

## Panel 2 - We spend how much? Institutional approaches to calculating Article Processing Charges (APC) expenditures

- Moderator: Maddie Hare, University of Ottawa
- Panelists:
  - [Stephanie Savage](#), University of British Columbia
  - [Mark Swartz](#), Queen's University
  - [Brienne Selman](#), University of Winnipeg
  - [Leigh-Ann Butler](#), University of Ottawa

### Summary

Institutions are grappling with the challenge of estimating and managing open access (OA) publishing costs. Speakers discussed various approaches, including estimating APC costs, collecting data, and leveraging proprietary and open publication data. They emphasized the importance of collaboration, stakeholder engagement, and considering factors like Tri-Agency policy mandates, language and institutional publishing behaviour. In communicating the insights of their reports, the panelists highlight the need for alignment with institutional reporting mechanisms and budgeting goals, and considerations of consortial commitments both current and future. Future directions for open scholarly publishing are considered in relation to 'following the money'.

### Chat Summary

- L Butler's [thesis](#) which discuss APC analysis process
- Panelists are seeking interest for national conversation and determination of a single process to determine APC costs: see the links in this [Google doc](#)
- Many acknowledged the comment that publisher-provided data is not always/often right
- One participant asked what/if persistent identifiers were used to disambiguate the data – response was DOIs and ISSNs
- One participant asked if the data was being used to further discussion in choosing diamond publications – the general response was yes, in so far as showing that the amount of money provided to commercial publishers, some could be used to sustain diamond initiatives

### Discussion Details

- Context and process of analysis: data sources, size of analysis etc.
  - A. A large university that is decentralized: multiple campuses, with many affiliated research institutions (uncertain definitions of 'belonging' to the university), different ways of capturing data and doing work. Analysis was done within the scholarly communication office of the library. Collaboration with collections unit of Library was necessary to obtain consortial contractual information. Process was based on course on how to measure APC costs. Web of Science (WoS) and Lens were used to generate the dataset. WoS was chosen for granularity of author rank (first, corresponding, collaborator), flexibility of retrieval, and future access. Unpaywall provided detailed information on OA status of a publication. An internal dataset of university authors was used to disambiguate search results. Postal codes in affiliation was used and preferred over email suffixes (email address does not resolve geographical 'belonging-ness'). APC information was obtained by publisher-provided lists, both publicly available and publisher contacts, as well as obtaining individual costs for journals together with information from collections

team. Dataset was formed using OpenRefine for data cleaning and Excel to perform cost calculations.

- B. Small university with small library support: analysis was performed by two librarians. Faculty list was easy to validate as faculty is known to the librarians. WoS was used as it is licensed by the university. Unpaywall was used to clarify OA status. Data cleanup was done manually using Excel and using institutional knowledge.
  - C. Medium university with many units/segmentation. Used special project manager within internal audit department together with librarian and PhD student. Scopus was used, employing a method by [panelist cited author that could not be clarified]. Dataset was created by searching on university authors by open access status. Using APC information, and the SNIP (source normalized impact per page) value, total costs were estimated. The auditor employed a second method using the institution's financial system. Each department was analyzed by reported or identified APC expense (some expenses were miscategorized and required disambiguation).
  - D. A large university which is decentralized. Librarian had predecessor which had compiled internal data on their open access agreements, and disciplinary publication analysis to inform future agreement decision-making. Additionally, a dataset was generated to estimate APC spending by university researchers in gold OA journals. This dataset contained price lists from the big 5 publishers in a three year period. A second open dataset based on DOAJ data was used and, with the 2 combined datasets, an estimate of APC spending was generated. This APC dataset was then combined with WoS publication data to analyze corresponding authors costs. This publication dataset was author disambiguated using postal codes. The postal code was also the identifier to match the publication list to the APC dataset. More recently, in addition to looking at gold OA APC spending, hybrid journals were analyzed as well. In this analysis, medians and not averages were used to account for large skew due to APC charges by publishers such as Nature.
- Key outcomes or insights from the analyses presented:
    - A. In a one year period, from approximately 13000 publications, 1600 articles were identified as belonging to the university author. In this set, 142 unique publishers were determined with the top 5 identified from largest to least: MDPI, Springer, Elsevier, Wiley and Taylor&Francis. The estimated APC spending range before accounting for discounts was \$4.9 and \$8.2 million. The estimated savings range between \$500,000 and \$900,000. With the discounts, the APC spending is estimated between \$4.2 and 7.5 million.
    - B. One of the analyses wished to investigate their authors' choices with respect to Canadian publishers. In the top 250 journals in the dataset, only 8% were Canadian. Also, 35% of the 250 journal offered green (self-archival) option without publication embargo. This is to say that 65% do not offer a cost-offset, immediate open access option. This publication dataset, demonstrating author choice, may be reflective of the disciplines at that institution (i.e. very social sciences oriented). 90% of this same set do charge some form of APC. With the current Tri-Agency OA policy mandate of 12 month embargo allowance, 56% permitted this but only 35% allowed immediate deposit.
    - C. In the first method, the analysis determined that the APC spending in a 5 year span was \$1.3 million for 901 articles. In the second method employing the auditor, the APC spending estimate was \$685,000 for 238 articles. Future recommendations included a method to internally track APC spending, which was actualized in the establishment of an

account code for APCs. This will result in more efficient and accurate reporting of internal APC expenditure and publication analysis. Other recommendations included was the assignment of APC tracking to a specific department; having mandated deposit in university IR; establish a campus-wide OA policy.

- D. From the analysis from 2022, 2142 gold and 554 hybrid OA publication from university authors. The total cost estimate of APC spending was \$7.3 million for any of the university authors (this may be an over-estimate as not every paper results in the associated university author paying the APC). Of the top 5 publishers in the dataset ranked according to expenditure include SpringerNature (\$1.7 million); Elsevier (\$1.2 million); MDPI and Frontier (expenses not mentioned). There was a higher percentage spent on gold v. hybrid journals. 2/3 of the \$7.3 million was spent on gold OA v. hybrid. But by median, hybrid journal costs were higher (\$4418/paper) compared to gold (\$3000). Examining the dataset by corresponding authors, the estimated cost of \$7.3 million is reduced to \$2.8 million (878 OA articles). Including APC discounts available to university authors, there is a further reduction of \$2.1 million.
- How is this data being applied:
    - A. The process and data illuminated how good (or bad) the data quality is, the validity of the cost estimates and publishing behaviour, and the complexities of the analysis process itself. Value is in the potential of trend analysis in year over year comparison and how factors such as whether transformative agreements inform where university authors choose to publish (for example). Also to use the data to signal the importance of OA policy on campus and the role/value the library has in publishing and research dissemination and the potential to save the university/authors money.
    - B. The analysis showed that university authors are choosing to publish in journals not included in the transformative agreements, so it causes questions as to what transformative agreements present a return on investment. Also, in analyzing publication behaviour in OA journals, it demonstrated that researchers are considering the funder mandates. This information provides predictive information if the mandate changes.
    - C. The data creates a scholarly communication conversation starting point across institutional units and faculties. Researchers speak to their own experiences to then determine what services/supports are required to better assist them in meeting the mandates.
    - D. The analysis has provoked more questions for future analysis: for example, what is the difference between a list price and cost average? Also, there will be further collaboration with collections colleagues to understand how the transformative agreements are communicated to possibly influence author choice.