

How small open access monograph presses can make the most of an increasingly rich data landscape



*Until relatively recently the ability to exploit new data for open access books was restricted to large publishers or content aggregators with the resources to invest in its collection, management, and analysis. However, **Lucy Montgomery, Cameron Neylon, Alkim Ozaygen and Tama Leaver** describe how barriers to engaging with data are falling, with open access monograph publishers now having growing access to data relating to usage and engagement. Such readily available data can help smaller OA publishers understand how individual titles are performing, where scarce promotion resources might best be deployed, and how a press is performing on its social mission.*

One of the many benefits of a shift from print to digital distribution is the growing availability of data about how research outputs are used. This data has the potential to help researchers and publishers understand the processes, audiences, and relationships involved in scholarly communication in new ways. But capturing and managing usage and social media data remains a real challenge for many publishers – especially for those working in the open access (OA) monograph space.

In May 2016 [UCL Press](#) invited [Knowledge Unlatched Research](#) and the [Centre for Culture and Technology](#) at Curtin University to engage 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 was 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. The data used in the study reflected the first year of UCL Press's operations.

The results of our study are [available here](#). They suggest that readily available data and low-cost approaches to its aggregation and analysis can provide useful development and strategy insights for OA monograph presses.

Why is data a challenge for OA monograph publishers?

In contrast to the natural and medical sciences, which are dominated by a handful of large publishing houses, humanities and social sciences (HSS) publishing is characterised by a lack of commercial concentration among publishers, with small to medium-sized presses playing a key role in meeting the research and communication needs of HSS communities. The involvement of many small mission-focused players in the production of the core outputs is a strength for HSS research communities. However, smaller presses often find the volume and diversity of data available to them a challenge.

While it is clear that sales data are an insufficient measure for the value and performance of presses publishing freely accessible books (or of the books themselves), identifying data relevant to the day-to-day work of an OA press, and how it should be captured, managed, and interpreted, remains complex.

As a result, the ability to exploit new data for OA books has been restricted to large publishers or content aggregators with the resources to invest in its collection, management, and analysis. However, barriers to engaging with data are falling.

What kinds of data are available to OA monograph presses?

OA monograph publishers have growing access to data relating to use and engagement. This data is available to publishers via their own websites, publicly available social media, and the hosting platforms they work with. In our study we focused on:

- Download statistics made available by platforms and repositories hosting books
- Google Analytics data
- Social media data

We cross-referenced this against key date and promotion information provided by the press; dates of book reviews published in major news outlets, or the launch of a MOOC relating to a title, for example.

How did we gather and interpret the data?

In this study usage data was collected and managed with R scripts. Social media data was gathered with a freely available open source tool, [DMI-TCAT](#), albeit one that requires the capacity to set up a server to use.

Graphs and time plots were created using [Tableau](#), a proprietary but relatively cheap tool for data management and visualisation. 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](#) and more limited but familiar tools like spreadsheets. Other easy-to-use tools for creating map-based visualisations are also available.

What kinds of questions did we ask of the data?

A press publishing OA monographs may have many different motivations and needs with respect to analysing and using data about books. These range from tactical issues of whether to invest time in social media promotion, or in hosting OA content on multiple platforms, to large-scale strategic questions of the balance of a title portfolio and its evolution. In our study we simplified these into three questions:

1. How is this book doing?
2. What promotion strategies are effective?
3. Is the press delivering on its mission goals?

What did we find out?

The data available provided useful insights into the three questions we asked in this study. We also identified relatively simple steps that presses can take to increase the richness of the data available to them, and the level of insight it can provide. These include using tagged social media links in promotion campaigns, and providing authors with tagged links they can use when promoting their own books via social media. Ensuring platform partners are aware of the value of granular usage data for presses is also important: monthly download figures are less useful than daily, or even hourly, download data.

What should presses do?

The data arising from digital distribution of OA monographs presents real opportunities to better understand how people find and access HSS books, and the most effective strategies for widening reach and impact. Readily available data has the potential to help publishers understand how individual titles are performing, where scarce promotion resources might best be deployed, and how a press is performing on its social mission. The challenges of making the most of this data are real; finding the resources to capture, manage, and engage with data that arrives in different formats may be especially daunting for smaller presses. However, these challenges need not be insurmountable.

Some university-based OA presses may be able to engage more effectively with data by working with campus partners including the library and IT department. For others there may be value in exploring the formation of cooperatives with like-minded or complementary presses in order to gain access to economies of scale associated with 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 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 blog post is based on the authors' article, "[Getting the best out of data for open access monograph presses: A case study of UCL Press](#)", published in *Learned Publishing* (DOI: 10.1002/leap.1168).*

*Featured image credit: [bady qb](#), via *Unsplash* (licensed under a [CC0 1.0](#) license).*

Note: This article gives the views of the authors, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our [comments policy](#) if you have any concerns on posting a comment below.

About the authors



Lucy Montgomery is Director of the Centre for Culture and Technology at Curtin University, Australia. She is also Director of Research for Knowledge Unlatched Research: a close-knit team of researchers and publishing industry practitioners working together to help realise the possibilities of digital technology and open access for specialist scholarly books and the communities that care about them. Her ORCID iD is: [0000-0001-6551-8140](https://orcid.org/0000-0001-6551-8140).



Cameron Neylon is Professor of Research Communications, in the Centre for Culture and Technology at Curtin University, and Director of Knowledge Unlatched Research. He was a founding Director of FORCE11 and a contributing author to the altmetrics manifesto, the Panton Principles for Open Data and the Principles for Open Scholarly Infrastructure. He has been a biochemist and a technologist, worked in scholarly publishing as an advocate for open access, and now focuses on studying the changing cultures and institutions of the academy. His ORCID iD is: [0000-0002-0068-716X](https://orcid.org/0000-0002-0068-716X).



Alkim Ozaygen is a Data Analysis Tool Developer at Curtin University's Centre for Culture and Technology, where he is completing a PhD exploring the uses of open access books. He is also a technical advisor to Knowledge Unlatched Research. His ORCID iD is: [0000-0001-6813-8362](https://orcid.org/0000-0001-6813-8362).



Tama Leaver is an Associate Professor of Internet Studies at Curtin University. His research interests include online identity, social media, digital death, infancy online, mobile gaming, and the changing landscape of media distribution. His ORCID iD is: [0000-0002-4065-4725](https://orcid.org/0000-0002-4065-4725).