Purdue University

Purdue e-Pubs

Charleston Library Conference

Time to Take New Measures: Developing a Cost-Per-Cited-Reference Metric for the Assessment of E-Journal Collections

Megan G. Kilb
University of North Carolina at Chapel Hill, griffma@email.unc.edu

Virginia B. Martin
Duke University, virginia.martin@duke.edu

Tessa L. Minchew
North Carolina State University, tlminche@ncsu.edu

Time to Take New Measures: Developing a Cost-Per-Cited-Reference Metric for the Assessment of E-Journal Collections

Megan G. Kilb, E-Resources Librarian, University of North Carolina at Chapel Hill

Virginia B. Martin, Head, Continuing Resource Acquisitions Department, Duke University

Tessa L. Minchew, Electronic Resources Librarian, North Carolina State University

Abstract

The current primary quantitative measure of e-journal subscription return-on-investment (ROI) is cost-per-use (CPU). While CPU is widely used, it also widely criticized and should not be relied on to the exclusion of other factors when assessing ROI. Because CPU is an imperfect measure, the authors developed a new, complementary metric for evaluating e-journal subscription ROI: Cost-per-cited reference (CPCR). CPCR assigns a dollar value to each citation of a particular journal by authors affiliated with the subscribing institution during a specified time period. By focusing on the content that researchers cite in their scholarly output, a CPCR metric assists in measuring the value of journal subscriptions to researchers and the institutions that support them. This article gives a very high-level overview of a collaborative project, conducted by librarians in the Triangle Research Libraries Network (TRLN), to develop a local CPCR metric and apply that metric to the evaluation of a consortial Big Deal. The authors explain CPCR, how they calculated and applied it to a particular shared Big Deal, and where they would like to take it in future. A more in-depth description of this project may be found in *Serials Review's* final issue for 2016 (Martin, Gray, Kilb, & Minchew, n.d.).

Introducing CPCR

Librarians have long struggled with assessing the value of their collections to their communities of interest. To address this struggle, a group of librarians in Triangle Research Libraries Network (TRLN) developed a quantitative measure, cost-per-cited-reference (CPCR), to better ascertain the value of the consortium's Springer e-journal package to the member libraries' users. The authors determined the number of times researchers within the consortium cited an article in each title in the Springer package, and then they compared that number to the itemized cost of the e-journal title to calculate CPCR for each title. This number allows librarians to look beyond annual download counts and cost-per-use to gain a better understanding of users' interactions with downloaded content, namely how often they use the content in the cited references of their publications and how that compares to the cost of the title. As institutions continue to whittle down their e-journal collections in response to budget cuts and other constraints, CPCR also provides an additional data point to help prioritize which resources and subscriptions will best meet local needs.

The CPCR metric, however, is not without its faults. Many of the pitfalls of traditional citation analysis and CPU analysis also plague the CPCR data point. For example, title changes and title mergers, all too common in serials with long publication histories, might lead to an artificially high CPCR and correspondingly low citation count, unless the previous title data is merged into the data for the current title. Additionally, the CPCR metric may heavily favor STEM fields that rely on journal literature to disseminate findings and publish research. Journals in fields with many published issues containing many articles per year will naturally have more content available for researchers to cite. Moreover, the composition of an institution's dataset of cited references can be affected by the extent to which journals in a field are indexed in citation indexing services used to compile data. Web of Science and Scopus, in particular, have historically indexed journals in the sciences more heavily than journals in the social sciences and humanities, which compounds the problem of favoring STEM titles when analyzing the value of a multidisciplinary package. Finally, as with traditional citation analysis, without examining the full-text of each locally produced publication, it is impossible to tell from a mere dataset how an article is being cited. Many of the articles cited might be used to

support and bolster a researcher's thesis, but in some cases, the researcher is citing an article to dispute and discount the claims being made therein. The inability to determine the author's intent in citing an article makes it difficult to definitively assign value to the e-journal that contains the article.

Methodology

The authors used Scopus to gather cited references for the consortial dataset. A standardized advanced search query was employed to identify all TRLN publications from calendar years 2013 and 2014. From there, cited references were downloaded for each set of publications. The member libraries' high-research outputs, coupled with protective measures put in place by the Scopus interface, made the data collection process somewhat complex and circuitous. The raw datasets were de-duped, merged, and compiled, first at the individual campus level and again at the consortial level. The merged dataset did not contain ISSNs, and, therefore, had to be matched against the Scopus title list of indexed content by title only. Unsurprisingly, this comparison yielded many titles with no matches. The authors manually reviewed the unmatched titles to identify and isolate problematic Springer titles, which were then added as alternate titles on the working copy of the Scopus title list of indexed content. At this point, the authors could use the Scopus title list to transfer

ISSNs to the raw datasets of cited references and identify titles in the Springer package that were not indexed in Scopus. Using these cleaner datasets, the authors calculated the number of articles that were cited by a TRLN researcher in a Springer e-journal during calendar years 2013 and 2014. To arrive at a CPCR for each title, the authors divided the 2013 and 2014 itemized subscription costs for each e-journal title by the number of articles cited by TRLN researchers in the corresponding years.

This newly calculated data point lends itself especially well to evaluating a consortially managed Big Deal e-journal package such as TRLN's Springer deal, which follows a shared ownership, curation, and access model, where participating libraries enjoy access to subscriptions curated by other member libraries. While this model allows for increased access to a wider array of content, it also requires member libraries to consider consortium-wide usage when making title-level decisions. The terms of the TRLN deal allow a small degree of flexibility in the form of cancellations, giving member libraries the incentive to develop a database tool that tracks costs and usage at the title level over the course of the deal. The authors could take data from the TRLN Springer ROI database tool and pull it into the existing CPCR dataset to examine title-by-title how the CPCR data compared to CPU data.
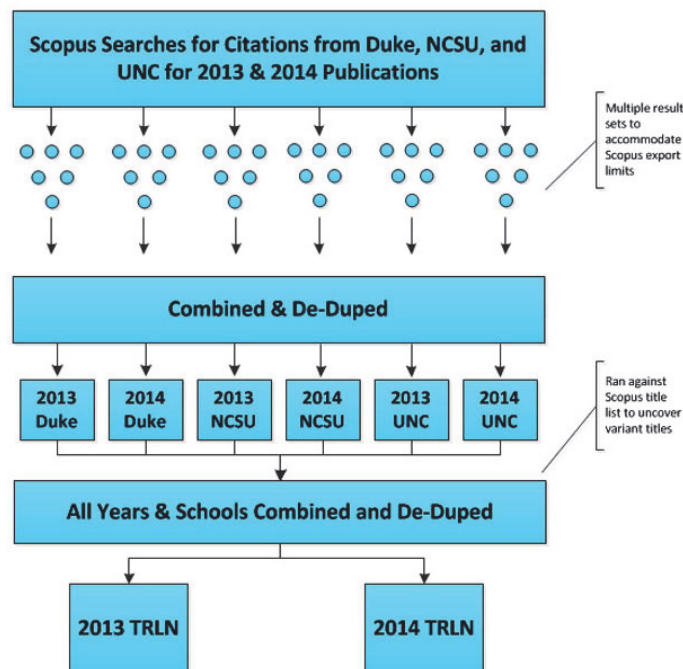

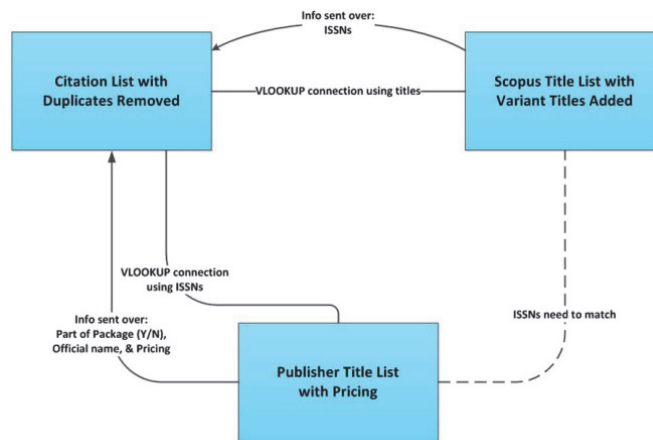**Figure 1. How cited references from Scopus were compiled for analysis.**

Figure 2. How cited reference data was cleaned up and merged with data from other sources.

Findings—Applying the Data to the Springer Deal

The authors first identified 40 titles with a high mean CPU in 2013 and 2014 and examined the titles' CPCR. In all but five cases, the CPCR was similarly high enough to confirm the authors' suspicions that these titles could safely be canceled. For the outlying five titles, the titles' CPCR was relatively low compared to their CPU. Additionally, the number of cited references had increased year over year, thereby reducing the titles' CPCR. The authors also reviewed the CPCR for the 63 titles that TRLN institutions had canceled for the 2016 subscription year and found that only three of these titles had a CPCR of less than $100, while 29 of the titles were not cited at all in 2013 and 2014.

CPCR data reinforced the vast majority of the consortium's actual cancellation decisions and also has the potential to influence decisions about titles that are not immediately clear cancellations or retentions. The COUNTER usage and CPU for *Wirtschaftsdienst*, for instance, suggested that it would be a good candidate for cancellation, but the CPCR data indicated that it may be of more value to researchers than the download data suggested. CPCR for *Wirtschaftsdienst* was actually lower than its CPU and had decreased considerably from 2013 to 2014, as TRLN scholars cited it more often in publications in 2014. In this case, having CPCR data available could potentially lead to the retention of this title when it otherwise might be cancelled based solely off usage and CPU data.

Next Steps

After developing an initial methodology for harvesting and analyzing the data, the authors are revising their data collection and analysis strategies for future iterations of the project. First, the dataset quickly became too unwieldy for Excel. Merging data and performing somewhat complex Excel formulas on files with hundreds of thousands of rows of data frequently crippled and crashed the machines the authors used to manipulate the data files. The authors are in the process of transferring the existing dataset to Access, which thus far, has been much nimbler in loading and manipulating the datasets. To further streamline data collection and reduce the chances of inconsistent and error-prone data retrieval, the researchers hope to take advantage of Scopus' API service to harvest further years of cited reference activity for the member libraries. Moreover, the authors hope to build up a dataset that covers a longer time frame. Adding more longitudinal-cited reference data to the dataset should give the consortium members a better idea of how their researchers' needs are evolving and how TRLN can shape the content of the deal to meet those needs. Finally, the researchers would like to explore how to apply the CPCR metric to the TRLN Big Deal package with Wiley, which follows a drastically different access model, relying less on shared curation and ownership.

References

Martin, V. B., Gray, T., Kilb, M. G., & Minchew, T. L. (in press). Analyzing consortial "Big Deals" via a cost-per-cited-reference (CPCR) metric. *Serials Review, 42*(4).