

Getting the best out of data for small monograph presses: A case study of UCL press¹

Authors: Lucy Montgomery, Cameron Neylon, Alkim Ozaygen and Tama Leaver

Abstract

Digital developments in scholarly publishing are giving rise to new data sources with the potential to provide insight into how OA monographs are being used and to support strategic decision-making by publishers. However, small OA monograph publishers face practical challenges in identifying relevant data, as well as in capturing, managing and interpreting it. This case study of UCL Press reports on a collaborative research project that sought to address some of these challenges. The project involved UCL Press, Knowledge Unlatched Research and the Centre for Culture and Technology at Curtin University in Australia (CCAT). Our goal was to identify the extent to which data that can be easily accessed by a small OA monograph press can be combined with low-cost tools for its analysis in order to provide useful insight into development and strategy; and to identify practical steps that can be taken by small OA monograph publishers to ensure that they are making the most of the data that they have access to.

Key Words

Open Access Monographs; Analytics; Business intelligence; Library based publishing.

Introduction

In contrast to the natural and medical sciences, which are dominated by a handful of large publishing houses (Lariviere, 2015), HSS publishing is characterized by a lack of concentration and the key role that small and medium-sized publishers play in meeting the research and communication needs of HSS communities (Prainsack, 2013). The emergence of a diverse array of born-digital OA monograph publishers therefore continues a broader trend of diversity across the Humanities and Social Sciences (HSS) publishing sector. As demand for innovative publishing services continues to increase and digital technology lowers barriers to experimentation and new entrants in the publishing space, the number of born-digital OA monograph presses is growing. Many are based within University libraries: the 2016 Library Publishing Directory lists 118 library-based publishing programs, located in the US, Canada,

¹ The authors gratefully acknowledge feedback funding for this project provided by UCL Press. Support and feedback provided by Dr Frances Pinter and Lara Speicher has been invaluable.

Brazil, the UK, Germany, and Australia as involved in the publication of original works by scholars, researchers and students, including 905 monographs in 2015-16 (Library Publishing Coalition, 2017). Independent, scholar-led operations, such as the Cambridge-based Open Book Publishers² and Israel-based Rounded Globe Publishing³ are also engaging actively with the needs of HSS research communities in order to push the open agenda forward for specialist scholarly books.

The involvement of many small mission-focussed players in the production of the core-outputs of the Humanities and Social Sciences is a strength for HSS research communities as they navigate complex processes of innovation and change associated with open and networked digital technologies. However, a consequence of this diversity in HSS monograph publishing is often a lack of resources within each publisher. Providing high quality publishing services for long-form writing is costly (Maron et al 2016). Funding sustainability is a perpetual problem for many HSS monograph publishers, particularly in the context of sharp declines in monograph sales over the past thirty-years (Jubb 2017). While it is clear that sales data is an insufficient measure for the value and performance of presses publishing freely accessible books (or of the books themselves), questions of what data is relevant, and how it should be captured, managed and interpreted, remain complex. Usage data and social media analytics can provide information not just about the volume of a book's use, but also about the audiences who are engaging with it, how and why. However, the ability to fully exploit the data available has generally been restricted to large publishers or content aggregators with the resources to invest in its collection, management and analysis.

Nonetheless, as this case study demonstrates, these new data sources have the potential support strategic decision-making by small OA monograph publishers. Furthermore, smaller publishers generally have access to rich data about the books that they publish and free and low-cost tools are helping to lower the barriers to its collection and analysis. This collaborative research project set out to engage with data that is readily available to a small OA monograph press and to explore the extent to which low-cost tools for data capture and analysis could provide useful insight into strategic decision making. Carried out between May and September 2016 and funded by UCL Press, the project involved researchers from Knowledge Unlatched Research and the Centre for Culture and Technology at Curtin University in Australia (CCAT).

The information challenges for a small press

A small press publishing Open Access monographs may have many different motivations and needs with respect to analysing and using data of the types described in this case study. Questions that a publisher seeks to answer may range from tactical issue of whether to invest time in social media promotion or hosting OA content on multiple platforms, to large-scale strategic questions of the balance of a title portfolio and its evolution. We divide these diverse questions into three broad categories:

² See: <https://www.openbookpublishers.com/>

³ See: <http://roundedglobe.github.io/diy-academic-publishing/>

1. *How is a specific book performing?* What is the audience for a book? How is the book used? How does this compare to other titles published by the press? To comparable titles from other presses?
2. *What promotion strategies are effective?* How effective are specific efforts by the press, authors, or other stakeholders to promote both specific books and the publisher's list overall. Are successful strategies consistent or do they differ substantially from book to book?
3. *How is the publisher delivering on its mission?* Are target audiences being reached? Is the content mix right? Is the institutional home being well served?

Usage data (downloads and page views), combined with social media analysis, have the potential to address these questions. Alongside simple measures of volume there is the possibility for demographic analysis. It is also possible to examine the effect of specific interventions such as promotional events, social and traditional media campaigns or the inclusion of books within online learning offerings. In this study we explored the extent to which data readily available to UCL Press could be combined with inexpensive or freely available tools in order to address these questions.

UCL Press and the Case Study

UCL Press launched in June 2015 as the UK's first fully open access (OA) university press (Ayrís & Speicher, 2015). The establishment of the Press, which sits within UCL Library Services, coincided with a significant drive at UCL to make research publications OA: reflecting the University's commitment to maintaining its position as a leader in Open Scholarship, as well as its desire to help shape the future of scholarly publishing. By investing in a fully OA University Press, UCL joined a growing number of Universities actively financing the provision of innovative publishing services needed to support the global open knowledge agenda. As UCL Press observes, such services have an important role to play in ensuring that research is able to tackle 'global Grand Challenges such as poverty, disease, hunger' (UCL Press, 2017).

Access to detailed usage data is an important benefit of University-led OA publishing initiatives. Digital distribution is making it possible to understand the processes, audiences and relationships involved in scholarly communication in new ways. At the same time, the ability to tell rich stories about how research is shared across global knowledge communities is becoming ever more important. This is particularly true in the context of a lean funding environment and the growing expectation that Universities can and should be able to demonstrate the impact of public investments in research (Allen, 2013). In May 2016 UCL Press invited Knowledge Unlatched Research and the Centre for Culture and Technology at Curtin University to engage with it in a collaborative research project, exploring the extent to which readily available data could shed light on how and why a global community of readers were engaging with UCL Press books. An additional goal of the project was to explore practical strategies for capturing and interpreting data arising from OA monographs.

Methodology

In late May 2016 UCL Press made the following data available to the project team for the purposes of the study:

- All of the books published by UCL Press. This was a total of 11 titles in May 2016. In August 2016 additional data was provided for a further 3 books;
- Details of print sales until the end of August 2016.
- Download statistics as collected by UCL Press from all known hosting locations;
- Details of key marketing and social media campaigns;
- Press cuttings, book reviews, and online articles by or about UCL Press;
- Google Analytics access for the UCL Press marketing website.

The press also introduced the project team to selected UCL Press authors, so that additional information about author-perceptions of the benefits of OA publishing, social media activity and the use of UCL Press titles within MOOC and online learning programs could be gathered.

The Knowledge Unlatched/CCAT team separately contacted The OAPEN Digital Library in order to request access to detailed usage logs for UCL Press titles.

The project team worked to integrate usage data into an interactive visualization dashboard using Tableau. The integration of usage data into an interactive visualization dashboard made it possible to explore trends in usage, make comparisons between download patterns across different platforms, and identify correlations between key dissemination activities and use of the books.

1.1 Repositories

During the study-period UCL Press monographs were hosted in the following locations:

- **The OAPEN Digital Library** a Netherlands-based foundation that provides hosting and preservation services for peer reviewed scholarly monographs.
- **UCL Discovery** UCL Discovery is UCL's open access repository, showcasing and providing access to UCL research publications.
- **The Internet Archive** A San-Francisco based nonprofit digital library that provides free public access to collections of digitized materials, including books.
- **Unglue.it** A crowd-funding initiative for open access Books.

Preliminary analysis revealed that the vast majority of usage was occurring via just three platforms: the UCL Press repository, The OAPEN Digital Library, and the Internet Archive. We therefore focussed our analysis on data from these three sites.

The OAPEN Digital Library, which hosts more than 3000 monographs, makes usage data available in COUNTER format for member publishers on a quarterly basis. In addition to these reports, the project team were also granted access to data sets relating to the larger OAPEN

collection, for the purposes of interrogating the relative performance of UCL Press titles against similar cohorts of English-language monographs. Other sources of publically available data, including download averages and geolocation-filtered usage for books made OA via Knowledge Unlatched, also provided comparative data against which the use of UCL Press titles could be measured. Access to OAPEN usage logs for UCL press titles made it possible for the project team to apply geolocation filters to this data.

UCL Discovery is UCL's open access repository and runs on an EPrints system. Download data relating to individual items hosted within UCL Discovery is routinely made available via a publically accessible dashboard. Because it was not possible to obtain direct access to UCL Discovery usage logs, we relied on the repository's publically available download data for the purposes of our project. At the time of this study publically available UCL Discovery data was not available in the COUNTER format. As a result, UCL repository data is not directly comparable to OAPEN data. UCL Discovery usage figures are made available for retrospective 12-month periods only. An aggregate number of downloads according to country for the 'last 12 months' is also provided. The format of the publically available UCL repository data makes it impossible to identify seasonal changes in download activity according to country. Low time-resolution for the country-specific download data also makes it difficult to correlate changes in country-specific activity with changes in the ways in which titles are being promoted and made available.

The Internet Archive does not provide access to detailed usage data. Rather, it publishes total download figures relating to individual items. UCL staff collate these figures within an excel workbook on a monthly basis. In the absence of more detailed information about usage via the Internet Archive, we used these manually collected records for the purposes of our analysis.

Two important limitations exist for this readily available usage data. The first is that because usage data is gathered and reported according to different standards, direct comparisons cannot be made across repositories. Although it was possible to identify general patterns in usage that were consistent with the promotional strategy of UCL press in this study, a lack of standardisation in platform specific data makes it impossible to infer the reasons behind small differences in download figures across platforms. The second limitation is time resolution: much of the data were often only available on a monthly basis. This means that very rapid events, such as a moderate spike in downloads following social media promotion or the publication of a review, may not always be discernible.

1.2 Google Analytics

Google analytics relating to the UCL Press marketing website provided an additional source of data for the project. Unique page view data and referrer data for pages devoted to marketing individual UCL Press titles proved especially valuable. Discrepancies between the numbers of users downloading the books and numbers of unique visitors to the UCL press site provided insight into the role of direct download links (for example, links to PDF downloads provided in online learning materials and MOOCs) in usage patterns. Google Analytics also provides information on referrers, that is the website from which a user arrived at UCL press. This data

added an important layer of information to our analysis of access and discovery pathways for UCL Press titles.

1.3 Key Dates

UCL Press provided us with comprehensive information about key marketing events relating to each title. This included the Press's media logs, detailing press coverage relating to the titles that it published; book reviews; the use of UCL Press titles in online learning and MOOCs. The project team plotted these key dates and events against download activity and sales in Tableau, in order to identify relationships between the volume of usage and efforts to promote titles.

1.4 Social Media Data

UCL Press provided the details of their social media accounts, and information about specific social media posts relating to the focal examples. UCL Press also provided the details of social media accounts managed by the authors of books examined. Using an open source social media analytics tool, the entire activity of the UCL accounts, and of key authors, in the time-range specified was extracted and mapped against the book download data. While the social media data thus included exact timestamps, the lack of hourly or even daily granularity in download data impeded attempts to reveal the correlations between social media activity and book downloads.

Results

We compared download figures for each book on the three platforms. Download figures for each title via individual hosting platforms are shown in Figure 1. We do not aggregate download figures from the three repositories because these platforms do not calculate download figures in the same way. We show average downloads per month over the lifetime of the book on each repository as the titles differ in age and each title is not necessarily loaded to all three repositories on the same date. Four of the top five most downloaded books broadly categorised as HSS monographs were part of the *Why We Post* series: a series of books arising from a major European Research Council funded digital anthropology project, led by UCL Professor Daniel Miller.⁴

⁴ <http://www.ucl.ac.uk/why-we-post>

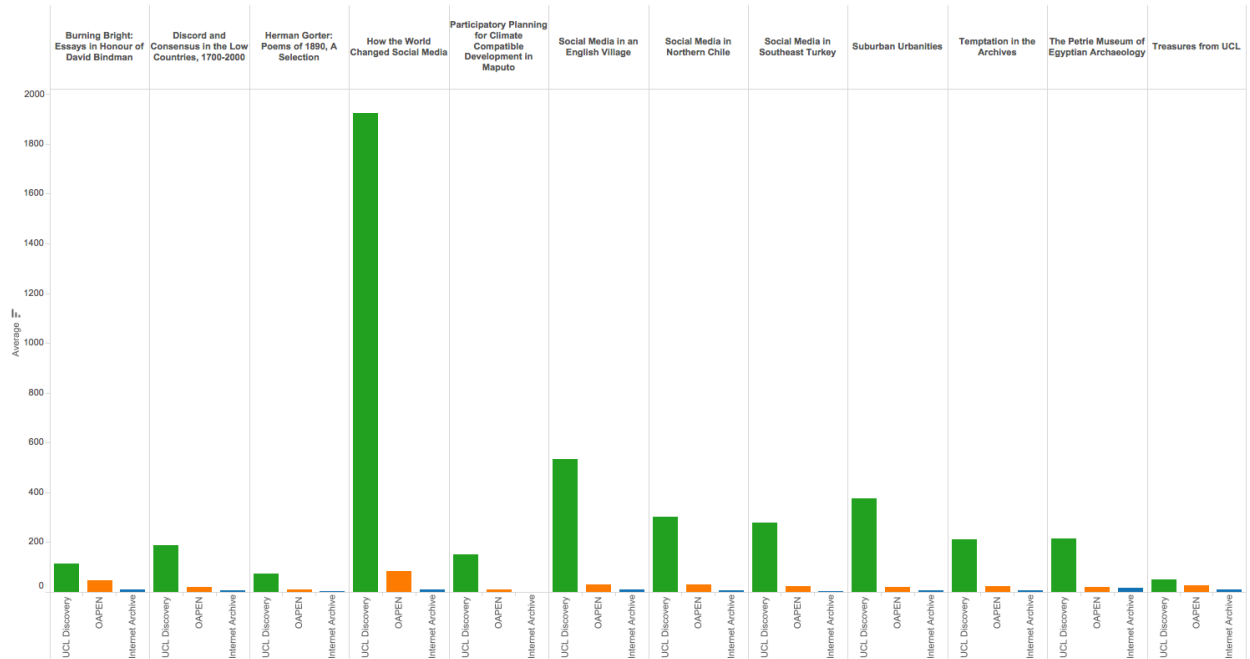


Figure 1: Monthly average downloads over the lifetime of each title on three repositories. The UCL repository shows higher rates for all titles. Three titles from the 'Why We Post Series' show substantially higher download activity, especially 'How the World Changed Social Media'.

UCL Press provided a list of significant events related to the promotion of their titles. In order to see if these events had an effect on usage for these titles, we created timeline graphs, plotting key events against titles downloads. In order to see the real effect of these events, it was necessary to cross-reference data with referring channels. Figure 2 shows one very large spike in activity correlated with an event: the launch of the UCL Massive Open Online Course (MOOC) 'Why We Post: The Anthropology of Social Media' program was associated with a significant increase in downloads of the title *How the World Changed Social Media* as well as two other titles associated with the *Why We Post* series. An increase in downloads for *Temptation in the Archives* could potentially be associated with reviews appearing in two important venues (THES⁵ and Bookseller⁶). However, the time resolution of the data limits drawing firm conclusions from this. Similarly a marked increase in downloads can be seen for *Participatory Planning for Climate Compatible Development in Maputo* in the same month that a review of the book appeared in *Environment and Urbanisation*.⁷ A higher time resolution of data would improve the confidence that could be given to conclusions drawn.

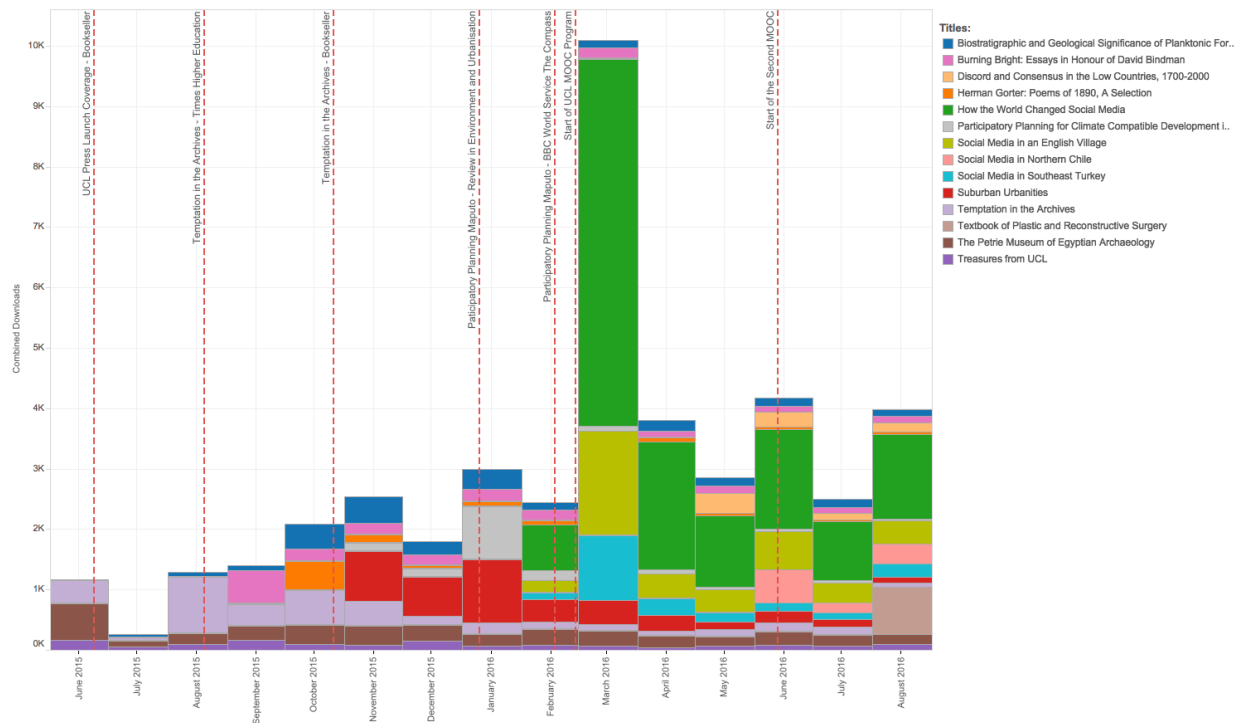


Figure 2: Key events timeline graph shows that after the start of UCL MOOC program, downloads for 'How the World Changed Social Media' title increase.

⁵ <https://www.timeshighereducation.com/books/review-temptation-in-the-archives-lisa-jardine>

⁶ <http://www.thebookseller.com/news/deaths-lisa-jardine-and-david-cesarani-315188>

⁷ <https://www.environmentandurbanization.org/participatory-planning-climate-compatible-development-maputo-mozambiqueplaneamento-participativo>

Downloads of *How the World Changed Social Media* peak after the launch of the MOOC program connected to the book. To further understand the reasons behind this peak we examined the top traffic mediums for the title’s download via the UCL discovery platform, where the majority of downloads occurred. The number of downloads for the period June 2015 and July 2016 was 13,476. This is significantly higher than the number of unique views for the *How the World Changed Social Media* page on the UCL Press website, which was only 8,396. This means that a majority of readers were using a direct download link for to access the PDF, bypassing the title's web page. This is most likely to have resulted from the launch of the *How the World Changed Social Media* MOOC, which provided a link directly to the PDF file in the UCL Discovery repository. A direct link to the PDF was also provided in the MOOC’s Week 1 module. High MOOC attrition rates over subsequent weeks may also help to explain why download rates for this title are higher than other titles in the same series, which were referred to in later weeks within the MOOC. MOOC participants are typically high at the beginning of a course, and decline as the course progresses. It is worth noting that the data cannot directly confirm that the MOOC is the source as direct links may be circulating on the web. If the MOOC materials had used a specially tagged link it would have been possible to disambiguate the source and confirm the discovery route more directly.

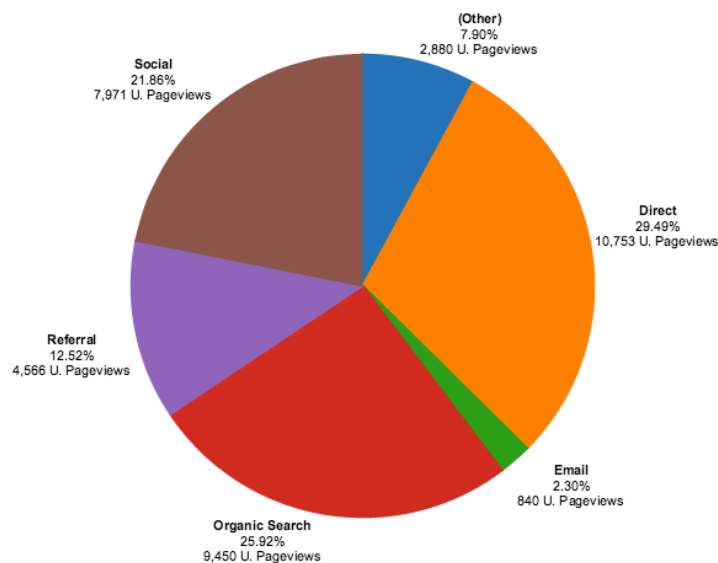


Figure 3: Referral channels for “How The World Changed Social Media” title page shows that Direct Links account for a high proportion of use during the MOOC period.

Source	Views Referred by Source	Proportion (%)
(direct)	3148	32.62%
google	2369	24.54%
h-announce	978	10.13%
air-l	493	5.11%
t.co	454	4.70%
facebook.com	378	3.92%
blogs.ucl.ac.uk	166	1.72%
m.facebook.com	149	1.54%
futurelearn.com	139	1.44%
UCL Press website	134	1.39%
Total Pageviews	9652	100.00%

Table 1: Top ten sources for the title "How the World Changed Social Media". Email lists play an important role in disseminating the information on "How the World Changed Social Media" title.

Further analysis of traffic sources to the UCL Press web page for *How the World Changed Social Media* reveals that email lists also played an important role in the dissemination of the title among researchers. H-announce, a mailing list from an interdisciplinary forum for scholars in the humanities and social sciences, generated 978 page views. This is the third-largest source of traffic, after direct links and the Google search engine. 493 page views are attributable to Air-L: an Association of Internet Researchers mailing list. Referrals from Facebook's web and mobile instances provided a combined 527 page views, while Twitter (using their URL shortener t.co) generated 454 views; social media thus generated 10% of the total referrals. Links to the book's page published in *The Economist* and *Wired Magazine* generated 85 and 64 unique page views respectively. The UCL MOOC program course introduction page on Future Learn was the source of only 65 page views of the book landing pages. It is likely that many of the direct links came from within the related MOOC. It is important to note that this data does not capture links to the PDFs .

Whole Repository (views)	Social Media in Southeast Turkey (views)	Participatory Planning...in Mozambique (views)
United Kingdom (8760)	Turkey (348)	United Kingdom (268)
United States (6863)	United Kingdom (345)	United States (166)

Germany (1480)	United States (247)	Germany (73)
Brazil (1370)	Italy (77)	India (55)
China (951)	Brazil (75)	Canada (48)
Italy (937)	Germany (74)	South Africa (42)
Australia (883)	Australia (49)	Portugal (35)
France (814)	Nigeria (47)	Mozambique (34)
Netherlands (725)	France (40)	Italy (30)
Turkey (655)	Netherlands (35)	Australia (30)
Canada (641)	India (34)	Kenya (29)
India (534)	Canada (34)	Netherlands (24)

Table 2: Listing of countries ranked by usage for the UCL repository as a whole and for two titles with a specific geographic focus; “Social Media in Southeast Turkey” and “Participatory Planning for Climate Compatible Development in Maputo, Mozambique”. In both cases we see increased usage of these books associated with geographic areas that are relevant. For “Social Media in Southeast Turkey” we see significant interest from Turkey. Indeed Turkey’s relatively high overall position in the ranking is largely due to this single title. For “Participatory Planning...” while the effect is not as strong we see interest from Mozambique itself, South Africa (an adjacent country with political interests in local development) and Portugal (Mozambique is a Lusophone country and former Portuguese colony).

One of the advantages of OA monographs is that they are accessible from developing countries, where print or subscription access is often limited. We have explored whether titles that relate to a specific geographical location receive more downloads from these places. Among these 11 monograph titles we identified two titles that displayed this pattern: *Participatory Planning for Climate Compatible Development in Maputo, Mozambique* and *Social Media in Southeast Turkey*. We examined usage of these titles according to country. For each title we ranked countries by download and compared this ranking to that for the repository as a whole. Table 1 shows that these two titles were viewed proportionally more in the country that they discuss when compared to the average across this corpus.

Discussion

In the introduction we raised three broad categories of question that are both of interest to a small monograph press, and challenging to answer due to resource constraints.

1. *How is this specific book doing?* What is the audience for a book? How is the book used? How does this compare to other titles published by the press? To comparable titles from other presses?
2. *What promotion strategies are effective?* How effective are specific efforts by the press, authors, or other stakeholders to promote both specific books and the publisher's list overall. Are successful strategies consistent or do they differ from book to book? *i.e.*
3. *How is the publisher delivering on its mission?* Are target audiences being reached? Is the content mix right? Is the institutional home being well served?

Our analysis shows that it is possible in principle to gather rich data from existing systems that can inform a press on these issues. The critical issue to address is how easy it is for a press to gather and use this information in practice.

The practicalities of gathering and analyzing data

The data utilized in this study is all either publicly available or should be available to a press through relevant hosting repositories, although in no case is it strictly open in licensing terms. Knowledge Unlatched Research has both established relationships and technical experience in knowing how to ask hosting repositories for the appropriate data. This is not the case for many OA presses, who struggle with identifying available data, as well as with strategies for gathering and interpreting it. The technical experience needed to support more effective data gathering and interpretation strategies is becoming easier to acquire and presses are becoming more aware of the importance of developing strategic relationships with repositories for effective OA monograph programs. We can also expect a gradual trend towards standardization of data availability from and amongst repositories in the future.

However, given the resourcing realities faced by many OA monograph presses, investments in shared infrastructure capable of helping presses to engage effectively with the growing availability of data relating to OA books are also needed. Networks like OPERAS (*Open Access in the European Research Area Through Scholarly Communication*), which is seeking to coordinate services, practices and technology among key players in the European SSH OA publishing space⁸ have an important role to play in enabling such developments. So, too, do groups like the US-based Library Publishing Coalition⁹ and posited (but not yet funded) initiatives like *Project Meerkat*,¹⁰ which is seeking to help monograph stakeholders to capture and engage with usage data via mechanisms that support continued experimentation, diversity and innovation in OA monograph publishing.

Usage data gathering was managed with R scripts. It would in principle be possible for a press to gather this data manually but this approach does not scale well as a press releases more titles. Social media data was gathered with a freely available Open Source tool, albeit one that requires the capacity to set up a server to use (Borra and Rieder, 2014). The capacity to deploy and use simple command line tools is likely to be increasingly important for a lean and agile press. The future development of new technologies for authoring and publishing, as well as

⁸ <http://operas.hypotheses.org/aboutoperas>

⁹ <https://librarypublishing.org/>

¹⁰ <https://educopia.org/research/meerkat>

data gathering analysis makes it probable that these skills will become an important resource in the continued development of OA monograph publishing. University Presses in particular may find that strong cooperation with internal repository technical managers provides a practical mechanism for accessing technical knowledge and capacity.

In contrast to data gathering and organisation, data analysis is generally straightforward, at least at the small scale we describe in the case study. In this study we used Tableau, a proprietary but relatively cheap tool for data management and visualization. Other free and widely available tools could be used ranging from highly scalable – if more challenging to learn – tools such as R and R Studio through to more limited but familiar tools such as spreadsheets. The analyses we have presented have generally been simple graphs, rankings and comparisons that are readily generated by generally available tools. Map visualisations are less familiar but a wide range of easy to use tools for generating map-based visualisations are available.

The data used in this study was fairly straightforward for our research team to source – although we readily acknowledge that our familiarity with data challenges specific to OA books, existing relationships with key platform partners, and the technical skills that we were able to devote to this task are not the norm for small OA monograph presses. As such, the availability of high quality, affordable services relating to the collection, management and analysis of data associated with the uses of OA books has the potential to add important value to the OA monograph space. So, too, does helping OA monograph publishers to grow their in-house capacity to engage effectively with new sources of data and new tools for capturing and managing it as they become available.

[Addressing the key questions for a press](#)

Gathering and analysing the data is one step. But we must also address whether these data can actually address the questions we originally posed. In general, the answer to this is a partial yes. The answers will not be complete, but investigating these data sources can provide valuable strategic insight for a small press.

[How is this book doing?](#)

A press's assessment of the success of a book has traditionally been judged on the basis of print sales. While sales may in some cases provide a proxy for interest among some specific audience sectors, the relationship between sales patterns and reader interest in Open Access books that are freely available online in digital form is complex. It is often assumed that page views and downloads provide a more reliable indicator of engagement with OA publications, although there are substantial challenges in comparing usage data across platforms, as well as serious questions as to how to interpret these figures. Nonetheless within a single publisher there is value in comparing aggregate views and downloads. In the UCL case stand-out successes can be identified, as can the relative importance of different download platforms.

Similarly, social media analysis can define some level of the degree of buzz around a book. There are however substantial challenges in gaining representative data with low numbers of social media posts linking in an identifiable way to specific books. Social media posts by third parties which refer to a book without linking to it directly (eg. Linking to a third-party announcement of a book's publication, book review, or including a photograph of a book but

no link), are not easily mapped. This may be partially overcome by searching for posts mentioning book titles. However, distinguishing between a book title and other uses of the same terms often proves challenging.

Both social media and usage data can also be of value in looking at demographic patterns of usage. In the UCL case geo-location of usage data could provide a sense of geographic reach. It also showed that at least in some cases books with a specific geographic context were gaining audiences in those locations. That is, it is possible to examine how diverse the publisher's list is at least in terms of the geography of the audience.

There is valuable information to be gained by comparing titles within a single publisher. Comparisons across publishers are much more challenging. There is a danger of assuming that since quantitative data can be obtained that these are comparable across publishers or even platforms. In practice usage data, even that compiled to a standard like COUNTER is not reliably comparable across platforms.

Comparing different publishers within a platform may also be unreliable. In the UCL case the predominance of usage at the UCL repository is driven by the promotion strategy of the press, which provides links to the UCL Discovery platform via its Marketing site and in other marketing related activities. However, a different publisher may direct traffic towards an aggregator like OAPEN. The location link to an OA resource provided within MARC records also varies, impacting on how the promotion of OA titles via library catalogues impacts on usage figures. Comparing UCL Press downloads to those of other publishers on OAPEN thus requires careful contextualisation. Furthermore, categorisation metadata for books is inconsistent, raising questions about the extent to which reliable comparative data for particular categories of books can be drawn from the information held by aggregators. Improved strategies for the categorisation of monographs will be an important issue for developing benchmarks in the future.

Simply ranking titles and comparing numbers is therefore not generally informative. However, it is possible to identify exceptional titles, and also to observe changes in performance over time, or differences between titles. This information can be useful in showing what promotion strategies are delivering results and where to invest resources. Finally there is the potential to gain valuable data on the performance of titles and collections in reaching specific audiences. Geographic usage patterns are the easiest to obtain. Google Analytics can provide other forms of demographic analysis although this may be of limited applicability. There is the potential in the future for using bespoke geo-location to identify on-campus and off-campus use or other classifications that are of specific interest to scholarly monograph publishers depending on the types of metadata tracked and made available by online tracking and analytics packages.

[What promotion strategies are effective?](#)

For a small monograph press promotion strategies are likely to be as much driven by authors as by the press itself. Promotion is likely to be episodic and opportunistic. Gaining a sense of what works, and when, for the press's target audiences, is therefore incredibly valuable. In the UCL Press case it is possible to gain significant insight by simply looking at how specific events lead to changes in usage or social media activity.

The title *How the World Changed Social Media* and the associated project are a good example to consider. The subject matter has a substantial potential for general interest, there were a series of titles with difference subject focus, and it was embedded in a considered dissemination plan. For this title there was clear evidence of specific events driving strong spikes in downloads and views of the book. In particular, it was clear that the start of the MOOC run by the project team was associated with a very large spike in usage. For other books there was evidence that reviews were driving increases in usage. By contrast some other events were not easily associated with detectable usage changes. It is important to note that a lack of signal does not necessarily mean a lack of effect. There are a range of reasons why successful promotions may not lead to detectable signals including them driving non-monitored distribution pathways, the strength of the signal or the time resolution of the data. Strictly speaking an association of an event with a spike does not necessarily show causation. It is possible to consider approaches to planning and monitoring promotions that could show causation, however these would involve significant resources. For a small press the value gained in tracking which promotion events lead to detectable spikes in usage can still be very valuable.

One important issue with connecting signal to event is time resolution. In this case study the time resolution of usage data was monthly. While large-scale events such as the *Why We Post* MOOC lead to detectable signals on this timescale, smaller-scale events, especially including social media posting, may lead to download increases that are of much shorter duration. Increases in usage that last a day or so can easily be lost in data with a time resolution of a month.

Increasing the time resolution of data aggregation is therefore an important area for future development and is a level of granularity that would have very minimal additional technical resource requirements. However, managing and utilising the data collection will require greater human resources. While there is much opportunity for automation here a small press may not have the expertise to manage this, and in many cases a significant manual component of management and data curation will be required. A press should expect each data collection cycle to require several hours of staff time to manage, after the data collection has been setup. Outages and other technical difficulties will add to this.

Another source of valuable data, and for analysing changes, is social media and referral data. If the promotional strategy for a book is reliant on social media, or targeted advertising, then success should be detectable in shifts in the proportions of referring services. In the case of *How the World Changed Social Media* we see direct referrals peaking with the launch of the MOOC. Targeted Google searches and academic list-servs also account for proportionally high levels of traffic. This pattern highlights the value of collaboration between OA presses and open courseware initiatives, as well as the continued importance of discipline-focussed academic list-servs for OA monograph audiences. In this context, the multi-platform strategy for dissemination adopted by the *Why We Post* team delivers an audience for the books.

[Is the press delivering on mission goals?](#)

Compared to traditional sales data a real advantage of usage and social media information, and the associated demographic data is its ability to inform on diversity of usage and audience.

Although the dimensions of that diversity assessment are currently limited with only geographical data being readily available there is an opportunity to consider how the range of titles on offer are being used by different audiences. Similarly, for a press with specific audience goals there are opportunities to use this data strategically to test delivery.

These kinds of analyses will only provide guidance on such high level strategic questions. They can provide new information that would not have been previously available. In many ways the richness of the information environment has great potential in supporting a much more sophisticated and diverse range of missions within monograph publishing, including institutional goals such as outreach and impact (JISC 2017). Equally that richness creates challenges in managing and effectively using the information. However, the potential is there.

[Making it easier. Building best practice on gathering, managing, and analyzing data for the future](#)

Our study shows that it is possible to gain valuable insights into the uses of OA books by collating and analyzing usage and social media data. We were also able to identify a series of relatively straight-forward steps that could be taken by a press to maximise the richness of the data captured. These fall into three categories: proactively gathering and storing data as frequently as is feasible; providing best practice advice to those engaging with promotion, particularly when promotion is via social media; and making it easier to track specific efforts to publicise books by generating and using tagged links.

Within our case study we noted a number of cases the suggestion that a specific event, a review, social media activity, or MOOC, has contributed to usage. We also noted a number of events that might have been expected to produce an increase in usage, but where none was visible. Much of our inability to draw firm links between particular promotion activities and changes in usage is attributable to the time resolution of the data available for this study. Monthly usage figures are not able to reveal changes that occur over much shorter time-periods: for example, a spike in downloads on the day that a link is Tweeted. Many repositories only report total aggregate usage or monthly records. However, there is scope for ensuring that key repository partners are made aware that a more granular level of usage reporting is valuable to presses. There is also scope for investment in the development of automated scripts for capturing and aggregating usage data from platforms on a weekly or even daily basis, as well as for the development of tools for ensuring that this data can be easily integrated into usage dash-boards that can help publishers and their communities to make use of it.

The data used for this study highlighted the potential of tagged links to make it easier to unpack relationships between social media activity and usage for OA books. Providing guidance to authors and Press staff on good practice in creating and sharing links is particularly important. Ideally, all promotion material issued by the Press, links on all online platforms, and material written for authors (that they can repost to their own online spaces, social media, list serves, and so forth) should include a standardised URL. Ensuring that standardised URLs includes the DOI of the book will make it easier to capture mentions of the book via social media, even if third parties are only sharing the link. Ensuring that Author Questionnaires ask where authors will promote their work online, and making efforts to track these platforms, will also help to ensure that an author's role in promoting usage can be identified, and that likely hot-spots of

social media activity relating to the book are not missed. Discouraging the use of direct links to PDFs and instead encouraging the inclusion of a link to a site that allows for referral data to be captured, such as a book's page on the press's promotion site, is also sensible.

Finally, presses might also wish to consider actively tagging links so that the promotional program or platform that they are associated with can be easily identified. In our case study it seems almost certain that the first *Why We Post* MOOC drove substantial direct download activity to the PDFs of the books. However, it is not possible for us to demonstrate a direct connection between the MOOC and the increase in downloads, as we are unable to rule out the circulation of direct links to the PDF in promotional materials that are not connected to the MOOC. It is possible to "tag" links associated with specific campaigns by adding an identifier to the tail of the link. Such links are common in email lists and social media promotional campaigns and could be expanded for instance by providing individualised links to authors and the press.

Conclusion

The data explored in this case-study captures a particular moment in the development of a rapidly evolving OA monograph publishing landscape, as well as in the growth of a newly established OA press. Investments in infrastructure to support OA publishing, including through Horizon2020 initiatives such as OPERAS-D¹¹ and HIRMEOS¹² projects, are providing new opportunities for the effective integration of OA monographs into scholarly communications ecosystems. Established players in the dissemination of subscription scholarly content, such as JSTOR¹³ and Ingenta¹⁴ are also taking steps to support access to OA monographs, increasing the dissemination options that OA monograph publishers might choose from, as well as the complexity of gathering and interpreting usage data.

Through a case study of UCL Press we have identified the ways in which readily available data and relatively low cost approaches to data aggregation and analysis can provide useful insights for the development and strategy of a small monograph press. Particularly with the development of small-scale presses with diverse and targeted missions, the new information environment afforded by Open Access online publication of research monographs offers substantial opportunities. Several of the key questions that will be asked by a press, whether on the performance of individual titles, the choice of investment in promotion activities, or gaining insight on performance against the press mission, can be addressed with readily available data.

There are a range of preparation actions that small presses may need to take to address these opportunities, including proactively considering their internal capacity for data gathering, developing and supporting best practice in online promotion for both staff and authors, and through applying tagged links to gain better data on referrals and sources of activity. The extent to which these approaches can be put into practice by individual monograph presses depends

¹¹ <http://operas.hypotheses.org/operas-d>

¹² <http://www.hirmeos.eu/>

¹³ <https://about.jstor.org/whats-in-jstor/books/open-access-books-jstor/>

¹⁴ <http://www.ingenta.com/products/content-management/ingenta-open/>

heavily on the resources that are available and the institutional environment that the press is operating within. For some university-based OA presses it may be possible to engage more effectively with the possibilities of data by working with on campus partners including the library and IT department. For others there may be value in exploring the possibilities of forming cooperatives with other like-minded or complementary presses in order to gain access to economies of scale that might be associated with the development of tools and strategies for gathering and presenting usage data from a range of sources. Because real strategic advantages are available for those that engage with the opportunities, continued investment in the development of the shared infrastructure needed to support a vibrant OA monograph landscape will be key to ensuring that small OA monograph presses are able to make the most of increasingly rich data landscapes. This infrastructure will be both technical and human. A Press engaging in data collection and management should expect staff time to be required on an ongoing basis to obtain the full benefits of data collection.

As with all data, the particular information provided by usage and social media data have their limitations. Absence of evidence is not evidence of absence necessarily. Careful work to ensure that the interpretation of available evidence supports the challenging resource allocations that a press needs to make, but does not drive them will be required. In particular it is clear that these data are not complete, missing potentially important dissemination pathways such as ad hoc personal file sharing as well as having systematic biases.

Nonetheless as part of an overall strategy that seeks to leverage the rich information environment to support the diverse missions of small monograph presses these data are valuable. It is not inevitable that the resource advantages of a large press will overwhelm the ability of a small expert-led press to target specific audiences and achieve specific goals. There is opportunity for diversity as well as opportunity in diversity.

References:

"About UCL Press." UCL Press. Accessed June 21, 2017. <https://www.ucl.ac.uk/ucl-press/about>.

Allen, Liz. "Opinion: Measure for measuring's sake?" *Wellcome Trust Blog*, January 31, 2013. Accessed June 21, 2017. <https://blog.wellcome.ac.uk/2013/01/31/opinion-measure-for-measureings-sake/>.

Ayris, Paul, and Lara Speicher. "UCL Press: the UK's 'first fully open access' university press." *Insights the UKSG journal* 28, no. 3 (2015): 44-50. doi:<http://orcid.org/0000-0002-6273-411X>

Borra, E., & Rieder, B. (2014). Programmed method: developing a toolset for capturing and analyzing tweets. *Aslib Journal of Information Management*, 66(3), 262–278. <https://doi.org/10.1108/AJIM-09-2013-0094>

Library Publishing Coalition (2017), *Library Publishing Directory 2017*, available at: <https://www.librarypublishing.org/resources/directory/lpd2017>

JISC. 2017. *Changing Publishing Ecologies: A Landscape Study Of New University-Led Presses And Academic Publishing*. London: JISC. <http://repository.jisc.ac.uk/6666/1/Changing-publishing-ecologies-report.pdf>.

Jubb, M. (2017), *Academic Books and their Futures: A Report to the AHRC and the British Library*, London.

Lariviere, V., Haustein, S., & Mongeon, P. (2015), The Oligopoly of Academic Publishers in the Digital Era, *Plos One*, published June 10, 2015. doi: <https://doi.org/10.1371/journal.pone.0127502>

Maron, N. L., Mulhern, C., Rossman, D., & Schmelzinger, K. (2016). *The Costs of Publishing Monographs: Toward a Transparent Methodology* (pp. 1-51, Rep.). Ithaca S R. doi:<https://doi.org/10.18665/sr.276785>

Wouters, P. (2011, March 7). Can metrics support open access monograph publishing? Citation Culture Blog. Retrieved June 19, 2017, from <https://citationculture.wordpress.com/2011/03/07/can-metrics-support-open-access-monograph-publishing/>

J. Priem, D. Taraborelli, P. Groth, C. Neylon (2010), [Altmetrics: A manifesto](http://altmetrics.org/manifesto), 26 October 2010. <http://altmetrics.org/manifesto>

Prainsack, Barbara, Marc Caball, Soulla Louca, and Roland Pochet. *Open Access: Practice, Opportunities and Challenges*. 2013. Accessed June 20, 2017. www.cost.eu/download/COST_Open_Access_Report.

Outsell, *Humanities and Social Sciences Publishing: Market Size, Share, Forecast, and Trends* (report), 16 January 2015