IJoL International Journal of Librarianship

ISSN: 2474-3542 Journal homepage: http://journal.calaijol.org

Breaking up the Big Deal?: A quantitative approach

Shawn Vaillancourt and Kerry Creelman

Abstract:

This paper explores the methodology used to assess the value of a "Big Deal" package during inter-institutional negotiations in 2020/21. As part of the Texas Library Coalition for United Action (TLCUA) collaborative negotiations with Elsevier, the University of Houston (UH) Libraries evaluated the value of their "Big Deal" journal package against alternative subscription options presented through the negotiation. Analysis was performed primarily with quantitative factors, including usage, cost, cost per use, and faculty citations. After identifying the highest and lowest performing journal titles in these categories, librarians were able to eliminate them easily from further deliberation. Weighted scoring was then applied to moderately performing titles. Weighting identified additional titles that could be sorted by specific factors, leaving a smaller pool for title-by-title deliberation performed by a team of librarians and a final pool of "must have" titles to be presented to library stakeholders for review and to inform decision-making. Ultimately, the analysis helped UH Libraries clarify evaluative values and informed the decision to retain the "Big Deal" in favor of alternative options that would have significantly reduced access to content used by our research community.

To cite this article:

Vaillancourt, S. & Creelman, K. (2024). Breaking up the Big Deal?: A quantitative approach. *International Journal of Librarianship*, *9*(4), 3-13. https://doi.org/10.23974/ijol.2024.vol9.4.396

To submit your article to this journal:
Go to https://ojs.calaijol.org/index.php/ijol/about/submissions

INTERNATIONAL JOURNAL OF LIBRARIANSHIP, 9(4), 3-13
ISSN: 2474-3542

Breaking up the Big Deal?: A quantitative approach

Shawn Vaillancourt and Kerry Creelman
University of Houston, Texas, United States

ABSTRACT

This paper explores the methodology used to assess the value of a "Big Deal" package during inter-institutional negotiations in 2020/21. As part of the Texas Library Coalition for United Action (TLCUA) collaborative negotiations with Elsevier, the University of Houston (UH) Libraries evaluated the value of their "Big Deal" journal package against alternative subscription options presented through the negotiation. Analysis was performed primarily with quantitative factors, including usage, cost, cost per use, and faculty citations. After identifying the highest and lowest performing journal titles in these categories, librarians were able to eliminate them easily from further deliberation. Weighted scoring was then applied to moderately performing titles. Weighting identified additional titles that could be sorted by specific factors, leaving a smaller pool for title-by-title deliberation performed by a team of librarians and a final pool of "must have" titles to be presented to library stakeholders for review and to inform decision-making. Ultimately, the analysis helped UH Libraries clarify evaluative values and informed the decision to retain the "Big Deal" in favor of alternative options that would have significantly reduced access to content used by our research community.

Keywords: Big Deals, Journal Packages, Quantitative Assessment

INTRODUCTION

The University of Houston (UH) is a large, urban Carnegie Research 1 institution with more than 46,000 students, 2,800 faculty, and 281 degree programs. Areas of research strength include energy, infrastructure, space, health, data sciences, and the arts. In 2020, we opened a new medical school and are encouraging community-based, interdisciplinary health research. UH aspires to be a Top 50 public university and Association of American Universities (AAU) eligible. It is essential to ensure that UH Libraries (UHL) provides access to the research and learning materials to support the research and academic enterprise of the university. We strive to carefully steward the collections funds entrusted to us in order to provide access to the materials that our users need. As the costs of our continuing resources continue to increase each year, this means carefully reviewing our collections strategies to see which makes the most sense to continue.

Like many academic universities, one collection strategy we employ is subscribing to "Big Deals" with academic publishers to provide access to as much research content as possible for our community. While common, the value of the "Big Deal" has been under consideration by academic

librarians for years as the price of the deals account for significant portions of collections budgets. UH Libraries has several “Big Deals” with publishers but had not previously considered breaking up any of our “biggest” deals and had therefore not engaged in systematic assessment of our deals. This changed in 2021 as we considered breaking up our “Big Deal” with Elsevier in favor of alternative subscription options.

In 2020, more than 40 universities in Texas formed the Texas Library Coalition for United Action (TLCUA) to examine collections models with academic publishers. They began collaborative negotiations with Elsevier to consider various models for providing access to journal content. Multiple issues (Fisher, 2021) were at stake during the negotiations, all with their own considerations and weightings by negotiators. As part of the negotiations, several options for access were identified, including continuing the “Big Deal.” To determine which option aligned best with institutional budgets and priorities, each university conducted its own assessments of existing agreements and proposed alternatives.

Leading into TLCUA negotiations, the UHL ScienceDirect subscription content with Elsevier included 517 journal titles with post termination access (PTA); 4000+ journal titles with backfile, but not current access; and 1802 titles in its “Freedom Collection,” a collection that other Elsevier libraries may be familiar with. The “Freedom Collection” model saw each library subscribe to an additional number of Elsevier titles to provide access only at a reduced rate. The “Freedom Collection” emphasizes access over ownership and thus only provides current and some archival access, but does not include any PTA.

UHL needed to determine which of the negotiated options would best support UH during its strategic research growth. In particular, UHL needed to determine:

- whether breaking up our "Big Deal” would impact access to research content needed by the campus population at an acceptable level
- how much access would be lost
- what cost savings would be realized and
- would the university administration be willing to accept that level of loss in access within the context of other campus priorities?

WHY ASSESS “BIG DEALS”?

Since the emergence of the “Big Deal,” librarians have been grappling with how to effectively assess their value. On the surface, these packages make a lot of sense. Libraries get access to far more material than they ordinarily would be able to provide to patrons for a fraction of the cost of subscribing to those titles individually. “Big Deals” also often come with added benefits like discounts on other products or titles, cost increase caps, and the ability to negotiate for author, open access (OA), or use rights (Hoeppner, 2018). But even at a discounted rate, these packages eat up a large portion of a library’s budget. In order to be responsible stewards of those funds entrusted to academic libraries by their institutions, librarians have a responsibility to examine the performance of these packages versus other possible access options and alternative uses for the funds.

While the "Big Deal" provides access to large collections, it comes with several issues that have caused libraries to reconsider the value of the "Big Deal." With the price tag associated with these deals, a significant percentage of a library's collections budget can be tied up in these large packages, which results in a lack of flexibility for the library to tailor and adjust their collection to meet user needs with their remaining budgets, and minimal ability to focus on the content that is actually used and/or needed from "Big Deal" publishers (Hoeppner, 2018). Meanwhile, significant portions of the content included in the "Big Deal" goes unused in what has been described as "increas[ing] serial expenditures to acquire a large number of unused journals" (Shu et al., 2018). Plus, with inflationary increases built into the pricing of these deals, there is no long-term sustainability for these arrangements. Without continually investing new, higher levels of funding into these deals, all of them will have to be broken or downsized at some point. The scholarly publishing community also raises issues about the ethics and impact of "Big Deals" on larger goals and movement towards making public research openly available to all, in that continuing to pay for and support "Big Deals" continues to encourage locking research behind a paywall. As a result, many institutions, where able, have been moving towards other arrangements like Transformative Agreements that are being tracked by the Scholarly Publishing and Academic Resources Coalition, or SPARC (SPARC, 2023).

WHAT FACTORS ARE CONSIDERED WHEN ASSESSING "BIG DEALS"?

As libraries seek to assess the value and viability of their "Big Deals," they consult many different combinations of data to reach their decisions. Generally, this involves a blend of multiple factors, quantitative and qualitative, that compensate for the shortcomings of any one measure (Ivanov et al., 2020). Usage is a primary factor in considering journal packages, whether this be actual usage of subscribed content through downloads or other usages of the journals by faculty through citations of, or authorship in, those titles (Chamberlain, 2022; Dawson, 2015; Gagnon, 2019; Shu et al., 2018). Frequently, librarians will apply or adapt Bradford's law in their analyses and unbundling attempts, attempting to identify the titles responsible for 80% of a "Big Deal"'s usage. However, disciplinary variations in citation behavior have led librarians to adjust their application to their own packages (Dawson, 2015; Gagnon, 2019). Pairing usage with cost results in the ubiquitous measure of cost per use (CPU) found in nearly every analysis (Johnson & Cassady, 2018; Lemley & Li, 2015). Cost considerations can also include costs of cancellation or additional content fees (Nabe & Fowler, 2015), or comparisons of CPU to the costs associated with interlibrary loan (ILL) or article on demand services (Lemley & Li, 2015). Librarians also make use of bibliometric scores as quantitative measures of journal quality, Impact Factor being the most common, though their overall validity for these purposes have been called into question (Ivanov et al., 2020; Johnson & Cassady, 2018). Applications like Unsub also help librarians factor in estimations of how much usage of a title could potentially be attributed to open access articles or those that would be available from years in which the library would have perpetual access post-termination (Chamberlain, 2022). Others have looked at qualities of the overall packages themselves to help evaluate their value, including such measures as the proportion of titles unused, the number of titles that account for most of the usage, as well as package level versions of usage, cost, and CPU (Blecic et al., 2013; Rathmel et al., 2015).

Librarians are cautioned not to rely on usage alone (Rathmel et al., 2015), and as a result many analyses also try to consider qualitative factors. Faculty input is the most common qualitative

data point considered in "Big Deal" analyses. This is frequently in the form of some sort of faculty survey or database, which is either conducted in anticipation of examining a "Big Deal," or information is collected on a regular basis and allows for faculty to identify titles that are important to their discipline or their own personal research (Dikboom, 2016). Dawson (2015) cautions that users do not like to be over-consulted, so cultivating a culture of consultation would be important for any attempts at regular user input. Discipline is also a factor in analysis, often used to calibrate quantitative factors like citation behavior (Gagnon, 2019), but also to identify titles that are important to major programs/curricular areas, institutional research focus areas, or accreditation needs. Subject librarians are often the source of this knowledge, though those with less experience or familiarity with the subject areas under examination tend to rely more heavily on quantitative factors (Johnson & Cassady, 2018).

Scalability of qualitative factors to larger analyses is a big question for librarians. Some, like Dawson (2015), recognize that their methods may be difficult to apply to larger packages. Of course, Johnson and Cassady's (2018) large working group, formed specifically to break up a package, can help bring these factors into the work of larger packages, though library staff will still have only so much time to go more in depth on a title-by-title basis. To strike a balance, librarians have been strategically applying their criteria, using quantitative factors to "see which journals are clearly essential, which clearly unneeded, and which require in-depth evaluation" (Blecic et al., 2013) using more qualitative factors. A similar stratified approach has been applied by UH Libraries.

METHODOLOGY

Due to the diversity of institutions comprising the TLCUA and range of needs, several options were negotiated, and any institution that was part of the TLCUA would be able to choose from options including: two shared title list (STL) options (either a comprehensive or medical STL; with or without adding in further subscriptions), skipping the STL entirely and making selections title-by-title, or maintaining existing product options. Recognizing that each institution has its own research specialties, programs, and focus areas, the deal included the ability to swap out a limited number of titles for others in Elsevier's catalog for the STL options. This cap was determined as 3% of the total value of the package for the institution. While the dollar amount of titles we could substitute in this way was capped, we could also add as many additional titles as desired beyond this through out-of-pocket subscriptions at the same TLCUA discounted rates.

This scenario had 2 major determinations for the group to identify:

1) Which titles from the STL could be removed (if any) with the least impact on library users.
2) Which titles not included in the STL need to be recovered (if any) in order to get the most value out of a financially responsible investment in subscriptions to Elsevier titles.

DATA POINTS CONSIDERED[1]

We used several data points to help inform our deliberation over titles.

- **Maximum usage out of 2 years**: Based on COUNTER compliant reports from 2019 and 2020, the unique item requests[2] for each title were examined and the highest usage from the 2 years was identified. Examining 2 years of usage was determined to account for variations in use from courses that may be only offered every other year, but sufficient to cover the minimum necessary course offering for a class to remain in the academic calendar.
- **Average usage over 2 years**: Based on COUNTER compliant reports from 2019 and 2020, the unique item usage for each title was averaged across the 2 years. Averaging the 2 years of data helped give a more predictable metric that evened out cyclical variation for a more predictable annual estimate of usage.
- **Cost**: The discounted cost offered to TLCUA participants, used for Cost Per Use.
- **Cost per use**: The cost divided by the average unique item usage over 2 years. Essentially, the amount it costs to download each article if the subscription price were spread over all the downloads. This is a metric commonly compared to the cost to obtain the article through other means, usually ILL, through fees, staff time, etc. to determine the value the subscription has offered.
- **Citations in publications**: Based on Microsoft Academic data, and obtained through our Unsub subscription, this is the 5-year average of citations to articles in a journal included in articles published by UH authors.

Knowing that we would easily surpass our STL substitution allowance with titles that would be desirable to our campus community (in scenario 2 above), and to allow the most flexibility, we removed as many of the lowest use titles as we could from the STL until we reached our maximum swap credit. 37 titles were swapped out of the STL to hit our swap credit (approximately $83,500). This allowed for analysis to focus on the remaining titles.

PHILOSOPHICAL QUANDARIES: WHAT DO WE VALUE?

[1] Authorship in each publication was also considered, but ultimately discarded. Since we have a robust institutional repository in which we may archive UH author work, and our primary imperative is to facilitate access to materials, we decided prioritizing titles our authors are publishing in was of lesser importance and if authors needed access to the title, that would be reflected in the usage.

[2] The current COUNTER standards that these reports are based on (Release 5, implemented 2019) uses a metric called "Unique Item Requests." In terms of articles, this is equivalent to downloading the full text of the article. The Uniqueness means that each article download will only be counted once, even if a user downloads an article multiple times in a sitting (e.g., views the HTML version of an article, then saves the PDF).

There were ~1200 titles not included in the STL at a value of well over $3M to consider. We obviously could not include all titles and make good use of our funding, but examining each of the titles individually would not be an effective use of our time. We took a phased approach to the data to gradually identify the more obvious titles to recover and save time and nuance for looking at the less obvious candidates to include.

In order to make a data informed decision, we examined each of the above metrics and considered the range of values that were in the data. Quartiles were calculated for each metric that allowed us to determine cutoff points for representing the top 25%, 50%, etc. of the possible values (see Figure 1 for an example). We also discussed how we valued each of the available metrics in relation to each other.

We decided that usage was our highest priority and reflected the metric that would have the highest impact on our users, so should be prioritized the highest. Cost per use was also important as a metric of value, but since we were accounting for usage with 2 metrics already, and CPU is partly impacted by high usage, this was of lesser importance. Citations were also important as particularly valuable kinds of usage, so while we wanted authors to have access to the materials that were having particularly high impact on their productivity, we expected much of this to be reflected in overall usage, and so citations were of lesser importance. Thus, whenever we scored these factors relative to each other in our weighted calculations, they were ranked Usage > CPU > Citations.

In the first phase, we took large rough swaths of the data and included the highest performing titles in each of Citations, Cost per use, and Average 2-year usage. If a title ranked in the top 25% of Average 2-year usage (>151) or Cost per Use (<$15.26), or greater than 30 average citations per year[3] we automatically included the title for reclamation. This ensured we captured titles with the highest raw usage, highest research usage, and/or the most efficient usage, in terms of return on subscription investment. Here, the cost of instant access via subscription is warranted, and attempts to save money by redirecting to ILL would be a disservice to users.

[3] Unsub uses a ratio of 1 citation equating to 10 downloads, which we could use as representing an industry standard. However, we wanted to be more conservative in how much we were automatically reinvesting in subscriptions, so chose to use 1 citation equating to 5 downloads. This would then give 30 citations in a year being equivalent to 150 downloads, roughly mirroring our metric for average downloads.

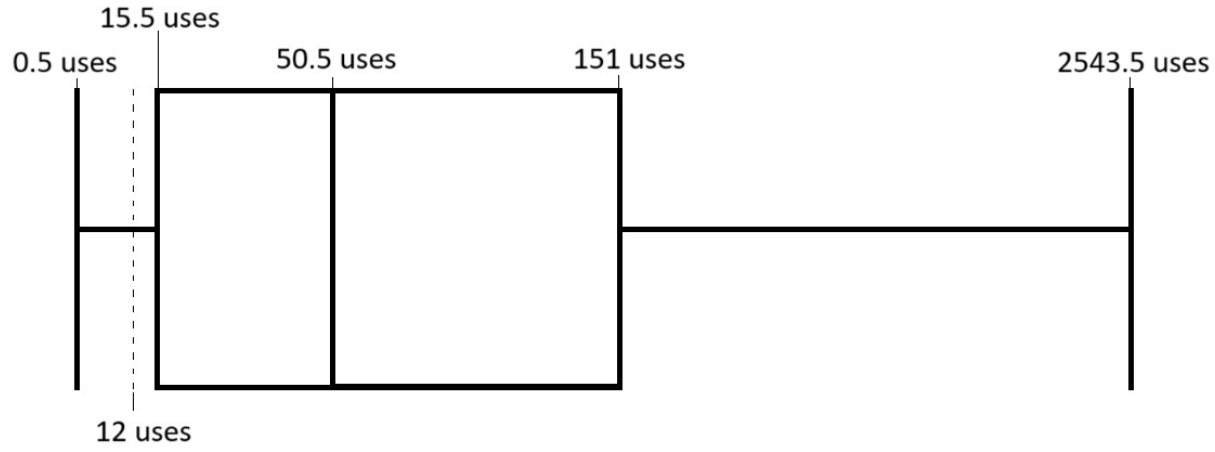

Figure 1. Quartile breakdown of 2-year average article downloads.

The other side of this, of course, was exceptionally poor performance. However, we didn't want to lose a particularly important title by being too aggressive in what we automatically excluded. To balance this, we only automatically excluded titles that had no downloads, or averaged less than a download per month (<12/year) as being of such insignificant use that their absence would not be as impactful to users (notated in Figure 1 above).

In our second phase we used a more nuanced approach that built upon our previously established value logic (Usage > CPU > Citations). Remaining metrics were divided into two groups based on whether they supported retaining the subscription (Positive) or supported dropping the title (Negative). These groups were then split into two tiers: the Positive into a higher tier for upper quartile scores and a lower tier for above average scores, with the Negative higher tier being the lowest quartile scores and the lower tier being below average scores (see Figure 2). This approach gave us a very rough weighting of our metrics by double counting the highest/lowest quartile; additional adjustments were made to individual metrics in the Excel functions as needed. The total of the positive and negative metrics can then be summed, providing a "calculated value" that we used to try to compare titles more holistically across all metrics. So, a journal might have, say, a really high level of usage, and get a positively weighted score for one or both usage categories, and it might also have a high CPU which would give it a negative score in that category. Since usage is weighted more heavily than CPU and could be represented in 2 different ways depending on the exact score, this would likely result in an overall positive score for the title.

Figure 2. Weighted tier calculations

(Note: A screenshot of the workbook used, where each column scores a title for its performance on a given metric based on the criteria for that tier. Totals for each tier are summed up to give the calculated value total (on the left). Columns for the Positive and Negative 'Tier 1' calculations have been hidden since they were mostly accounted for in Phase 1.)

This overall approach paralleled that of Blecic et al. (2013) in their work with journal metric data. They were able to rank journals by scoring each journal on each criterion and calculate a combined ranking score. They point out that librarians must use judgment to determine where any cutoff points may be that define where a middle ground of journal titles exists, requiring further and closer examination.

For us, if materials had a positive overall weighted score in the top ~⅔ of possible remaining scores (>5), we included these titles without further examination, assuming that there were enough overall factors or heavily weighted factors to pull the score into those upper levels.

A positive score below 5 would be in the lowest third of possible positive scores, and likely to have more middling or mediocre values such that the scoring could have gone either way. These titles were designated for manual review.

If materials had an overall negative score, and no redeeming positive ratings, we discarded the title from consideration. However, if there was even one redeeming positive rating, we checked the title manually to see if there was any compelling reason to include the title.

A small group of librarians with collections responsibilities reviewed the titles designated for manual review, using their professional judgment and qualitative factors to determine whether there were compelling reasons to keep the title. This included pulling in additional historic data, comparing title scope to priority institutional research areas, consideration of overall impact on users and budget, etc. The final determination on these titles was decided by consensus of the participating librarians based on the assessments made in this final review.

The resulting lists were shared with internal library stakeholders for review. They received lists of titles projected to be retained via STL, subscribed to specially, or dropped. Their task was to determine if there were titles that the larger analysis had missed and categorized improperly. For example, they could share information about titles on the retention list that were no longer relevant because of a program or staffing change. They could also share information about compelling reasons to retain a title like a new program focus, new researcher focus area, or demonstrated curricular need/embedment in the curriculum that the data hadn't captured.

During this process, faculty were not consulted directly about titles in the analysis. Our assessment and recommendation to UHL administration was to retain existing packages, making a need to consult faculty on loss of access redundant. However, throughout the negotiation period, faculty were regularly apprised of possible outcomes through a variety of communication venues.

CONCLUSION

Ultimately, this analysis allowed UHL to consider the viability of breaking up the “Big Deal” and pursuing a shared title list alternative subscription model. The data-informed analysis indicated a significant loss of access to titles being regularly used. The potential financial impact on ILL and subscription budgets to continue to provide access to that content was significant. After presenting potential lost access versus possible cost savings, we opted to preserve existing campus access by maintaining the “Big Deal.”

Given the number of students and researchers in the UH community, this form of analysis for "Big Deal” packages is ideal. It allows for consideration of important data points across a large scale of available titles without the time investment of comprehensive title by title examination. Titles that easily fall into the “keep” or “cancel” categories can be sorted readily and allow for qualitative data, where available, to be brought in for consideration of titles that need the additional input to solidify a decision.

This same approach can be applied to evaluate alternatives to a “Big Deal” in a variety of scenarios, including budget scarcity, contract renewals, or a desire to explore alternative strategies. There may also be other factors or services available that would impact analysis, results, and decision thresholds. For example, different institutions may apply a different ILL threshold against which to determine the value of instant access, impacting their cost-savings analysis. Alternatively, the implementation of an article on demand service could change how we consider cost per use or treat titles with both a high use and high cost per use. Article on demand services add an alternative to subscriptions that could make a-la-carte options preferable to ILL as it provides near-instant access yet may be cheaper than subscriptions for high cost per use cases. Combining article on demand with existing PTA and aggregator-supplied content may also present a viable alternative to subscriptions, at least in the short term, to ensure continued access while not having to maintain high subscription costs that may duplicate aggregator content and PTA coverage. Future applications of this methodology would have to account for local contexts, other factors, and services, but we believe this approach to be transferable to most scenarios. The core of the analysis remains sound regardless of the adjustments to variables that would be made to account for new contexts, factors, and services.

Acknowledgements

The authors would like to acknowledge with gratitude, the work of Ian Knabe. His collaboration and insights on this analysis were invaluable to its success. We also owe thanks to Nancy Linden for her time and efforts in the final title reviews.

References

Blecic, D. D., Wiberley, S. E., Jr, Fiscella, J. B., Bahnmaier-Blaszczak, S., & Lowery, R. (2013). Deal or no deal? evaluating big deals and their journals. *College & Research Libraries, 74*(2), 178-193. https://doi.org/10.5860/crl-300

Chamberlain, S. (2022). Lessons learned from reevaluating big deals with unsub. *Serials Review, 48*(3-4), 234-237. https://doi.org/10.1080/00987913.2022.2132090

Dawson, D. (2015). A triangulation method to dismantling a disciplinary "big deal". *Issues in Science and Technology Librarianship,* (80). https://doi.org/10.5062/F4610X9H

Dikboom, M. (2016). Tackling big deals: The experience of maastricht university. *Interlending & Document Supply, 44*(3), 93-96. https://doi.org/10.1108/ILDS-02-2016-0008

Fisher, E. (2021, March 21). Elsevier Negotiations – March Update, *UH Libraries News*. https://web.archive.org/web/20240222111258/https://libraries.uh.edu/blog/2021/03/25/elsevier-negotiations-march-update/

Gagnon, S. (2019). Journal publishers' big deals: Are they worth it? *Against the Grain, 29*(2), 22. https://doi.org/10.7771/2380-176X.7747

Hoeppner, A. (July 12, 2018). *Big deals and why I like them* [Presentation Slides]. *Showcase of Text, Archives, Research and Scholarship,* https://stars.library.ucf.edu/ucfscholar/860

Ivanov, A. O., Johnson, C. A., & Cassady, S. (2020). Unbundling practice: The unbundling of big deal journal packages as an information practice. *Journal of Documentation, 76*(5), 1051-1067. https://doi.org/10.1108/JD-09-2019-0187

Johnson, C. A., & Cassady, S. (2018). How librarians make decisions: The interplay of subjective and quantitative factors in the cancellation of big deals. *Collection and Curation, 39*(1), 6-14. https://doi.org/10.1108/CC-05-2018-0013

Lemley, T., & Li, J. (2015). "Big deal" journal subscription packages: Are they worth the cost? *Journal of Electronic Resources in Medical Libraries, 12*(1), 1-10. https://doi.org/10.1080/15424065.2015.1001959

Nabe, J., & Fowler, D. C. (2015). Leaving the "big deal" ... five years later. *The Serials Librarian, 69*(1), 20-28. https://doi.org/10.1080/0361526X.2015.1048037

Rathmel, A., Currie, L., & Enoch, T. (2015). "Big deals" and squeaky wheels: Taking stock of your stats. *The Serials Librarian, 68*(1-4), 26-37. https://doi.org/10.1080/0361526X.2015.1013754

Shu, F., Mongeon, P., Haustein, S., Siler, K., Alperin, J. P., & Larivière, V. (2018). Is it such a big deal? on the cost of journal use in the digital era. *College & Research Libraries, 79*(6), 785-798. https://doi.org/10.5860/crl.79.6.785

SPARC. (2023). *Big deal cancellation tracking.* https://sparcopen.org/our-work/big-deal-cancellation-tracking/

About the authors

Shawn Vaillancourt is currently the Collections Analytics Librarian at the University of Houston Libraries. He has been a professional librarian for over 15 years, and obtained his MLIS at the University of Western Ontario.

Kerry Creelman is an Associate Librarian and the Associate Dean for Collections Strategies & Discovery at the University of Houston Libraries. Her professional experience and interests cover teaching and learning, collections and discovery services, organizational structures and personnel development, and leadership and management.