Gone, but not Forgotten: An Assessment Framework for Collection Reviews

By: Kristin Calvert & Whitney Jordan

Abstract

Assessing the effectiveness of previous cancellation decisions is often overlooked in favor of current competing priorities. The authors outline their approach for evaluating journal, journal package, and database cancellations; how they evaluated whether the library could adequately support the needs of academic programs after substantial cancellations; and how to ensure the evaluative criteria used for collection review are useful and meaningful. They provide a framework for assessing database cancellations and the criteria used in their collection reviews.


GONE, BUT NOT FORGOTTEN:
An Assessment Framework for Collection Reviews

Kristin Calvert and Whitney Jordan*

BACKGROUND
Following the university’s recent reaffirmation of accreditation, the campus culture of assessment focused on continuous improvement. There has been a greater emphasis on connecting previous assessment results to new initiatives to demonstrate that assessment informs decision-making. Within the library, technical service librarians began applying the principle of continuous improvement to our collection assessment activities. At the end of a major cancellation project, librarians would regularly debrief and discuss how well the process went and what could be improved. For subscription cancellations, the library regularly reviewed interlibrary loan requests for cancelled titles; however, there has not been an obvious way to measure the impact of database cancellations.

The authors decided to dig into the idea of assessing collection assessment. They began with reviewing available studies and were able to find a significant body of literature on journal cancellations; a fair amount published specifically on Big Deal cancellations; but almost nothing on assessing database cancellations. In fact, most formal descriptions of collection assessment describe it as a cycle, but not in the same way continuous improvement frameworks would consider it. For instance, Pesch’s electronic resources life cycle is to acquire, provide access, administer, support, evaluate, renew. The basic steps for a collection assessment project are metric selection, data collection, soliciting feedback, making decisions, sharing results, and record keeping. They almost inevitably end with a decision to cancel or renew, with the cycle repeating for the renewed resources.

Critically reviewing the process, and whether the outcomes of the assessment were met, is an underdeveloped area of study for collections assessment. The authors were unable to find any significant investigations into the decision itself outside of the journal case. Furthermore, the discussion or reflection portion of many articles consists of anecdotal evidence from the librarians responsible. The authors wanted to have a more formal process for assessing cancellation decisions.

Over the past ten years, the library has undergone four major cancellations and several smaller reviews. There are very few “easy-to-cancel” subscriptions remaining. Everything is either highly used, needed for accreditation, or contains unique materials unable to access through other means. The next time a resource is canceled, it will be missed and will undoubtedly impact the library’s ability to support academic programs on campus. It is crucial for the authors to identify a way to review the decisions made beyond the

*Kristin Calvert, Head of Discovery and Technology Services, Western Carolina University, kcalvert@wcu.edu; Whitney Jordan, Former Acquisitions Librarian, Western Carolina University; whitney.p.jordan@gmail.com
gut feelings surrounding those decisions. This research was intended to improve local processes and develop a framework for assessment that was adaptable and holistic. The authors sought to avoid purely numeric assessment (such as use or turnaways) and capture the qualitative impact collections and cancellations have on reference, instruction, and research. Furthermore, they hope their assessment can also be used to make a case to the university for additional funding.

LITERATURE REVIEW

Much of the literature on collection review focuses on the process used when undertaking a collection review or assessment project. Articles outline the thought process and steps taken throughout the project including factors leading to the project, the scope, data collected, criteria applied, feedback received, communicating results, and offer reflections on the overall project. The project ends with communicating the results of the assessment and, more often than not, there is not a plan outlined to follow up on the outcomes of the collection review in order to assess whether the decisions enacted were good decisions. When a follow-up assessment on collection review decisions does happen, it tends to revolve around journals and journal packages where librarians have access to quantitative data, which they can use to determine whether the decision(s) made in a cancellation of this type had a negative impact. However, similar research on conducting an assessment for database cancellations in order to determine whether those decisions limited the impact on the campus community could not be found. Nixon offered a reflection of missed opportunity and what could have been done differently in their journal cancellation project; however, it is presented more as a 'lessons learned' rather than as an assessment of the review process.

The assessment of a collection review must focus on whether outcomes were met. These outcomes may be determined by the reason for undergoing a collection review process. Often they are undertaken in response to an external pressure, such as space needs, budgetary restrictions, increased subscription costs, or evolving campus programs, etc. The outcome in these situations are often specific and easy to measure, for instance a set dollar amount or recovering a certain amount of linear feet of shelving.

Recently, the literature on collection assessment has shifted towards making them more sustainable and strategic. In their article on building a sustainable collection review process, Keisling and Sproles state that “[a]lthough the literature addresses many aspects of creating and applying data and rubrics to collection assessment, there remains a dearth of discussion on creating a sustainable, on-going process that records and retains decisions, data, and input over time.” An advantage to regular, on-going collection reviews is that they ensure librarians have a better understanding of the collections strengths and weaknesses. Furthermore, Lantzy et al state that “[t]his dramatic increase in the development of formalized collection assessment processes reflects one way academic libraries are responding to pressures to create and demonstrate more value in their collections with less funding.” Measuring outcomes when the impetus for collection review focuses on the intrinsic and extrinsic value of collections will require new metrics and a more nuanced approach.

With a formal, continuous process for collection review, it is important to iteratively improve the process. Assessing the quantitative data provided for the collection review needs to be assessed to remain relevant to decision-makers. Librarians often make as much data available as possible to decision-makers, which can result in data fatigue and cause those reviewing resources to disregard data when making their recommendations. Keisling and Sproles note in their conclusion that "for some reviewers, volumes of data overwhelmed and occasionally disregarded or neutralized the data." While many studies show that subject liaison and faculty input is helpful in making final decisions and getting buy-in, it is vital to have a process to ensure the data being provided is helpful in the decision-making process.

RESEARCH QUESTIONS

When undertaking an assessment of the decisions made during past reviews, the authors decided to take a three-part approach. Part one focused on journal and journal package cancellations. Part two focused on database evaluation and cancellations. Lastly, the overall process was reviewed.
With journal cancellations, there was a bottom line to meet and the outcome set was a fixed dollar amount of savings. To assess that outcome, the library needed to know whether the cost savings justified cancelling these titles. This was also the library’s first experience cancelling a large journal package and the expectation is that more packages will be cancelled in the future. It was important to track whether there was still a demonstrated need for the cancelled journal titles because no individual pick-ups were made. There were quite a few high use titles, especially in the cancelled journal packages, and tracking demand was a condition for many accepting the cancellation decision. Anecdotally, the authors had heard from teaching faculty and subject liaisons about the importance and necessity of specific journals that had been identified for cancellation but struggled with how to balance that professed need with the available data. With all that in mind, the study set out to determine

1. What were the actual cost savings?
2. Was there patron demand for cancelled titles, particularly for the higher use titles in the journal packages?

In thinking about how to assess database cancellations, it was necessary to take a step back and examine how librarians review databases more generally. The goal for this part of the study was to identify any common criteria used to evaluate databases and use these common values to create assessment metrics. The secondary goal was to explore what a “bad” decision might mean. If the wrong database was cancelled, what would that look like? Patrons needed to be more than just inconvenienced. It would be necessary to demonstrate cancellations have harmed the library’s ability to provide service to its patrons. This portion of the study sought to answer the following questions.

1. How do subject librarians assess databases for cancellation? Are there common values?
2. What effect have cancellations had on student learning, reference services, and research?

DEVELOPING AN ASSESSMENT FRAMEWORK

Journals

In 2016/17, Hunter Library underwent a collection review to address a significant budget shortfall. This review cycle resulted in the cancellation of two journal packages, seventy journal subscriptions, and a selection of databases. The decisions made during the collection review needed to bring the budget back into balance by reducing costs while minimizing negative impacts on the campus community as much as possible. In our approach for assessing these outcomes for journal cancellations, three methods were used: 1) reviewing post cancellation use, 2) calculating hypothetical cost, and 3) calculating the actual expenses incurred.

For post-cancellation demand, turnaway reports were reviewed for the journal package titles as well as any use that incurred a charge. This use included journal subscriptions, article requests from on-demand services, like Get It Now, and interlibrary loan (ILL) requests. From this, ongoing demand and post-cancellation costs could be measured.

For hypothetical costs, the library’s previous subscription cost was used and then projected using an estimate of inflation. Individual journal price inflation was estimated using EBSCO’s five-year price analysis. Journal package inflation was estimated using EBSCO’s historical price analysis (the annual inflation rates were not consistent or fixed by the contract). This was done once for all journals and journal packages to estimate expenditures had these cancellations not taken place. Next, the cost to subscribe to individual journals which had been requested through ILL or another on-demand service was calculated using the pricing listed in EBSCOnet.

Then, the actual ILL and on-demand services expenses incurred for articles from the cancelled titles was calculated. These costs included Get It Now article fees, subscription costs for one resubscribed journal title, and the actual ILL costs over the post-cancellation period. ILL costs included ILL fee management (IFM) transaction costs ($0.25 per request), IFM library-to-library borrowing charges, and ILL overhead cost ($13 per request). The overhead cost-per-transaction utilized the total overhead cost divided by the number of borrowing, lending, and document delivery transactions for filled and unfilled requests. The ILL overhead cost for FY19 were calculated from information in Table 1.
Kristin Calvert and Whitney Jordan

The analysis found that the actual expenditure over two years was $2,837. In comparison, the library would have spent $153,487 had these subscriptions not been cancelled or $263,780 had we elected to subscribe to any title with an ILL or Get It Now request. The cancellations represented a significant cost avoidance as the actual expenditure reflected only 1% of the hypothetical costs.

Based on the data gathered, the authors believe the decision to cancel these individual journals and two journal packages did not have an adverse impact on ILL or on the campus community. High-use titles in the journal packages did not translate to ILL and document delivery requests. The same has held true for high turnaway titles. The authors speculate that this disconnect between turnaways and ILL requests could be a result of casual browsing, patrons not considering ILL to be an acceptable replacement service, or patrons procuring the articles through other means.

**Databases**

Journal cancellations can be assessed and ongoing demand tracked; however, databases do not have comparable options. Anecdotal evidence suggested individuals had regretted cancelling certain resources. The authors were interested in finding a way to evaluate those regrettable decisions and the process that produced them. They began thinking about ways to assess the effectiveness of the criteria provided to help guide those decisions. A review of the literature yielded no shortage of literature on how libraries have conducted their collection review and what criteria they employed during the review process. Nothing was found which covered how to determine whether or not a good decision was made or whether the criteria were helpful to others.

To get a better understanding of their process for collection review, the first step in the study was to informally interview two subject liaisons. The information gained in this interview was developed into a survey to be distributed to the rest of the library. These interviews asked each subject liaison about how they reviewed databases, both within their discipline and outside it, and what was most important to them when considering a database for cancellation. They were asked about the criteria provided by technical services librarians: the usefulness of the criteria, whether they added their own criteria, and if/how they involved teaching faculty. The survey consisted of three parts, which explored the criteria provided for the collection review, the review process itself, and the impact of the cancellation decisions.

The first section of the survey focused on criteria used to evaluate databases and was open to everyone in the library who had some involvement with collection development. In addition to reviewing criteria previously used when evaluating databases, a literature review was conducted to look at other characteristics and factors to consider. The criteria collected was consolidated into a single list. An exhaustive list of criteria was compiled,
then was grouped into smaller and smaller categories, until a list emerged that was short enough for people to rank meaningfully. The survey began by asking everyone to rank six broad categories of collection development factors: program or institutional needs, cost/cost-per-use, availability of alternative content, intended audience and scope, faculty feedback, and database features, interface and content.

A total score was calculated for each ranked category of criteria by multiplying each score by the number of people who assigned that ranking, and then summed them. Specific program or institutional needs was the most important factor and faculty feedback was ranked the least important. However, a statistical test on the data, called a Friedman rank sum test, was applied to be more rigorous. This test is like an ANOVA, or analysis of variance test, but can accommodate sample sizes that are too small to make assumptions about normality. The test tells us whether at least one of the categories is more or less important than at least one other category. With a p-value of 0.29, there is insufficient evidence in the data to say there is a difference in importance between the categories. Each person ranked them too differently.

The survey asked everyone to review the full list of twenty-four specific criteria and pick up to ten that were the most important to them. The most selected criteria were usability/interface, full-text availability, uniqueness of content, the program size or number of users served, faculty feedback, accreditation requirements, and cost/cost-per-use. The full list is available in Appendix B.

The second approach taken to understand how librarians evaluate databases was to look into decision-making theory. During the preliminary interviews, the subject librarians were asked to reflect on decisions they made to cancel databases. There were several dichotomies that came up in the interviews. Some librarians preferred access to full text and others prioritized indexing, as full text can always be requested through ILL. Some were comfortable knowing that the information was available in another database (or freely available online) and others felt the ease of use or other features made the resource extremely valuable. There are some people who prioritize databases that are used by a large number of students and some who focus more on databases that serve a purpose that cannot be replaced or replicated elsewhere. When the choice of cancelling databases comes down between two very different products, there needs to be a way to choose between apples and kumquats.

The analytic hierarchy process is a structured decision-making technique often utilized for group decisions. In AHP, the group identifies a set of criteria they will use to make decisions. It applies relative numeric weights to each criteria using pairwise comparisons. The alternatives are reviewed based on these weights and the outcome is then decided by comparing the overall scores. A common example is with car buying. A family needs to buy a car and have identified price, mileage, and horsepower as the criteria. They would then compare each criteria side-by-side. Is mileage more important than price? If so, by how much? They would discuss each pair in turn. Then they would look at their potential cars: a hybrid sedan, an SUV, and a minivan. Each alternative would be given a relative score in each criteria. A hybrid sedan would score high in mileage but low in horsepower. The SUV would score higher in horsepower but might be more expensive.

This approach was applied with our selectors and these database dichotomies. Using the feedback gathered in the survey, the authors created a model that would be used to decide between three databases the library had cancelled in the last round: LexisNexis, PrivCo, and America: History and Life. LexisNexis was highly used by a large number of people, but the content was duplicated in several other databases. PrivCo is a database of private company data used by business students. It was used by a small percentage of our students and faculty, had overall low use, but consisted of content and information not available anywhere else. Finally, America: History and Life was a highly used abstract and indexing database available in EBSCOhost with a fair amount of the content available in other databases.

The criteria were:

- **Added Value**: The content may be available elsewhere, but this database provides added value through a better search experience.
- **Number of patrons**: the database serves a large number of patrons or academic programs. Cancelling it would impact many people.
- **High Use**: The database has very high use. It’s well-utilized and cost-effective.
Full Text: The database has full text. Other databases may only have citations or abstracts.

Uniqueness: The database contains unique content. We'd be unable to get the information anywhere else, whether through interlibrary loan or otherwise.

For true application, a single, consensus model would be built. Here, an AHP model was built based on each response (n=10) to look for differences and similarities between individuals. A mean response model to aggregate the group's perspectives was also calculated. From here the authors performed two analyses. First, each model was looked at to identify clusters of similar priorities. While there were no consistent patterns, three general groupings were apparent. One group of people highly prioritized databases with high use and that are used by a large number of patrons. Another group highly prioritized unique content above any other factor. The last group had fairly equal weights across all categories; these people are the 'It's all important/It all depends' group.

The second analysis was to see which database in the scenario would be kept or cancelled by the model. Despite the very wide range in weights, every model kept LexisNexis and all but one model cancelled America: History & Life. This may be a weakness in how weights were assigned to each characteristic of these three databases by the authors. A step which was not performed, but which would be interesting, would be to compare AHP model decisions to liaison opinions. Each person would be asked which databases they think they would keep or cancel in this scenario and compare their stated decision versus their model's decision.

The middle survey section continued asking about criteria but focused on the process employed in the most recent collection review. Instead of asking about criteria in general, respondents were asked about how the criteria and guidance supplied by technical services librarians was used in practice. In this part of the survey, the authors hoped to gather critical feedback about how their colleagues viewed the process as a whole and whether their views on criteria used to evaluate databases had shifted after the decisions were made. This part of the survey was only open to librarians who had participated in the previous collection review.

The final part of the survey was only open to librarians with liaison responsibilities and looked for anecdotal evidence about whether or not the decisions made during the collection review had an impact on the teaching, learning, and research needs of the campus community (see Appendix 1). The goal for this part of the survey was to learn about how effective the process employed was, what went well, and whether there had been any regrets. It looked specifically at each participant's experience, on the reference desk, in reference interviews, in their own teaching, etc. and asked whether there was a resource the library no longer had that would have been critical to the success of that interaction and for a brief justification for why. It also asked whether there were any resources cancelled that had not been missed.

As with other areas of the survey, the feedback varied widely. Overall, they were satisfied with our process and our criteria. Many librarians described how they used their knowledge of students, faculty, and their subject area to help them make decisions. It wasn't much of a surprise to see that about half expressed difficulty with evaluating databases outside their area. Without that first-hand knowledge, many feel unequipped to judge a database. Likewise, the databases they missed having tended to be ones in their discipline and the ones they didn't miss were specific to other disciplines. Overall, nearly everyone had an example of when a database would have been useful, though most were able to find workarounds.

There were several inconsistencies in the survey responses. Questions were intentionally included that tried to get at the same information multiple times to try to cross-validate the responses and catch these sorts of discrepancies. First, there was no clear consensus among respondents. The authors would be interested in looking at a wider dataset, though they suspect that in this area, decisions will always be highly individualized and likely never something that can be generalized across all libraries. Second, although nearly everyone said faculty feedback was crucial, it's consistently ranked very lowly by everyone. And third, there are criteria like indexing and abstracting, and high use, which are ranked highly, but the feedback received from the open-ended questions really highlighted that some of the biggest regrets are from the opposite side of the same coin. And so where the authors struggle, and will continue to, the next time databases are evaluated, is because it is not possible to evaluate on just a few criteria. It's rarely possible to use the same criteria to judge every type of database.
TRANSLATING RESULTS INTO ACTION

Figure 1 illustrates the steps of a collection assessment project including the assessment of the assessment. In this section, the authors offer the reader some guidance for designing a collection review through the lens of continuous improvement.

The first step is to set clear outcomes. The authors recommend the outcomes not be limited to a single financial target to hit. Outcomes should be measurable and followed up on to demonstrate the success or lack of success of the review process.

The next step is to identify criteria and metrics. Table 2 offers examples and suggestions to consider, some focus on journal and database cancellations, and others on the overall collection review processes. Criteria should be specific to both the library and to the assessment. A smaller, subject-specific collections project requires different criteria than a comprehensive review project. The criteria employed need to be tailored to the resources and desired outcomes.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Sample Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost Savings</td>
<td>Reduce subscription costs to 60% of package costs</td>
</tr>
<tr>
<td></td>
<td>Net savings of $50,000</td>
</tr>
<tr>
<td>Cost Efficiency</td>
<td>Subscription CPU vs. ILL CPU</td>
</tr>
<tr>
<td>Low Post-Cancellation Demand</td>
<td>Filled article requests do not exceed copyright limits</td>
</tr>
<tr>
<td></td>
<td>Few faculty subscription requests</td>
</tr>
<tr>
<td>Stakeholder Satisfaction</td>
<td>70% of faculty members were ‘satisfied’ or ‘highly satisfied’ with</td>
</tr>
<tr>
<td></td>
<td>…the collection review process</td>
</tr>
<tr>
<td></td>
<td>…opportunities to provide feedback</td>
</tr>
<tr>
<td></td>
<td>75% of librarians ‘agreed’ or ‘somewhat agreed’ that</td>
</tr>
<tr>
<td></td>
<td>…the collection review criteria were helpful</td>
</tr>
<tr>
<td></td>
<td>…the cancellations were fair and equitable</td>
</tr>
<tr>
<td>Equitable Across Disciplines</td>
<td>Number of departments affected by cancellations</td>
</tr>
<tr>
<td></td>
<td>Current strategies and cancellations do not disproportionately affect</td>
</tr>
<tr>
<td></td>
<td>departments that were targeted in past cancellations</td>
</tr>
<tr>
<td>Support Student Learning</td>
<td>50% of students are ‘always’ or ‘often’ able to find a relevant information</td>
</tr>
<tr>
<td></td>
<td>resource</td>
</tr>
<tr>
<td></td>
<td>Unmet information needs in reference, consultations, and assignments</td>
</tr>
</tbody>
</table>

Figure 1. Collection review assessment framework
The authors recommend developing consensus around criteria and metrics. The list of criteria used in this study appears in Appendix 2. Take this list of criteria, add to it, and then ask the decision-makers in your library to rank them. Do a small local study to determine what is most important to you, right now. Likewise, it can be a good exercise to try something like an AHP. The model should not be used to make every decision for you (the models are neither that robust nor infallible), but these exercises can be used to structure discussions around criteria and values.

Determine what metrics you will use when measuring the outcome. Table 2 has several options to get you started. For journal reviews specifically, the authors recommend calculating the true cost to fill ILL requests at your library every three to five years using the information in Table 1.

Lastly, and before you begin the collection review, create a follow up plan. Often by the end of cancellation projects, no one has the time or energy to plan an assessment. Decide early how you want to track the potential impact these decisions will have on your patrons. Make a plan to check in on demand from ILL and document delivery services, if applicable. Think about who will see that impact most directly and determine if there is a simple way for you to gather that information. Plan to continue monitoring demand for a few years. With perpetual access being a term of many license agreements, it can often take a few years for these cancellations to be felt. Then schedule a time to review the decisions and see how you did. You might not financially be in a position to resubscribe to anything, but you will have the data that can support the request which can help you lobby for additional resources.

In applying the principles of continuous improvement to the collection review process, the authors have gained a better understanding of how others in the library approach collection review. While the authors recognize that criteria employed to evaluate a resource will change from resource to resource and even from review to review, this process has illuminated the aspects of a resource that liaisons find most beneficial in supporting the teaching and research needs on campus. It has also highlighted the criteria subject liaisons found most useful in the last review and what could be beneficial to consider in future reviews.
APPENDIX 1
PAST REVIEW FEEDBACK QUESTIONS

Question 1. To what extent did you use the above criteria to rank databases?

Question 2. What prevented you from using the criteria to a greater extent?

Question 3. Did you use any other criteria?
   If yes, What other criteria did you use?

Question 4. Would you recommend using these criteria in future database reviews?
   If yes, What changes would you make?

Question 5. It is crucial that I involve my teaching faculty in future database decisions. (Likert)

Question 6. I feel confident I can evaluate databases outside of my subject areas. (Likert)

Question 7. Student learning has suffered by not having access to one or more of these databases. (Likert)

Question 8. Have you had interactions, either at the reference desk or in a research consultation, where one or more of the above resources would have been integral to helping the patrons?
   If yes, Which resource(s) would have been beneficial? Why? Please provide as much information as you can recall about the interactions.

Question 9. Have there been any course-assigned projects where use of one or more of the above resources would have been necessary for the students to be able to complete the assignment?
   If yes, Which resource(s) would have been necessary? Why? Please provide as much information as you can recall about the assignment and any workarounds you found.

Question 10. Have you received a request from a faculty member to reinstate any of the above resources?
   If yes, Which resource(s)? What was the justification?

Question 11. Are there any databases cancelled in the collection review that you have not missed?

Question 12. If we could resubscribe to two resources, which would you choose? Why?
APPENDIX 2

RESOURCE EVALUATION CRITERIA

- Usability / interface
- Quality of indexing / relevancy rankings
- Indexed content availability (citations)
- Subject thesauri
- Full-text availability
- Self-directed use (you don’t need a librarian to use it)
- Technical / authentication issues
- Intended use (research-oriented, classroom use, course assignments)
- Audience (undergraduates, graduate students, faculty, community users)
- Scope (subject-specific, interdisciplinary, multidisciplinary)
- Uniqueness (duplicated content / overlap)
- Scale (program size / no. of users served)
- Raw use numbers (high or low)
- History of resource cuts to programs
- Supports a new or growing program
- Product name recognition among patrons
- Faculty feedback / requests
- Accreditation requirements
- Sufficient number of resources for a program
- Cost/Cost-per-use
- Free alternatives on the web
- Available through ILL or on demand purchase
- Downgrades / smaller packages
- Compatibility with discovery systems
- Other

NOTES


15 Dawson. "A Triangulation Method."