



# How Can We Use Social Media Data Related to OA Monographs

Alkim Ozaygen, Lucy Montgomery, Cameron Neylon, Katie Wilson, Richard Hosking, Karl Huang

## ► To cite this version:

Alkim Ozaygen, Lucy Montgomery, Cameron Neylon, Katie Wilson, Richard Hosking, et al.. How Can We Use Social Media Data Related to OA Monographs. ELPUB 2020 24rd edition of the International Conference on Electronic Publishing, Apr 2020, Doha, Qatar. 10.4000/proceedings.elpub.2020.18 . hal-02544911

**HAL Id: hal-02544911**

**<https://hal.archives-ouvertes.fr/hal-02544911>**

Submitted on 16 Apr 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# How Can We Use Social Media Data Related to OA Monographs

Alkim Ozaygen, Lucy Montgomery, Cameron Neylon, Katie Wilson, Richard Hosking and Chun-Kai (Karl) Huang

---

## Introduction

- 1 Until the past decade, most studies exploring the quality, reach and impact of research outputs have focused on data relating to journal articles (Torres-Salinas et al. 2014). Relatively little work has been done to explore the extent to which altmetrics approaches are capable of producing meaningful information about the role of specialist scholarly books in the research and knowledge sharing practices of scholarly communities (Neylon et al. 2018). However, digital traces of online interactions relating to scholarly books have the potential to shed light on how people use, discuss, and share monographs, as well as the role that individuals, institutions and platforms play in this process. These traces may also make it possible to better understand the effects of OA status on patterns of digital dissemination and use for scholarly books. As the authors, publishers and funders of long-form research publications search for strategies to support successful transitions to open access for books, such information is needed.
- 2 This paper explores the extent to which social media data is capable of shedding light on the digital lives of OA scholarly books. The paper begins with a discussion of key differences between monographs and journal articles and considers the implications of these differences for efforts to identify, capture, and analyse data associated with the use of OA scholarly books. It then goes on to report on a study of data related to 28 OA monographs in the humanities and social sciences. The study captured mentions of the study-set of monographs via Twitter, Facebook, Wikipedia and online blogs; as well as user ratings on Google Books, Amazon and Goodreads. Information relating to the ways in which the books were bookmarked and cited was captured via the online reference managing platform Mendeley. Strategies used to both capture and interpret these data

are discussed, before the paper goes on to take a deeper dive into social media data associated with a single OA monograph title: examining the ways in which discussions and mentions of the book unfolded across the Twittersphere, and exploring correlations between traditional measures of use and impact, and altmetrics. The study demonstrates the potential of social media data to provide useful insights into the ways in which OA monographs are being used by communities; and provides practical lessons for researchers interested in developing robust altmetrics approaches for scholarly books.

## What is different about monographs?

- 3 Monographs and journal articles are both key forms of research publication. Both forms are used by researchers to communicate new knowledge, theories, and insights. Books and journal articles each also play an important role in research evaluation: serving as evidence of a researcher's achievements, and as a reference point for assessments of the 'impact' of scholarship. In spite of these similarities, important differences between books and journal articles also exist. These differences have implications for efforts to capture the data needed to understand the digital lives of scholarly publications in general, and books in particular. Key differences between books and journal articles are summarized below.

## Monograph Publishing is Characterised by Diversity

- 4 In contrast to journal publishing, which has been a major site of commercial consolidation over the past three decades, monograph publishing continues to be characterised by a very high level of commercial diversity. Monographs remain an area of scholarly communication that is dominated by many smaller players, rather than a few big publishers (Larivière et al. 2015). The key role of many smaller players in the monograph publishing space has created a landscape in which the transition from print to digital publishing formats has occurred later, and in less uniform ways, than has been the case for journals. Furthermore, as print sales have declined and new digital possibilities for publishing books have emerged, business model experimentation, library-based publishing, and scholar-led presses have become defining characteristics in this area of scholarly communication (Crossick 2015; Speicher et al. 2018).
- 5 In this context, there is a growing demand for new tools capable of helping publishers, authors, and research funders to understand the impact and effectiveness of different approaches to publishing specialist scholarly books. Both publishers, and authors, increasingly operate in funding landscapes in which the capacity to demonstrate the reach and impact of research is valuable. Citation and usage data relating to books are one aspect of this. Altmetrics data – particularly data relating to the reception, circulation, and use of books, also represents a potentially valuable source of information for stakeholders in OA monograph publishing.

## Readers interact with books and journal articles in different ways

- 6 There are also differences between the ways in which readers interact with books in digital landscapes, and the ways in which they interact with journal articles. These differences create potential sources of altmetrics data for books that simply do not exist for journal articles. Online reviews of books are one example of this. In contrast to journal articles, which are rarely if ever reviewed by readers on commercial retail sites, online reviews are a popular resource for potential readers of long-form publications. Reviews provide readers with insight into whether an investment of time in a book will be worthwhile. Because reading books is time consuming, reviews remain important, even when books are available in OA at no financial cost to a prospective reader. Online retailers of books- such as Goodreads, Amazon, and Google Books, allow users to rate and comment on books. As a source of altmetrics data, the availability of online reviews of books, including monographs, creates the possibility that review platforms may provide additional, potentially useful, information about relationships between community reviews of scholarly works, and downloading patterns for OA monographs (Wu and Zheng 2012).

## Monographs are available in many different formats

- 7 Another key difference between books and journal articles is the fact that books are often made available to readers in a range of different formats. Some books are available as fully downloadable PDFs, as well as in epub. Others are only made available in web view formats. Some books are made available for download as individual chapters ('chunking'). Some are only downloadable in whole-book format. Experimentation with the possibilities of digital technology for specialist scholarly books also means that 'books' may include audio, video, and animated content<sup>1</sup> raising questions about what it means to 'read' a book and how usage should be defined and measured.

## Identifiers for books can be complex

- 8 A single title, which may be made available in any combination of the formats mentioned above, may also be hosted in more than one location, and made available via a range of different dissemination platforms. This is particularly true for OA books - which can be re-hosted by third parties without permission from the publisher. Tracking books hosted in multiple locations can be particularly challenging because the digital identifiers attached to books can also vary (Montgomery et al. 2018). The best established identifier for books, the International Standard Book Number is the (ISBN). It is common for publishers to assign different ISBNs to each of the formats that the book is made available in: hardcover, softcover and electronic. Some publishers have applied the same logic to the assignment of ISBNs to different digital formats, meaning that the PDF, EPUB, MOBI, and web version may all carry different identifiers. A growing proportion of publishers are also assigning DOIs to books. Some publishers also assign DOIs to individual chapters - and no clear convention on the way that this

should be handled has yet emerged across the industry. Individual repositories also have their own approaches to assigning identifiers to content - and many assign their own unique identifier to books, or chapters, made available via their platform. As a result, a single book may have a multitude of identifiers (Neylon et al. 2018).

- 9 The range of identifiers attached to digital books makes tracking, and capturing, social media mentions and online interactions relating to monographs particularly challenging. However, in contrast to journal articles, which are generally identified with a DOI on social media, monograph readers often tweet a book's title and the author's name. They are also likely to provide a URL link to a page that contains information about a book (for example, a link to a book's marketing page on a publisher website). While there are techniques that can be used to identify tweets that mention an author or a title, complications arise when social media mentions occur in languages other than english.

## Relevance of altmetrics for monographs?

- 10 Although there is no widely accepted definition of the term, "altmetrics" is generally used to refer to metrics that are concerned with the influence of a publication on social media, using indicators of visibility and awareness such as mentions of publications (Galligan and Dyas-Correia 2013; Holmberg 2014). Altmetrics remains an emerging, and rapidly evolving, area of bibliometric analysis. Much of the scholarly literature exploring the potential, and limitations, of altmetrics has, so far, focussed on journal articles rather than books. As discussed above, important differences exist between books and journal articles, and these differences have implications for the application of altmetrics approaches to scholarly books.
- 11 Furthermore, altmetrics approaches are a response to the growing volume of potentially useful data associated with sharing and use of digital scholarly publications; however, what this data can, and cannot, be used to measure remains controversial (Bornmann 2016). To date, the majority of empirical investigations of altmetrics have focused on correlations between citations and altmetric data (Bornmann 2015; Costas et al. 2015; Torres-Salinas et al. 2013). Controversy also surrounds the extent to which altmetrics are capable of providing insight into the scholarly 'impact' of a publication. According to Sugimoto, the term "impact" is often misappropriated by the altmetrics community, because this term connotes broader engagement and a more transformative effect than is currently obtainable using altmetrics data. For example, tweets or Mendeley saves do not indicate that a monograph has had a strong effect on the user (Sugimoto 2015). Sugimoto suggests that rather than being understood as a proxy for impact, altmetrics should be used to complement existing metrics and therefore to expand the tools available to provide insight into the dissemination of research. Thus, with knowledge of the strengths and limitations of the available tools and data, it is possible to construct richer narratives of the ways in which scholarship is diffused and the impact it has on society (Konkiel et al. 2016).
- 12 Social media data sources constitute the main source altmetrics data for books. Altmetrics data sources may also include: access data (views and downloads); citation data (scholarly citations, as well as citations to a work in patents and policy documents); mentions on blogs, news sites, and Wikipedia; social media mentions (Twitter, Facebook, etc.); appraisal data on book review and rating sites; usage data

from social bookmarking and reference management services (i.e. Mendeley); and activity on annotation platforms (Hypothesis Project). Regardless of whether altmetrics measure “impact” or not, these social media data sources have the potential to offer new insights into the usage and diffusion of scholarship.

- 13 In spite of the controversy surrounding the development of critical language, frameworks and methodologies to interrogate, and interpret, altmetric data, there is a high level of demand for new insights into the ways in which scholarly publications are being used, shared and discussed online. One result of this demand is the emergence of commercial providers of altmetrics services.<sup>22</sup> Although these services previously focussed on journal articles, altmetrics services are gradually being expanded to include scholarly books. However, each altmetrics service provider uses their own limited set of altmetrics data sources. The vertical integration of content related services (Chen and Chan 2019) is also linked to a growing trend that sees commercial altmetrics data providers relying predominantly on data from their own platforms. Gaps remain in the community’s knowledge of the best strategies for capturing altmetrics data relating to scholarly books; as well as the approaches to interpreting this data that are most likely to produce useful insights.

## Methodology and datasets

- 14 The study data set comprises the 28 titles included in the first pilot collection of books made OA via Knowledge Unlatched in 2014. These titles, published by 13 scholarly publishers, are in the fields of anthropology, history, literature, media and communications, and politics.
- 15 Data relating to how often each OA book was accessed included data arising from access of the books via the OAPEN Library. Country-based download reports for the titles hosted by OAPEN were provided by Institutional Repository Usage Statistics UK (IRUS-UK)<sup>3</sup>. IRUS UK captures and cleans OAPEN usage data using the COUNTER methodology for identifying and removing bot-usage. These reports were downloaded and analysed using a script.
- 16 To analyse mentions, usage, appraisal and citations, data were also collected from five social media platforms (Twitter, Facebook, Wikipedia, Goodreads and Mendeley) and queried on three citation databases (Scopus, WoS, and Google Scholar) in which the 28 title names were mentioned. For each platform, the number of titles covered was examined.
- 17 Using the Salesforce Radian6 platform<sup>4</sup>, public mentions of the 28 titles on Facebook and Twitter were collected. The findings from Facebook were analysed according to post type for each title. The number of tweets for each title was also examined, as well as the number of Twitter users mentioning this title. The number of tweets mentioning a title over varying periods from the release of the title was analysed. In addition, the Twitter capture tool<sup>5</sup> (TCAT) software was used to track how Twitter users disseminated information about a title. TCAT software tracks Twitter mentions along with the visibility of a user by the number of mentions they receive, how many followers they have, who follows them. It captures data using Twitter's live streaming traffic. Because TCAT software cannot access Twitter's historical data to track the connections between users in the Twittersphere, a newly published OA title from UCL Press was chosen as an additional case study. Nodes disseminating information about

the title to other groups were visualized and identified using the network analysis and visualization software Gephi<sup>6</sup> 0.9.1.

- 18 Mentions of books in the study set on peer-reviewed research blogs were tracked by querying the ResearchBlogging platform. The science-only aggregator ResearchBlogging.org (Shema et al. 2014) is a platform that was in operation until April 2017, direct readers to blog posts that refer specifically to peer-reviewed research. Wikipedia articles mentioning the titles were manually queried using the Google search engine.
- 19 Reviews and ratings for each title on the Amazon, Google Books and Goodreads platforms were inspected through these sites' APIs using a script. We identified whether or not titles were sold via Amazon and Google Books. In addition, the format in which the books were provided on the Amazon and Google Books platforms was noted. Using the Mendeley platform API we determined which users of this platform had collected the titles into their libraries and how the titles were being used. To identify how these titles were being read and to determine which titles were attracting more attention, the annotation platforms Hypothes.is and PaperHive were manually examined using each title name and also by using the PDF file of each title.
- 20 Citations of the titles were also investigated using a script to connect to and query the citation databases Scopus, WoS, and Google Scholar. Subsequently, the yearly aggregated citations for each title on WoS were checked in order to determine whether patterns in usage could be identified. Later, a correlation analysis was conducted to check how these data sources were related to one another.

## Datasets

- 21 COUNTER-compliant access reports were collected from the IRUS-UK website for the period of March 2014 to June 2017. A social network dataset was compiled for the period between 1 January 2014 and 1 July 2017. This database included 493 records from Twitter that related to all of the KU pilot collection titles and 96 records from Facebook that referred overall to 20 of these titles. In addition, the UCL Press title "How the World Changed Social Media" was chosen to be the subject of a case study investigating the dissemination of a scholarly monograph across the Twittersphere. Between 1 July 2016 and 1 April 2017, 181 tweets from 103 users referring to this title were captured.
- 22 The Wikipedia dataset also included 23 articles referring to 13 monographs from the KU pilot collection. All of the 28 titles' pages were collected from Google Books, Amazon and Goodreads. The Goodreads dataset also includes 49 ratings for 16 titles. The Mendeley dataset includes 20 titles, which were bookmarked by 288 readers, whose academic status is visible. The citation data includes eight indexed titles from Scopus, 28 indexed titles from Google Scholar and 27 indexed titles from WoS. The datasets are summarized in Table 1.

Table 1: Summary of collected datasets.

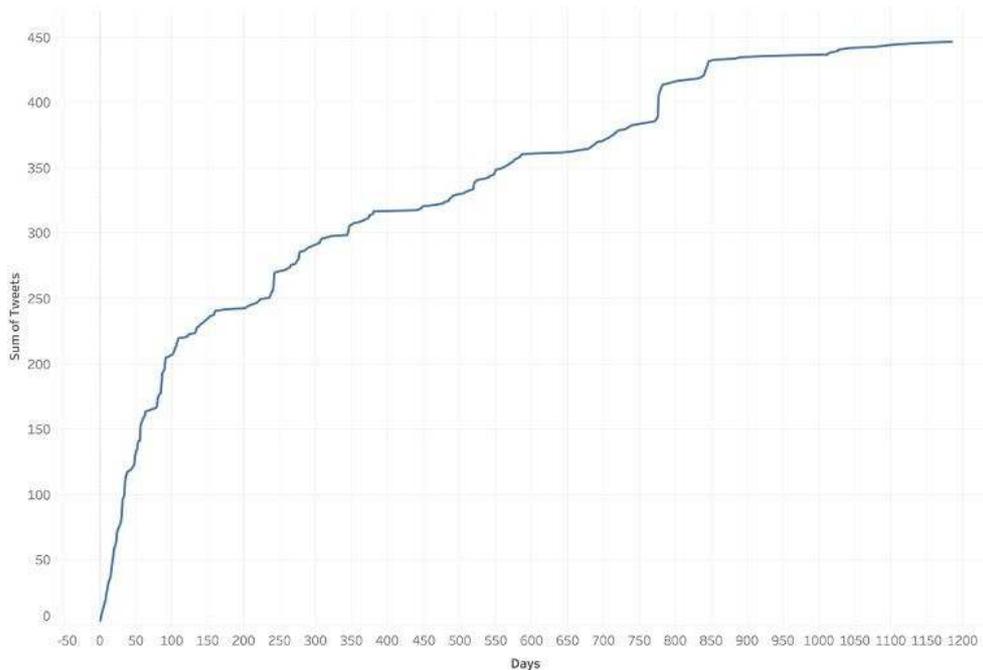
Data source	Date	Data
OAPEN repository	March 2014 – June 2017	Access data for 28 titles
Twitter	1 January 2014–1 July 2017	493 records for 28 titles
Facebook	1 January 2014–1 July 2017	96 records for 20 titles
Twitter for one UCL title	1 July 2016–1 April 2017	181 tweets for one title
Wikipedia	Until March 2018	23 articles referring to 13 monographs
Goodreads	Until March 2018	49 ratings for 16 titles
Mendeley	Until March 2018	20 titles bookmarked by 288 readers
Scopus	Until March 2018	Citation data for eight titles
Google Scholar	Until March 2018	Citation data for 28 titles
WoS	Until March 2018	Citation data for 27 titles

## Findings

### Social Networks

- 23 A total of 96 relevant Facebook posts were identified, in which 64 distinct Facebook users mentioned 21 of the KU pilot collection titles between the day the titles were uploaded to OAPEN and 22 June 2017. Using the Radian6 service, 493 tweets from 309 different authors were identified which mentioned KU pilot collection titles between the day they were uploaded and 22 June 2017. There were more tweets than Facebook posts mentioning these titles, which is in accordance with the findings of Xia and colleagues (Xia et al. 2016).
- 24 Figure 1 shows the number of tweets mentioning a title with respect to the period that had elapsed since the title was made OA. All 28 titles were made OA for the first time when uploaded to the OAPEN repository. Half of the tweets were produced in the first four months, with 35.35% produced in the first two months. In Figure 1, a plateau in tweet numbers can be seen after the first 850 days following publication. Eysenbach (2011) refers to the period from the first tweet until the plateau (the first 850 days) the “network propagation phase”, during which the new information is propagated through the Twitter network. He refers to the period following this phase as the “sporadic tweetation phase”, where mentions only occur sporadically.

Figure 1: Number of tweets mentioning the KU titles on a time scale. Most mentions occur in the first four months following publication.

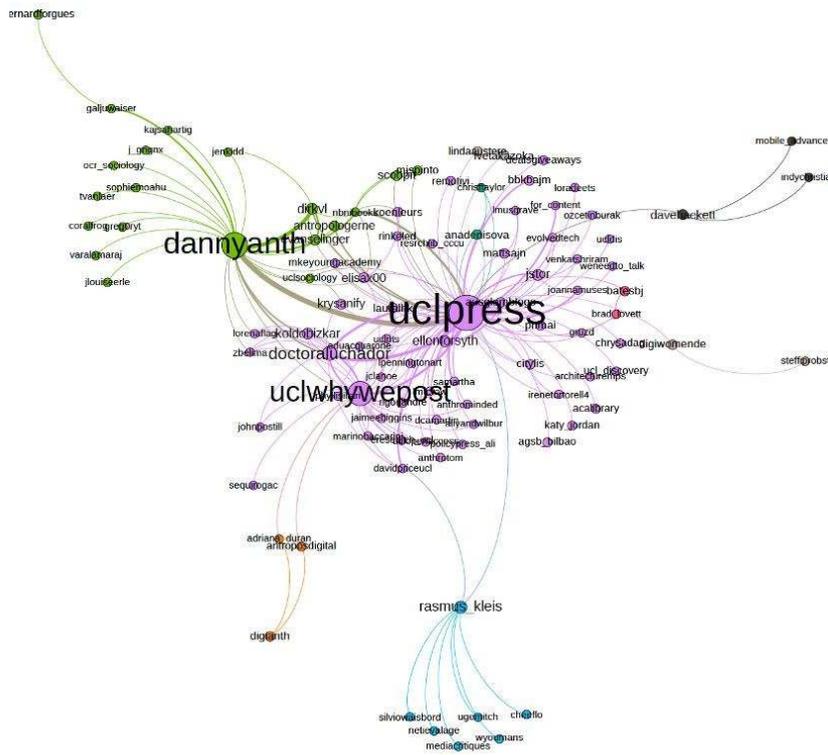


(Image source: authors)

- 25 Because this study was started two years after the KU pilot collection was made OA, it was not possible to track Twitter mentions of the titles by using the TCAT software to graph the network structure. Instead, as a case study, tweets mentioning the title “How the World Changed Social Media” were tracked for nine months (from 1 July 2016 until 1 April 2017) as part of a study involving titles from the OA monograph publisher UCL Press (Montgomery et al. 2018). Tweets for this title were captured five months after the title’s publication. The title was mentioned in 181 tweets from 103 distinct users, and 82.3% of these tweets contained links. During the network analysis of these tweets and their users, the largest connected subnetwork was investigated, which consisted of 100 nodes (or Twitter accounts) and 185 edges. Although nine different groups were identified (which were each assigned different colours, as shown in Figure 2), there were three main large groups. The biggest group is shown in purple and contains the UCL Press (‘uclpress’), ‘UCL Why We Post’ book series (‘UCLWhyWePost’), and JSTOR (‘jstor’) accounts; the green group contains the book’s author Daniel Miller’s (‘dannyauth’) account; and the blue group contains the account of Rasmus Kleis Nielsen (‘rasmus\_kleis’), a professor of communication at Oxford University.
- 26 In Figure 2, the size of a node is proportional to the node’s number of mentions, and the direction of tweets is clockwise. For example, in the green group, the user varalamaraj, situated on the far left, mentioned ‘dannyanth’ in their tweet. Thus, the direction is from user ‘varalamaraj’ to ‘dannyanth’.
- 27 The PageRank algorithm can be used to identify important nodes. The algorithm works by counting the number and quality of links to a node to determine a rough estimate of the node’s importance (Google 2011). However, important nodes do not necessarily receive the most mentions. A node can be mentioned by most mentioned nodes as well, thus having a high quantity of links. In other words, According to this algorithm, a

Twitter user may have few mentions, but some may be from popular users who themselves have many mentions. This makes the Twitter user important. Therefore these important nodes are the ‘uclpress’, ‘dannyanth’, ‘uclwhywepost’, ‘doctoraluchador’, ‘koldobizkar’, and ‘rasmus\_kleis’ Twitter accounts.

Figure 2: Twitter network graph for the title “How the World Changed Social Media” published by UCL Press.



(Image source: authors)

- 28 In addition to the use of the PageRank algorithm, a betweenness centrality analysis was conducted to identify the nodes' centrality in the network. Betweenness centrality is equal to the number of shortest paths between all nodes that pass through one node. This node can be seen as a bridging node for reaching other nodes. The bridging nodes for “How the World Changed Social Media” title are ‘uclpress’, ‘uclwhywepost’, ‘doctoraluchador’, ‘elisax00’, ‘ellenforsyth’, and ‘lauralhk’.
- 29 This shows the nodes that are responsible for pushing or receiving information from different areas of the larger network. These nodes are more engaged in information sharing. In other words, these are the Twitter accounts that are influential in the dissemination of information. Betweenness centrality analysis is helpful in identifying Twitter accounts that are key for the titles' dissemination across the Twittersphere. Therefore, the Twitter accounts that are identified using the PageRank algorithm and betweenness centrality analysis are crucial for publishers and repositories' promotion of their books.

## Scholarly Blogs, Wikipedia

- 30 Among the few altmetrics studies conducted on scholarly blogging (Priem 2010; Shema et al. 2014) only a small percentage of articles are covered on blogs. Costas, Zahedi and Wouters (2015) found 1.9% of the 500,229 articles they examined were mentioned on blogs. In this study, none of the 28 monograph titles were cited in blog posts collected from ResearchBlogging. This is likely due in part to the selective coverage of ResearchBlogging.
- 31 A total of 23 Wikipedia articles, which referred to 13 different monographs, were identified. Nine of the 23 Wikipedia articles contained no link to the PDF file of the titles. It is possible that the authors of these Wikipedia articles did not know that these books were OA.
- 32 In six Wikipedia articles, the references to the KU titles directed users to Google Books, and only one of these six articles directed users to the freely accessible content in Google Play. The remaining five articles directed users to Google Books pages, which contained only a presentation page of the title, or to Google pages where users had to pay for the content. To overcome this issue, which may prevent users from accessing OA content, Wikipedia can create a tool to provide a link to the OA content of the title once its authors enter an ISBN or DOI to refer to in their article.

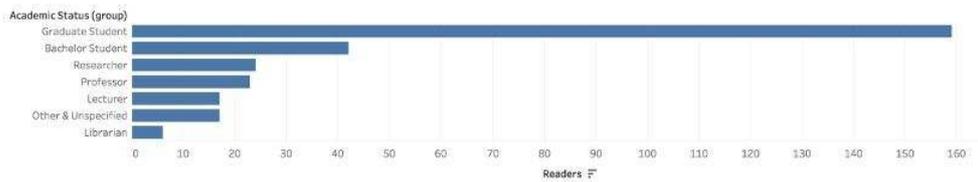
## Rating and reviewing platforms

- 33 Twenty-seven of the 28 titles were registered on Google Books, and only one book review was found on the platform. Access to the content on Google Books was also investigated. In spite of all the books being OA, only three monographs (from the same publisher) had free content access from Google Books. Further, Google Books was selling 11 titles through its Google Play platform.
- 34 Only 20 of the 28 titles' e-book versions were available on the book-selling platform Amazon, and only two of them (from the same publisher) had free access. On the other side, all of the 28 KU pilot collection titles were registered on the book review site Goodreads. Forty-nine ratings were given for 16 different titles.

## Reference manager and annotation services

- 35 On the bookmarking and reference manager platform Mendeley 28 different entries had been made for 20 titles. Twenty-six of these were for the entire book, and two were for a chapter of the book. These 28 different entries had a total of 288 readers.
- 36 Most bookmarks were made by graduate students (159), followed by undergraduate students (42). The number of researchers, professors, and lecturers bookmarking these titles was small compared to the number of students, as shown in Figure 3.

Figure 3: Academic status of Mendeley readers of the full set of KU pilot collection titles.



(Image source: authors)

- 37 The analysis using the annotation services revealed only one discussion related to one title on PaperHive and none on Hypothes.is.

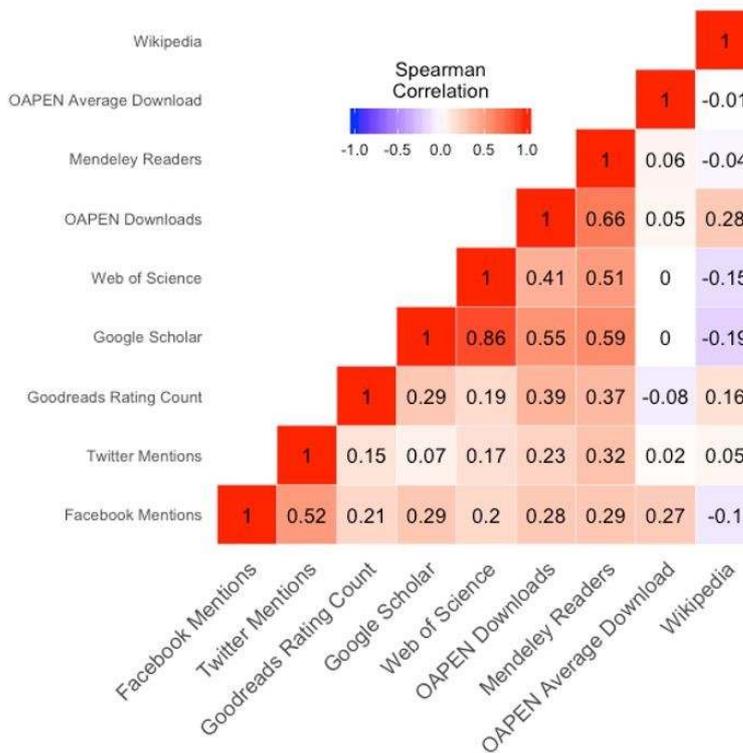
## Citation databases

- 38 Only eight titles were found to be indexed by Scopus. Six of the eight titles had citations. WoS indexed 27 titles, all having citations. Google Scholar indexed all 28 titles, and showed a higher number of citations. In the correlation analysis section Google Scholar and WoS were chosen as a proxy for the citation databases because of the number of titles they covered.

## Correlation analysis

- 39 A Spearman correlation analysis examined the relationships between the following metrics: total downloads of the titles from the OAPEN platform, number of domains mentioning these titles, number of citations from WoS and Google Scholar, number of citations on Wikipedia, number of mentions on Twitter and Facebook, number of bookmarks on Mendeley, and number of ratings on Goodreads. The results of this correlation analysis are shown in Figure 4. In the figure, negative correlations are in mauve and positive correlations in red.

Figure 4: Heatmap showing the correlation between numbers of events from the different data sources.



(Image source: authors)

- 40 Figure 4 shows that the highest correlation (0.86) is between Google Scholar citations and WoS citations. The second highest correlation is between Mendeley bookmarks and OAPEN downloads (0.66), and the third highest between Mendeley bookmarks and Google Scholar citations (0.59). The strong correlation between Mendeley bookmarks and citations from citation databases is not surprising, since researchers, in particular graduate students, bookmark titles in order to cite them. Mendeley bookmarks are also moderately correlated with the number of OAPEN downloads, as well as with the number of domains mentioning a title and socialnetworks mentions. The average monthly downloads from the OAPEN repository and the number of citations on Wikipedia are not strongly correlated with other metrics.
- 41 The following correlations had p-values above 0.05: OAPEN average downloads, Twitter, Wikipedia, and Facebook with all data sources. The only exception for Facebook was the correlation between Facebook mentions and Mendeley bookmarks, which had a correlation coefficient of 0.29 with  $p = 0.03$ . Only p-values below 0.05 were accepted as significant for the correlations between Mendeley bookmarks, Goodreads ratings, OAPEN downloads, number of domains mentioning titles, and citation numbers from Google Scholar and WoS. In summary, strong correlations are observed between citations on Google Scholar, citations on WoS, Mendeley bookmarks and OAPEN downloads. In addition, OAPEN downloads correlated with the number of domains mentioning titles, and number of Goodreads ratings.
- 42 Among the book review and rating sites Goodreads had the most ratings of the titles (49), covering 16 books. These data displayed a weak correlation with citation databases (0.19 with WoS and 0.29 with Google Scholar, both having  $p < 0.01$ ) and the Mendeley

data (0.37,  $p < 0.01$ ). A similar correlation was obtained by Zuccala and colleagues (0.212,  $p < 0.01$ ) between citation counts and Goodreads rating counts for history books (Zuccala et al. 2015).

- 43 In Mendeley entries were made for 20 titles. Most bookmarks for these titles were made by graduate students (159) which is in accordance with Haustein and Larivière's work (2014) on readership counts on Mendeley, where they examined 1.2 million documents published in journals from four disciplines (biomedical research, clinical medicine, health and psychology). They found that approximately two thirds of these documents were bookmarked by at least one user on Mendeley. The majority of these users were PhD students, postgraduate students and postdoctoral researchers.
- 44 The citation database Google Scholar and Web of Science showed high coverage of the 28 titles, and high coverage with each other. This indicates that Google Scholar, as a free citation database, is a viable data source in monograph research. This study also showed a significant correlation between Mendeley and citation databases (0.59 with Google Scholar and 0.51 with WoS, both having  $p < 0.01$ ). These findings are in accordance with a study conducted by
- 45 Bar-Ilan and colleagues, where they sampled 1,136 unique papers authored by 57 presenters who attended the 2010 Leiden Science and Technology Indicators (STI) conference (Bar-Ilan et al. 2012). Of these 1,136 documents indexed in Scopus, they found a significant correlation ( $r=0.45$ ) between an article's number of bookmarks in Mendeley and number of citations in Scopus. In another study, which involved 1,613 papers published in *Nature* and *Science* in 2007 (Li et al. 2012), positive correlations ( $r=0.60$  and  $r=0.54$ , respectively) were also found between the articles' bookmark counts in Mendeley and their citation counts in WoS.
- 46 In summary, Goodreads presents a good proxy for downloads, where we observed the number of comments on this platform affecting the number of access to books. This study also showed that Google Scholar had the best coverage for citation analysis. As with journal articles Mendeley platform is a good proxy candidate for citations.

## Conclusion

- 47 This study demonstrates the potential for altmetrics approaches to provide useful insights into the ways in which OA monographs are being used by communities. However, in order to access these benefits, altmetric studies of books need to engage with book specific platforms; adopt a title and author name based, rather than identifier or URL based, approach to searching for social media mentions related to titles.
- 48 Currently, most social media platforms make their usage data publicly available. Since these platforms use either DOI or ISBN as identifiers, there is a need to have both identifiers to collect data. Crossref metadata API (<http://api.crossref.org/>) can be used to find and convert ISBN and DOI identifiers. Using the DOI identifier it is possible to collect social media and citation data from Crossref or altmetric.com platforms. Open source tools are also available to download data from these platforms, such as the tools provided by rOpenSci initiative (<https://ropensci.org/>).
- 49 A key challenge in engaging with repository and platform data related to inconsistent approaches to classifying the subject of individual titles. Subject classification

approaches range from the very general to the very specific, making comparisons between platforms difficult. Using OCLC<sup>7</sup> subject classifications provides a practical strategy for overcoming this issue. OCLC provides Dewey Decimal Classification (DDC) and Library of Congress Classification (LCC) numbers for books identified via ISBN. The platform offers these classifications from their Classify API service<sup>8</sup> based on bibliographic records they collect from libraries.

- 50 Finally, although it is not possible to generalize findings because of the small sample we used, this study serves as a practical demonstration of strategies that can be used to successfully identify, track and analyse altmetric data related to books from social media and citation databases. The correlation analysis shows the extent to which these data sources are related.
- 51 Since each data source captures a different kind of interaction with an OA monograph, further work is needed in order to develop a sophisticated, critical understanding of the digital journeys taken by OA monographs; as well as how and why individual interactions with books occur (Haustein et al. 2015).

---

## BIBLIOGRAPHY

- Bar-Ilan, Judit, et al. "Beyond Citations: Scholars' Visibility on the Social Web." *ArXiv:1205.5611*, May 2012. *arXiv.org*, [arxiv.org/abs/1205.5611](https://arxiv.org/abs/1205.5611).
- Bornmann, Lutz. "Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics." *Scientometrics* 103, no. 3, 2015: pp. 1123-1144.
- Bornmann, Lutz. "Scientific revolution in scientometrics: the broadening of impact from citation to societal." *Theories of informetrics and scholarly communication*, 2016: pp. 347-359.
- CHEN, George, & Leslie CHAN. "Vertical Integration in Academic Publishing: Implications for Knowledge Inequality." Essay. In *From Projects to Sustainable Infrastructure: The 22<sup>nd</sup> International Conference on Electronic Publishing - Revised Selected Papers [Online]*, edited by Alejandro POSADA. Marseille: OpenEdition Press, 2019. DOI:10.4000/books.oep.9068.
- Costas, Rodrigo, Zohreh Zahedi, & Paul Wouters. "Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective." *Journal of the Association for Information Science and Technology* 66, no. 10, 2015: pp. 2003-2019.
- Cronin, Blaise. "The need for a theory of citing." *Journal of documentation* 37, no. 1, 1981: pp. 16-24.
- Crossick, Geoffrey. *Monographs and Open Access. A report to HEFCE*, 2015.
- Eysenbach, Gunther. "Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact." *Journal of medical Internet research* 13, no. 4, 2011: pp. e123.
- Galligan, Finbar, & Sharon Dyas-Correia. "Altmetrics: Rethinking the way we measure." *Serials review* 39, no. 1, 2013: pp. 56-61.

- Google. "Facts about Google and competition", 2011. [web.archive.org/web/20111104131332/https://www.google.com/competition/howgooglesearchworks.html](http://web.archive.org/web/20111104131332/https://www.google.com/competition/howgooglesearchworks.html).
- Haustein, Stefanie, & Vincent Larivière. "Mendeley as a source of readership by students and postdocs? Evaluating article usage by academic status", 2014.
- Haustein, Stefanie, et al. "Interpreting 'Altmetrics': Viewing Acts on Social Media through the Lens of Citation and Social Theories." *ArXiv:1502.05701*, Feb. 2015. [arXiv.org, arxiv.org/abs/1502.05701](http://arxiv.org/abs/1502.05701).
- HIRMEOS. "Metrics Services Specification". Accessed March 17, 2020. [www.hirmeos.eu/wp-content/uploads/2017/11/HI61-Metrics\\_Service\\_technical\\_specification-final.pdf](http://www.hirmeos.eu/wp-content/uploads/2017/11/HI61-Metrics_Service_technical_specification-final.pdf)
- Holmberg, Kim. "The meaning of altmetrics", *Proceedings of the IATUL Conferences*, Paper 1, 2014, [docs.lib.purdue.edu/iatul/2014/altmetrics/1](http://docs.lib.purdue.edu/iatul/2014/altmetrics/1).
- Konkiel, Stacy, Cassidy R. Sugimoto, & Sierra Williams. "What constitutes valuable scholarship? The use of altmetrics in promotion and tenure." *Impact of Social Sciences Blog*, 2016.
- Larivière, Vincent, Stefanie Haustein, & Philippe Mongeon. "The oligopoly of academic publishers in the digital era." *PLoS one* 10, no. 6, 2015: pp. e0127502.
- Li, Xuemei, et al. "Validating Online Reference Managers for Scholarly Impact Measurement." *Scientometrics*, vol. 91, no. 2, May 2012, pp. 461–71. doi:10.1007/s11192-011-0580-x.
- Mohammadi, Ehsan, et al. "Can Mendeley Bookmarks Reflect Readership? A Survey of User Motivations." *Journal of the Association for Information Science and Technology*, vol. 67, no. 5, May 2016, pp. 1198–209. doi:10.1002/asi.23477.
- Montgomery, Lucy, Cameron Neylon, Alkim Ozaygen, & Tama Leaver. "Getting the best out of data for open access monograph presses: A case study of UCL Press." *Learned Publishing* 31, no. 4, 2018: pp. 335-344.
- Montgomery, Lucy, Cameron Neylon, Alkim Ozaygen, Frances Pinter, & Neil Saunders. "The Visibility of Open Access Monographs in a European Context: A Report Prepared by Knowledge Unlatched Research", 2018. doi.org/10.17613/M6156F
- Priem, J. "I like the term# articlelevelmetrics, but it fails to imply\* diversity\* of measures. Lately, I'm liking# altmetrics." *Twitter*, September 2, 2010.
- Shema, Hadas, Judith Bar-Ilan, & Mike Thelwall. "Scholarly blogs are a promising altmetric source." *Research Trends* 37, 2014: pp. 11-13.
- Speicher, Lara, Lorenzo Armando, Margo Bargheer, M. Eve, Sven Fund, Delfim Leão, Max Mosterd, Frances Pinter, & Irakleitos Souyioultzoglou. "Operas open access business models white paper." *Zenodo*, July 30, 2018.
- Sugimoto, Cassidy. "'Attention is not impact' and other challenges for altmetrics." *Discover the future of research: Wiley Exchanges*, 2015.
- Sugimoto, Cassidy R., et al. "Scholarly Use of Social Media and Altmetrics: A Review of the Literature." *Journal of the Association for Information Science and Technology*, vol. 68, no. 9, Sept. 2017, pp. 2037–62. doi:10.1002/asi.23833.
- Torres-Salinas, Daniel, et al. "Altmetrics: New Indicators for Scientific Communication in Web 2.0." *Comunicar*, vol. 21, no. 41, June 2013, pp. 53–60. [arXiv.org](http://arxiv.org), doi:10.3916/C41-2013-05.

Torres-Salinas, Daniel, et al. "Coverage, Field Specialisation and the Impact of Scientific Publishers Indexed in the Book Citation Index." *Online Information Review*, vol. 38, no. 1, Jan. 2014, pp. 24–42. doi:10.1108/OIR-10-2012-0169.

Weissmann, Jordan. "The simple reason why Goodreads is so valuable to Amazon." *The Atlantic*, 2013, [theatlantic.com/business/archive/2013/04/the-simple-reason-why-goodreads-is-so-valuable-to-amazon/274548/](http://theatlantic.com/business/archive/2013/04/the-simple-reason-why-goodreads-is-so-valuable-to-amazon/274548/).

Wu, Weifang, & Rong Zheng. "The impact of word-of-mouth on book sales: review, blog or tweet?" *Proceedings of the 14th Annual International Conference on Electronic Commerce*, 2012: pp. 74-75.

Xia, Feng, et al. "Bibliographic Analysis of Nature Based on Twitter and Facebook Altmetrics Data." *PLOS ONE*, edited by Wen-Bo Du, vol. 11, no. 12, Dec. 2016. doi:10.1371/journal.pone.0165997.

Zahedi, Zohreh, et al. "How Well Developed Are Altmetrics? A Cross-Disciplinary Analysis of the Presence of 'Alternative Metrics' in Scientific Publications." *Scientometrics*, vol. 101, no. 2, Nov. 2014, pp. 1491–513. doi:10.1007/s11192-014-1264-0.

Zuccala, Alesia A., et al. "Altmetrics for the Humanities: Comparing Goodreads Reader Ratings with Citations to History Books." *Aslib Journal of Information Management*, edited by Dr Stefanie Haustein, Dr Cassidy R. Su, vol. 67, no. 3, May 2015, pp. 320–36. doi:10.1108/AJIM-11-2014-0152.

## NOTES

1. See, for example, the University of Michigan Press's Fulcrum initiative: <https://www.fulcrum.org/>.
2. Examples include Altmetric: <https://www.altmetric.com/>; and Plum Analytics: <https://plumanalytics.com/>.
3. <https://irus.jisc.ac.uk/>.
4. Radian6 is a social media monitoring platform designed to help marketing professionals study customers' opinions of their products in real-time.
5. <https://github.com/digitalmethodsinitiative/dmi-tcat/>.
6. <https://gephi.org/>.
7. OCLC is a global library cooperative that provides shared technology services, original research and community programs for its membership and the library community at large.
8. [http://classify.oclc.org/classify2/api\\_docs/index.html](http://classify.oclc.org/classify2/api_docs/index.html).

---

## ABSTRACTS

This paper reports on a study of social media events relating to 28 Open Access (OA) monographs, published between 2014 and 2015. As with citations (Cronin 1981) social media events represent the frozen footprints of the journey that monographs take as they move through digital landscapes. The study captured mentions of the study-set of monographs via Twitter, Facebook, Wikipedia and online blogs; as well as user ratings on Google Books, Amazon and Goodreads.

Information relating to the ways in which the books were bookmarked and cited was captured via the online reference managing platform Mendeley. The benefits and limitations of different altmetrics approaches to capturing and analyzing this data are discussed. Practical suggestions for researchers interested in the application of Altmetrics approaches to studies of monographs are also provided.

## INDEX

**Keywords:** open access, monographs, books, altmetrics

## AUTHORS

### **ALKIM OZAYGEN**

Curtin University, Australia

E-Mail: [Alkim.ozaygen@curtin.edu.au](mailto:Alkim.ozaygen@curtin.edu.au).

(corresponding author)

### **LUCY MONTGOMERY**

Curtin University, Australia

### **CAMERON NEYLON**

Curtin University, Australia

### **KATIE WILSON**

Curtin University, Australia

### **RICHARD HOSKING**

Curtin University, Australia

### **CHUN-KAI (KARL) HUANG**

Curtin University, Australia