EXPLORING USAGE OF OPEN ACCESS BOOKS VIA THE JSTOR PLATFORM

Lucy Montgomery, Neil Saunders, Frances Pinter & Alkim Ozaygen

Exploring Usage of Open Access Books via the JSTOR Platform

Table of Contents

List of Figures

List of Tables

1. Introduction and Project Brief

This document is a report prepared for the JSTOR Presses project "Exploring Usage of Open Access Books via the JSTOR Platform".

JSTOR's Open Access Books platform launched in October 2016. The first four publishers to submit content to the platform were UCL Press, University of Michigan Press, Cornell University Press and California University Press. Usage of the OA books made available via JSTOR by these publishers has been far in excess of the usage that each publisher has previously recorded via other distribution channels.

This report is the outcome of research commissioned and funded by the four presses. It engages with usage data made available by JSTOR relating to OA books in order to assist publishers in understanding how their OA content is being used; inform strategic decision making by individual presses in the future; and shed light on the potential for data relating to the uses of OA books to support the potential of open access books to reach wide audiences.

Additional key aims of the research are to help inform JSTOR in the development of the JSTOR OA Books platform; and to inform the development of JSTOR usage reporting. Ensuring that JSTOR usage reporting reflects the needs of OA publishers is also an important goal of the project. All four publishers have contributed to a discussion of the role and practicalities of usage reporting services provided by JSTOR.

The project focuses primarily on data collected by JSTOR and made available to the research team for the purposes of this study. The data considered in the report relates to the period between 12 August 2015 and 7 August 2017. This data has been augmented by a short questionnaire and interviews, carried out by phone with some of the publishers. It is important to note that books considered in this report became available via the JSTOR platform at different times. Some of the books included in the data set are also available in both OA and gated formats via the JSTOR platform. A summary of the first date for which activity relating to each publisher can be detected in our data set is below:

publisher	mindate	maxdate
Cornell University Press	2015-08-13 09:08:52	2017-08-07 23:57:28
UCL Press	2016-10-13 19:29:46	2017-08-07 23:55:10
University of California Press	2015-08-12 21:39:55	2017-08-07 23:09:01
University of Michigan Press	2016-08-24 10:52:08	2017-08-07 23:56:05

The questions that the study addresses include those listed below. However, as the report shows, some questions could not be answered satisfactorily within the limits of the data available to the team. There is scope for additional work tackling these questions.

- Where do the readers come from; i.e. are they already on JSTOR or do they come from other sites such as Google Scholar? What are the percentages?

- Where do readers come from in the world, and can institutional affiliation also be traced?

- What are the most popular subjects across the JSTOR platform?

- Are the most popular books on the OA platform similar in subject matter to the subjects already popular in JSTOR?

- What do readers do when they download, ie are they typically downloading just one chapter or multiple chapters?

- Are stats for the four initial publishers similar?

- Are there multiple downloads of the same content (eg. Chapter) from the same institution?

- What is the readers' behaviour in terms of the proportion of who downloads chapters and the proportion who just view?

- What can JSTOR data tell us about how long readers spend on each book that they view?

The four participating publishers were also asked to provide their perspectives on the following questions with the aim of informing JSTOR in the the creation of a more robust reporting program for publishers of open access books:

- How do publishers currently use the data that is provided by JSTOR?

- How is this received and who uses it?

- What data is most important for publishers to receive?

- With whom do publishers share (want to share in future) the data (internally and externally)?

- In light of the preliminary data analysis and results what feedback do the publishers want in the report.

2. Datasets

JSTOR usage report

The primary dataset used for quantitative analysis in this report was provided in the form of a CSV file describing 655072 events at JSTOR. The dataset covers the time period between 12 August 2015 and 7 August 2017. Of these events, 535888 are views or downloads of items under the Open Access Books license subtype.

A summary table of the first date for which activity relating to each publisher can be detected in our data set is below:

publisher	mindate	maxdate
Cornell University Press	2015-08-13 09:08:52	2017-08-07 23:57:28
UCL Press	2016-10-13 19:29:46	2017-08-07 23:55:10
University of California Press	2015-08-12 21:39:55	2017-08-07 23:09:01
University of Michigan Press	2016-08-24 10:52:08	2017-08-07 23:56:05

Table 1 – Date of first and last recorded activities within dataset, according to publisher

The file was read into R and the columns named as follows:

```
##  [1] "event_timestamp" "event_type"      "event_id"
##  [4] "book_id"         "book_doi"        "isbn"
##  [7] "eisbn"           "book_title"      "chapter_id"
## [10] "chapter_doi"     "publisher"       "session_id"
## [13] "user_id"         "referrer"        "license_subtype"
## [16] "session_start"   "session_end"
```

OA country summary

Summaries of usage by country were extracted from the standard JSTOR usage reports, supplied as Excel files, combined and labelled with the publisher to give a single dataset with the following columns.

```
## [1] "Month / Year"        "Country"
## [3] "Chapter downloads"   "Chapter views"
## [5] "TOC views"           "Total Chapter Views & Downloads"
## [7] "publisher"
```

OA institution summary

Summaries of usage by institution were extracted from the standard JSTOR usage reports, supplied as Excel files, combined and labelled with the publisher to give a single dataset with the following columns. Note that this dataset also includes countries.

```
## [1] "Month / Year"          "Name"
## [3] "Institution Country"   "Chapter Downloads"
```

```
## [5] "Chapter Views"                 "Total Chapter Views & Downloads"
## [7] "ToC Views"                      "publisher"
```

Non-OA institution summary

Summaries of non-OA usage by institution were extracted from the standard JSTOR usage reports, supplied as Excel files, combined and labelled with the publisher to give a single dataset with the following columns. Note that this dataset **does not** include countries.

```
## [1] "Month / Year"                   "Name"
## [3] "Institution Country"            "Chapter Downloads"
## [5] "Chapter Views"                  "Total Chapter Views & Downloads"
## [7] "ToC Views"                      "publisher"
```

JSTOR OA export

This dataset was read from the supplied Excel file and contains metadata about JSTOR OA titles, including subject headings using several classification schemes (JSTOR Discipline, BISAC and LC).

EISBN mapped to subject headings

The JSTOR dataset contains 2542 unique EISBNs. Each of these was submitted to the OCLC Classify API and subject headings were retrieved for 2321 titles.

Definitions

In this document a *user* typically means an anonymised ID representing a machine, as defined by JSTOR. JSTOR relies on IP authentication in order to identify users and to associate users with institutions.

Cookies are used by JSTOR to identify individual browsers. Multiple visits from the same browser can be linked to the same user ID. If an IP address is within an IP address block associated with an institution within the JSTOR database then a user can be linked to an institution.

An *unknown user* is a user whose IP address in not linked to a known institution within the JSTOR database.

3. Initial Exploration

Activity

An overview of activity, visualised as sessions by date for the 4 publishers.

The usage increase corresponding to the launch of the OA books platform in October 2016 is readily apparent.
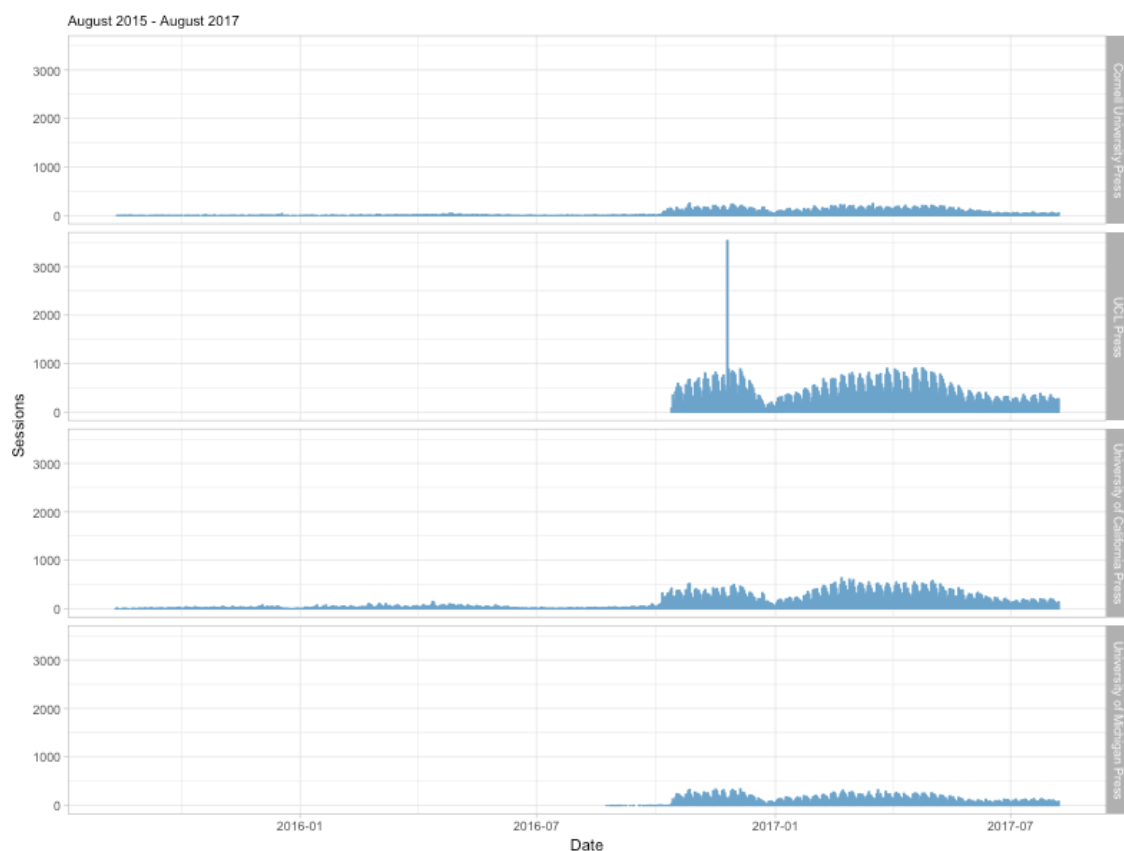
Figure 1 – Session per day by publisher (all licenses)

Books by publisher (all licenses)

A count of unique book IDs by publisher.

Publisher	Books
Cornell University Press	697
UCL Press	43
University of California Press	1391
University of Michigan Press	411

Table 2 – Books by publisher (all licenses)

Books by publisher and type of purchase

A count of unique combinations of publisher, license type and book ID. Note that the same title can appear in more than one non-OA category, so the sum of titles per publisher is more than the previous table.

Publisher	License	Books
Cornell University Press	Consortium_Book_DDA_Purchase	44
Cornell University Press	Consortium_Book_Purchase	2
Cornell University Press	Institution_Book_DDA_Purchase	544
Cornell University Press	Institution_Book_Purchase	434
Cornell University Press	Open_Access_Books	20
UCL Press	Open_Access_Books	43
University of California Press	Consortium_Book_DDA_Purchase	142
University of California Press	Consortium_Book_Purchase	158
University of California Press	Former_Consortium_Book_DDA_Purchase	5
University of California Press	Institution_Book_DDA_Purchase	898
University of California Press	Institution_Book_Purchase	1061
University of California Press	Open_Access_Books	29
University of Michigan Press	Consortium_Book_DDA_Purchase	24
University of Michigan Press	Consortium_Book_Purchase	2
University of Michigan Press	Institution_Book_DDA_Purchase	353
University of Michigan Press	Institution_Book_Purchase	106
University of Michigan Press	Open_Access_Books	21

Table 3 – Books by publisher and type of purchase

Table 3 can be simplified by aggregating the "Purchase" license types as the category "purchased", as in Table 4 below:

Publisher	oa	purchased
Cornell University Press	20	677
UCL Press	43	NA
University of California Press	29	1362
University of Michigan Press	21	390

Table 4 – Books by publisher and license type

4.Usage Data – organized by questions posed in the Brief

The questions defined in the project brief were used as the basis for data exploration and visualization. Comparisons between publishers are made where appropriate throughout the report.

Where do the readers come from?

i.e. are they already on JSTOR or do they come from other sites such as Google Scholar? What are the percentages?

As per Figure 2, the JSTOR platform accounts for the largest number of referrals to the OA books included in the study (34.1% of referrals). That is, 34.1% of readers are already on the platform when they access the OA books. Google.com (10.8%) and google.co.uk (2.8%) are also significant sources of referral to the books.
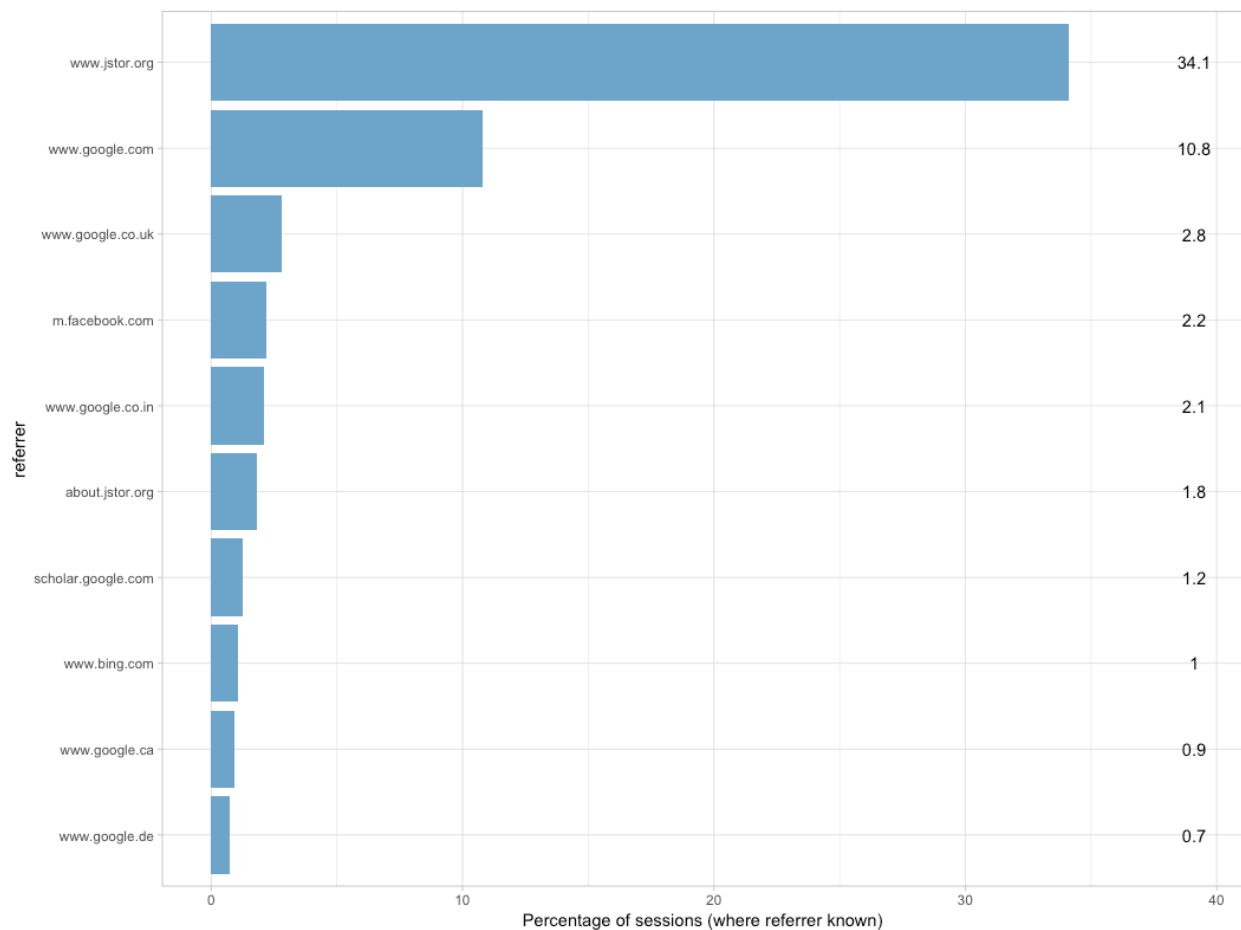
Figure 2 – Top 10 referrers by session – OA books

As Figure 3 shows, there is little variation in the referral pathway pattern associated with each of the four publishers.
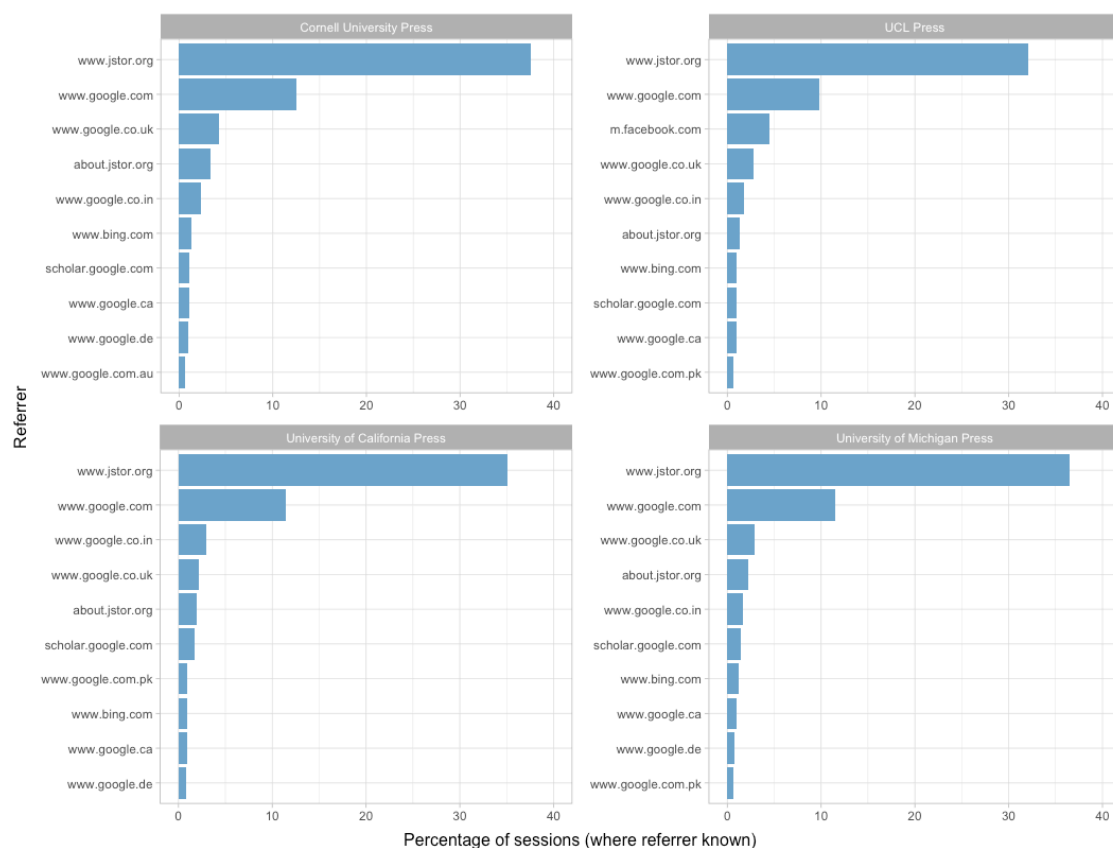
Figure 3 – Top 10 referrers by session and publisher: OA books

OA books: what do readers do when they download?

i.e. are they typically downloading just one chapter or multiple chapters?

A log-log plot of download count frequency versus download count (Figure 4, below) is close to the classic power law shape, indicating that most downloads are of 1-5 chapters per session, with far fewer downloads of large numbers of chapters. 78.57% of sessions involve the download of a single chapter. 97.12% of sessions involve downloads of 5 chapters or fewer.
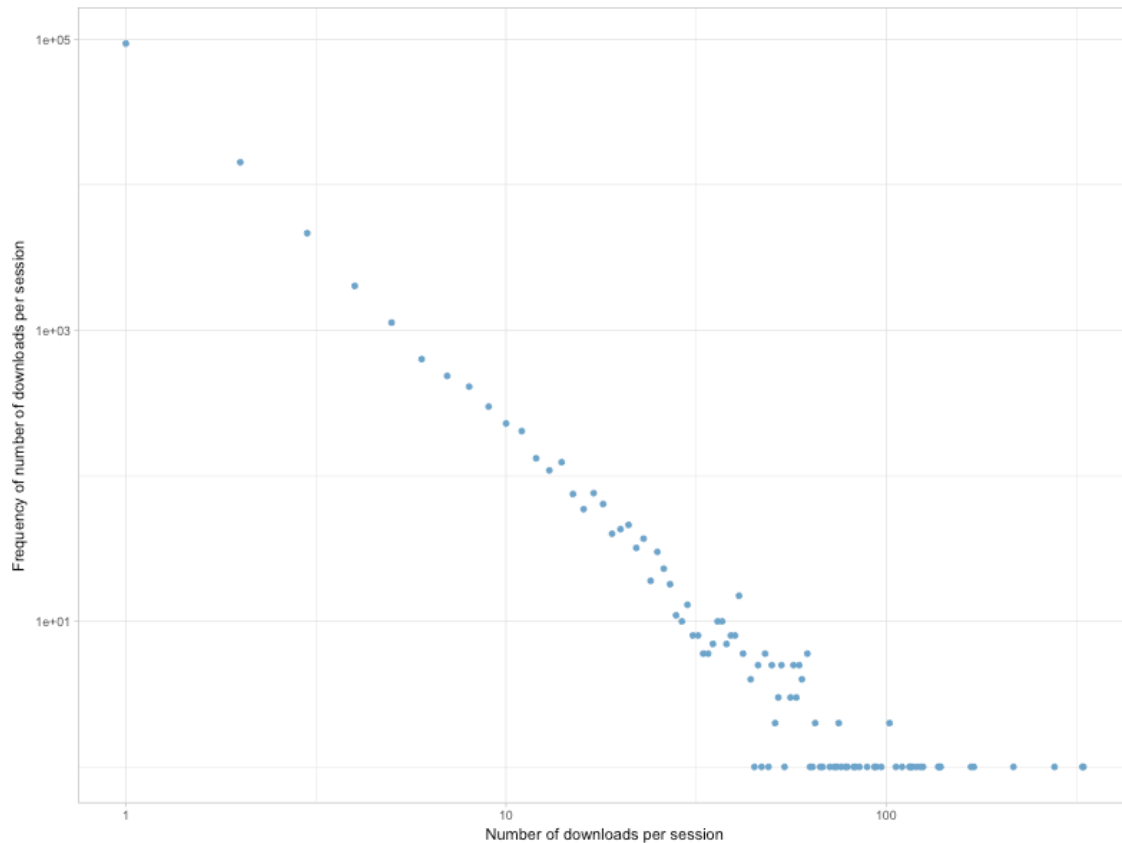
Figure 4 – Number and frequency of OA chapter downloads per session

Downloads/session	Count	Percentage	Cumulative percentage
1	93382	78.57	78.57
2	14259	12	90.57
3	4644	3.907	94.48
4	2016	1.696	96.17
5	1129	0.9499	97.12
6	633	0.5326	97.66
7	486	0.4089	98.06
8	410	0.345	98.41
9	299	0.2516	98.66
10	229	0.1927	98.85

Table 5 – Number and frequency of downloads per session with percentages

OA books: are there multiple downloads of the same content from the same institution?

As Figure 5 illustrates, when the data is interrogated on the basis of institutional user ID the pattern is similar. In this context a JSTOR user ID relates to a computer at an institution. Once again the log-log plot indicates a power law, *i.e.* most 'users' download a chapter 5 times or less, with far fewer multiple downloads of the same chapter. This means that 164050 users downloaded only one chapter across the entire period of the dataset, while just 27 users downloaded ten chapters.
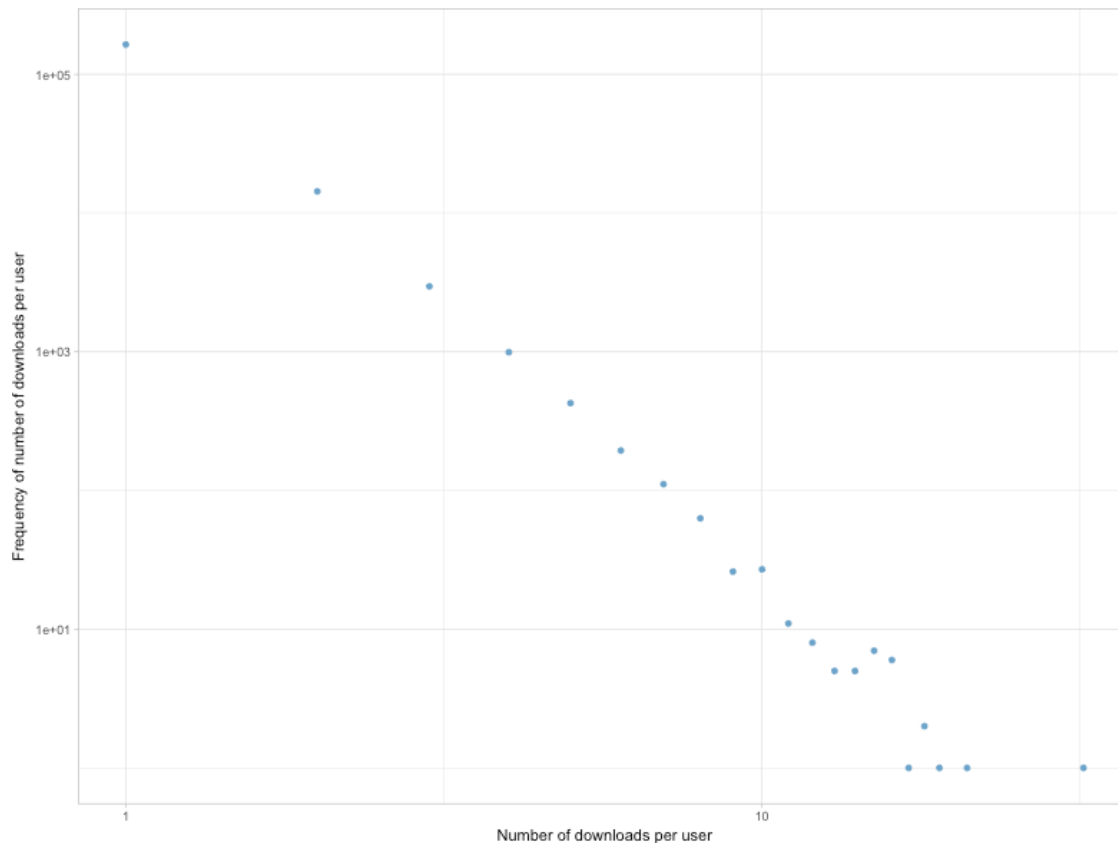
Figure 5 – Number and frequency of OA chapter downloads per user

Downloads/user	Count	Percentage	Cumulative percentage
1	164050	89.54	89.54
2	14326	7.819	97.35
3	2960	1.616	98.97
4	993	0.542	99.51
5	426	0.2325	99.74
6	194	0.1059	99.85
7	111	0.06058	99.91
8	63	0.03438	99.94
9	26	0.01419	99.96
10	27	0.01474	99.97

Table 6 – Number and frequency of downloads per user with percentages

OA books: What is the reader's behavior in terms of the proportion who download chapters and the proportion who just view?

We begin with a broad overview in Figure 6: views and downloads as a percentage of all events, by publisher. More than half of readers view, rather than download, chapters. A view is an "view_item" event in the file supplied by JSTOR. It indicates only that the page in question was opened in the user's browser. A download is an event of type "pdf_download" in the file supplied by JSTOR.
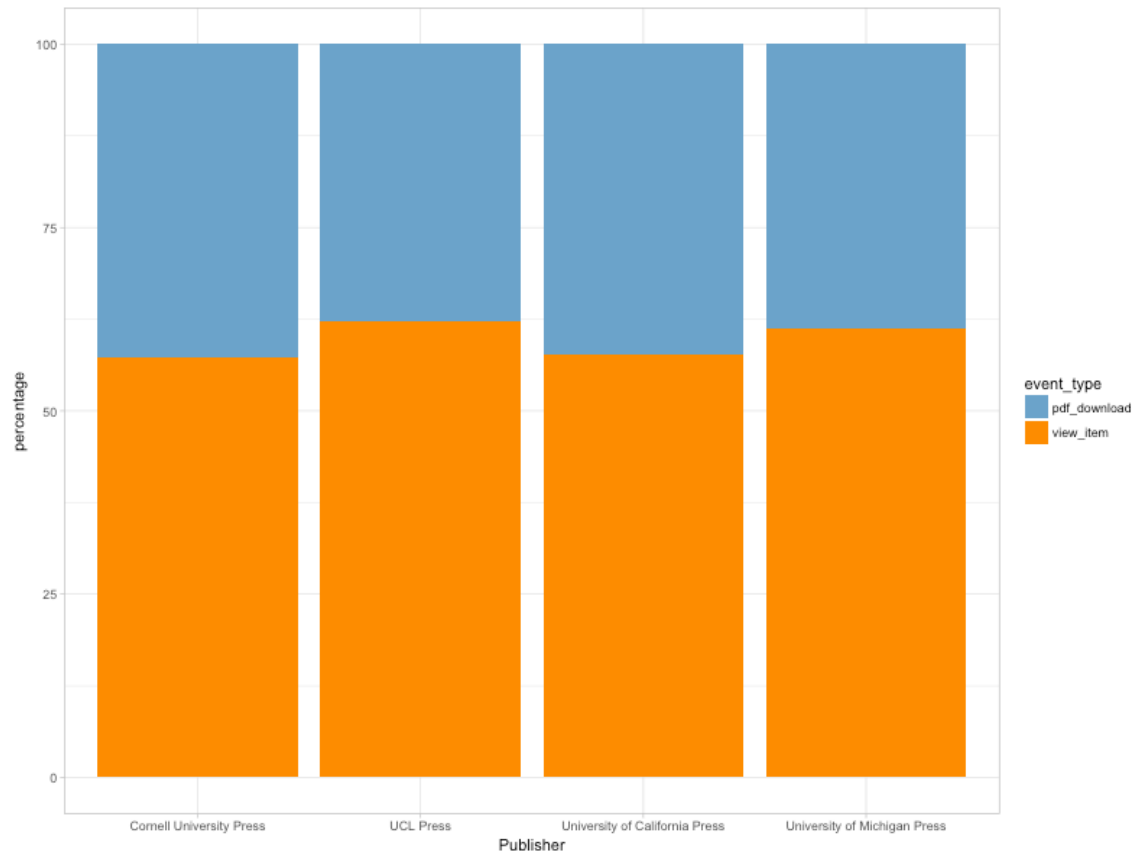
Figure 6 – Event type as percentage of events by publisher

Figure 7 breaks down the same data according to event type by user, showing a comparison between the percentage of users who only viewed, only downloaded or both viewed and downloaded:
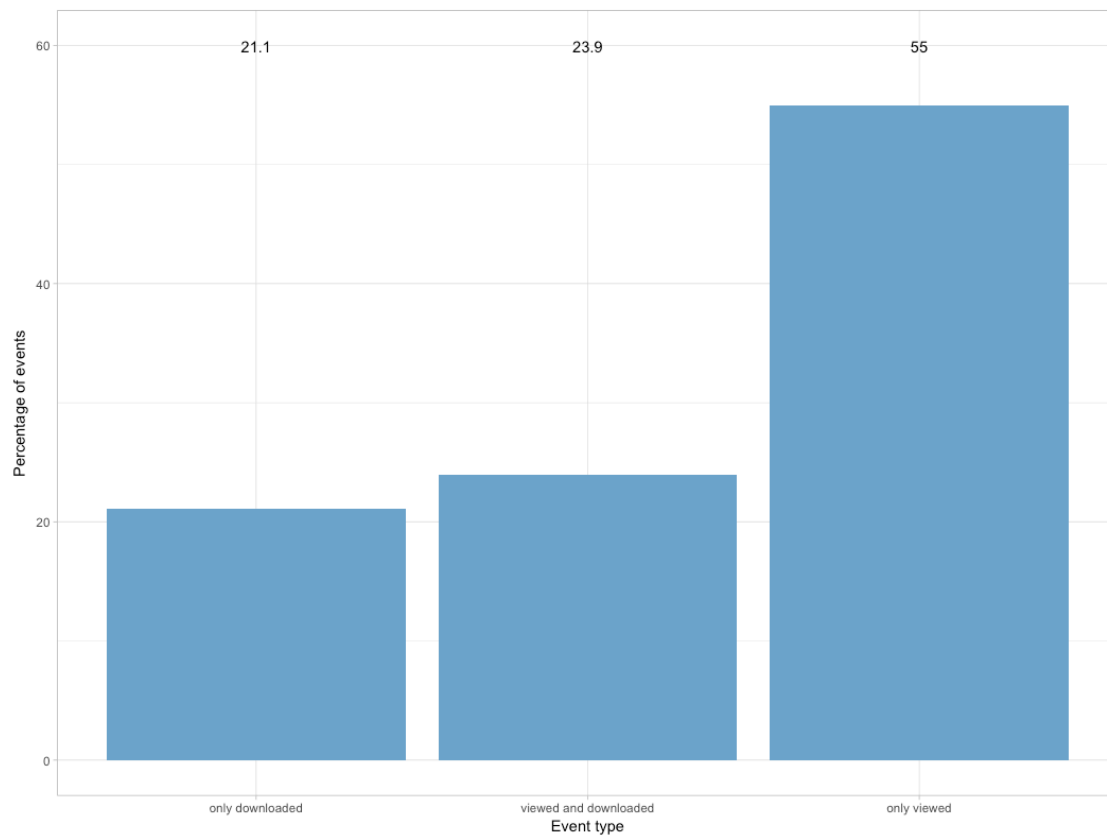
Figure 7 – Event type by user

In Figure 8 event type by user is displayed on a publisher by publisher basis.
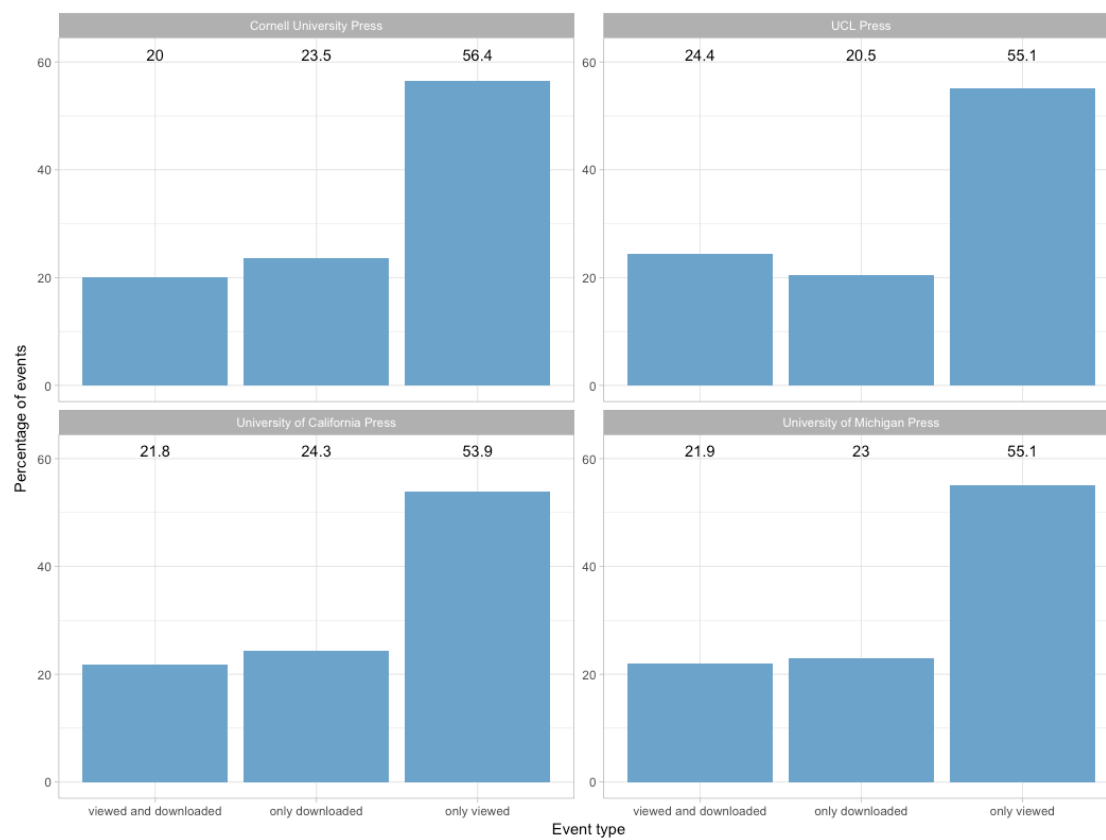
Figure 8 – Event type by user and publisher

Figure 9 displays the same data, further broken down into the percentage of individual users who only viewed, only downloaded or both viewed and downloaded chapters. 55% of readers only view chapters. 23.9% of readers both view and download during their session. 21.1% of readers only download and do not view. When broken into the percentage of sessions (rather than the percentage of individual users) and the pattern was very similar.
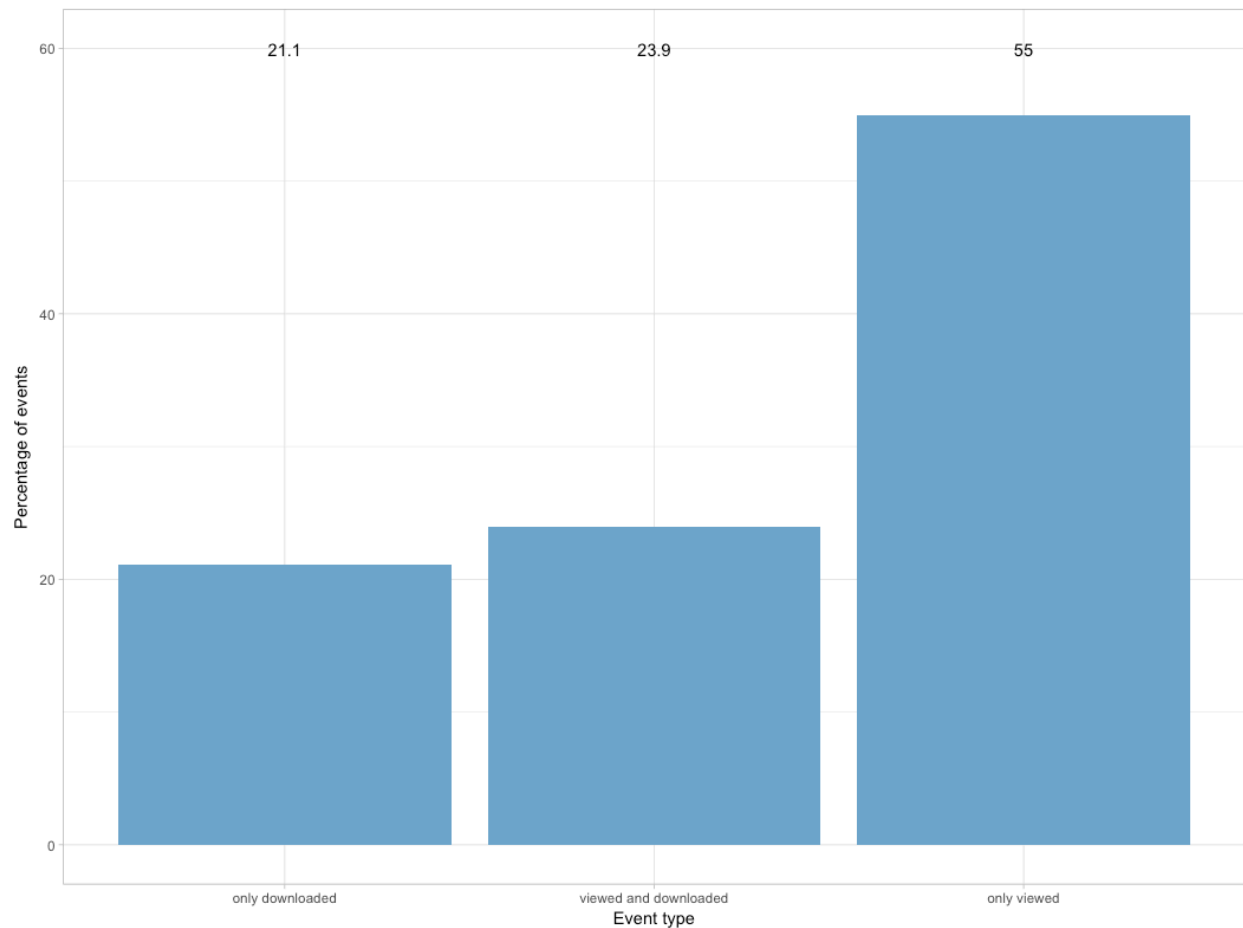
Figure 9 – Event type by user

OA books: what can JSTOR data tell us about how long readers spend on each book that they view?

Time spent viewing an item can only be calculated for the cases where within a session, a view event is followed by a second event (either another view or a download).

There are 159784 such view item events for OA books in the JSTOR dataset. This represents 24.4% of total events (655072) within the data set. As Figure 10 hilights, view time per item is highly-skewed.

View Duration	Value(seconds)
Mean	199.2
Median	38
Min	1
Max	32720

Table 7 – View duration in seconds

A cruder measure of time spent is session duration (note log scale for the x-axis).
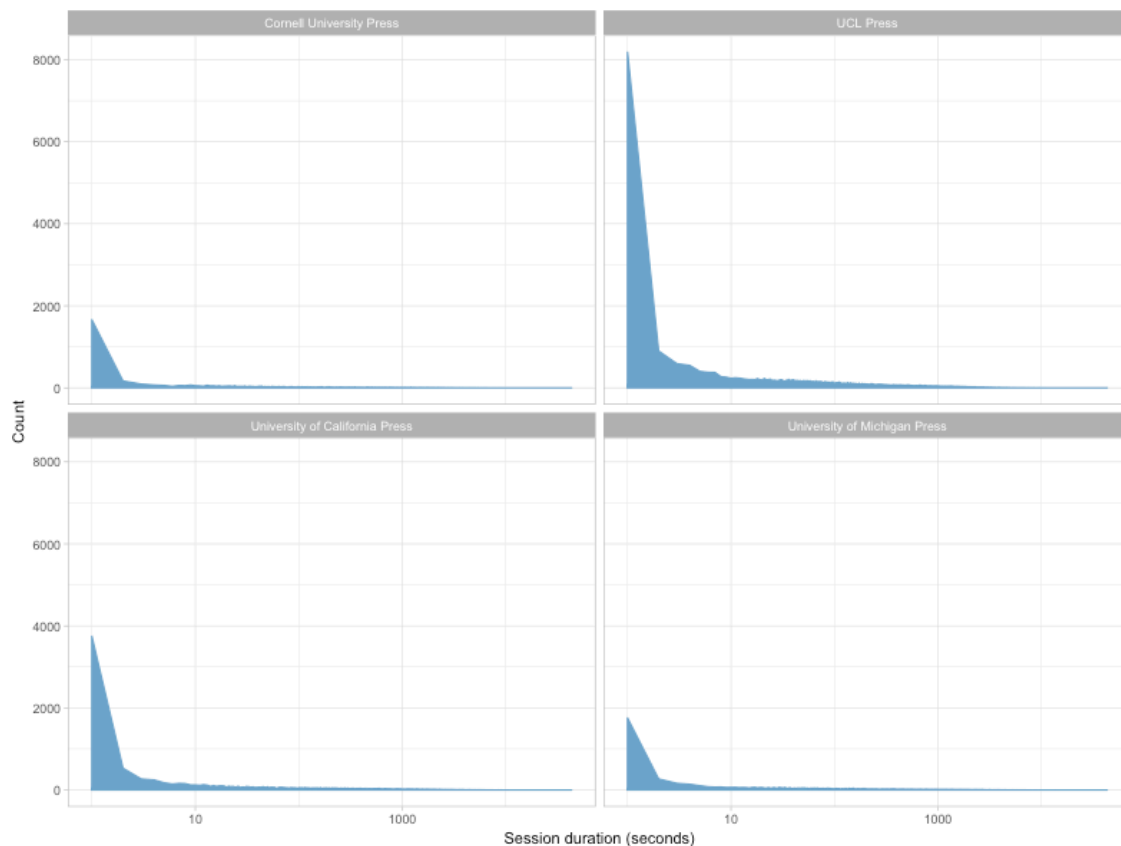
Figure 10 – Distribution of session duration by publisher

Mean session durations by publisher, shown in Table 9, are similar.

Publisher	Mean session duration (seconds)
Cornell University Press	1475
UCL Press	1297
University of California Press	1526
University of Michigan Press	1477

Table 8 - Mean session duration according to publisher

How usage of OA books compares to gated access of books from the same publisher on JSTOR?

Analysis of all events in the dataset according to publisher and license type, in Table 10, reveals that the majority of events involve engagement with OA books. Note that all UCL Press titles are OA.

| publisher	license subtype	pdf download	view item
Cornell University Press	Consortium Book DDA Purchase	844	971
Cornell University Press	Consortium Book Purchase	NA	4
Cornell University Press	Institution Book DDA Purchase	9928	7905
Cornell University Press	Institution Book Purchase	3783	3189
Cornell University Press	Open Access Books	22226	29717
UCL Press	Open Access Books	107387	177433
University of California Press	Consortium Book DDA Purchase	4001	4196
University of California Press	Consortium Book Purchase	4917	1759
University of California Press	Former Consortium Book DDA Purchase	68	9
University of California Press	Institution Book DDA Purchase	23790	21710
University of California Press	Institution Book Purchase	11515	9501
University of California Press	Open Access Books	54227	73952
University of Michigan Press	Consortium Book DDA Purchase	265	393
University of Michigan Press	Consortium Book Purchase	19	8
University of Michigan Press	Institution Book DDA Purchase	4820	4249
University of Michigan Press	Institution Book Purchase	751	589
University of Michigan Press	Open Access Books	27471	43475

Table 9 – All events by publisher and license type

Table 11 provides a breakdown of all events in the dataset according to publisher and license type. However, license type has been simplified into two categories: either gated or OA. Categorised as either gated or OA the breakdown is as follows:

publisher	event type	gated	oa
Cornell University Press	pdf download	14555	22226
Cornell University Press	view item	12069	29717
UCL Press	pdf download	NA	107387
UCL Press	view item	NA	177433
University of California Press	pdf download	44291	54227
University of California Press	view item	37175	73952
University of Michigan Press	pdf download	5855	27471
University of Michigan Press	view item	5239	43475

Table 10 – All events by publisher and license type (gated, oa)
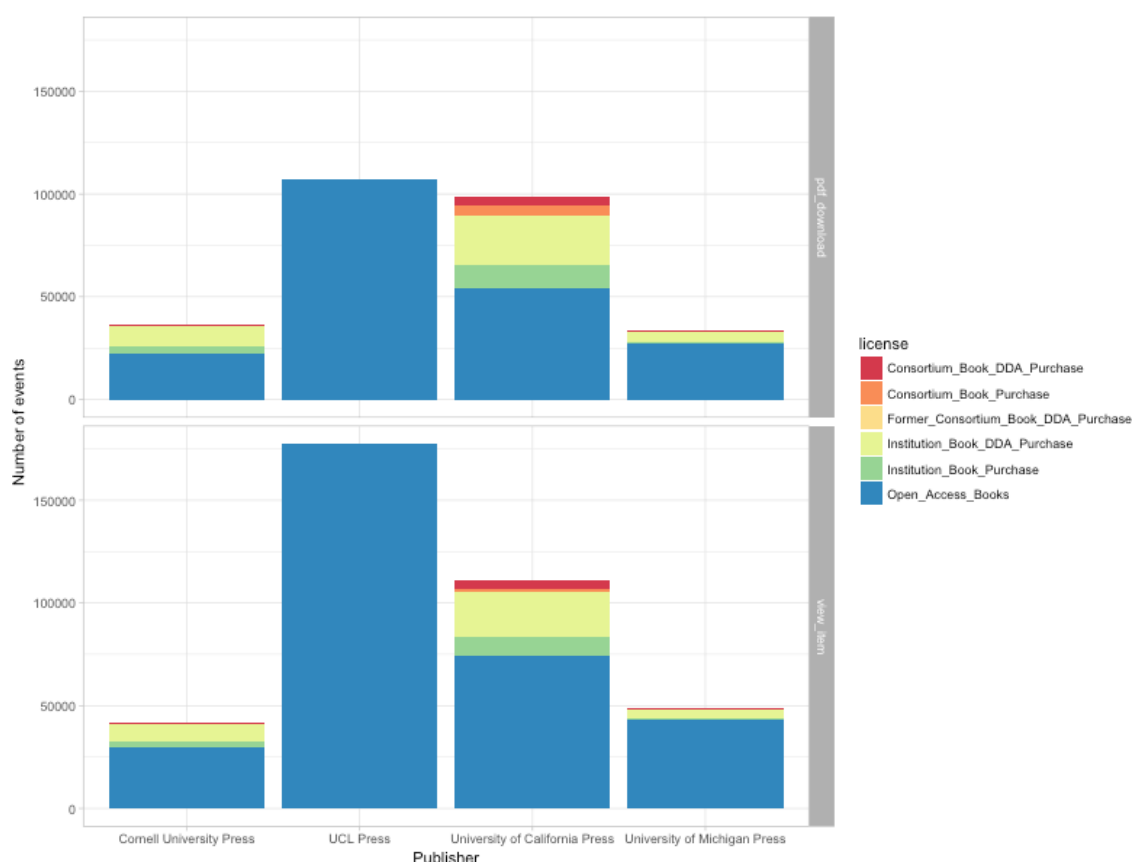
Figure 11 – Events by publisher and license type-bar chart

Analysis of each publisher's downloads according to license type confirms that open access book chapters account for the majority of downloads, as per Figure 12.
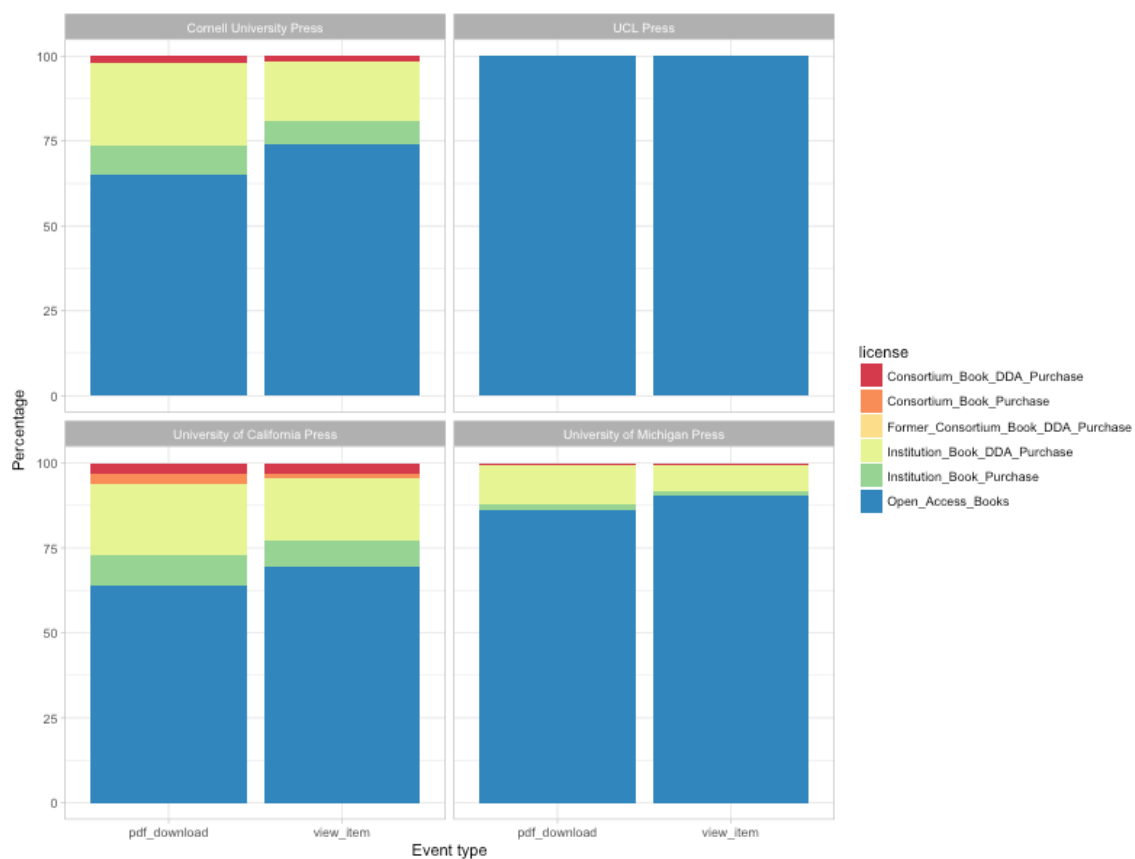
Figure 12 – Percentage of event types by publisher and license type – bar chart

Where do readers come from in the world, and can institutional affiliation be traced?

We begin by looking at the top 20 countries by percentage of total chapter views + downloads as per figure is below. The United States accounts for 42.6% of all views and downloads of OA chapters made available via the JSTOR platform. The United Kingdom accounts for just 6.8% of downloads and Canada and India account for 4% and 3.9% respectively.

OA events by country (all publishers)
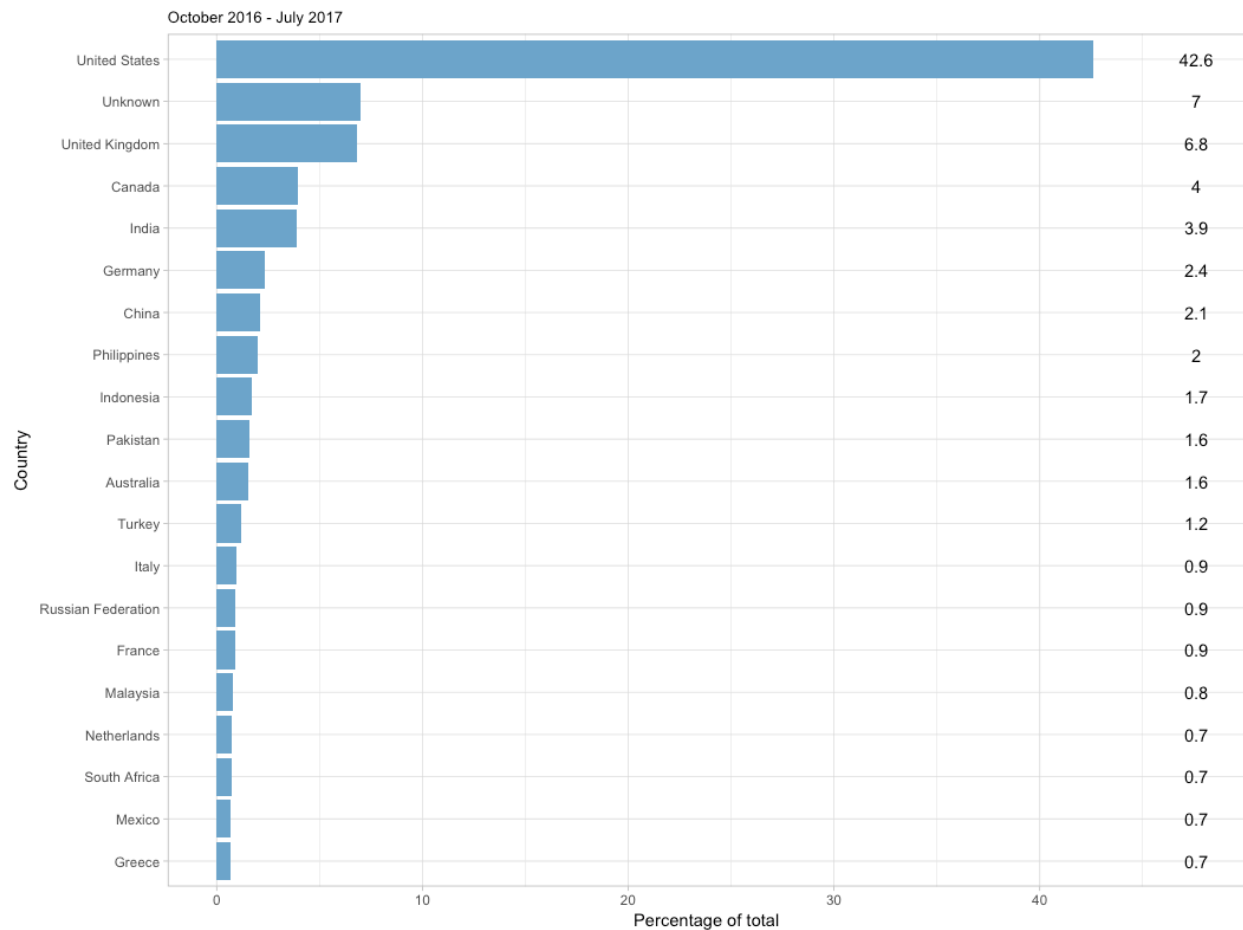
Figure 13 – Top 20 countries by total chapter views and downloads

Geographic usage is presented as a choropleth map in Figure 14 below:
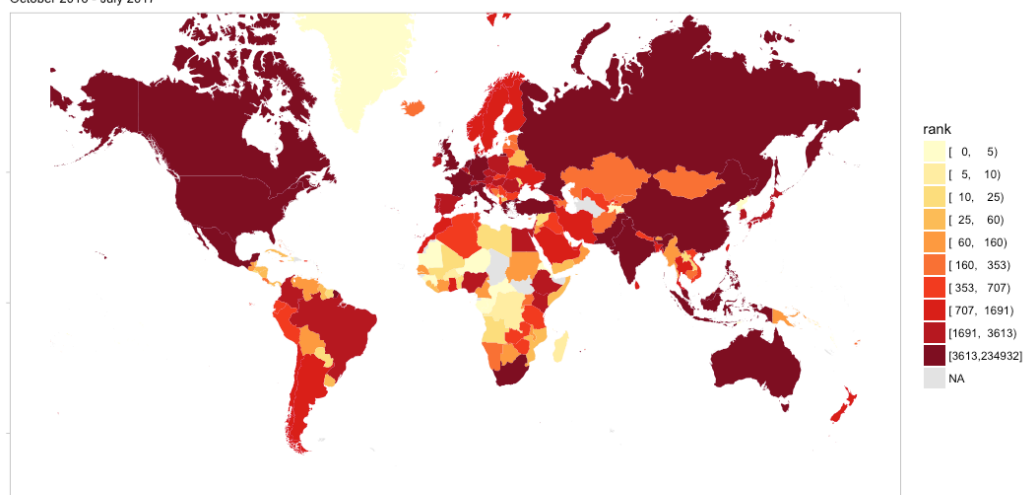
Figure 14 – Total chapter views and downloads by country

In Figure 15, when the data is broken down according to individual publisher (top 10 countries for each) the pattern is similar.

OA events by country and publisher
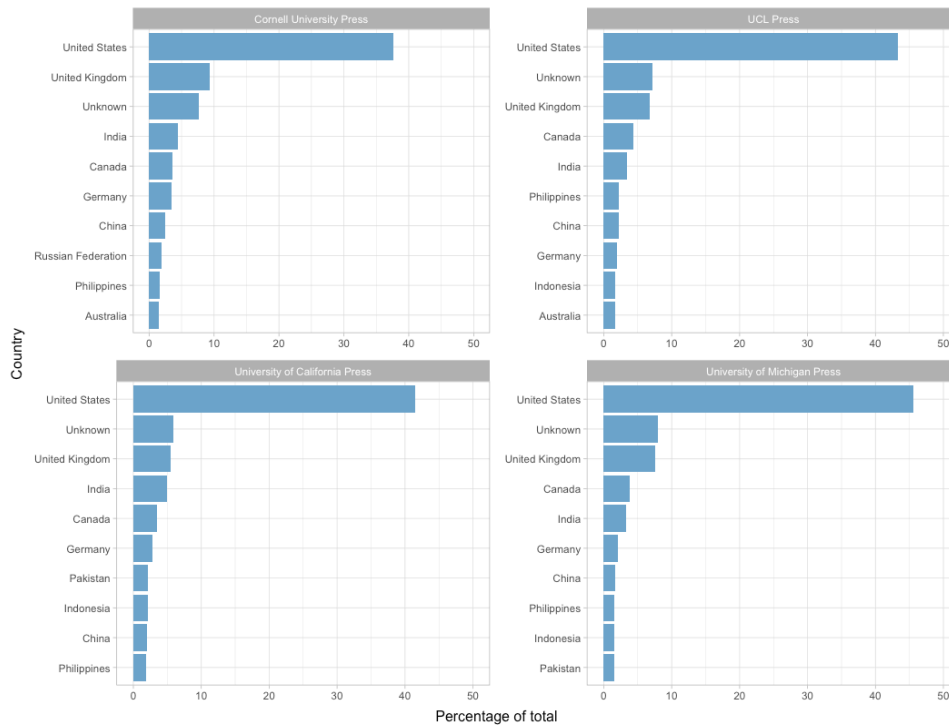
Figure 15 – Top 10 countries by total chapter views and downloads by publisher

Choropleth maps displaying chapter views and downloads on a country-by-country basis for individual publishers are provided in Figures 16, 17, 18 and 19.

Cornell University Press
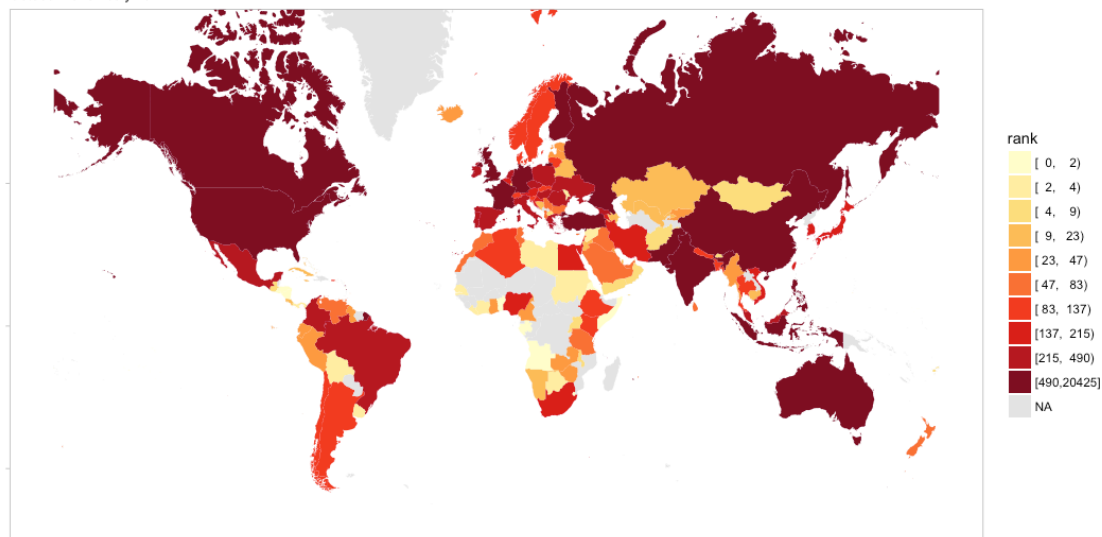
Figure 16 – Cornell University Press total chapter views and downloads by country

UCL Press
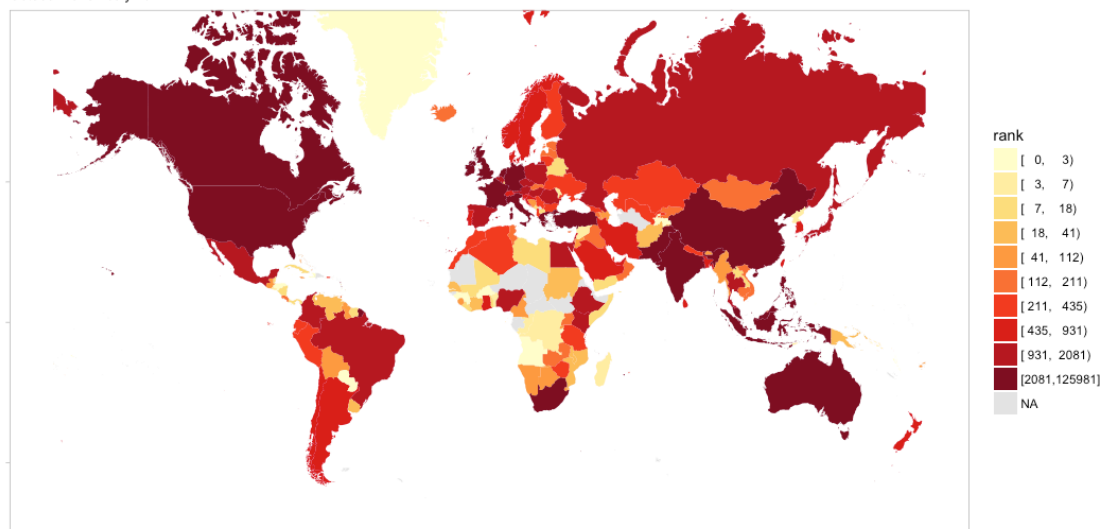
Figure 17 – UCL Press total chapter views and downloads by country

University of California Press
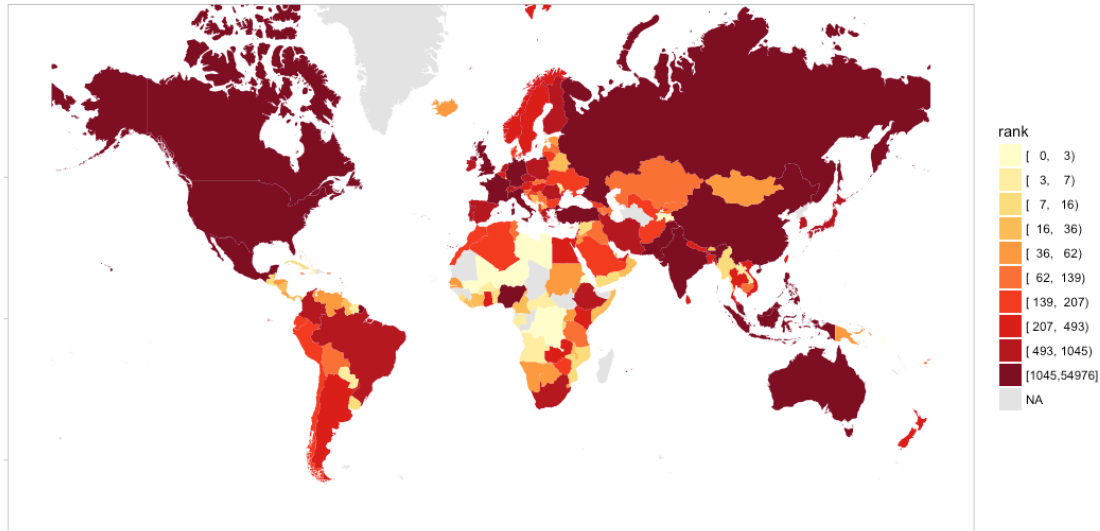
Figure 18 – University of California Press total chapter views and downloads by country

University of Michigan Press
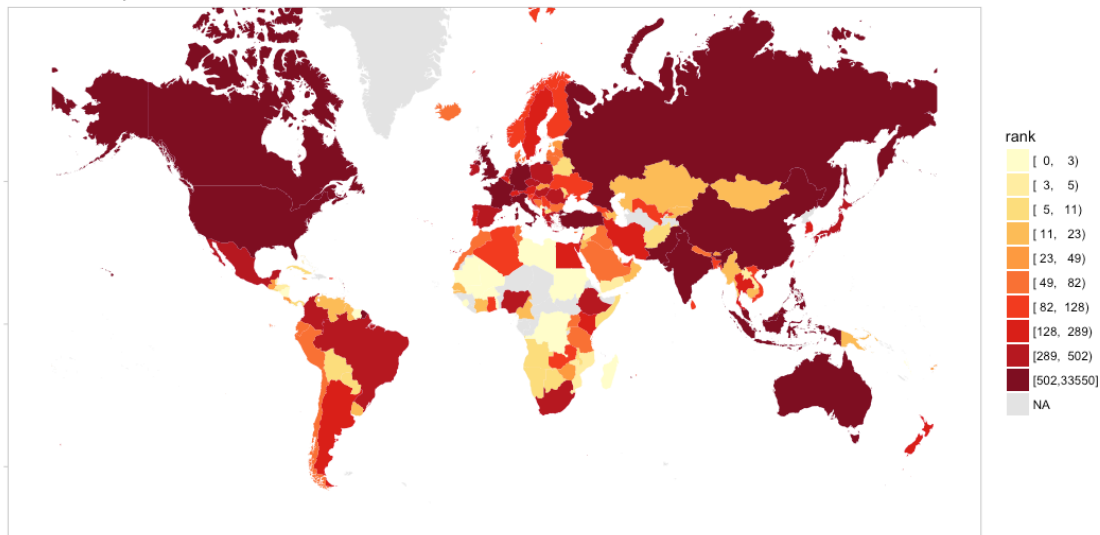
Figure 19 – University of Michigan Press total chapter views and downloads by country

Understanding use of JSTOR OA books in the global south

It might be argued that the country-based usage patterns identified in the previous charts reflect global patterns of academic population. That is, more visits to the JSTOR OA books platform occur in countries with larger academic populations. In an effort to correct for this we have used the dataset that contains both country and institution to calculate "total views and downloads per institution per country" in Figure 20 below. All records not linked to a named institution within the JSTOR dataset were excluded. This approach produces quite different results. It appears that institutions located in the global south are relatively high users of OA books made available via the JSTOR platform when compared to institutions located in the US, UK and Western Europe.

Note: the totals by country alone in this dataset differ to those in the previous charts, so these results must be treated with caution.

OA events by country, per institution
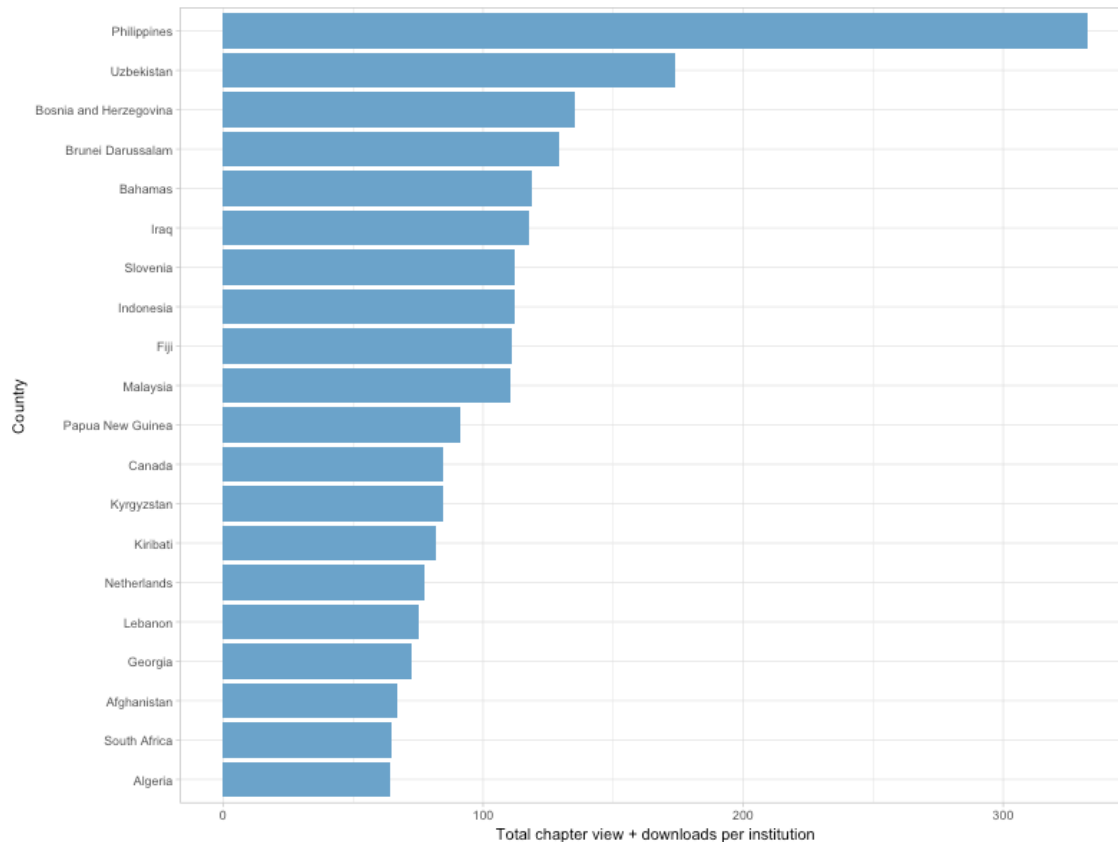
Figure 20 – Top 20 countries by total chapter views + downloads per institution

When the same approach to normalizing for the number of Universities in a country is applied to usage data associated with individual publishers, in Figure 21 below, differences become apparent. UCL Press titles are being accessed with particular frequency in the global south, when compared to the other three publishers. Because all of the UCL Press titles are OA it is not possible to discern the extent to which increased use of UCL Press titles in the global south is a result of the subject matter of the titles, or of their accessibility.

OA events by country and publisher, per institution
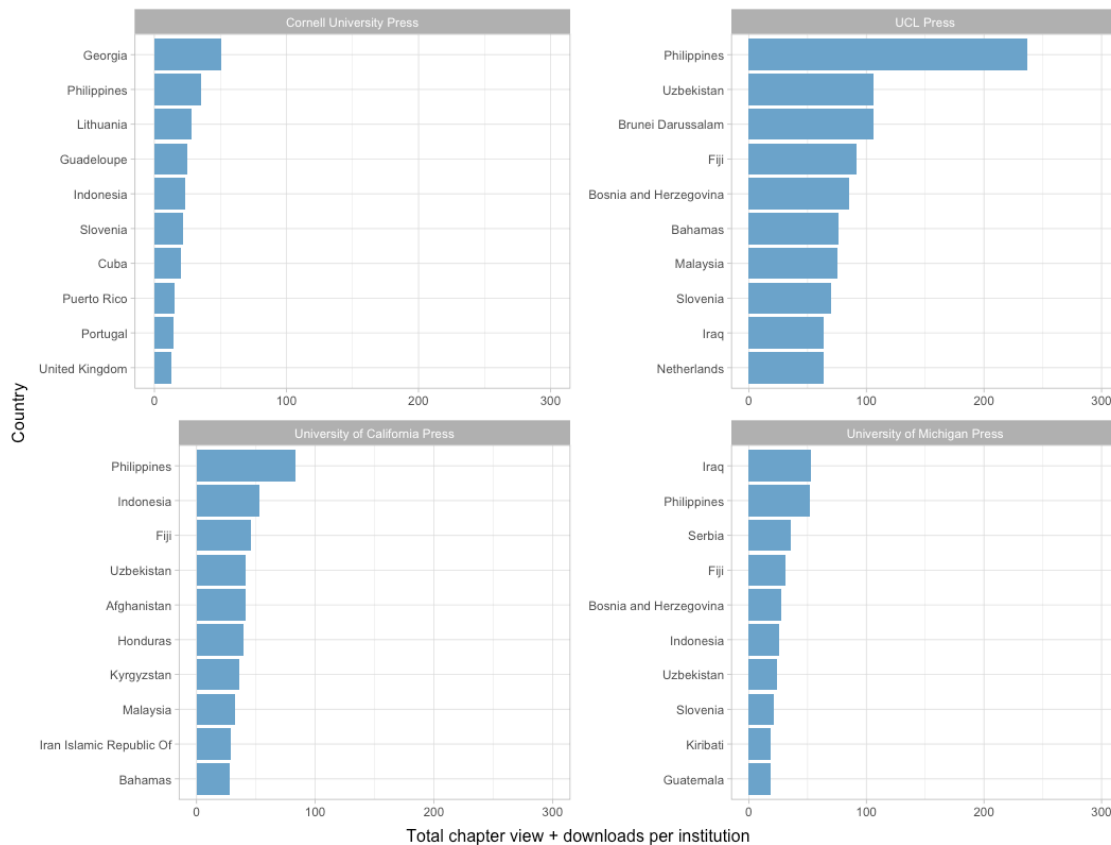
Figure 21 – Top 10 countries by total chapter views and downloads per institution, by publisher

Finally, Figure 22 provides a breakdown by institutions accessing open access titles (normalized by academic population) and reveals some overlap between publishers.

OA events by institution and publisher
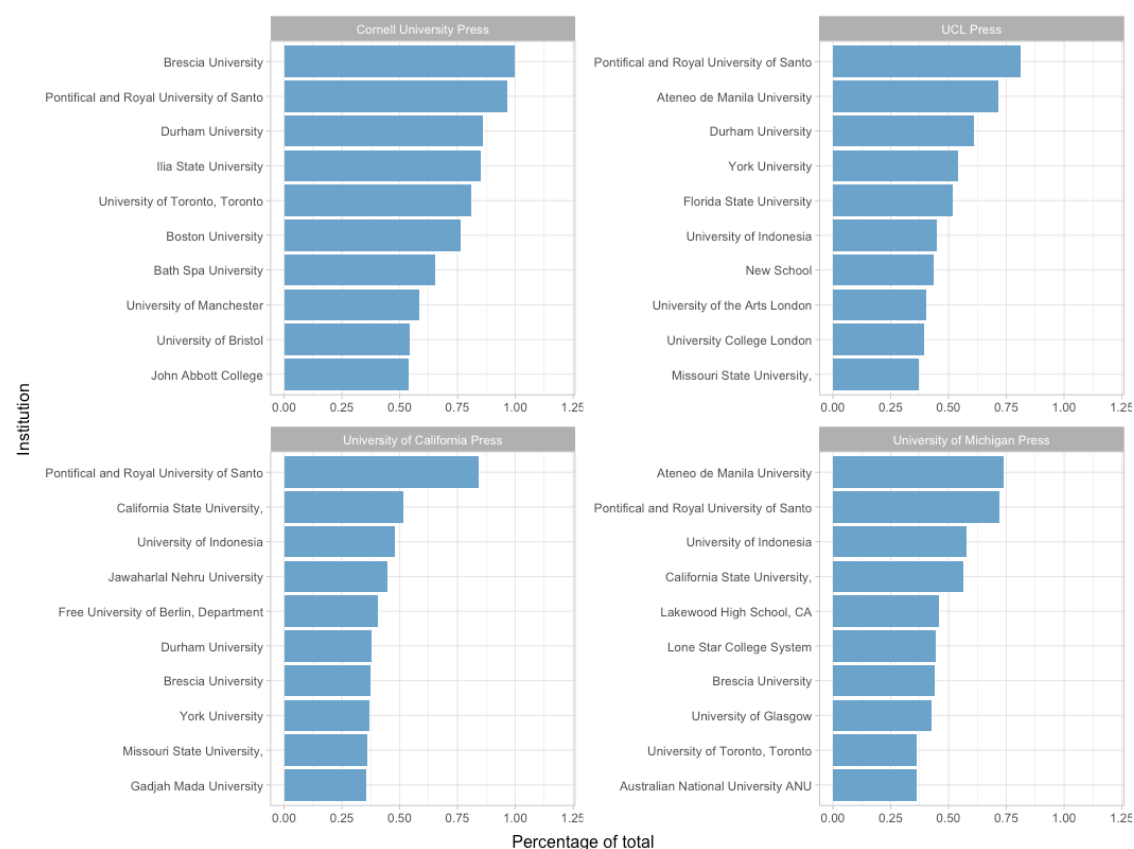
Figure 22 – Top 10 institution by total chapter views and downloads, according to publisher

What are the most popular subjects across the JSTOR platform?

Are the most popular books on the OA platform similar in subject matter to the subjects already popular in JSTOR?

As Figure 23 illustrates, titles classified in BISAC as Social Science are the most frequently downloaded group of both OA and non-OA books. There are other significant differences in the frequency of downloads when analysed according to OA or non-OA. The relative popularity of OA books classified as 'anthropology' stands out. Books classified as 'history' account for a relatively small number of downloads

Given the relatively small number of titles investigated in this report these results should be viewed with caution. There is potential for a single highly successful book or series to skew results.

Using JSTOR BISAC classifications

Figure 23 – Top OA downloads, minimum text frequency = 2

The same procedure has been used to generate Figure 24: top non-OA downloads, joined with the non-OA catalogue (JSTOR-Outreach-Catalog-Export). Of non-OA downloads, 'Social Science', 'History' and 'General' account for the largest proportion of downloads.

Figure 24 – Top non-OA downloads according to BISAC classification.

Finally, in Figure 25 we have constructed a comparison cloud to compare OA with non-OA terms. OA terms are presented in green, while non-OA terms are presented in red. These are the top 100 BISAC terms found in both OA and non-OA titles. Figure 25 shows that a relatively large proportion of OA books are classified as 'Political Science' when compared with non-OA titles. This may reflect the subject classifications of books that are currently being made available in OA. For example, the Knowledge Unlatched program identified target subject areas including Political Science for participating publishers, impacting on the number of Political Science titles made available in OA during the period covered by the dataset. Furthermore, some subject areas produce more books, whether open or closed, than others. Differences in the types of books being made available in OA and non-OA formats are important to take into account when interpreting subject related differences in the popularity of open and closed books made available via the JSTOR platform. As the number and diversity of monographs made available in OA increases it may be possible to gain more accurate insights into relationships between OA availability and download frequency of books in particular subject areas.

Note: the size of a subject's text is in proportion to its frequency within its category (OA or non-OA). The sizes of the words are not comparable across categories.

Figure 25 – Comparison cloud – top 100 BISAC terms found in OA and non-OA titles

5. Engaging With Data About OA Books: Publisher Perspectives

The qualitative component of this project explored the different ways in which individual publishers currently engage with data on the uses of OA books made available by JSTOR and other platforms. Usage data about OA books is providing presses with valuable information about the relative impact of their efforts to ensure that books reach the readers that value them. However, at a practical level, interpreting newly available data and successfully integrating it into day to day workflows remains a challenge.

Considerable variation in the ways in which usage data is ingested and used exists between each of the four presses engaged in this study. UCL Press differs from the other three in that it is an OA first Press launched in 2015, while the others have much longer histories. UCL Press has focussed first on building its digital marketing and distribution strategy and been more sensitive to short term changes in usage that might be linked to social media promotion, although it has also found it necessary to engage with traditional marketing approaches. The press's focus on digital distribution and access to its own distribution platform have influenced the role of usage data in the day-to-day operations of the press. The three US based University presses that also publish non-OA books are pioneering the incorporation of OA within established workflows. They are keenly aware of opportunities to engage effectively with usage data made available by aggregators in this context. This includes developing strategies for incorporating OA usage data into the individual workflows and established marketing processes of presses that are also publishing content in other formats.

The availability of usage data for OA books is a relatively new phenomenon. Thoughtful approaches to understanding what usage data means for an individual press, or an individual title, are needed. Variations in usage patterns may be a result of different subject behavior, reactions to events, brand presence, money invested in marketing or simply that the dataset is not large enough and time period under study is too short. Furthermore, usage data is made available to presses in different formats. The usage data made available to publishers reflects the format in which an OA book is presented to users via the platform (for example – as HTML on-screen view only content, as a whole book download, or as individual chapter downloads) - as well the approach of an individual platform to collecting, cleaning and sharing data. As such, it is often difficult to make direct comparisons between usage figures across different platforms.

The availability of individual chapters of OA books means that usage data is collected and presented to publishers on a chapter-by-chapter basis. Publishers have welcomed this additional level of granularity when it comes to understanding how book content is being used and have indicated that additional analysis of usage patterns within the JSTOR data would be valuable to them. But the flip side of this is that data from other platforms such as OAPEN, which make books available as whole-book downloads, is not directly comparable with JSTOR data. Ensuring that publishers have access to contextual data (for example JSTOR subject averages, OA book averages etc) thus has the potential to increase the usefulness of usage data to publishers making books available via JSTOR. JSTOR's usage data is likely to be uniquely useful for publishers of both OA and gated books because it is

the only aggregator platform where usage of OA and non-OA books can be directly compared.

5.1 Publisher Responses to Written Questionnaire

Publishers were invited to respond to a series of eight questions relating to the role that usage data plays within their organization, as well as the specific challenges they face in engaging with it. Questions, along with collated comments, are below. We draw attention to the areas where enhancements or improvements in the way that JSTOR data is made available or presented would bring benefits to publishers. We are aware that it may not be practical to deliver the whole of any wish list but usage data is the currency of impact for open access publishers and the more that platforms such as JSTOR can deliver robust and meaningful usage data the better for authors, institutions, funders, and publishers that care about open access books and need to justify their investments.

1. How does your press use the JSTOR data?

Assess and monitor usage and trends

Aggregate with other download stats

Share internally with select staff and then discuss at staff meetings

Establish usage benchmarks

Provide relative impact of various usage-driving campaigns

Look at how people arrive at our OA content

Bolster case to board and others about OA benefits

2.How is this received and who uses it internally in your Press? Does it have any impact on decisions made in editorial, marketing or other departments?

Marketing and Distribution Manager receives it.

No editorial or marketing decisions made on basis of data, though if a chapter-by-chapter content breakdown was available, this would help inform our social media/ content marketing activities/strategies.

Reports of OA and non-OA books are compared and shared with a few stakeholders

This information is important for funders and to publishers who can demonstrate the wide use of the research that is being funded.

Looking for patterns

3. What other sources of OA usage data relating to your titles is available to you (i.e. from which platforms)? Do you have a process for gathering and managing this data? If so, what is it?

University Platform – checked weekly, OAPEN, Internet Archive, World Reader, Own platform

Reports aggregated manually into an Excel spreadsheet (though developing an automated approach – work in progress)

KU/OAPEN (COUNTER compliant), Own Google Analytics for titles on own platform (but not systematic – usually only done if requested)

Ubiquity (but sporadic), Google Analytics, OAPEN COUNTER reports plus enhanced versions at book level on a monthly basis

Altmetric data, Unglue.It survey results, KU surveys, sales data

No systematic way of gathering this data

4. To whom do you send OA usage data now (from JSTOR and other sources). For example: funders, authors, parent institutions, others?

Parent institution, authors and potential authors mainly, but we also make public our overall stats and averages quite widely.

We do not systematically or consistently send OA usage to external stakeholders

Google Analytics unsatisfactory

Google Data Studio presents an interesting option for more customized reports, but haven't yet successfully generated a combined author (or other stakeholder) report

Data mostly used internally

Authors have generally not requested data

Would like to be more proactive and use data as actionable business intelligence for making promotion and acquisition decisions.

5. Do you promote any good news stories of OA usage? If so, how (website, social media, annual reports, other)?

Website, social media, annual report, blog, award entries, our catalogue

Mostly internally

Social media mentions are forwarded to our marketing dept

Twitter

Ubiquity publishes some usage data on its site

Information pushed out in email newsletters, social media

6. What information about OA usage is most important for you to receive?

Basic information, such as number of downloads per title

Countries each title is downloaded in

Institutional level data

Where are the users

Category of users

How they found the work

Total Usage

Global distribution patterns

User engagement (number of pages etc)

Usage by Institution

Usage by title

7. What do you find useful about the current format of the JSTOR usage reports?

I like the format of the JSTOR data- it makes the data easier/clearer to handle and manipulate (some other platforms report with a different file per title).

Institutional data is useful to help us to identify trends.

Country details

Institutional details

Country and Institution details (and comparing between the two tabs)

Sheer volume of data is useful

Distinction between views and downloads is useful

But less known where an individual is not affiliated with an institution

8. What would make the JSTOR usage reports more useful?

Chapter-level reporting (i.e., which chapters are being read)

Key word reporting - which words are people using to discover which content?

How are people finding the content- e.g. is it via libraries, search engines etc?

Which chapters are being downloaded – 'one 50 times or 50 one time.'

JSTOR chapter counting makes it difficult to combine with data from other sources.

JSTOR doesn't have a stats portal which is pretty much industry standard

Summarise more easily time by period, title, and geographical region

Define what is a 'view'

Interested in chapter level downloads and where they are coming from around the world

Create measures of user engagement with the content

Better visibility into what the lack of institutional affiliation actually means

Find ways of comparing OA with pay-walled content

Present graphs, charts, dashboards, period-over-time comparisons as Google does.

6. Conclusions and Recommendations

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, publishers face practical challenges in identifying relevant data, as well as in capturing, managing and interpreting it. This is a particular challenge to publishers who rely on third-party aggregators to distribute their ebooks since the degree and type of reporting varies very widely. As an established platform in the scholarly communications space, JSTOR has an important role to play both in ensuring the widest possible dissemination of OA books, as well as in providing publishers with key information about how and where OA books are being used. It is currently one of the only platforms to display OA and non-OA books from the same publisher in a comparable way and therefore has a key role to play in illuminating the differences in usage between closed and open content.

As this study illustrates, data collected by JSTOR is rich and has the potential to shed important light on how users find and engage with Open Access books. The willingness of JSTOR and the publishers involved in this study to allow the research team to analyse data on a comparative basis makes it possible to identify patterns in use that are consistent across publishers, as well as to discern individual differences in usage that might be associated with an individual press's content offering or approach to marketing and distribution.

The power of JSTOR to connect OA books with readers in the United States is particularly apparent, suggesting that JSTOR is playing a valuable role in the integration of OA books into pathways of discovery and use that North American researchers are familiar and comfortable with. When usage data is corrected for academic population signs, there are also encouraging signs that OA books are being used in the global south – where readers are less likely to have access to content if it remains behind paywalls. Also encouraging are the hints that users at institutions who might not otherwise afford access to publishers' books (remembering that JSTOR customers subscribe to a wide range of different journal and book collections) are using OA books. These appear to include high schools and community colleges.

More so than journals, the book business has been driven by intermediaries throughout its history. Even in the transition to ebooks intermediaries continue to be important in the widespread distribution of book content. Thus, having book content available through the full range of discovery outlets is critical to ensuring access to research communities. However, the high proportion of readers originating in North America and already on the JSTOR platform when they access the books examined in this study hints at the continued importance of multiple distribution pathways for OA books as a mechanism for ensuring that the key outputs of the Humanities and Social Sciences make their way beyond the walled gardens of academia. Furthermore, embracing multiple platforms remains an important factor in the ability of publishers to make books available in a range of formats, as well as in the capacity of publishers to experiment with different approaches to OA distribution and marketing. JSTOR offers readers access to PDF chapter downloads, while Ubiquity, for example, offers whole books for download in an ePub format. One publisher said that neither of these solutions is ideal and that their preference would be to embed the full HTML text on the page and allow users to select an appropriate reader. This is especially important for multimedia content.

Although challenges associated with integrating usage data into established workflows remain, JSTOR's willingness to make comprehensive usage data available and interest in working with publishers to ensure that it is presented in ways that are useful, are key advantages of the JSTOR platform for publishers. Generally speaking the usage data from commercial vendors is opaque and so comparisons between platforms are hard to come by. Publishers would welcome more harmonization of data at source, though they expect to further manipulate the information to meet their specific needs. More work needs to be done to explore the value of usage standards like COUNTER in the creation of data sets that might make comparability between data sets a possibility – particularly as Release 5 of the COUNTER code of practice accommodates license type identification in usage reports. There may also be an opportunity for JSTOR to work with publishers to call for and develop industry standards in usage reporting for OA books.

JSTOR might consider providing additional data and analysis in order to ensure that publishers are able to derive information that is valuable to them from usage reports. Publishers involved in this study were particularly interested in data about views vs downloads. The visualization of headline data – along with basic contextualization – might also add important value to publishers faced with challenges of communicating the value of Open Access within their organizations, as well as to authors and funders. In this context, OA publishers may be particularly interested in data that could support claims for extra reach to users at less well funded institutions, in the Global South, or outside the academy. The detailed analysis of users unaffiliated with a subscribing institution that might shed light on the extent to which OA is widening access to content beyond academic communities was beyond the scope of this report. It, would, however, be worthy of further study. This is likely to require inviting users of OA content via the JSTOR platform to provide information about themselves and their purpose: data that is not currently collected by JSTOR, but which might be gathered on a voluntary basis using a simple questionaire. Additional analysis of institutional patterns of usage of OA content, and access to gated versions of the same content at the same institution, would also be valuable

in helping publishers to gain a more detailed understanding of the extent to which open access is widening user communities; as would an analysis of usage that sought to identify course adoption, particularly in high-school systems. This analysis was beyond the scope of the current study.

Self service portals as are already available for usage data relating to some gated content – and would be a helpful addition to JSTOR's OA books offering.

Most of the publishers involved in this study independently mentioned the desire to know more about the users of their books. The proportion of authenticated vs unauthenticated users, as well as patterns of authentication across locations, were identified as particular points of interest. The pronounced trend towards users downloading individual chapters, rather than multiple chapters from books, was identified as a particular point of interest – and might serve as the starting point of a future research project.

Publishers are also interested in comparisons between the uses of OA vs closed content. This study has carried out some analysis on this basis – although results must be viewed with caution, given the limited number of publishers and books involved in the study. As the number of OA books on the JSTOR Open platform increases the value of such comparisons is likely to increase. The University of California Press suggested that for them a random comparison of around 100 titles per subject would be a meaningful number. For smaller Presses reaching this figure might not be feasible.

All publishers felt that regardless of their budget more could and should be made of engaging OA authors with the promotion of their books. Making usage data more visible is seen as part of that campaign.