



OAPEN-UK matched pairs pilot: Final Report January 2016

Ellen Collins, Research Information Network, Caren Milloy, Jisc

"If the objection is that the pilot wasn't realistic, I think I'd say that in fact it was, as life is like this. The big service providers like Google will, in real life, change their rules half-way through one's project; people will, in real life, struggle to find the free copies of things that are also charged for; in real life, data is messy and attempts to level the playing field often fail."

(OAPEN-UK steering group member)

Overview

This report outlines the main findings from the OAPEN-UK matched pairs pilot. The pilot ran from September 2011 – August 2014. Five publishers submitted pairs of titles to the pilot, matched as closely as possible on a number of areas. The project team randomly selected one title from each pair to be made open access, and the other title was used as a control. Over three years, we gathered sales and usage data to try to understand what happens when you make a monograph open access. A sixth publisher joined the project in September 2013.

About the pilot

The OAPEN-UK matched pairs pilot attempted to understand – broadly speaking – what happens when you make a book available in open access. This covers the effect on sales and usage of the book, but also how publishers and their supply chains are affected when trying to make an open access monograph available. The experiment was designed to operate in the real world. It tells us about what happens when you try to make a book open access in a system which is designed to sell, and where open access publishing is a very small proportion of overall business. The results are therefore time-limited. It is unlikely that we would find the same results in a system where open access for monographs was a bigger proportion of overall business; where it was familiar to publishers and users (both individuals and libraries) and where systems and services were designed to accommodate its needs.

The real-world environment for monograph publishing meant that it would be impossible to collect the data we would need to undertake meaningful statistical analysis of the two groups with the sample sizes available to us; nor would we ever be able to control for all the variables within the project. This report therefore presents descriptive data from the experiment, and also reflects upon what the publishers have learned from their participation; in particular, we share some lessons for other publishers who might be considering an open access monograph publication stream. Readers seeking statistical analysis could look at the results of the OAPEN-NL project, which learned from many of the challenges faced by OAPEN-UK.

Set-up





We selected a matched-pairs experimental design, applying the OAPEN business model. We invited publishers through an open tender to submit matched pairs of titles for inclusion in the pilot. The pairs had to match as closely as possible on subject area, publication date, age to print sales ratio, price and format history. The titles submitted could be older titles (to explore impact on back list sales) or new titles (to explore impact on front list sales) but they had to have been published between 2006 and 2010 when the tender was released. The publisher was responsible for discussing the pilot with the author and had to agree that any of the titles submitted could be made available under a creative commons licence (CC-BY NC ND at the minimum). The publishers were informed

OAPEN's business model is based on a hybrid approach to open access books, publishing both an open access edition and conventional editions that are offered for sale. The cost of the open access edition is calculated as the first copy costs of a book, based on all the costs that go into producing the digital file of the publication. Publishers charge a publication fee for the open access edition based on the first copy costs and recover all other costs through sales.

that one title from each pair would be randomly assigned to the experimental group (open access) and one titles to the control group. All titles had to have been through the publisher's traditional peer review process.

Five publishers were successful and the steering group selected 29 pairs for inclusion in the project. Another publisher joined in September 2013, increasing the total number of pairs to 47. The five

The initial steering group consisted of Jisc, AHRC, ESRC, librarians and researchers. The successful publishers joined the steering group upon completion of contract.

original publishers received a fee of £6,000 per pair of titles included for their participation in the project. This fee was based on the original OAPEN project: it covered the changed licence of the book, and also the work publishers had to undertake to collect and supply data to us, work through issues as they arose in the matched pairs experiment, and to participate in other strands of the

project. The sixth publisher received a participation fee and opted to give most of this funding directly to participating authors.

Open access titles were randomly selected from each pair. These titles were made available as PDFs for download from the OAPEN library platform. They were also made 100% visible in Google books with a PDF download (the control titles were 10% visible) and the publishers made them available to download as PDFs from their own websites. The Creative Commons licence meant that authors, universities and any other interested parties could make the full text PDF available from anywhere else they chose – for example, personal websites or institutional repositories. Publishers were free to continue to sell alternative electronic formats, or to give them away for free. These decisions reflect OAPEN-UK's ambition to work within the real world, recognising that it would not be possible to control the availability of a book once it is made open access.

About the titles

Most titles had already been published upon entry into the project, with the majority being 1-2 years old (Table 1).





Figure 1: Publication year of titles

Year	2011 entry	2013 entry
2006	6	
2007	7	
2008	5	
2009	6	
2010	32	
2011	2	1
2012		18
2013		14
2014		3

Monograph sales typically decline rapidly after a book has been published: publishers in the project estimate that around 70-80% of sales occur in the first year of publication, with most of the rest happening in the second year and a handful of copies being sold each subsequent year. Most pairs in the project had fewer than six months difference in publication dates for the two titles (Table 2). The pairs with longer differences were among the older titles. For this reason, the steering group were satisfied that the declining sales that most monographs experience over time would not affect the data collected. Any future replications of this experiment might want to consider whether they can match the pairs more closely on publication date in order to avoid this issue.

Table 2: Age difference between pairs of titles

Difference	2011 entry	2013 entry
0-6 months	20	16
6-11 months	5	1
12-17 months	3	0
18-23 months	0	0
24-29 months	1	1

Data collection

Primary data sources for analysis were the OAPEN library (for open access titles only), Google Books, and publishers' own sales data. Usage, in particular, happened on a much wider range of platforms than those for which we were able to collect data. Publishers were able to supply a limited amount of data about usage on their own platforms, but such use is a very small proportion of overall use — the majority happens on third party sites such as library aggregators. Despite its partial nature we have used Google Books as our main source of usage data, since both control and experimental titles should be equally discoverable there.

Data were collected when titles entered the project (baseline data) and after the first, second and third years of the project. Table 3 shows the data collection schedule for the project.

Table 3: Data collection

Platform	Data collected	Definition (via	Time
		https://support.google.com/books/partner/answer/3	period





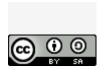
		323499?hl=en-GB)	
OAPEN library	PDF downloads		Monthly
	Geographical origin of PDF downloads		Annual
Google Books	Book visits	A Book Visit is registered each time a unique user views one of your books on Google Books. This number includes informational page views (such as the "About this book" page), as well as preview content page views.	Monthly
	Book visits with page views	The number of Book Visits in which users accessed preview pages of your book. This doesn't include visits where a user accessed only informational pages for your books.	Monthly
	Pages viewed	The total number of unique preview content pages that a user viewed in a given session (counted as a 24-hour period). If a user views the same page of your book twice during a session, only a single page view is registered.	Monthly
Publisher	Print sales		Annual
	E-book sales		Annual
	Total sales		Annual

Challenges

The OAPEN-UK pilot was a real time exploration, situated within a supply chain that is set up to sell, not to support the discovery of freely available content. We (the project team, publishers, librarians, researchers, funders) encountered many obstacles and learnt many lessons along the way. This section sets out some of the main issues that we encountered. They can be grouped into two categories: first, issues that arose because of the way that books are currently supplied, marketed and sold, and second, changes made by companies that affect mechanisms for delivery of our open access titles.

Current supply, marketing and sales model for books

- Most of our publishers' websites were based on a global template used across the company. This template was not designed to make books available without payment and using a Creative Commons licence, so publishers had to invent workarounds to make the PDF version of the open access titles available. In some cases, this meant that the open access PDF was not obvious or easy to find. To increase visibility, we devised best practice guidance which was implemented by all the publishers, but the open access titles likely remained secondary or less visible to users than paid for content.
- Academic libraries have agreements in place with aggregators for large collections of
 ebooks, including monographs. Library ebook aggregators were not set up to deal with a
 small number of open access titles and at the time of the pilot set up, were mainly selling
 collections and packages of ebooks to libraries. The administration involved in reimbursing a
 library or swapping in a different title to account for a title now being available in open





- access, was not deemed cost effective. Aggregators were often also unable to set a zero price and so even the frontlist titles suffered from this problem, especially as sales usually begin up to a year before publication. Since a great deal of use happens on aggregator platforms it is likely that many users never encountered the open access version.
- Library vendors and online bookshops were also not set up to deal with open access or to sell zero priced books and publishers have very little control in this area. Amazon, for example, is visited by academics and librarians to check new releases, compare pricing and order titles yet there was no way to set the price of the open access title to zero or to highlight that an open access version was available. Once we had identified this issue, we worked to implement a line of text in the ONIX feeds of publishers that would surface on sites like Amazon, to highlight the availability of the fee PDF on the OAPEN Library. However, as this text was not always displayed prominently, we are not certain that it will have made a big difference.

Changes affecting the pilot

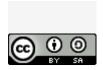
- Over the three years of the pilot, publishers made changes to their global systems and services. In several cases especially the larger publishers a spot check on the titles would reveal that charges for the PDF of an open access book were re-instated, or the link to the version on the OAPEN platform had been removed. This was usually down to well-intentioned correction of what seemed to be errors by staff who had little knowledge of the project, or changes to publisher systems that were carried out without regard to the project (for most publishers, the titles that were made open access for OAPEN-UK represented a very small proportion of their overall lists).
- External companies also made changes to their systems and services which affected the
 pilot. For example, in 2012 Google implemented Google Play, changing the availability of
 several of the open access titles from 100% to 10% and removing links to download the free
 PDF instead, directing users to purchase the title from Google Play. We responded by
 creating a workaround, including re-setting the price on Google Play to 'zero' where
 publishers were making the e-book available for free. But this will have affected usage.
- Initially the OAPEN Library only provided MARC records to libraries but in 2012 the OAPEN
 Library become discoverable through library discovery services such as Primo and Summon.
 This increases the potential for the open access titles to be found and we would expect this
 to increase usage on the OAPEN platform.

There are a third set of issues which relate to data collection.

- Publishers were not universally able to give us the level of detail we would have liked about sales and usage. Often their sales data is not broken down by platform, and their ability to extract usage data from third parties depends upon their contracts with those third parties.
- Publishers had specific problems with some titles and periods for data collection. These are set out in Table 4 below.

Table 4: Issues with data collection

Source	Problem	Solution
Google Books	One publisher unable to	This publisher excluded from all
	provide Google Books data	Google Books analysis





	from October 2013 onwards	
	One publisher unable to provide Google Books data for one title in first three months of project	This pair excluded for first three months of Google Books analysis
Sales	One publisher did not sell electronic versions of the open access books	This publisher excluded from electronic and total sales analysis
	One publisher started selling e- books in June 2013	Electronic sales data included in analysis from June 2013 onwards

Findings

Throughout this section we sometimes use ratios to illustrate the difference between experimental and control group sales and usage. The number in the figures shows, for each sale or use of a control group title, how many sales or uses of the paired experimental group title occurred. If the number is greater than one, the experimental title outperformed its control group partner; if the number is less than one, the control group partner outperformed the experimental title.

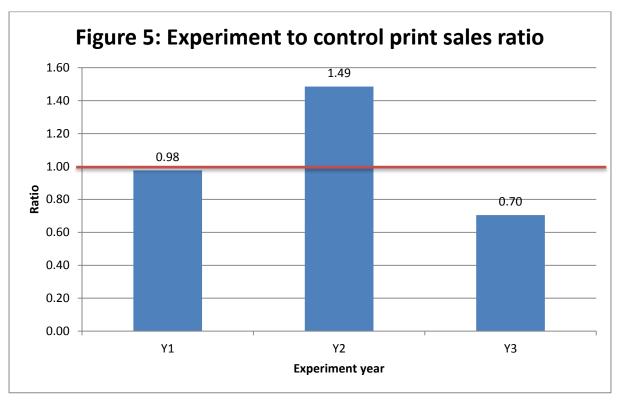
This is a helpful method of providing easy visual comparisons, but it is susceptible to problems where the raw data consists of very small numbers. For monographs after the first year of publication sales numbers tend to fall and are generally quite low so we do have some small numbers which have affected how the ratios appear. Findings should therefore be treated with caution.

The effect of open access on print sales within our sample was insufficient to overcome the known variability of monograph sales

Figure 5 shows the ratio of experimental to control group title sales across all five original publishers, and Figure 6 shows the raw total sales across all five original publishers. Figure 7 shows the ratios broken down by publisher. There is no clear pattern over the three years of the project, suggesting that open access did not affect sales sufficiently to overcome the generally noisy print sales data.







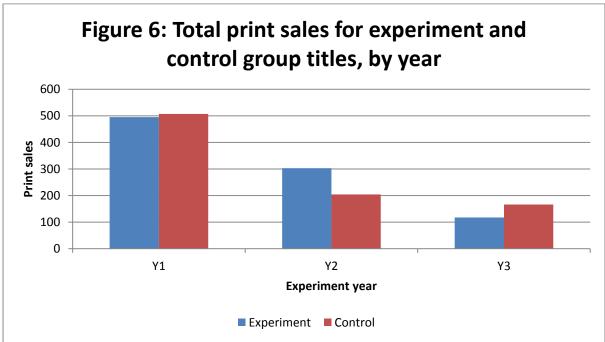
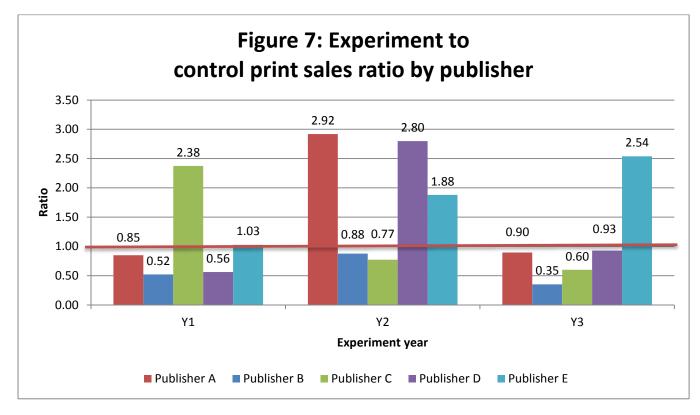


Figure 7 illustrates the caution with which publishers should treat data emerging from their own open access experiments. Looking just at their data, Publisher C might conclude that, over the long term, open access has a very negative effect on print sales, while Publisher E might conclude the exact opposite. Neither would be wise to base a long-term business plan on such a small sample of data, but in the current environment they are unlikely, on their own, to have significantly larger numbers of books to compare.



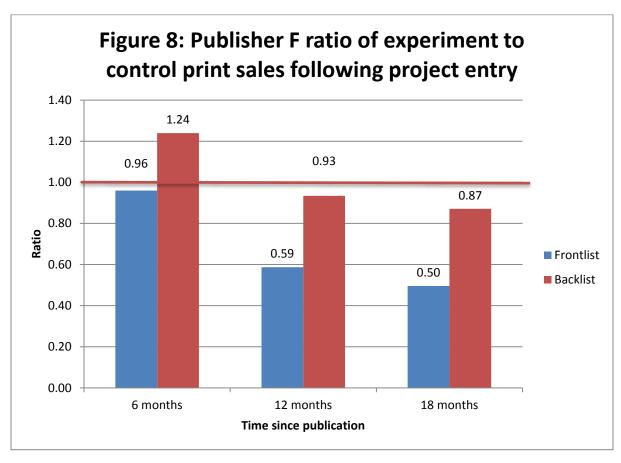




The sixth publisher was able to supply data on a monthly basis. They included six pairs of frontlist titles within the project – these titles entered the project immediately upon publication. They also included six pairs that we consider to be 'backlist' titles – more than twelve months old when they entered the project. Figure 8 shows the ratio of experimental to control print sales for the frontlist and backlist groups at 6, 12 and 18 months after project entry. This shows that for both frontlist and backlist titles experiment group sales declined relative to print group sales over the time period, suggesting that perhaps there might be an effect on open access for print sales, more pronounced for frontlist titles. But again the very small sample means further investigation would be needed before any business models were based upon this finding.







Electronic sales were not large for either the experiment titles or control group titles and do not make any difference to overall findings here.

Usage on the OAPEN platform has increased during the project, with considerable variability

Figure 9 shows average PDF downloads from the OAPEN library from September 2011 to August 2015 for all five original publishers. Figure 10 shows the same data broken down by publisher, and

also shows the data for Publisher F. Over the project period, and the year after it ended, PDF downloads of the titles has increased, although there is considerable variability. This data is not COUNTER compliant, but OAPEN implemented COUNTER standards in July 2013. Figure 11 shows the COUNTER and non-COUNTER compliant data for the period after COUNTER standards were implemented, and show similar overall patterns of use for both collection methods, but numbers were lower and variability less pronounced for the COUNTER data.

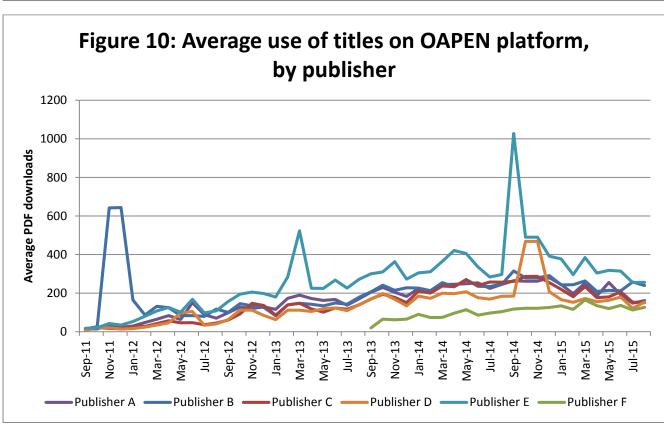
COUNTER (Counting Online Usage of Networked Electronic Resources) is an international initiative serving librarians, publishers and intermediaries by setting standards that facilitate the recording and reporting of online usage statistics in a consistent, credible and compatible way.

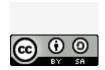
http://www.projectcounter.org/abo ut.html



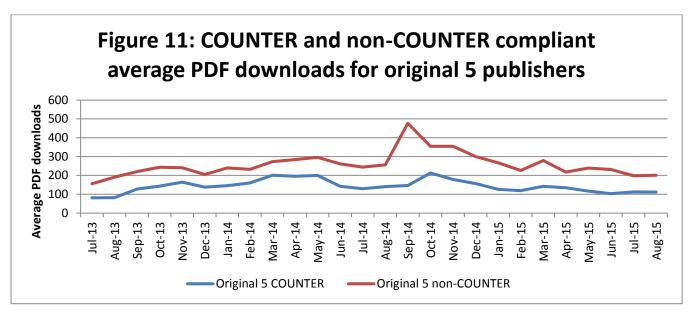


Figure 9: Average usage of titles on OAPEN platform: original 5 publishers 600 Average PDF downloads 300 200 100 0 May-15 Jan-13 Mar-13 Mar-15 Nov-11 Mar-12 May-12 Jul-12 May-13 Sep-13 Nov-13 Mar-14 Sep-14 Nov-14 Sep-12 Nov-12 Мау-14 Jul-14



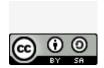




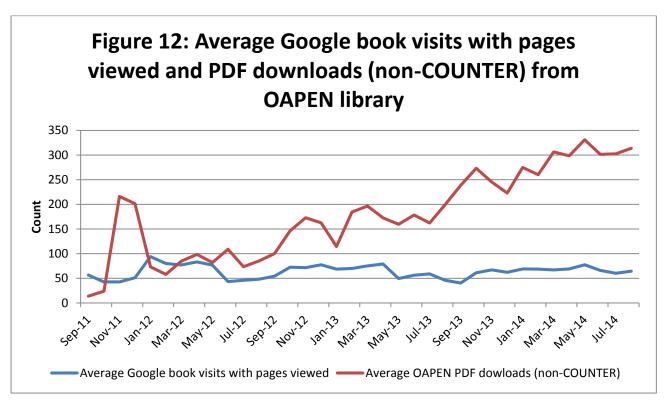


Usage on the OAPEN platform grew relative to usage on Google Books for most publishers during the project

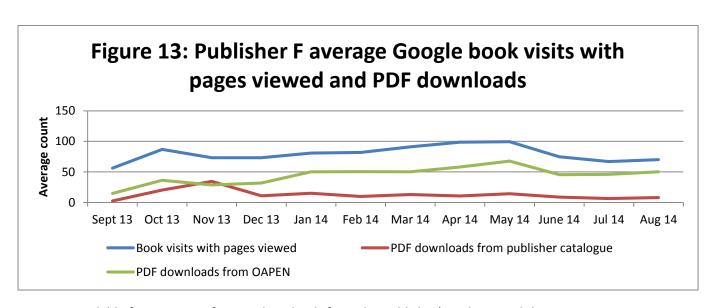
Figure 12 shows average non-COUNTER PDF downloads from the OAPEN library and average Google Book visits with pages viewed over the three years of the project. We have only included titles for which we have both sets of data (see Table 4 for exclusions). We use Google Book visits with pages viewed as the most directly comparable measure to the OAPEN PDF downloads because for both measures the user must click through from the title's landing page. The data shows steady growth in use on the OAPEN platform, and much flatter use on Google Books over the course of the project.







The data for Publisher F in Figure 13 looks rather different. Comparable data for Google Books, OAPEN and the publisher's own website are available for one year of the project, and they show that average use is consistently higher on Google books than on either OAPEN or on the publisher's own sales catalogue. This publisher also made the title available on its library e-book platform; it appears to have had much more usage here, but the data are not sufficiently reliable to include within this analysis.

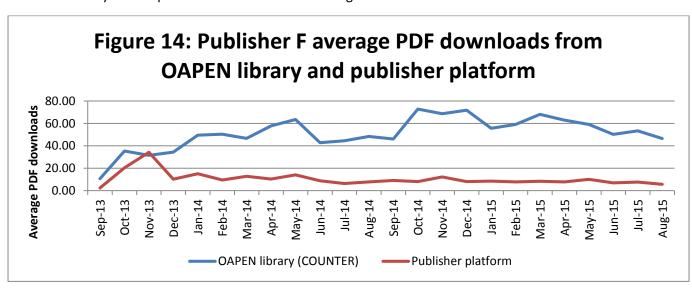


Data are available for two years for PDF downloads from the publisher's website and the OAPEN library. Figure 14 shows that average use on the OAPEN website grew relative to use on the





publisher's own platform over this period. The spike in use in November 2013 on the publisher's website is largely due to the exceptional performance of one title; this title was heavily promoted on various social media by its authors via a link to the publisher website. Again, usage on the publisher's dedicated library e-book platform seems to have been higher.



Open access does not seem to have an effect on whether a book is discovered, but it may have an effect on how much of the book is accessed

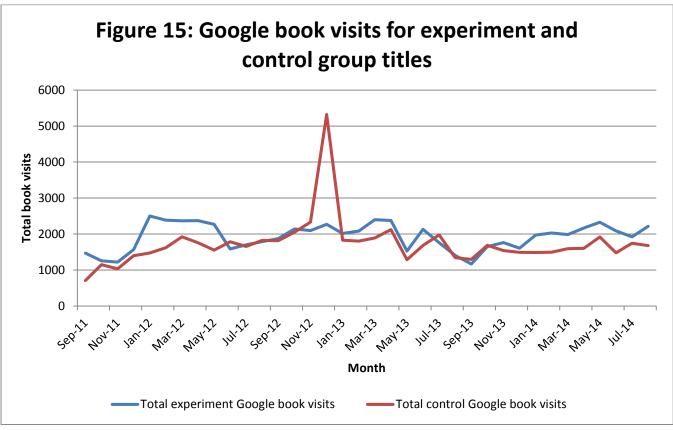
Figures 15, 16, and 17 show how titles were used on Google Books. We have only included data from the four publishers who could supply it for the entire three year period (one pair was excluded for the first three months).

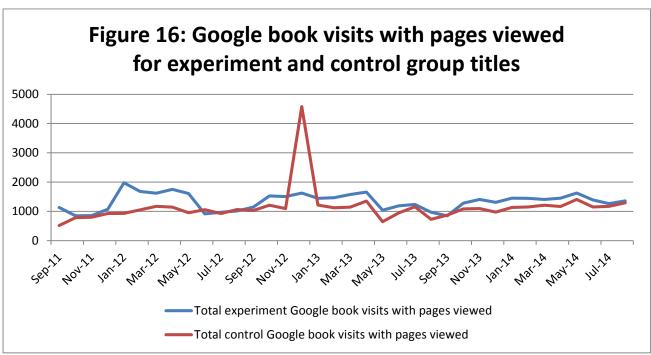
Figures 15 and 16 show very little difference between the experiment and control group titles, apart from the big spike for the control group in December 2012 which relates to a single title – the publisher of the title was not sure what might have caused this spike. But Figure 17 shows a much more noticeable difference; the total number of pages viewed for experiment titles over the course of the experiment were around 1.8 times the number of pages viewed for control group titles.

The patterns of total page views in particular seem to show some synchronisation with the academic year, with usage rising each September – April and falling over the summer months. This might indicate that students are important users of titles on Google Books; certainly this would tally with librarians' perceptions that students favour the convenience of and make high use of electronic content.



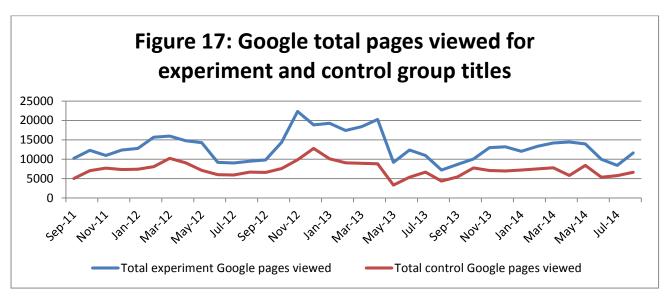












Conclusions

It is well known that usage studies based upon online activity data cannot always tell us much about what a user is actually doing with content. This study is no different: we do not know if users read the PDFs they downloaded or the pages that they viewed, or even that all the users recorded are human.

This, along with the limitations outlined earlier in our report, means that we would caution against taking these results as universally-applicable evidence of what happens when a book is made open access. But bearing the limitations in mind, we conclude that:

- in the current environment offering a small number of titles in open access is unlikely to affect print sales, although this may change as open access for monographs develops and becomes more common;
- use of open access books appears to be growing on the OAPEN library, relative to other sites;
- open access does not seem to increase the discoverability of monographs, but it does seem to increase the amount of a monograph that is accessed;
- students may be an important user group for open access monographs.

Overall, given that open access is likely to remain a small proportion of most traditional publishers' business and that the effect of open access on sales appears to be low in the current environment, we believe that this is an opportune time for publishers to experiment with a variety of open access business models for monographs. A number of funders, including research councils, the EU and the Wellcome Trust, are offering funding to support open access monograph publishing, and support is also available from cross-institutional projects such as Knowledge Unlatched. Given the increasing emphasis on open access for research outputs in the UK and internationally, academic publishers should find it possible, and valuable, to explore this area.





Further research

Our findings indicate some areas where further research, both quantitative and qualitative, might be useful.

- Where are users discovering and using open access monographs? A study could look at a corpus of titles on, for example, the OAPEN library and work with publishers, aggregators and, if possible, vendors to collect information about user behaviour. Working with a greater range of suppliers would help test our initial finding that use seems to be increasing on dedicated OA monograph sites rather than in other destinations. The study should look at front and backlist titles, and also at titles which have been available on the open access site (for example, OAPEN) for different amounts of time. It should also consider whether developments such as the ability of Google Play to display books at zero price have affected the discoverability of open access content.
- How do students use monographs? This study could take a number of approaches, but the objective would be to understand how important monographs are for teaching and learning in the arts, humanities and social sciences in order to consider how open access for monographs might benefit students. Components might include an analysis of course reading lists; student surveys; a survey of course conveners and leaders; analysis of library borrowing data; analysis of usage data held by library suppliers etc. The work should build on that undertaken by the Jisc e-books observatory project, by KB+ for e-books and COPAC CCM.

Publisher lessons learned

At the end of the project, publishers who participated in OAPEN-UK shared some of their experiences and advice that they think would benefit other publishers considering a move into open access for monographs. We present some of these thoughts below.

"The sales and usage data required for the project had to be sourced from a number of places and involved a number of colleagues/departments to collate. Combined with some issues with the Google Books data, this was a more challenging aspect of the project than expected."

"For established publishers the biggest challenge is adapting systems and processes – which have historically been designed and built to deliver paid-for and printed content – so that they can deliver open access content. A major part of the work involves identifying each aspect of the book publishing process that will be affected by OA – from commissioning, third party rights and contracting, to production, distribution, marketing and sales – and adjusting systems and processes accordingly. Metadata must also be updated for each platform on which monographs are available."

"For larger publishers, there is a need to consider staff training so that representatives from all relevant departments are aware of process differences for OA monographs, which are likely to be a small proportion of content overall for the foreseeable future."





"A major challenge was increasing author awareness of the implications of publishing third party materials in open access works published under Creative Commons licences and the difficulty of obtaining the necessary permissions to do so."

"We would recommend that other publishers take time to consider the end-to-end processes and how they impact on sites and systems. It is worth thinking issues through beforehand rather than implementing fixes later on. We also recommend keeping authors well informed on aspects of a "Gold" open access model, particularly around Creative Commons licensing and exactly what this means for their research so that they can make informed choices, with the support of their editor. Rely on the skills and wealth of knowledge of your commissioning editors, production and digital teams to work through the practical issues, and make sure you have adequate data reporting in place so you can build that evidence and knowledge base."

"On a practical level we would advise other stakeholders to take time to discover and communicate specific requirements for open access publishing: funders and institutions can support authors by making their requirements clear and upfront. For authors, we would hope there is sufficient information out there to understand options and requirements for publishing your research; particularly on issues of licensing, copyright and the implications of these for reuse and sharing. Any remaining questions on open access should be raised with the editor early on – certainly before you sign any contract to publish. As publishers, we will endeavour to offer as much support as we can to authors, but there is very much a role for more local provision from dedicated OA advisors at institutions, particularly around issues of contracting. There is certainly a need for all stakeholders to communicate well and collaborate on the issues of OA monographs to find sustainable long-term solutions"

"We would recommend that you inform yourself as widely as possible of the implications and issues from the perspective of the publisher/author/institution/to inform judgments and choices, and what is realistic from a business model perspective."

"Consider OA as a viable alternative that can work alongside traditional monograph publishing models, but remain flexible when considering policies and processes."

"The digital supply chain was shown not to cater adequately for OA books. Many vendors (including library aggregators) are technically unable to host books at zero price and there was a lack of appetite for OA books on those platforms in general. If wider dissemination is one (of many) desired outcomes of OA publishing, the wider digital supply chain needs to be engaged more effectively in future."



OAPEN-UK Matched Pairs Pilot: Final Report by OAPEN-UK is licensed under a <u>Creative</u> Commons Attribution-ShareAlike 4.0 International License.