key questions:

- Do OA books and non-OA books show different patterns of geographic usage?
- Is there evidence of wider usage particularly from countries and areas that are not high users of non-OA books?
- Does such performance vary, depending on the form (e.g. monograph, brief, contributed volume) of the book or its disciplinary area?
- Is there robust evidence that OA books outperform non-OA books on various proxy measures of usage?

Analysis and discussion

Open access books show more overall usage

In the first instance, we compare the average number of downloads, citations and unique domains (referencing the books) across OA and non-OA books as two groups. We also draw parallel comparisons of the two groups across book types and discipline clusters. These are summarized in Figure 1.

OA books as a group have on average ten times more downloads than non-OA books (first pair of bars in the top panel). There are more than double the number of citations for OA books. To a lesser extent (proportionally), there is also on average a higher number of unique web domains referencing OA titles. Higher levels of usage (via all three proxies) for OA books are also observed across each of the groups by book type and discipline cluster. It can be seen that the magnitudes of difference for each metric across OA and non-OA books vary across the different groups. For example, the difference between downloads of OA books and non-OA books seems to be amplified for the biomedical sciences. Given our sample these differences could be specific to Springer Nature or may reflect differences in disciplinary usage. Further work will be required to identify the driving factors. However, there is a consistent pattern across the different groups that OA books are seeing more usage.

'there is a consistent pattern across the different groups that OA books are seeing more usage'

The number of citations is a useful proxy of academic usage, while web visibility provides insight into how the books are being used on the web: either as linked text or references. These are additional proxies of usage to the number of downloads for books. The presence of higher levels of usage signalled by all three proxies suggest that OA books are not only being downloaded more often than their non-OA counterparts, but are also being read, used, referenced and attracting attention in different ways. This strengthens the case for a usage effect that is related to OA specifically.

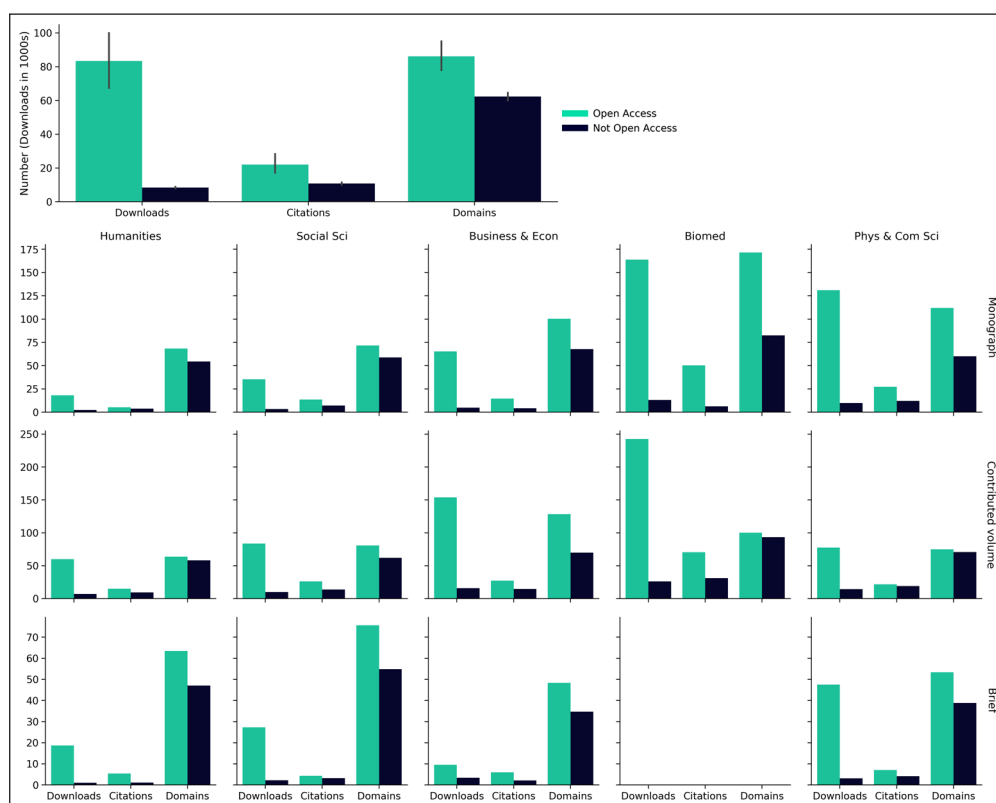


Figure 1. Open access books show more usage and attention through average numbers of downloads, citations and web domains

The top panel compares average numbers of downloads, citations and web visibility across all OA books and all non-OA books in our sample. Download numbers are provided in 1,000s and vertical lines are the 95% confidence interval for each metric. Parallel comparison across book types and discipline clusters are given in the subsequent panels. Each panel includes relevant books published in 2015, 2016 and 2017. There are no OA books in the brief category for biomedical sciences.

Open access books show more usage over time

We are also interested in comparing the download levels over the lifetimes of the books. Using the times recorded for downloads and the publication dates of each book, we are able to provide the time series of downloads per month over 40 months from the date of publication. These are summarized in Figure 2, showing overall trends comparing all OA to all non-OA books in the corpus, and for each variation of book type and discipline cluster.

The number of downloads is displayed in log-scale, meaning the magnitude of differences are even greater than shown. The 95% confidence bands for downloads are shown in the top panel, which compares downloads over time for all books in the sample. Each panel includes relevant books published in 2015, 2016 and 2017. Some books have usage data from prior to the official release date so show usage prior to zero months. We do not have specific explanations for the spikes in usage evident in some panels, although high measured usage in the first month of release is common (see e.g. Business and Economics, and Biomedical Monographs). Spikes in usage later in the life of books could be a result of high usage of a specific book (e.g. as a result of its use as a reading in a massive online open course, or MOOC).

The top panel shows clear evidence of OA books having an advantage in the number of downloads over time. For all 40 months in the analysis, OA books have recorded significantly more downloads than their non-OA counterparts. That is, not only do OA books have a higher number of downloads to begin with, but this effect is also persistent over time.

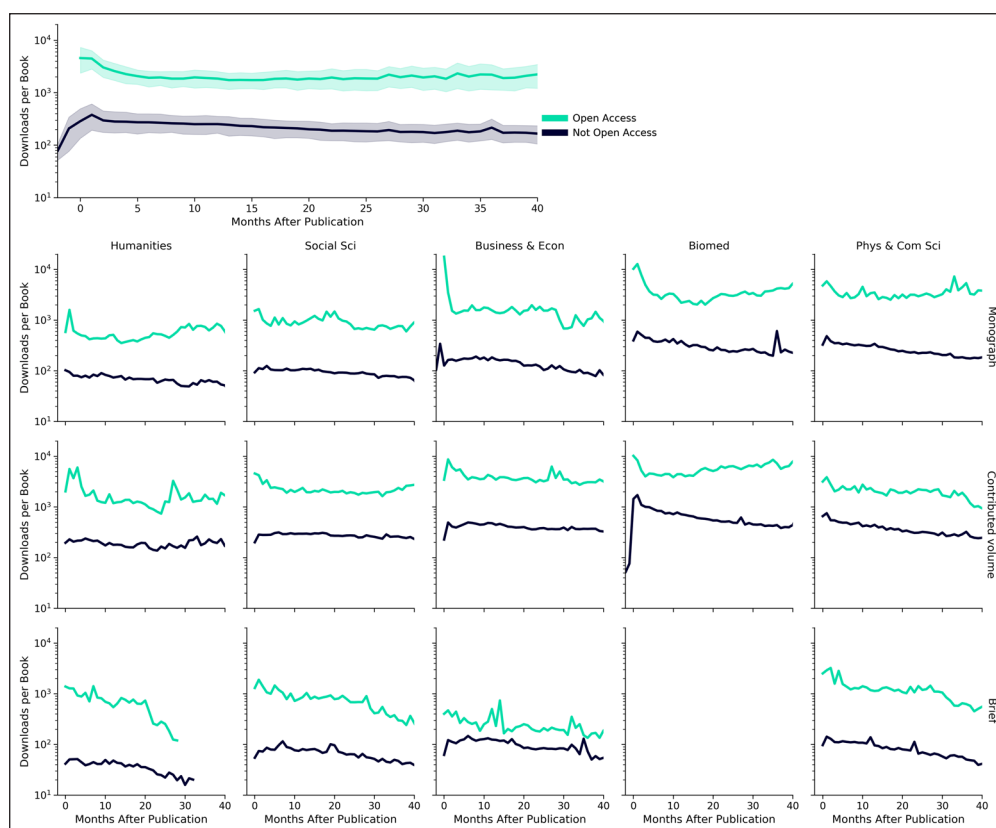


Figure 2. Open access books depicts more downloads per book for every month since date of publication

This general pattern carries over to subsets of books by book type and discipline clusters. Again, there are variations across groups but persistently higher downloads for OA books. It is also interesting to note that many groups of OA books seem to enjoy a more impactful starting point (noting the sudden shock of downloads at time zero with log-scale number of downloads displayed in many of the panels in Figure 2).

Open access books show usage in a wider range of countries

By analysing the geolocation of downloads, we are able to provide comparisons of book usage across countries. For each book, we record the number of downloads from each country. Subsequently, we can calculate the number of downloads per book for any specific country. We do this for the set of all OA books included in our study and for the non-OA books. The results are visualized in Figure 3.

Evidently the usage (via downloads) of both non-OA and OA books is international and spans the globe. Usage of non-OA books is identified in 118 countries (by definition these are all countries with licensed access to Springer Nature content). In contrast, usage of OA books is identified in 201 countries. For both OA and non-OA books the highest levels of usage are seen in the USA, UK, Germany and mainland China. With only a few exceptions, OA books see higher levels of usage across the globe. There is also evidence that OA books see more average downloads from some African and Latin American countries that otherwise had very little access to non-OA titles.

'By analysing the geolocation of downloads, we are able to provide comparisons of book usage across countries.'

An important aspect to consider is the role of population size in the geographical patterns of usage. One challenge with the analysis of academic usage is to identify a good proxy of 'academic population size'. Examples of this include normalization by the country's number of people in tertiary education, overall academic output size and total number of citations. We provide one such example through normalizing the downloads by total number of publications (see Supplementary Materials for details on this data). This is visualized in Figure S1 of the Supplementary Materials.

7 While total publication size is not a perfect proxy for the number of potential academic readers of the books, this normalization suggests that 'usage per academic work' for both OA and non-OA books is fairly consistent across North America and North Western Europe, as well as mainland China, South Africa and Australia. Focusing on other geographical regions, and consistent with the overall download counts (see Figure S1 of the Supplementary Materials), Egypt emerges as a heavy user of non-OA content relative to overall academic outputs, along with Uruguay, Ethiopia and Uzbekistan. Relatively heavy users of OA content relative to academic output size include Somalia, Afghanistan, Bhutan, Niger and South Sudan.

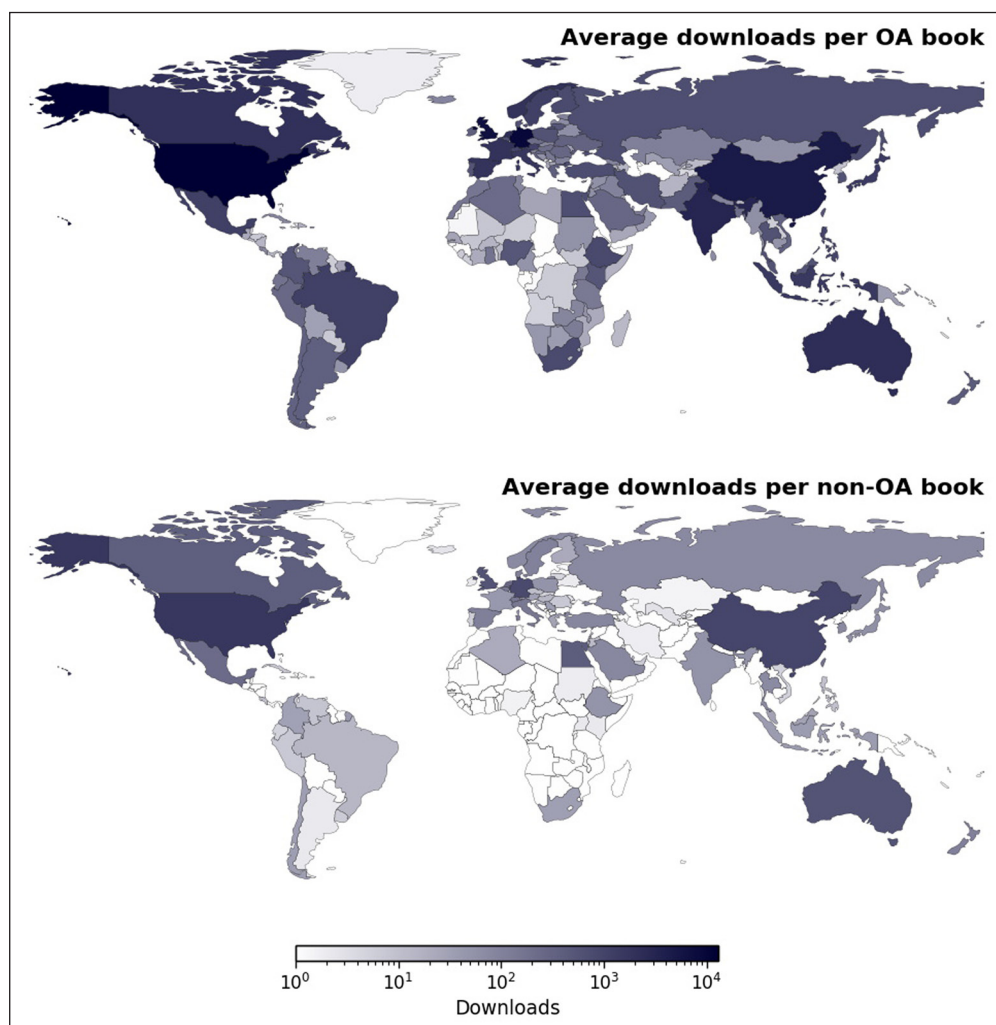


Figure 3. The geographic distribution of downloads for OA and non-OA books

Average usage by country across the whole corpus for OA (top) and non-OA (bottom) books. Several countries show a greater concentration of usage for OA books and these countries are predominantly in the southern hemisphere. Both maps are on the same colour scale (in log-scale).

Open access books show increased usage for underserved populations and low to middle income countries

Overall, we find OA books having to reach more countries globally than non-OA books. Figure 3 also presented some evidence that OA books see improved usage for traditionally underserved countries and low to middle income countries. To explore this in more detail we focus on the anonymized downloads, which are solely attributed to OA books (as non-OA books can only be accessed if subscribed to, i.e. logged access). In particular, we can

8 examine usage in countries that do not otherwise have access to Springer Nature books in digital formats. Usage of OA books was identified in a wide range of countries that recorded zero usage of the non-OA books in the dataset. Of these countries where only OA books recorded usage, over 20 were from Africa. Usage of OA books from countries that do not otherwise purchase Springer Nature e-book titles totalled 118,247 downloads, representing 1% of the total anonymous usage of OA titles.

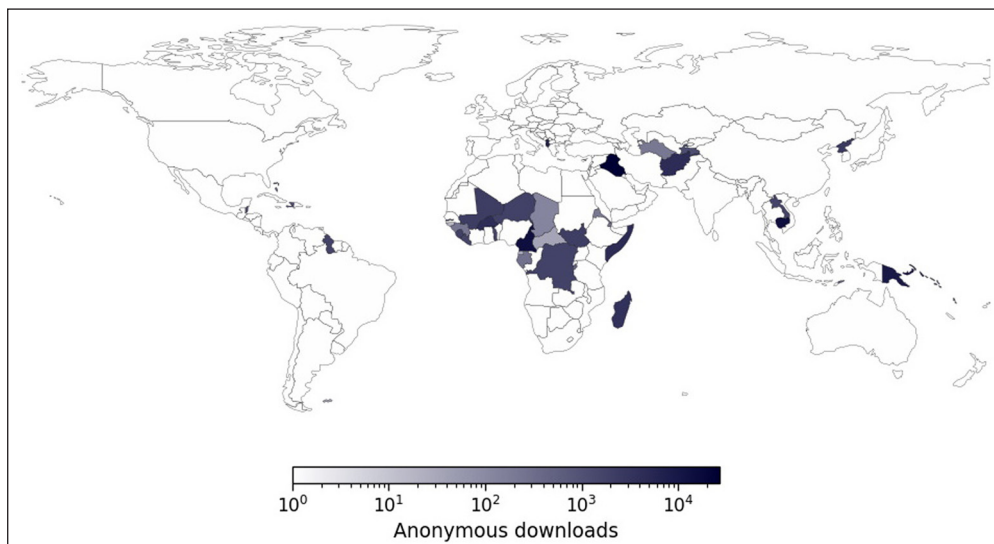


Figure 4. Anonymous usage from countries with no logged usage

A wide range of countries with zero logged usage show anonymous usage (see Figure 4) with a total of 118,247 downloads representing 1% of the total anonymous usage. Of those countries, more than 20 were from Africa, with others mostly in the Middle East and South East Asia. Countries with low GDP and development are significantly represented in this group.

Open access books show higher diversity of usage

While we have provided evidence for wider usage of OA books (i.e. more countries download OA books), it is also important to understand the level of disparity amongst country usage. We can provide a more quantitative measure of this effect by examining a disparity index. That is, how much usage deviates from the situation where all countries show even usage. The Gini coefficient is a disparity index that is often used to define levels of income inequality. We can use the same calculation to measure inequalities in geographical usage and use this to compare OA and non-OA books. A lower Gini coefficient indicates more diverse usage. That is, a lower number means lower inequality, or if preferred, greater equality. The Gini coefficient is calculated for every book in our sample. Figure 5 shows a summary of these results for OA and non-OA books overall, and for different combinations of book type and discipline cluster.

For the overall corpus, and for every individual category, the usage by country is substantially more diverse for OA books. This form of analysis may be useful in identifying books that have significant potential to reach diverse geographic audiences. It might also be interesting to examine whether a low Gini coefficient for a non-OA book suggests potential for substantially enhanced usage if the book were converted to OA. It should be noted that there are outliers amongst both OA and non-OA books in terms of their values in the Gini coefficient. In particular, there are exceptional books in both categories that may have broad interest (e.g. *A Theory of Philosophical Fallacies*) or narrow geographical focus (e.g. *A History of Male Psychological Disorders In Britain, 1945–1980*).

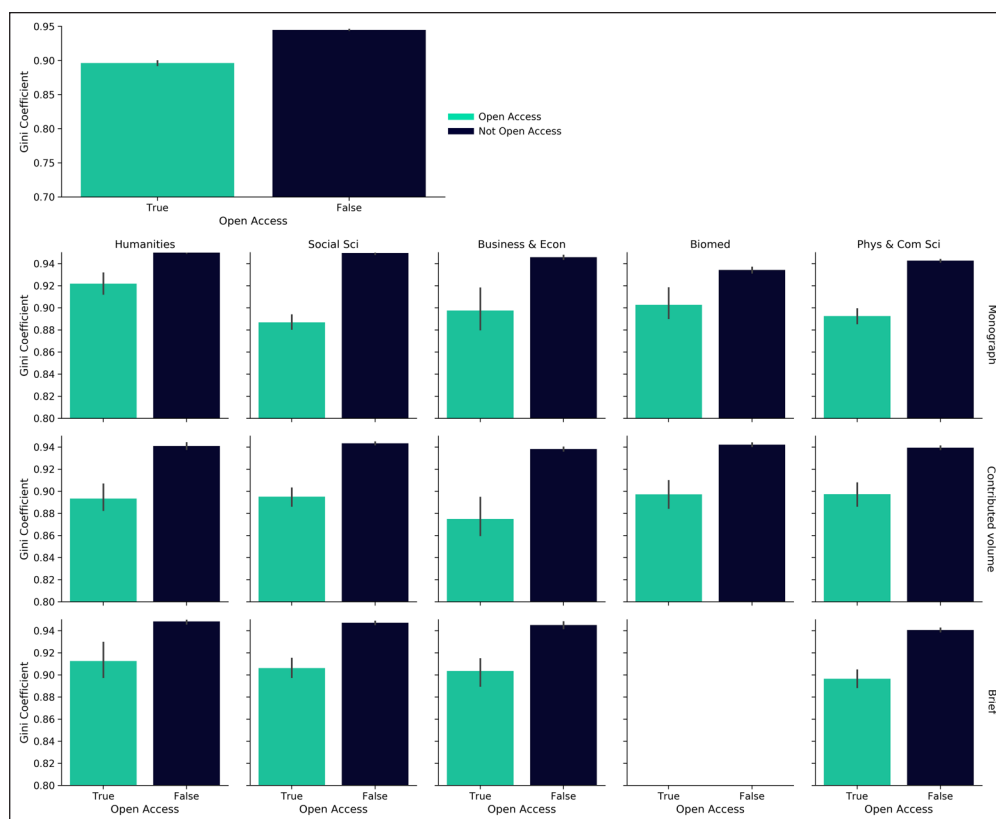


Figure 5. Diversity of book usage amongst countries as measured by the Gini coefficient

The Gini coefficient is a statistical measure of inequality. Here the coefficient is calculated for the contribution of each country to the overall usage of each book. A lower Gini coefficient means more diverse usage. The median Gini coefficient and the 95% confidence interval are shown. For the corpus as a whole and for every category, the median Gini coefficient of OA books is lower, meaning that the geographical usage of OA books is more diverse (i.e. less unequal).

Open access books associated with a larger title effect on geographic usage

Approximately 16% of the whole corpus of books in this study had a region or country name in their title or subtitle. The proportion of OA versus non-OA books with a geographic reference in the title is approximately the same. We hand coded title-references to countries or regions, including variations referring to language (e.g. 'Chinese') and regions ('Africa', 'Latin America' as well as possessives such as 'sub-Saharan' and 'Latin American') and examined the usage from those regions, focusing on Africa and Latin America as examples. We are interested in discovering whether there is a 'title' effect, i.e. a book with a title referencing a region shows greater usage in that region, and if this effect differs across OA and non-OA books.

In Figure 6 we show the relative increase in downloads for books referencing Africa in their titles. We calculate the expected downloads per book for each country from the whole dataset and divide the actual usage for those with 'Africa' in the title by the expected downloads. A value of five indicates that there are five times more downloads for books referencing Africa in the title in a given country than is expected based on the whole dataset (or the set of non-OA or OA books respectively). The three panels show download patterns for all books, OA books and non-OA books, respectively. Parallel visualizations for books referencing 'Latin America' are provided in Figure S2 of the Supplementary Materials.

'For Africa and Latin America, we see a substantial increase in usage overall from those regions featuring in the title'

For Africa and Latin America, we see a substantial increase in usage overall from those regions featuring in the title. In both cases there is also some evidence of increased usage in some parts of the other region (e.g. increased usage in a small number of African countries for titles that have Latin America in the title). In the case of books with 'Africa' in the title there is increased Latin

10 American usage in Guyana, Suriname, Venezuela and Panama. Other countries with increased usage are Laos, the Solomon Islands and Timor-Leste. It should be noted that many of these countries have relatively low usage so the changes may not be significant in statistical terms.

The size of this geographic title effect is substantially larger for books that are OA. OA books about Africa are widely read beyond Africa as well as in Africa. There is a very large increase in usage compared to the whole sample across the African continent. By contrast, non-OA books with 'Africa' in the title show a usage increase only in South Africa, Uganda, Ethiopia and Sudan. Not only is OA associated with increased usage in countries under-represented in global scholarship, but it is also associated with increased global usage of scholarship about under-represented countries. See the case study in the Supplementary Materials for an example of the enhancement of usage for a book with a specific country in the title.

Similar effects are seen for Latin America, although the enhancement is not quite as localized as in the case of Africa. Usage in Latin America is substantially greater for all books with 'Latin America' (including variations and possessives such as 'Latin American') in the title. The size of this increase is substantially larger for OA books. In addition, there is broader usage internationally for the OA books. The overall effect is larger for books with 'Latin America' (maximum of 100 times more usage) than it is for 'Africa' (maximum increase of fivefold).



Figure 6. Increases in usage for books with 'Africa' in the title

11 Overall usage for each country of books referencing 'Africa' in the title was divided by usage for that country for the full corpus (or for the set of OA or non-OA books respectively). Increases in usage are shown, with a value of five meaning there are five times the number of downloads for books referencing Africa in the title. Countries showing unchanged or decreased usage are displayed in the lightest shades. OA books show a greater increase across a range of countries, with the increase concentrated in Africa.

Anonymous usage versus logged usage

Recall that usage (via downloads) for OA books can be categorized as logged access or anonymous access. In fact, we see significant differences in usage numbers across these two types of access. The overall amount of anonymous usage is always greater than logged usage for each book: generally, twice as much (with exceptions for only a small number of books in our study).

We cannot directly ascribe anonymous usage to 'general public' or 'non-academic' usage because a proportion of this will be off-campus or personal device usage by scholars. However, there are differences in the patterns of usage at the country-level. Anonymous usage is higher in Kenya, Brazil, India and Iran. Logged usage is comparatively higher in Egypt. We have already noted that there is substantial anonymous usage in countries for which there is no logged usage (see Figure 4). There is also a wide range of countries with a very high proportion of anonymous usage, despite these countries including institutions that have access to non-OA content. This list includes Syria, Ukraine, Georgia, Guatemala and Sri Lanka. The distribution of downloads across countries for each type of access is displayed in Figure S3 of the Supplementary Materials.

'The overall amount of anonymous usage is always greater than logged usage for each book'

Web visibility across top level domains (TLDs)

In this study we use the number of unique domains determined by webometrics analysis as a proxy for web visibility. While OA books display higher levels of web visibility overall and for various categories of books (see Figure 1), these differences are proportionally much less than those seen for downloads and citations.

We can examine geographical aspects of this web visibility by looking at the top level domains (TLDs), the part of a URL after the last 'dot' in the domain name. While we cannot ascribe a geographical location to common TLDs such as .com and .org, many domain names can be associated with countries. Broadly speaking the geographical representation of pages that refer to books in the corpus is consistent with the usage via downloads and citations, with European (.uk, .de, .it), North American (.edu, .ca) and Australian (.au) TLDs (Note 6) dominating (see Table S1 of the Supplementary Materials). The top ten TLDs constitute 80% of all web pages identified for the whole corpus of books. Overall, the difference between the number of websites and the range of TLDs between the OA and non-OA books shows a 39% increase in unique domains referencing the OA books versus non-OA titles (with the increases for each of the top ten TLDs displayed in Figure S4 of the Supplementary Materials). This is a relatively small increase compared to the tenfold effect on downloads and more than doubling of the number of citations.

Springer Nature has well-established and effective pathways for marketing and digital dissemination which are applied to both OA and non-OA titles. The small effect of OA on the number of websites referring to these titles is most likely to be a result of the fact that both OA and non-OA titles benefit from these processes.

Limitations and further work

The primary limitations of this study are that it only examines books from a single publisher, and only examines usage of OA books via a single platform: SpringerLink. Springer Nature OA books are also made available via a range of other platforms, including the OAPEN Digital Library, the Directory of Open Access Books (DOAB), and from Google Books, Apple Books, Amazon and funders' own platforms. Where appropriate, OA books are also indexed in Web of Science, Scopus, PubMed's NCBI Bookshelf, PMC, Medline, as well as more than 200 more abstracting and indexing services and Google Books. Direct usage on those platforms is not captured in this study, although traffic referred from these sites may be included in our data. We also only capture digital usage and reach of e-books and do not consider print sales and distribution which may also show different trends for OA and non-OA books.

'The primary limitations of this study are that it only examines books from a single publisher'

Springer Nature is a large publisher with an experienced and effective sales and marketing team and online infrastructure. We would predict that the overall effect of this would be to reduce the difference between OA and non-OA books on the metrics we can measure. On the other hand, the dataset used in this study is significantly larger than the sets used to inform other published studies, which have used data from small OA monograph publishers with a limited number of titles. To our knowledge this is the largest analysis comparing usage and visibility of OA and non-OA books ever conducted. The size of the dataset increases confidence in the study's conclusions.

A significant statistical limitation is that the study was conducted on a retrospective stratified sample. We therefore cannot completely rule out confounding effects resulting from variables beyond our control. Specifically, we have not controlled for affiliation or the prestige or fame of authors. There is some risk that there is a correlation between the wealth of an institution (and therefore its ability to fund OA publication), the prestige and reach of authors and therefore the downloads and citations of books. However, the nature of our stratified sample and the consistency of positive effects across all groups, for all types of books, for all disciplines, for all three years of publication and for all times after publication provides confidence that the effects of OA are credible.

In some cases, usage numbers are small and this can exaggerate the effects seen when seeking to normalize usage. The precise size of geographic effects and to some extent the ordering should therefore not be relied on. However, the broad patterns of change and directions of effect are robust and the broad geographic patterns of changes in usage are consistent across various subsamples and for individual books. Overall, we are highly confident of the claim that OA is associated with increased usage in countries which suffer exclusion from scholarly discourse.

'OA is associated with increased usage in countries which suffer exclusion from scholarly discourse'

The webometric analysis is reliant on construction of a search term that combines the title with author names. This can be expected to experience false negatives (not all web pages referring to the book will contain this information) and some false positives (particularly for short titles and common author names). Nonetheless the broad pattern of visibility should be reliable and is supported by its concordance with the geographic usage data.

Supplemental Files

The supplemental files for this study can be found as follows:

- Alkim Ozaygen, Lucy Montgomery, Cameron Neylon, Chun-Kai (Karl) Huang, Ros Pyne, Mithu Lucraft and Christina Emery, Supplementary Information for "More readers in more places: the benefits of open access for scholarly books," (2021), <https://doi.org/10.5281/zenodo.5571123>.

Data Accessibility Statement

Data and code are provided at Zenodo. As the book-level usage data is proprietary to Springer Nature we do not share this detailed data. However, we do provide full code for audit and hashes to ensure the provenance of both raw and processed data.⁸ Data for webometrics, citations, normalization, and the full set of titles analysed are provided.⁹

Acknowledgements

Springer Nature provided funding to COARD to conduct this research and also provided the data.

Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other *Insights* articles can be accessed here – click on the URL below and then select the 'full list of industry A&As' link: <http://www.uksg.org/publications#aa>.

Competing Interests

This work was funded by Springer Nature through a research project by COARD (previously Knowledge Unlatched Research). Springer Nature also provided the usage and citation data and were involved in the design of the open access sample set (which includes all English language open access books published across 2015–2017 by Springer Nature). The comparison set of books that are not open access was selected from the full list of comparable books by the COARD team. Several of the co-authors are from Springer Nature and were actively involved in discussion of the analysis and preparation of the article narrative.

Notes

1. COUNTER is an international non-profit membership organization of libraries, publishers and vendors. COUNTER publishes a widely accepted standard for calculating the usage of electronic resources, as well as a Code of Practice for handling and cleaning usage data for scholarly publications. See <https://www.projectcounter.org/> (accessed 19 October 2021).
2. Springer Briefs and Palgrave Pivots are Springer Nature's short-form book types, typically 50–125pp/25,000–50,000 words. For brevity, both formats are referred to as 'briefs' throughout the report.
3. The IP2Location lite DB9 was used for this study, <https://lite.ip2location.com/database/ip-country-region-city-latitude-longitude-zipcode> (accessed 19 October 2021).
4. Webometrics aims to measure the impact of a research object across the web by examining numbers and types of hyperlinks and employing bibliometrics approaches to examine usage patterns. Tomas C. Almind and Peter Ingwersen, "Informetric analyses on the World Wide Web: Methodological approaches to 'webometrics'," *Journal of Documentation*, 53, no.4 (1997): 404–426, <https://doi.org/10.1108/EUM000000007205> (accessed 19 October 2021).
5. A domain name is an address that people use on the internet, whether for websites or for email. It is a string of characters which usually spells out a word or the name of a company, organization or person. For the URL <http://ccat.curtin.edu.au/about-us.html> the domain name is curtin.edu.au.
6. TLD refers to the last segment of a domain name, or the part that follows immediately after the period. TLDs are classified into two categories: generic TLDs (gTLD) and country-code TLDs (ccTLD). Examples of some common TLDs include .com (commercial businesses), .org (organizations), .net (network organizations), .gov (U.S. government agencies), .edu (educational facilities like universities), .ca (Canada), and .au (Australia).

References

1. Philip M. Davies et. al, "Open access publishing, article downloads, and citations: randomised controlled trial," *BMJ*, 337 (2008): a568, DOI: <https://doi.org/10.1136/bmj.a568> (accessed 19 October 2021).
2. Xianwen Wang et. al, "The open access advantage considering citation, article usage and social media attention," *Scientometrics*, 103 (2005): 555–564, DOI: <https://doi.org/10.1007/s11192-015-1547-0> (accessed 19 October 2021).
3. Kim Holmberg et. al, "Do articles in open access journals have more frequent altmetric activity than articles in subscription-based journals? An investigation of the research output of Finnish universities," *Scientometrics*, 122 (2020): 645–659, DOI: <https://doi.org/10.1007/s11192-019-03301-x> (accessed 19 October 2021).
4. Jonathan P. Tennant et. al, "The academic, economic and societal impacts of Open Access: an evidence-based review," *F1000 Research*, 5 (2016): 632, DOI: <https://doi.org/10.12688/f1000research.8460.3> (accessed 19 October 2021).
5. Alkim Ozyaygen, "Analysing the usage data of open access scholarly books: What can data tell us?," PhD thesis, Curtin University, (2019), <http://hdl.handle.net/20.500.11937/79585> (accessed 19 October 2021).
6. Ronald Snijder, "Do developing countries profit from free books? Discovery and online usage in developed and developing countries compared," *Journal of Electronic Publishing*, 16, no. 1 (2013), DOI: <https://doi.org/10.3998/3336451.0016.103> (accessed 19 October 2021).
7. "The OA Effect," Springer Nature, 2017, <https://www.springernature.com/gp/open-research/journals-books/books/the-oa-effect> (accessed 19 October 2021).
8. Alkim Ozyaygen, Lucy Montgomery, and Cameron Neylon, Data for: "More Readers in More Places: The benefits of open access for scholarly books," *Zenodo*, (2020), DOI: <https://doi.org/10.5281/zenodo.4018842> (accessed 19 October 2021).
9. Alkim Ozyaygen, Cameron Neylon, and Karl Huang, Code for: "More Readers in More Places: The benefits of open access for scholarly books," *Zenodo*, (2020), DOI: <https://doi.org/10.5281/zenodo.4019215> (accessed 19 October 2021).

Article copyright: © 2021 Cameron Neylon, Alkim Ozaygen, Lucy Montgomery, Chun-Kai (Karl) Huang, Ros Pyne, Mithu Lucraft and Christina Emery. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use and distribution provided the original author and source are credited.



Corresponding author:

Cameron Neylon

Professor of Research Communications

Centre for Culture & Technology

Curtin University, AU

E-mail: Cameron.Neylon@curtin.edu.au

ORCID ID: <https://orcid.org/0000-0002-0068-716X>

Co-authors:

Alkim Ozaygen

ORCID ID: <https://orcid.org/0000-0001-6813-8362>

Lucy Montgomery

ORCID ID: <https://orcid.org/0000-0001-6551-8140>

Chun-Kai (Karl) Huang

ORCID ID: <https://orcid.org/0000-0002-9656-5932>

Ros Pyne

ORCID ID: <https://orcid.org/0000-0002-7310-3153>

Mithu Lucraft

ORCID ID: <https://orcid.org/0000-0003-0355-6576>

Christina Emery

ORCID ID: <https://orcid.org/0000-0003-3903-6724>

To cite this article:

Neylon C, Ozaygen A, Montgomery L, Huang C-K (Karl), Pyne R, Lucraft M and Emery C, "More readers in more places: the benefits of open access for scholarly books", *Insights*, 2021, 34: 27, 1–14; DOI: <https://doi.org/10.1629/uksg.558>

Submitted on 14 June 2021

Accepted on 21 July 2021

Published on 01 December 2021

Published by UKSG in association with Ubiquity Press.