

# Pennsylvania Libraries: *Research & Practice*

Practice

## Using a Workflow to Reduce Database Costs without Affecting Collection Needs

Karen Sheldon

*Karen Sheldon is Electronic and Instructional Services Librarian at Delaware Valley University, [karen.sheldon@delval.edu](mailto:karen.sheldon@delval.edu)*

A small university with limited staff and resources created a workflow to help streamline decision making during renewal of subject databases. A literature review of common assessment methods found cost per use, content overlap analysis, and relevance to the curriculum to be the most common metrics. However, each individually provided an incomplete portrait of usage. A flowchart documented the decision process and incorporated multiple metrics. Implementing this workflow reduced the library's materials budget by 3% for the following fiscal year. This process may be implemented at other libraries to lower overall budget spending, while still providing content aligned with the local curriculum.

### Introduction

Postsecondary enrollments of four-year private nonprofit institutions continued to decline in 2016 (National Student Clearinghouse Research Center, 2016). Delaware Valley University, a small interdisciplinary institution outside of Philadelphia, was no exception. As such, the library reexamined its e-resource subscriptions to comply with reduced budgets. Having already trimmed the obvious “low hanging fruits” in previous years, it was important that any future cuts affected students and faculty as little as possible. Without a large library staff or faculty input, though, it was difficult to have a clear understanding of how a subject database was being used and its impact on campus. Using elements from several studies, it became possible to develop a process and workflow for assessing subject databases at a small institution.

### Developing A Workflow

Database usage and full-text download cost per use were important metrics for assessing e-resources over the last decade. Wichita State University Libraries had been using Excel spreadsheets to track e-journal and database usage since 1998 (Walker, 2009). Excel was the tool most commonly mentioned for libraries without an electronic resources management system. Maintaining a multiyear data set was important at the University of Nevada as well. Similar types of resources such as individual e-journals, e-journal packages, full-text databases, and aggregator databases were

assessed within their own category. The importance of evaluating like against like informed the scope of this workflow, limiting it to subject databases only.

A download cost per use threshold was often determined based on a categorical average, about \$5.00 for aggregator databases (Hiott, 2004; Tucker, 2009). Meanwhile, discussions reminded libraries that platform design and amount of content available were only two potential factors which could influence cost per use. Another important consideration was the usefulness of content, which could not be assumed based on full-text downloads (Bucknell, 2012). With this in mind, additional metrics were incorporated into the workflow following the initial \$5.00 cost per use threshold.

Popular metrics included top-used journal titles and content overlap. Top-used journals were identified using Pareto's Principle that 80% of usage comes from 20% of titles (Enoch & Harker, 2015). Having identified top used journal titles along the same criteria provided a manageable list for assessing overlap in the workflow, which could be run through Excel easily. Overlap was analyzed by quantity and quality, using factors such as citations and impact factor (Sutton, 2013). Librarians at University of North Texas explored several freely available and paid tools which could act as faster alternatives to Excel. However, inconsistencies in the results showed that multiple tools were necessary to get a true picture (Harker & Kizhakkethil, 2015). Add-ons to Excel such as AbleBits allowed tables to be merged easily. However, this required using a unique identifier for all journal titles. Not all e-resources had an ISSN; therefore this tool eliminated some titles from the packages (Kimball, 2016). Excel contains default options which can highlight duplicate values as part of its Conditional Formatting menu, but this was overwhelming when using a large data set. By limiting overlap within the workflow to top-used journal titles only, it was possible to determine content uniqueness broadly while only utilizing available tools. However, this did not account for content usefulness to an institution.

A cascading model at the University of Ottawa simplified the assessment of content usefulness by asking subject liaisons to only consider e-journals whose price had increased by more than 6% (Moisil, 2015). Many large university libraries focused on reduced e-journal individual subscriptions and "Big Deal" packages, using input from subject liaisons to develop cancellation lists (Pedersen, Arcand, & Forbis, 2014). Texas A&M University Libraries were tasked with reducing a financially unstable ProQuest collection while still supporting research needs. Content analysis data was paired with knowledge of curriculum and institutional research needs (Tabacaru et al., 2016). By taking the analysis a step further, this study was not only able to demonstrate which content was being used, but also to estimate how and by whom.

All articles found included libraries with either collection management teams or subject liaisons. For libraries without these resources, faculty can serve as subject experts. Faculty input was solicited at Salisbury University after subject liaisons had developed a short e-journal list for cancellation based on the previous year's usage (Hardy, Zimmerman, & Hanscom, 2016). In developing the workflow, it was apparent that not every academic department was open to a relationship with the library. Therefore, available information within the library from reference and instruction departments informed content usefulness in conjunction with faculty input.

A combination of these metrics was used, as shown in Figure 1. A \$5.00 cost per use, based on local averages, constituted the first red flag. Using Pareto's Principle, an overlap analysis of holdings within EBSCO identified unique content. Finally, content usefulness was determined based on the subject expertise of the reference and instruction librarians, as well as consultations with faculty. In this way, the workflow answered many of the same questions asked at large institutions without the financial or time burden.

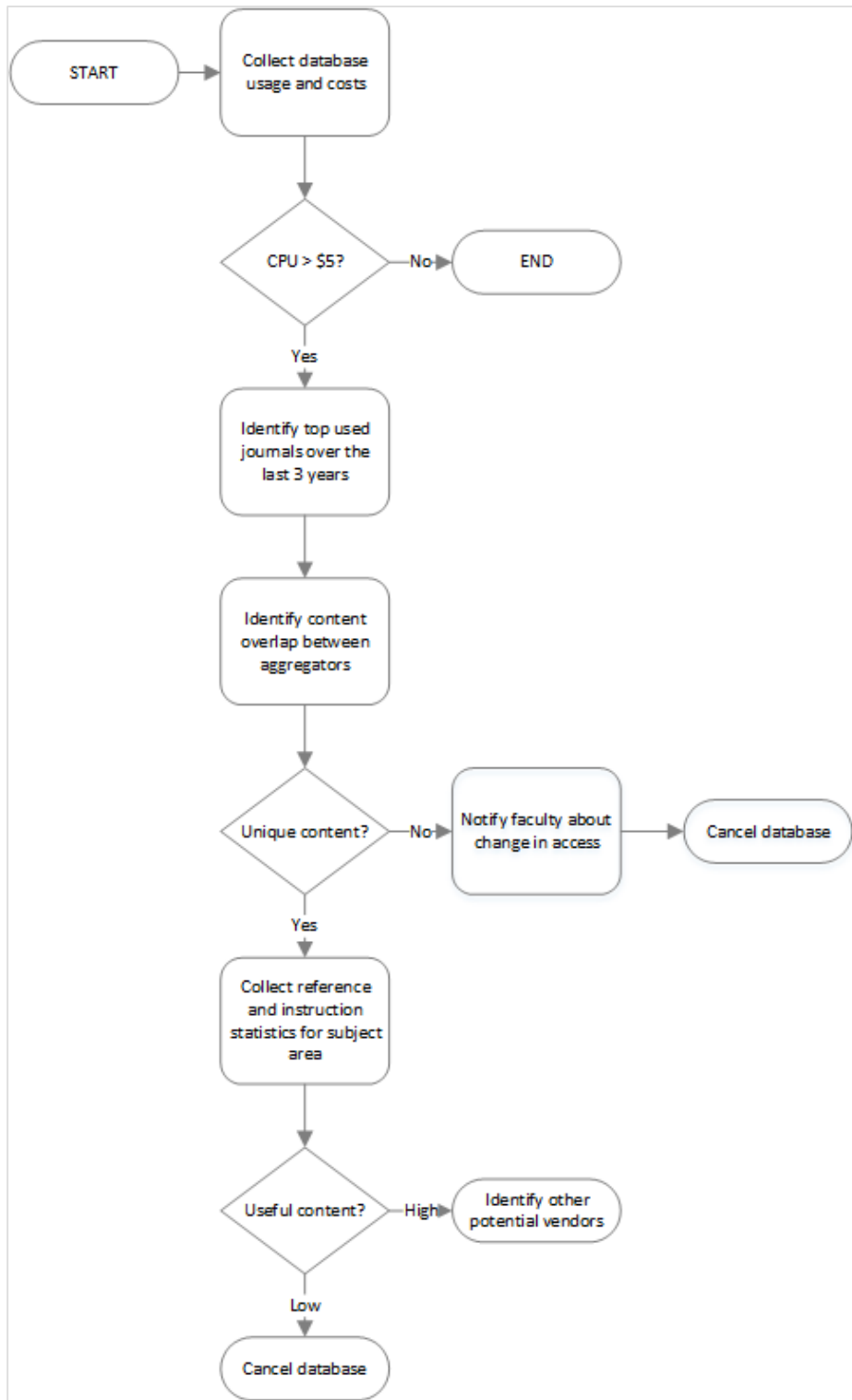


Figure 1  
Database renewal decision flowchart

## Workflow Decision #1: CPU > \$5?

All usage statistics reports were run annually to match the University’s fiscal year, beginning July 1st of the previous year and ending June 30th of the current year. Two different COUNTER 4 reports, Journal Report 1 (Number of Successful Full-Text Article Requests by Month and Journal) and Database Report 1 (Total Searches, Result Clicks and Record Views by Month and Database), were utilized when available. Multi-database platforms such as ProQuest, EBSCO, Gale, etc. did not offer Database Report 1. Standard or Platform Reports were used instead.

Data was entered into an Excel workbook, consisting of 3 sheets, which calculated cost per use and cost change for each database (see Figure 2). The first sheet, named Costs, included a list of all databases purchased by the library with the annual subscription cost. The second sheet, Usage, included the number of searches and full-text downloads per year by database. The third sheet, CPU Change, divided the annual cost by the number of searches and full-text downloads by database. The Cost Per Use (CPU) Change spreadsheet template may be downloaded at [bit.ly/2KRIP07](http://bit.ly/2KRIP07).

	2015-2016					2016-2017				
Databases	\$	S	D	\$/S	\$/D	\$	S	D	\$/S	\$/D
Academic OneFile										
Access Science	\$ 870.00	347	217	\$ 2.51	\$ 4.01	\$ 922.00	275	195	\$ 3.35	\$ 4.73
APA	\$ -	0	0			\$ -	0	0		
PsycArticles	\$ 5,500.00	6653	3510	\$ 0.83	\$ 1.57	\$ 5,775.00	4598	6677	\$ 1.26	\$ 0.86
PsycInfo	\$ 4,675.00	4884	0	\$ 0.96		\$ 4,909.00	4458	0	\$ 1.10	
PsycTests	\$ 2,420.00	2213	288	\$ 1.09	\$ 8.40	\$ 2,541.00	1201	462	\$ 2.12	\$ 5.50
PsycTherapy	\$ 6,037.00	240	102	\$ 25.15	\$ 59.19	\$ 6,325.00	36	123	\$ 175.69	\$ 51.42
Biological & Agricultural Index										
BioOne Combination	\$ 2,783.00	872	1018	\$ 3.19	\$ 2.73	\$ 2,868.53	706	984	\$ 4.06	\$ 2.92

Figure 2  
Database usage tracked annually in Excel

The CPU Change sheet referenced cost, search, and download data from the previous two sheets. The cost per search and cost per download columns were formatted using a Highlight Cells rule from the Conditional Formatting menu to identify values over the previously established \$5.00 cost per use threshold. This allowed for the easy identification of databases with low usage. If the database had costs per use over \$5.00 for 3 years, it became subject to further assessment. Three years of data accounted for variations in assignments, course offerings, and faculty. After having entered data for the year 2015-2016, only two subject databases had usage costs over \$5.00 for the last three years: Communication and Mass Media Source (CMMS) from EBSCO and Criminal Justice Periodicals (CJP) from ProQuest.

## Workflow Decision #2: Unique Content?

Journal Report 1 was run for both databases over the same three-year period. From these two reports, the top-used journal titles were identified. Twenty percent of journal titles with usage were included, following Pareto’s principle. Using this criteria, CMMS had 50 journals which received high usage. The top used journal title was *Language in India*, followed by the conference papers of the International Communication Association. CJP had a list of 22 journals to compare. *Reclaiming Children and Youth* and *Criminology* were tied for top usage, followed by *Policing*. These lists were then compared to the library’s complete journal holdings.

This required creating a second Excel workbook containing the entirety of the library’s journal holdings, named Database Overlap (Figure 3). All holdings were entered in the first sheet of the workbook, named All Holdings. Holdings Report from EBSCO was used to identify all holdings. The report included information such as hosting platform and journal coverage dates. On the second sheet, named Without CMMS, all holdings except those from the database being assessed, in this case CMMS, were included. A third sheet, named CMMS Comparison, was added to the same workbook and the top used journal titles from CMMS were entered. An IF logical test was entered into a blank column next to the journal titles on the second sheet, which identified duplicates. These duplicates were examined further for coverage differences. Of top used journal titles from CMMS, 30% were duplicated in other subject and general databases owned. The remaining 60% had no overlap, indexed or full-text.

1	Title	Full DUPLICATE?	CMM Start	CMM End	Source	Start	End
2	Language in India	132 Duplicate	3/1/2006	present	OA EBSCO	1/1/2001	present
3	Conference Papers -- International Communication Association	32					
4	Quarterly Journal of Speech	29					
5	COMMUNICATION RESEARCH REPORTS	22					
6	Publishing Research Quarterly	20 Duplicate	3/1/1991	1 yr delay	ABI	3/1/1997	1 yr delay
7	COMMUNICATION EDUCATION	19					
8	Journal of health communication	19					
9	JOURNAL OF COMPUTER-MEDIATED COMMUNICATION	18 Duplicate	11/1/2005	1 yr delay	OA EBSCO	1/1/1995	present
10	Media Psychology	18					

Figure 3  
Identifying duplicates titles from Communication and Mass Media database

Another list with all holdings except CJP was made and the comparison was repeated. Of the CJP journals used in the last three years, 40% had more coverage in Academic Search Complete. Five of the remaining journals were unique, including the most used journals, *Reclaiming Children and Youth* and *Criminology*. This analysis found that CMMS and CJP primarily had unique content.

## Workflow Decision #3: Useful Content?

Relevance of database content was determined using reference and instruction statistics. Reference statistics were tracked in a shared Excel workbook. Entries were broken down by time of day, date, weekday, and question type with optional fields for teaching faculty, course name, and notes. This notes area was used to provide additional details about the topic searched, available resources, etc. Reference trends were determined at the end of each semester using Pivot Tables and through discussions with library faculty. Pivot Tables within Excel allow users to extract significance from a large data set by summarizing data in another table. In this instance, Pivot Tables were used to determine at

which times of day questions were asked, what types of questions were asked, and which courses were most represented. Instruction statistics were also kept in Excel, recorded by semester and by subject area, which helped document other research trends and assignments.

Searches replicating known assignments showed that the majority of communications topics were well served by more general databases, such as Academic Search Complete. Specialized topics such as dialect history, oral traditions, and instant information societies had fewer resources, but were also less of a focus in the Media and Communication program. CMMS no longer supported the institution's needs as it once had.

The university's Criminal Justice program was much larger than Communication and Mass Media program. There were no reference statistics and few instruction statistics to provide guidance to the research done in that area. CJP was also a ProQuest product and, therefore, not included in EBSCO's Discovery System (EDS). This tool was a cornerstone of the mandatory freshman information literacy sessions. As all students had some familiarity with EDS, preference was typically given to content which is readily indexed and accessible through it. CJP gave the appearance of not supporting the institution's needs.

## Results

Based on the logic of the workflow, CMMS was prime for cancellation. This information was shared within the library and the database was cancelled for the year 2017-2018. Discussions with the English and Communication faculty were productive, especially as majority of the content used was available through other databases. CJP should be cancelled based on the same criteria of cost per use and unique content. Low engagement and interaction from faculty and students led to few purchases in the subject area. However, the Criminal Justice program had one of the highest enrollments at the university and a master's program was soon to follow in fall 2018. An emphasis was placed on improving the library's relationship with faculty and increasing awareness of resources. If usage does not improve in the upcoming years, the possibility of purchasing access to the top utilized journals individually or through other vendors will be revisited.

## Limitations

This study was limited to subject databases rather than individually purchased electronic journals or generalized databases. Subject databases were expected to have less usage than more general collections. Individually purchased journals, likewise, were not comparable. Many were purchased for a specific course or faculty member and had lower usage. Metrics used will also likely change in future years. A \$5.00 cost per use may change as the value of \$5.00 decreases or if the threshold no longer remains the average cost per use. Acceptable amounts of overlap will also vary with budgetary demands and discovery system integrations. Content needs will change as university programs develop and shift focus. While this workflow is not a perfect system, it does have the flexibility to adapt. Additional metrics such as impact factor may be added in the future in order to strengthen the assessment of content quality, which is currently only measured by usefulness to the institution.

## Applications for Other Libraries

One of the goals of this process was to develop a workflow which would be able to evolve as well as allow other libraries to adapt it to their own needs. This workflow includes that flexibility. COUNTER Release 5 will become effective in January 2019 and will become the first important adaptation libraries will have to make. Database Report 1 will become part of DR\_D1 Database Search. Journal Report 1 will be replaced by TR\_J1 Journal Requests. This will

expand upon Journal Report 1 by including Gold Open Access and Archive articles. Inclusion of these titles may affect cost per use.

Adaptation may also be required in establishing the cost per use threshold if \$5.00 is inappropriate based on local averages. The current workflow is designed primarily to assess the databases rather than reduce budget spending. If that were the objective, a lower cost per use threshold would be more appropriate.

Delaware Valley University, as an EBSCO subscriber, has easy access to download all the library's journal holdings in order to assess content uniqueness. Libraries without this service could still download holdings from individual platforms. Some initial difficulties in developing consistent formatting should be expected. However, unless there were significant title changes, the holdings would not need to be updated often.

The metric with the most opportunities for local adaptation is content usefulness. Libraries with collection management teams or subject liaisons would do well to incorporate their input on the appropriateness of content based on the local curriculum. Surveys of students and faculty could also provide an additional layer of information to assess user preference. Libraries with subscriptions to services such as Journal Citation Reports from Web of Science could include citation reports and journal impact factors in their workflow to assess the value of databases to their field.

## References

- Bucknell, T. (2012). [Garbage in, gospel out: Twelve reasons why librarians should not accept cost-per-download figures at face value](#). *The Serials Librarian*, 63(2), 192-212. doi:10.1080/0361526X.2012.680687
- Enoch, T., & Harker, K. R. (2015). [Planning for the budget-ocalypse: The evolution of a serials/ER cancellation methodology](#). *The Serials Librarian*, 68(1-4), 282-289. doi:10.1080/0361526X.2015.1025657
- Hardy, B. B., Zimmerman, M. C., & Hanscom, L. A. (2016). [Cutting without cursing: A successful cancellation project](#). *The Serials Librarian*, 71(2), 112-120. doi:10.1080/0361526X.2016.1196635
- Harker, K. R., & Kizhakkethil, P. (2015). [The quick and the dirty: The good, the bad, and the ugly of database overlap at the journal title level](#). *The Serials Librarian*, 68(1-4), 249-254. doi:10.1080/0361526X.2015.1016858
- Hiott, J. (2004). [Collecting and using networked statistics: Current status, future goals](#). *The Library Quarterly: Information, Community, Policy*, 74(4), 441-454. doi:10.1086/427414
- Kimball, R. (2016). [Journal overlap analysis of the GeoRef and web of science databases](#). *Science & Technology Libraries*, 35(1), 91-98. doi:10.1080/0194262X.2015.1128374
- Moisil, I. (2015). [Renew or cancel? Applying a model for objective journal evaluation](#). *Serials Review*, 41(3), 160-164. doi:10.1080/00987913.2015.1065466
- National Student Clearinghouse Research Center. (2016). [Current term enrollment estimates - fall 2016](#). Retrieved from <https://nscresearchcenter.org/wp-content/uploads/CurrentTermEnrollment-Fall2016.pdf>
- Pedersen, W. A., Arcand, J., & Forbis, M. (2014). [The Big Deal, interlibrary loan, and building the user-centered journal collection: A case study](#). *Serials Review*, 40(4), 242-250. doi:10.1080/00987913.2014.975650
- Sutton, S. (2013). [A model for electronic resources value assessment](#). *The Serials Librarian*, 64(1-4), 245-253. doi:10.1080/0361526X.2013.760417
- Tabacaru, S., Pickett, C., Stephens, J., Ugaz, A., Ramirez, D., Xiao, D., . . . Reddy, D. E. (2016). [Bundled database subscriptions: Challenges, opportunities, and assessment](#). *Serials Review*, 42(1), 31-38. doi:10.1080/00987913.2016.1142928
- Tucker, C. (2009). [Benchmarking usage statistics in collection management decisions for serials](#). *Journal of Electronic Resources Librarianship*, 21(1), 48-61. doi:10.1080/19411260902858581

Walker, M. (2009). [E-resource statistics: What to do when you have no money](#). *Journal of Electronic Resources Librarianship*, 21(3-4), 237-250. doi:10.1080/19411260903466426