

# Use of Institution Data Analysis for Publisher Negotiations

---

**Niamh Malin**

*University of Cambridge*

## **Introduction**

In recent years there has been an increase in transformative deals being negotiated between publishers and academic institutions. This is in response to a combination of factors, including the considerable focus on open access policies and the desire to make the transition to open access as seamless as possible for researchers. Publisher negotiations in the United Kingdom have been supported by Jisc, a UK agency focused on tertiary education, research, and innovation (Vernon, 2022), however institutions can individually negotiate with a publisher where they feel necessary. Quantifying how negotiation outcomes could impact the institution can empower library staff to confidently negotiate towards a transformative agreement. Data is powerful, and the data exchanged between institutions and publishers is ever growing - such as publishing quantities, read/usage, and financial data, and these can be assessed over time, against one another, or in comparison to other publishers. This paper documents the journey of the author as the data analyst at the University of Cambridge library, who was responsible for analysing the data relationship between the institution and the publisher Springer Nature as part of the 2022/23 negotiations.

The author researched the background of publisher negotiations, before gathering and presenting relevant data between the University of Cambridge and Springer Nature. Once the negotiations were underway the author began exploring the repeatability of their work, how it could be applied to additional publisher negotiations, and how the approach could be adapted by other institutions. This paper will explain the background, methods, and future of the data explored as part of the University of Cambridge's negotiations with Springer Nature.

## **Background**

As negotiations become more prominent, so too does the body of literature to support librarians undertaking this process. Institutions have shared their journey, conference talks have addressed the topic, and organisations who support education and libraries have created resources to support these negotiations, many of which will be discussed.

The open access advocacy organisation SPARC has developed resources for supporting publisher negotiations, including 'Data Analysis for Negotiation' which was created by their community data analysis working group (SPARC, 2021). It assesses a wide variety of data sources, both open and subscription databases, to provide an understanding of the data and databases available. It describes appropriate applications of usage statistics, and explores the multitude of sources, aggregators, and library management systems which may provide these statistics. The data visualisation resource collates a variety of research and guidance created by institutions and their libraries to visualise library data. This resource is extensive in scope and provided an excellent starting point for understanding usage data and journal analysis.

ESAC is an open community of information professionals dedicated to putting the vision of open access to research into practice, and the ESAC data analytics working group published 'Uncover the publishing profile of your institution' (ESAC, 2020). This resource provides a breakdown of questions to explore, where to collect the data from, how to clean the data, how to enrich it with other sources, methods of analysing data, and finally how to communicate the analysis with stakeholders. This accessible approach, with additional resources as needed, provided extensive support for tackling publications data and terminology.

For those working in smaller libraries, or with smaller publishers, cOAlition S and the Association of Learned and Professional Society Publishers (ALPSP) partnered to develop a toolkit specifically for 'Enabling Smaller Independent Publishers to participate in OA Agreements' (Information Power, 2021). While not used for this project, the toolkit includes a data

template to help publishers collect the information needed to inform negotiations and provides librarians with an insight into the data relevant to publishers, and thus relevant to their work too.

The Office of Scholarly Communication at the University of California has been prominent in sharing their publisher negotiations experience and insights, publishing multiple articles and resources to support others undertaking a similar journey (Office of Scholarly Communication, 2020). In May 2019 the University of California created a toolkit for 'Negotiating with Scholarly Journal Publishers' as a direct response to their 2018-19 negotiations with Elsevier (Office of Scholarly Communication, 2019). Broader in scope than the ESAC and SPARC's resources, this toolkit discusses the negotiation goals and strategy, the communications planning and execution, the role of data analysis, and how to plan a walkaway (alternative access). Focusing on the role of data analysis, the toolkit provides seven recommendations to approaching the data, including "verify publisher-provided data with local analyses", alongside valuable insights from their Elsevier negotiations.

Following from their work at the University of California, Mathew Willmott and Jeffrey MacKie-Mason presented at the SANLic 2019 Conference, as it had a "strong focus on the transition to open access and transformation agreements" (South African National Library and Information Consortium, 2019). MacKie-Mason spoke about 'How the University of California prepares for negotiations', focusing on the formation of teams, the division of roles, and the preparation for Elsevier negotiations (MacKie-Mason, 2019). Whereas Willmott presented 'Tools for Transitions: Data Analysis for Negotiations', demonstrating methods and strategies employed to support multiple publisher negotiations (Willmott, 2019). Both talks provide practical insight to the University of California's work, which, alongside the aforementioned resources, was applied to the University of Cambridge's data approach to negotiations with Springer Nature.

In 2020 Jisc supported UK institutions to negotiate the Elsevier ScienceDirect Journal Agreement, and the University of Cambridge formed an internal working group to engage in the negotiations, assess the relationship with the publisher, and understand the impact any new agreement would have on the staff and students at Cambridge. Dominic Dixon, then Research Librarian at Cambridge University Libraries, was responsible for exploring and presenting the data, and published an internal Tableau dashboard to answer questions of the relationship with Elsevier (Dixon, 2021). The graphs and data produced enabled staff to understand the university's use of and contributions to Elsevier, to understand the magnitude of the relationship with Elsevier, and demonstrated the importance and impact of the negotiations. After successful negotiations with Elsevier, the working group shifted focus to the Springer Nature read and publish negotiation.

The negotiations between the publisher Springer Nature and UK institutions began in April 2022 (Jisc, 2022), and the University of Cambridge's negotiations working group divided the workload into three strands: communications (internal and external), 'Plan B' (preparing alternative access), and data analysis. The author supported the working group with a variety of data tasks and analysis, such as answering similar questions to the Elsevier dashboard to establish the relationship with Springer Nature.

- What percentage of research published by Cambridge authors is published by Springer Nature?
- Who is funding the Cambridge research published by Springer Nature?
- Over five years, what is the open access breakdown of Cambridge publishing?
- How frequently are Springer Nature publications cited by Cambridge researchers? Are they open access?
- What percentage of usage by Cambridge individuals is within Springer Nature journals?
- How does Springer Nature compare to other publishers financially?
- Which Cambridge schools could be most impacted by a walkaway scenario?

These questions, and others, were presented in a Microsoft PowerPoint to the working group, academic staff, and senior management, to provide an understanding of the universities relationship with Springer Nature will be explored through this paper.

## **Methodology**

The negotiations working group required the relationship between the University of Cambridge and Springer Nature to be quantified using data, to provide an understanding the impact that walking away from Springer Nature could have on the university.

### *Question the Question*

The author developed a workflow based on the cross industry standard process for data mining (CRISP-DM) to address each question as a data report, cycling through phases of understanding, preparation, modelling, evaluation, and deployment (Hotz, 2023). First the author established a set of questions to understand the data report.

1. What are the data elements?
  - Is the data point quantitative (usage, financial, time, quantity) or qualitative (feedback, interactions).
  - What relationships could be explored.
  - What should be ignored/filtered.
  - How will it be compared/contextualised.
2. Why is this data being gathered?
  - Is there an expected outcome.
  - Who will it be for, and how should it be presented.
  - Where will the output be stored and/or shared.
3. What type of data will it be?
  - Does the data need to be live, continually updated, or static.
  - Is it internal, external, or open data.
  - Is it a one-off report, or will it be repeated.

These three questions (and their related sub-questions) ensured an understanding of the data report, with flexibility to evolve alongside project timelines, priorities, budget, and staffing. It proved vital to understand the initial requirements of a data report and made it easier to adjust with changes along the journey.

Showing the quantity of published outputs with Springer Nature in 2022 provides no meaning as a single number, it requires comparison to previous years or other publishers within the same time frame, to contextualise the data. For the negotiations, Springer Nature was often compared to Elsevier for two reasons, first the publishing and usage of Springer Nature and Elsevier by the University of Cambridge is similar, and second, because the Elsevier negotiations had happened the year prior, the Elsevier data was significant to the audience. When contextualising publisher-institution relationships it could compare the time frame, similar publishers, the subject area, or financial commitments.

### *Data Sources*

Libraries and publishers each produce extensive quantities of data, and this project focuses on publishing outputs of the university with the publisher, and usage data by the university of the publisher's material. After researching the limitations and possibilities of a variety of data sources, and with special regard to uncovering the relationship between the University of Cambridge and Springer Nature, the use of Dimensions for publications data and COUNTER through JUSP for usage data was established.

An initial barrier when preparing the data was a lack of comprehensive metadata for the Springer Nature journals. Metadata could be found through platforms such as Unsub<sup>i</sup>, Portico<sup>ii</sup>, Knowledge Base Plus (KB+)<sup>iii</sup> and various publisher websites but combining the metadata with Dimensions<sup>iv</sup> and COUNTER reports<sup>v</sup> often proved difficult. During a task to manually check each Springer Nature journal editorial board to identify University of Cambridge academics (as Open Editors could not provide the

information<sup>vi</sup>), the author documented the journal title, imprint, and relevant URL. The resulting list of journals was combined with metadata from open sources (such as Portico and KB+) to create a Springer Nature dataset, totalling 4,676 journals. It was openly published on the University of Cambridge repository to aid other librarians in the publisher negotiations with Springer Nature (Malin, 2022). It is important to note that the published Springer Nature dataset does not include agreement information, as this will vary per institution and over time, and therefore was held within a local copy of the dataset for analysis, but it does include active years, (alternative) title(s), (Unsub) ISBN(s), title ID and subject area.

Dimensions is “the world’s largest linked research database” (Digital Science & Research Solutions, 2023), to which the University of Cambridge have a paid analytics service subscription, and it was used to export large publications datasets for this project. When creating datasets initially, there was often a focus on ‘article’ quantities, and therefore datasets were filtered to only article outputs. However, it was later noted that arts and humanities research is more likely to produce book publications, whereas STEM research favours article outputs (Michalikova, 2022). Filtering to ‘article’ types caused an unintentional bias which was not required or intended, and thus each data report was refreshed to include all outputs with the publisher, which could then be filtered intentionally during analysis if appropriate (e.g., to exclude preprints).

Limitations to the Dimensions analytics service include exports of maximum 50,000 rows, set columns exported, and it does not filter on every available aspect of the publication. Filtering by number of authors or if an output was cited by another is unavailable within Dimensions analytics but can be explored using the Dimensions API service<sup>vii</sup>. The API service does require a subscription as it utilises coding (dimensions search language) to query the Dimensions database directly through Jupyter notebooks, the GitHub repository or Google Colab notebooks (Digital Science & Research Solutions, 2020). As part of the publisher

negotiations, two API calls were developed using Google Collab, to extract nuanced datasets from Dimensions.

- Publication Analysis: Create a publications dataset within a combination of customisable variables (including research field and funder) (Malin, 2023).
- Citation Analysis: Create a dataset of publications cited in a custom publication’s dataset (Malin, 2023).

Publishing these API calls was supported by Dominic Dixon (University of Cambridge), Eric Schares (Iowa State University), and the extensive Dimensions API guidance available. These API calls were developed to be approachable by any (data) librarian seeking publications datasets, and thus can be used with no coding experience or manipulated further by those with experience. Additionally, to support (data) librarians further, the API calls include a variety of table summaries for single-element questions, such as “Which are the most popular journals?” (Figure 1).

4.3 Which are the most popular journals?

Export the journal name with the number of publications, and publisher.

```
[ ] journals = pub_list.value_counts(['journal.title', 'journal.id', 'publisher'])
journals = journals.to_frame().reset_index().rename(columns= {'citations': 'journal' : 'title' })
journals.index.name = 'index'

#Export data to CSV
datetime_object = datetime.datetime.now() #this will add the current time to the file name
save_as_csv(journals, f"4.3_Pub_Journals_{GridID}_{datetime_object}")

#preview top 10
journals.head(10)
```

===  
Saved: API\_CSV\_Extraction/3.2\_Pub\_Journal\_grid.5335.0 2023-04-04 14:28:20.192752.csv

index	journal.title	journal.id	publisher
0	bioRxiv	jour.1293558	Cold Spring Harbor Laboratory
1	Research Square	jour.1380788	Research Square Platform LLC
2	medRxiv	jour.1369542	Cold Spring Harbor Laboratory
3	Cancers	jour.1043163	MDPI
4	PLOS ONE	jour.1037553	Public Library of Science (PLoS)
5	Nature Communications	jour.1043282	Springer Nature
6	Frontiers in Immunology	jour.1045166	Frontiers
7	Communications Biology	jour.1300829	Springer Nature
8	BMJ Open	jour.1044693	BMJ
9	Frontiers in Physiology	jour.1044459	Frontiers

Show 25 per page

Figure 1. Publications Analysis API: Quantifying the most popular journals from a publication’s dataset.

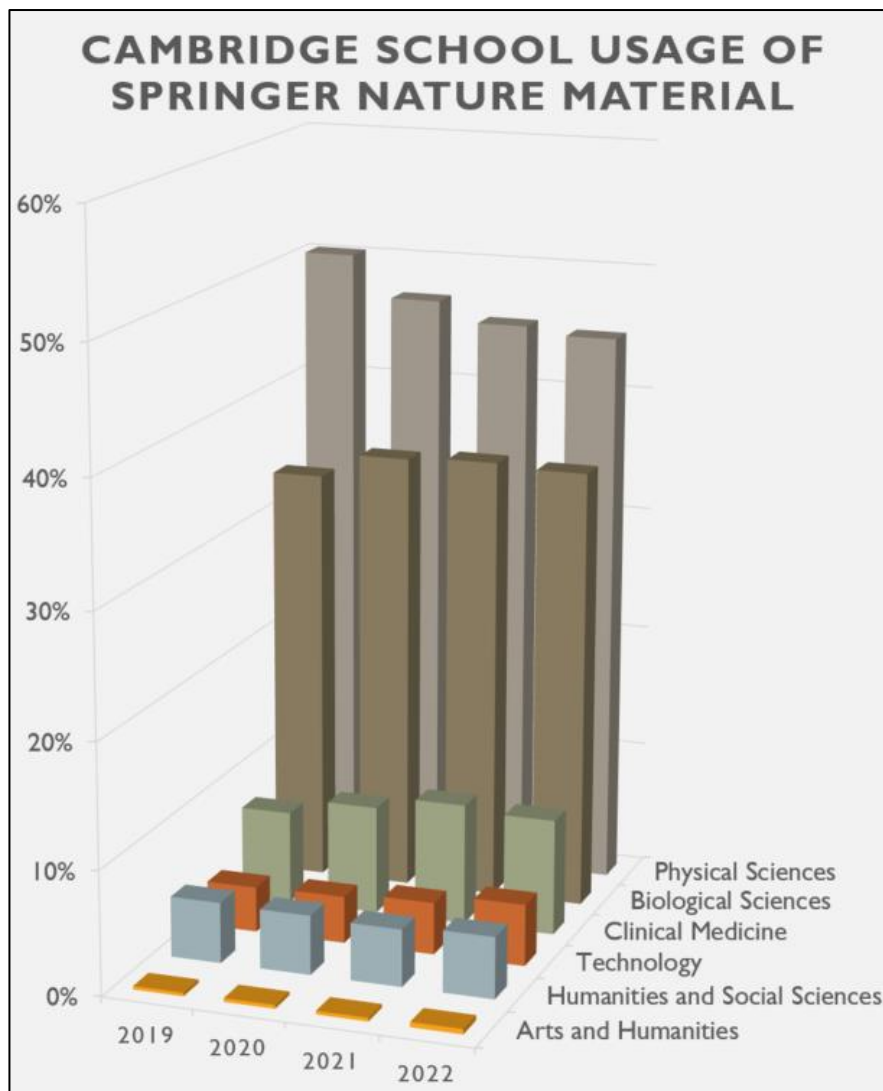


As for usage data, Jisc's Journal Usage Statistics Portal (JUSP) provides COUNTER compliant usage data and reports<sup>viii</sup>. Employing the COUNTER Master Report allows for "highly detailed and granular" datasets, versus the pre-filtered standard reports, and encourages a greater understanding of the data by exploring the data types, access methods and metric types (Jisc, 2020). COUNTER data does not distinguish based on the individual(s) using the material, therefore it does not separate staff, academic or student usage. COUNTER data can distinguish per publisher and/or platform, however the Springer Nature Group encompasses multiple brands/publishers, and therefore to download all relevant data for the negotiations, reports for each of the following was exported; 'Nature Journals', 'Scientific American', and 'SpringerLink (including Biomed Central)'.

### *Cleaning and Analysing*

Microsoft Excel was the predominant tool used to prepare and model data for this project, through the creation of tables, clearing excess quotation marks and space, uniform formatting, and deleting or hiding superfluous columns (such as abstract data). Utilising formula and tools effectively enabled the datasets to be explored and modelled more efficiently.

To gain a deeper understanding of the data from Dimensions or COUNTER, the formula XLOOKUP was utilised to connect to the Springer Nature Dataset. Combining two or more datasets enabled an analysis of the publisher-institution relationship, through quantifying outputs per journal in the deal being negotiated, or the journals used per discipline, and thus understand which University of Cambridge school could be most impacted by a walkaway (Figure 2). Understanding the impact a walkaway would have on STEM schools enabled the working group to focus communications and engagement with the relevant cohort.



*Figure 2. Combining datasets enabled a deeper understanding of usage per University of Cambridge school.*

Pivot Tables within Excel summarised data within chosen filters, rows, or columns, and therefore, instead of assessing a 50,000-row list of research outputs, a pivot table could count annual quantities per open access type, in a concise adaptable table – which could easily be turned into a graph. While possible to create a table using COUNTIF formulas, this can be tedious, manual, and would not provide space for further manipulation. Using pivot tables allowed for a greater exploration and analysis of the data, as comparisons and filters could be easily adjusted, and multiple pivot tables could relate to a single dataset.

After moving the superfluous metadata from a COUNTER report to a summary tab, and the usage data transformed into a table, the data was multiple columns wide when broken down per month, which makes it difficult to analyse different journals across multiple months (Figure 3). To change the COUNTER data into a longer table, the Unpivot function (via Excel power query editor) was used. Unpivot (or flattening) transforms transform columns into attribute-value pairs, where columns become rows. For example, if a table was 13 columns wide (journal name and twelve months), using unpivot transforms it to three columns of journal name, month (as text) and the usage data. Once COUNTER data had been unpivoted, it became easier to manipulate and analyse as part of the data reports (Figure 4).

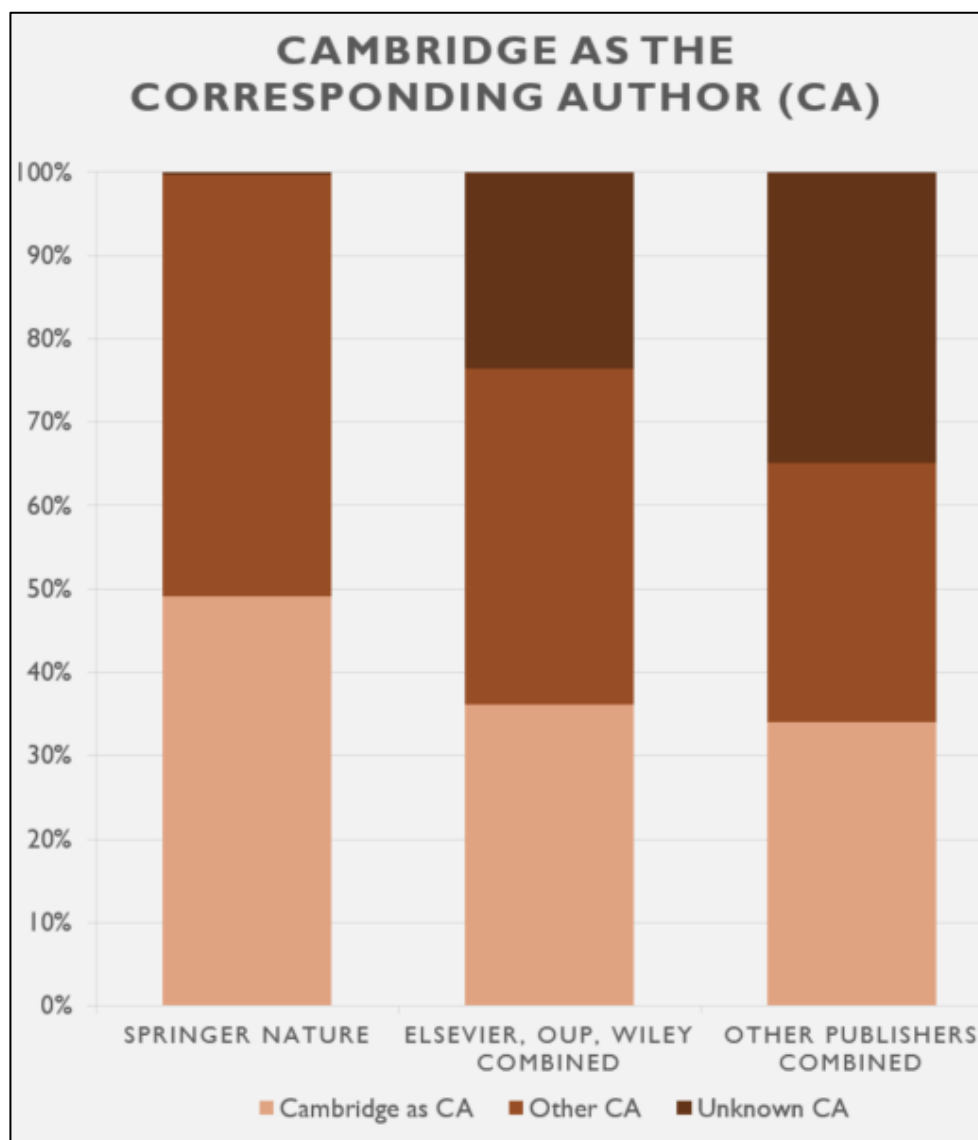
Title	Metric Type	Jan-22	Feb-22	Mar-22	Apr-22	May-22
Journal A	Total_Item_Requests	0	0	0	0	0
Journal B	Total_Item_Requests	0	7	0	1	0
Journal C	Total_Item_Requests	6	14	31	17	42
Journal D	Total_Item_Requests	0	0	0	0	3

Figure 3. COUNTER data in a table format.

Access Type	Month	Total Item Requests
Journal A	Jan-22	0
Journal A	Feb-22	0
Journal A	Mar-22	0
Journal A	Apr-22	0
Journal A	May-22	0
Journal B	Jan-22	0
Journal B	Feb-22	7
Journal B	Mar-22	0
Journal B	Apr-22	1
Journal B	May-22	0
Journal C	Jan-22	6
Journal C	Feb-22	14
Journal C	Mar-22	31
Journal C	Apr-22	17
Journal C	May-22	42
Journal D	Jan-22	0
Journal D	Feb-22	0
Journal D	Mar-22	0
Journal D	Apr-22	0
Journal D	May-22	3

Figure 4. Unpivoted COUNTER data.

While each dataset was a direct result of a data report, there was space to explore the data to understand if other questions could be answered (or if the dataset would correctly answer the intended data report). Time for experimentation enabled a greater understanding of the datasets and supported the creation of additional visualisations to support the negotiations, such as the likelihood of a Cambridge author being a corresponding author for Springer Nature publications (Figure 5).



*Figure 5. Understanding how often Cambridge authors are the corresponding author within Springer Nature publications provided an understanding on the use of funding.*

## Outcome

The purpose of this project was to create visualisations which would support the University of Cambridge during the publisher negotiations with Springer Nature. Visualising the relationship with Springer Nature through usage and publishing, supported the negotiations working group when communicating across the university, and when considering the proposed agreements. This included understanding the need for post-cancellation access rights (McCracken, et al., 2023), and visualising the use of material per year of publication (Figure 6). To communicate the impact of the negotiations, select visualisations were shared via a PowerPoint to the entire academic cohort, which was updated in January 2023 to represent four calendar years of data. The visualisations output of this project did not use live data, it did not need to be regularly updated and it was be stored internally, therefore changes to the data were adjusted manually when relevant. The PowerPoint was successfully shared with academic staff, it was used to present the negotiations to senior staff, and it received positive feedback during and after the project.

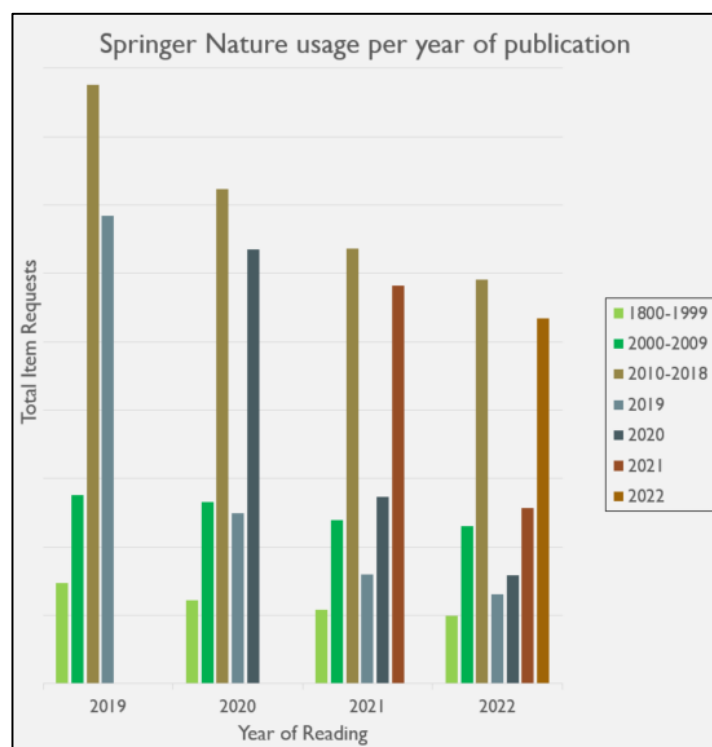


Figure 6. Exploring usage through year of publication provided understanding of the access needs to older publications.

Translating each data report into a visualisation required careful consideration of the audience, which included the working group, senior leadership, and/or all academic and library staff. As most visualisations would be shared with a wide audience, they were required to be accessible, concise, and clear, and many of the visualisations produced contained elements of approachable bar charts or column charts. Avoiding complex or ambiguous visuals (such as pie charts) prevents confusion about the data, as did trimming visualisations to show only relevant information, labels and titles (visualisations included in this paper have been edited to remove data labels). Key information was provided in text format to the left of the visualisation, further details were provided through legends, titles and alt text, and colour was used for consistency and comparison. The use of colour was powerful, for example using green to represent the quantity published using Green OA was informative and accessible (Figure 7).

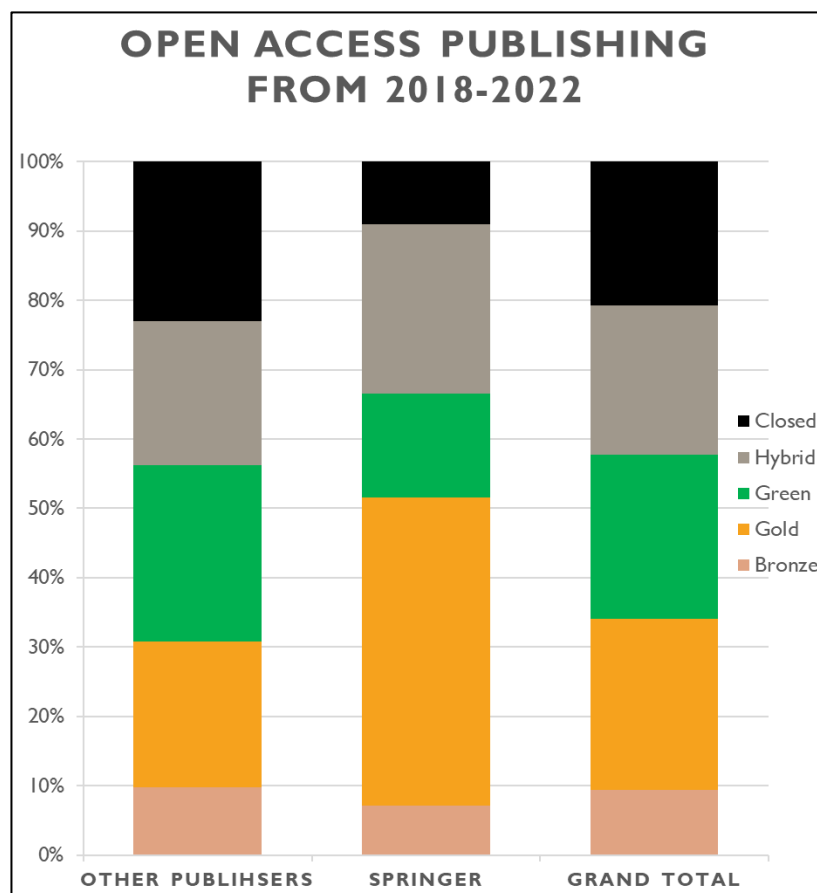


Figure 7. Open Access Publishing - Use of colour.

Alongside other data visualisations and reports, the PowerPoint was a successful outcome of the authors contribution to the publisher negotiations between the University of Cambridge and Springer Nature. Evidenced with its regular use by the negotiations working group, it enabled a cohesive understanding of the publisher-institution relationship, it supported meetings with senior staff to explain the significance of the Springer Nature negotiations, and it was shared to all university staff to aid their understanding of the role of Springer Nature within the University of Cambridge (and the potential impact of a walkaway).

### **What Next?**

The author recognises the privilege they have had exploring the possibilities within publishing and usage data to uncover the publisher-institution relationship, and when combined with the Office of Scholarly Communications focus on open access, the next stage of this project will be an open resource. The intention is that resources and processes by the authors and other data librarians will be compiled and made available openly for any institution to adapt to support their publisher negotiations.

Making the processes adaptable and repeatable became a key factor towards the end of the Springer Nature negotiations as the conversation turned to the next 'Big Deals' to be negotiated, with tighter timeframes and resources. Managing varying priorities on multiple timelines would not enable efficient work by the author as similar tasks would be repeated, therefore after consulting with the negotiations working group and relevant staff, key questions which could support any publisher negotiation were compiled, and these will form the basis of a resource to support (data) librarians during negotiations.

### *Microsoft Power BI*

Microsoft Power BI is comparable to Excel with its ability to manipulate and clean raw data, however it can also create dashboards. These dashboards can have interactive visual elements with filtering by users

and can import from multiple data sources. Dashboards have been used by IT and others to understand ticket progression or calls completed, and now data dashboards are becoming popular within institution libraries.

Power BI has multiple advantages over Excel for the preparation, modelling, and visualisation of data in relation to the publisher negotiations. First, an Excel pivot table is limited to summarising data from a single dataset, and as such XLOOKUP has been used to combine dataset columns. However, the XLOOKUP formula is limited to returning one column of data, and therefore must be repeated for each column being transposed into the dataset. Both processes are negated when using Power BI, as it can link data sets based on a single column using the Relationships feature, making it possible to link data across multiple datasets quickly and with dual cardinality (Microsoft, 2023). The unpivot feature available through Excel power query editor is also available using Power BI and is more accessible during the data cleaning process. Power BI unpivot provides greater flexibility in the selection of columns, it tracks each action making them easier to undo or repeat, and columns can be separated into a new connected report.

Power BI also offers additional customisation of visualisations, not available in via PowerPoint or Excel, and it enables the use of dashboards where users can apply filters to visualisations. There is an extensive range of visualisations available through Power BI, including the ability to impact custom visuals from the marketplace – such as gauge, card and KPI visuals. Power BI dashboards have slicers, filters per visual and/or page, and drill through filters, enabling additional exploration by the end user. Power BI dashboards also have 'Q&A' visuals which use the dataset to create editable FAQ which could interest the audience and support their exploration further. While further skills are required to utilise Power BI, it provides a wider range of tools and details during the creation of visualisations, but also interactivity for end-users.

The Power BI dashboard developed by the author uses exports from Dimensions and COUNTER, to create relevant visualisations for publisher negotiations. It is the author's hope that a template dashboard, along



with guidelines for usage, could provide each institution with the ability to thoroughly assess their relationship with a publisher. Understanding the relationship between the institution and a publisher (through read and/or publish data), can empower institutions when accepting or rejecting publisher deals.

### *Data*

When creating the Dimensions API calls, it was clear that the approach and structure of the questions could be useful to other librarians and institutions engaging in publisher negotiations. The authors API calls were developed with clearly labelled code and processes, therefore being adaptable by those with coding experience, but also legible by those with none. Dimensions API calls require a Dimensions subscription to run, however Open Alex is an open access database of publications and funding, with its own API<sup>x</sup>. A next step in this project is to translate the authors API calls into Open Alex API coding, thus extracting the same dataset as Dimensions but from an open-source database. Sharing the coding for the publisher negotiation questions used by the University of Cambridge hopes to give institutions of any size confidence in tackling data questions.

As for COUNTER data, the SUSHI API is available to query usage data directly, and will be explored to answer nuanced usage questions. COUNTER update 5.1 will become effective from January 2025 (Project Counter, 2023), and as such explorations with the data will hold until the update new reports are implemented. In the meantime, COUNTER exports are uniform across all institutions, and a set of guidelines for cleaning and presenting the data could therefore be of use. Importing cleaned COUNTER data to a Power BI dashboard template which would generate visualisations for a dashboard could be a powerful tool to support publisher negotiations (Figure 8).

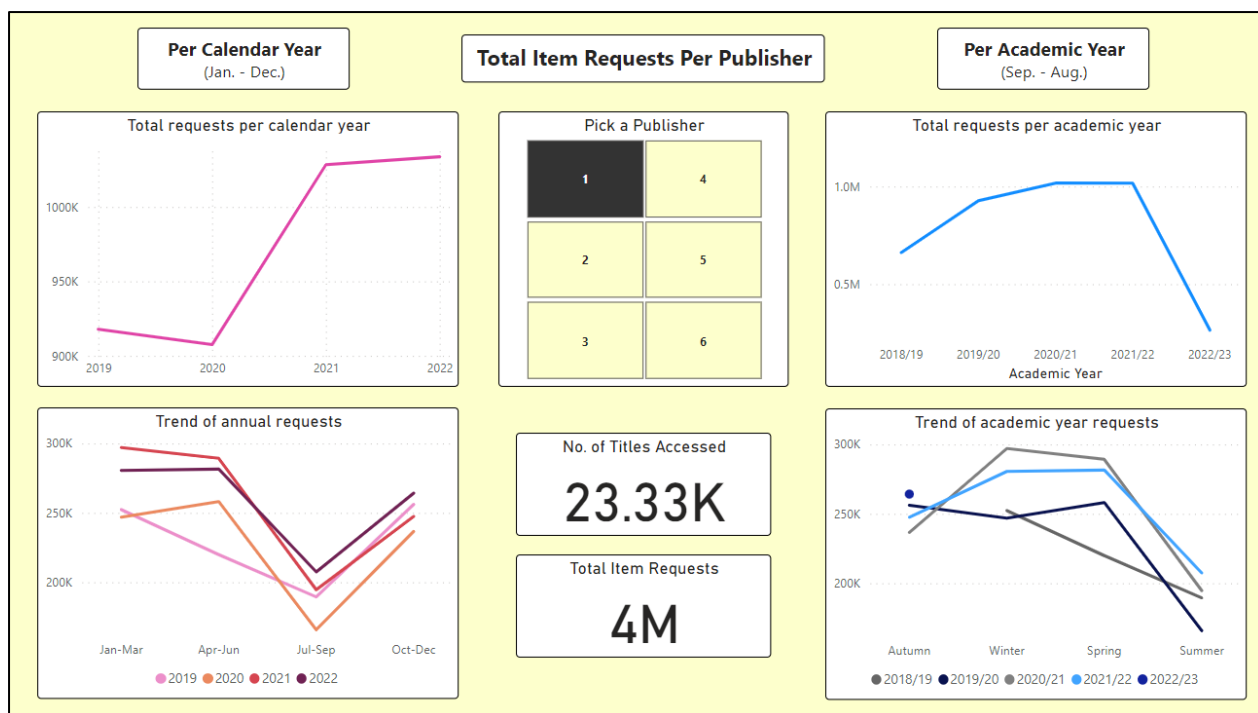


Figure 8. Trial Power BI Dashboard of usage data for six publishers.

### Call for Action

Creating a Power BI template which any institution could import raw data to create a dashboard of visualisations to understand the relationship with a publisher is the goal of this project. Importing publishing and reading data (financial datasets will differ considerably for the sake of a generic template), along with guidelines, could be combined to create a publisher negotiations resource for (data) librarians. Open access and equity are pillars for the publisher negotiations and as such should be the backbone of the work conducted to support the negotiations.

Sharing the API calls used to inform the visualisations is the first step to ensuring all institutions can be confident in their use of data to support publisher negotiations. Creating a resource to support (data) librarians will be a collaborative effort, the approach discussed here is only one perspective and as such should be a cause for opportunity to share and develop. While this project remains ongoing at the time of publication, it reflects the efforts of universities during the publisher negotiations, and the author hopes that negotiations can become empowering act for the institutions involved.

## Conclusion

Understanding the relationship between an institution and a publisher can be powerful for librarians during publisher negotiations. While Jisc supports the negotiations with sector-level visualisations and data reports, each institution can analyse their data for an individual understanding. However, given the time, resources, and expertise needed to manipulate the data and create visualisations, it is not often possible for every institution to analyse their relationship with a publisher. During the publisher negotiations with Springer Nature, the author took an exploratory journey to understanding the range of databases available, the usefulness of different data questions, and the repeatability of the analysis for themselves and other institutions, with the outcome to develop a resource to assess any publisher-institution relationship. The future resource hopes to become a combination of API coding, guidelines on exporting and cleaning data, and dashboard templates for presenting the data. Understanding usage, publishing patterns, open access breakdowns, and funding support, is a powerful tool for librarians when aiding the publisher negotiations. Confidence in the data enables librarians to reject an unacceptable deal, to walk away from a publisher, and/or seek alternative access when the need arises.

It's a lot like the cable bundle - they tell you you're getting 250 channels, but if you look inside your heart, you know all you want is ESPN and AMC.

Director of Information Policy at the University of Virginia, Brandon Butler, compares the Springer Nature bundle to cable TV, because of the four thousand journals in the deal a third were never accessed (Resnick & Belluz, 2019), which highlights the need for institutions to understand their relationship with a publisher. Big publishers force 'big deals', non-disclosure agreements, swallow up independent publishers, and use institutional data for profit, eating up library budgets and providing no compromise or reasoning to their ever-increasing subscriptions (Lamdan, 2023). Therefore, this project of data analysis for publisher negotiations hopes to provide confidence to (data) librarians to make data decisions.

Using data to negotiate with big publishers (who in some cases are now data brokers (Lamdan, 2023)), can empower university libraries to fight against the unsustainable pricing, restrictions to their publications, and secrecy of contracts.

## Bibliography

Digital Science & Research Solutions, 2020. *Welcome to the Dimensions API Lab*. [Online]

Available at: <https://api-lab.dimensions.ai/>

[Accessed 20 06 2023].

Digital Science & Research Solutions, 2023. *Dimensions Homepage*. [Online]

Available at: <https://www.dimensions.ai/>

[Accessed 20 06 2023].

Dixon, D., 2021. *Informing the Elsevier negotiations: Dominic Dixon on the work of the Data Analysis Working Group*. [Online]

Available at: <https://unlockingresearch-blog.lib.cam.ac.uk/?p=3018>

[Accessed 19 06 2023].

ESAC, 2020. *Uncover the publishing profile of your institution*. [Online]

Available at: <https://esac-initiative.org/about/data-analytics/publishing-profile/>

[Accessed 19 06 2023].

Hotz, N., 2023. *What is CRISP DM?*. [Online]

Available at: <https://www.datascience-pm.com/crisp-dm-2/>

[Accessed 20 06 2023].

Information Power, 2021. *Toolkit to foster Open Access Agreements*. [Online]

Available at: <https://www.alpsp.org/OA-agreements>

[Accessed 19 06 2023].

Jisc, 2020. *JUSP Guides*. [Online]

Available at: [https://jusp.jisc.ac.uk/guides/Step-by-step\\_guide\\_to\\_TR\\_filter.pdf](https://jusp.jisc.ac.uk/guides/Step-by-step_guide_to_TR_filter.pdf)

[Accessed 23 06 2023].

Jisc, 2022. *Springer Nature negotiations aim to secure a transitional open access agreement*. [Online]

Available at: <https://www.jisc.ac.uk/news/springer-nature-negotiations->

[aim-to-secure-a-transitional-open-access-agreement-25-apr-2022](#)

[Accessed 19 06 2023].

Lamdan, S., 2023. *Data Cartels*. Stanford, California: Stanford University Press.

Mackie-Mason, J., 2019. *How the University of California prepares for negotiations*. Cape Town, SANLIC Conference.

Malin, N., 2022. *Dataset of Springer Nature Group related journals*.

[Online]

Available at: <https://doi.org/10.17863/CAM.90061>

[Accessed 29 06 2023].

Malin, N., 2023. *Citation Analysis: Cited by a Publication Dataset*. [Online]

Available at:

[https://colab.research.google.com/drive/171EOZ9iF\\_dAIFXbRFvW-zHKSmdthVYpI?usp=sharing](https://colab.research.google.com/drive/171EOZ9iF_dAIFXbRFvW-zHKSmdthVYpI?usp=sharing)

[Accessed 29 06 2023].

Malin, N., 2023. *Publications Analysis: Analysing a Dataset*. [Online]

Available at:

[https://colab.research.google.com/drive/1b1WTR4e0Wy\\_wcbFfktbLj6\\_hjoKDTgo9?usp=sharing](https://colab.research.google.com/drive/1b1WTR4e0Wy_wcbFfktbLj6_hjoKDTgo9?usp=sharing)

[Accessed 29 06 2023].

McCracken, P., Levine-Clark, M., McDonald, J. & Price, J., 2023.

*Indispensable or unnecessary?: a data-driven appraisal of post-cancellation access rights*. [Online]

Available at: <https://insights.uksg.org/articles/10.1629/uksg.601>

[Accessed 30 06 2023].

Michalikova, M., 2022. *Differences in academic writing & publishing between STEM and humanities*. [Online]

Available at: <https://writingscientist.com/stem-humanities/>

[Accessed 20 06 2023].

Microsoft, 2023. *Create and manage relationships in Power BI Desktop*.

[Online]

Available at: <https://learn.microsoft.com/en-us/power-bi/transform->

model/desktop-create-and-manage-relationships

[Accessed 21 06 2023].

Office of Scholarly Communication, 2019. *Negotiating with scholarly journal publishers: A toolkit from the University of California*. [Online]

Available at: <https://osc.universityofcalifornia.edu/uc-publisher-relationships/negotiating-with-scholarly-journal-publishers-a-toolkit/>

[Accessed 19 06 2023].

Office of Scholarly Communication, 2020. *UC-Publisher Relationships*.

[Online]

Available at: <https://osc.universityofcalifornia.edu/uc-publisher-relationships/>

[Accessed 19 06 2023].

Project Counter, 2023. *Release 5.1*. [Online]

Available at: <https://www.projectcounter.org/release-5-1/>

[Accessed 21 06 2023].

Resnick, B. & Belluz, J., 2019. *The war to free science*. [Online]

Available at: <https://www.vox.com/the-highlight/2019/6/3/18271538/open-access-elsevier-california-sci-hub-academic-paywalls>

[Accessed 29 06 2023].

South African National Library and Information Consortium, 2019. *SANLIC 2019 Conference*. [Online]

Available at: <https://sanlic.org.za/sanlic-2019-conference/>

[Accessed 19 06 2023].

SPARC, 2021. *Data Analysis for Negotiation*. [Online]

Available at: <https://sparcopen.org/our-work/negotiation-resources/data-analysis/>

[Accessed 29 06 2023].

Vernon, A., 2022. *Supporting the transition to open access publishing - an update*. [Online]

Available at: <https://www.jisc.ac.uk/blog/supporting-the-transition-to->

[open-access-publishing-an-update-01-apr-2022](#)

[Accessed 20 06 2023].

Willmott, M., 2019. *Tools for Transition: Data analysis for negotiations*. Cape Town, SANLiC Conference.

---

<sup>i</sup> <https://unsub.org/>

<sup>ii</sup> <https://www.portico.org/>

<sup>iii</sup> <https://www.kbplus.ac.uk/kbplus7/>

<sup>iv</sup> <https://www.dimensions.ai/>

<sup>v</sup> <https://www.projectcounter.org/>

<sup>vi</sup> <https://openeditors.ooir.org/index.php>

<sup>vii</sup> <https://www.dimensions.ai/products/all-products/dimensions-api/>

<sup>viii</sup> <https://jusp.jisc.ac.uk/>

<sup>ix</sup> <https://openalex.org/>