Charleston Library Conference

**Book Usage Is Rollin’ Down: Multifaceted Assessment of Monograph Collection Performance to Optimize Purchase Decisions**

Tricia L. Clayton  
*Georgia State University, tclayton3@gsu.edu*

Skye C. Hardesty  
*Georgia State University, skye@gsu.edu*

Follow this and additional works at: [https://docs.lib.purdue.edu/charleston](https://docs.lib.purdue.edu/charleston)

Part of the [Library and Information Science Commons](https://docs.lib.purdue.edu/charleston)

An indexed, print copy of the Proceedings is also available for purchase at: [http://www.thepress.purdue.edu/series/charleston](http://www.thepress.purdue.edu/series/charleston).


[http://dx.doi.org/10.5703/1288284316443](http://dx.doi.org/10.5703/1288284316443)

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Abstract

Georgia State University’s print circulation has declined over 50% since 2010. Collection development librarians made several small-scale adjustments to address this trend, yet the drop off in use continued unabated. We had to totally rethink the book budget. To make changes strategically and responsibly, we needed answers to a variety of questions: Are there disciplines that do not need firm order allocations or even monographs? Does format matter? Does it matter how the titles are acquired with regard to approval versus firm order versus demand-driven acquisition (DDA)?

This paper discusses the multifaceted data-driven analysis we developed in order to provide a detailed and holistic picture of monograph collection performance and buying patterns. We share how we developed our analysis, what our data revealed, and the action items generated by our activities. This paper details how to combine large data sets from multiple sources for assessment, and how combining use and acquisitions data of print and electronic monographs helped us make improved purchasing decisions at Georgia State University.

Background

Georgia State University had a full-time equivalent (FTE) of 29,000 in FY16 and has over 1.5 million volumes. The fiscal year materials budget was $5 million, $800,000 of which was devoted to monographs, which included the approval plan, firm order budget, and DDA. The library uses a subject librarian model with 15 librarians doing title-by-title selection for 48 academic departments, each with its own firm order budget. The Collection Development Department Head allocates the overall monograph budget including the firm order budgets. The Collection Development department consists of two faculty and two staff members.

Historically, the library’s materials budget has been flat, with additional cuts in some years. For firm order allocations, a formula was used in the past but was discontinued; however, those allocations were used as a starting point going forward. Now, adjustments are made each year based on conversations with the Collection Development Department Head and subject librarians about department and program changes. The department head then makes the final allocation decision. Typically, if there was not a major change or a cut, most departments would get a 5% increase for inflation.

This firm order budget allocation process was usually sufficient, but within the past few years, the process needed to change for a variety of reasons: The university’s growth in science, technology, engineering, and mathematics (STEM) areas, increased cost of, and demand for continuing resources, implementation of demand-driven acquisitions, and in our library, some selectors were increasingly concerned that they were spending money for the sake of spending it. In addition to all those circumstances, there was a significant decrease in the circulation of books in the library.

The circulation of the library’s print collection dropped 55% from FY10 to FY16, which was significant and cause for concern. Circulations by patron group showed that faculty checkouts were fairly steady, but both graduate and undergraduates had a very large decrease. Given the size of undergraduates versus graduate students, we feel that undergraduates are the main force behind the decreasing circulation. The department made several attempts to address the issue such as adding DDA in 2009, extensive changes to the library’s approval plan in 2010, and adjusted firm order budgets, but none of that seemed to work.
Creation of the Materials Budget Review Group

Given the various changes within the university and the library, it was necessary to start making major adjustments to the materials budget, which would change the level of financial and resource support in certain subject areas. Because of the politically sensitive decisions we would be making, it made sense to use as much data as we could to devise a different looking collection and budget allocation that was both defensible and best supported the university. In late fall of 2014, the Associate Dean of Collections asked that a group be charged with devising a primarily data-driven materials budget allocation process to better support the research and teaching activities of Georgia State.

The group, called the Materials Budget Review Group, consisted of the Collection Development Department Head, the Collection Services Librarian, two subject librarians, and the Assessment and User Services librarian. The original deadline for the implementation of the group’s decisions was FY16, which was a tight deadline. After initial meetings, the group decided that this was a multiphase project. Because electronic resources were well used and print circulation had fallen so dramatically, the decision was made to focus on the book budget exclusively as the first priority.

Working on the book budget turned out to be the group’s only phase. In the middle of the project, Georgia State University consolidated with Georgia Perimeter College, a two-year college, which heavily impacted the library. Full implementation of the group’s recommendations, as well as further work on the budget, were put on hold while we grappled with going from a 29,000 FTE single campus to 43,000 FTE and five additional branch libraries. However, the group was able collect and analyze data that was useful in making collection and budget decisions regarding books and collection management.

The group began with the idea that we needed to establish where all departments were on the spectrum of monograph use in research and teaching, then define what level of support the library should provide to each department. We spent a lot of time collecting data that we thought would help us answer the question of who needed books and who did not. During that phase, the group realized that focusing on who was using the books at a department level was not really possible. Trying to find out why the books were not circulating should be a later consideration and investigation. What
made sense to the group as both a practical move, and possible to achieve within the deadline we were given, was to find out how we were spending the monograph budget and how was what we bought performing. Once we had answers to those questions, we could make decisions on how best to optimize our purchasing decisions moving forward.

**Project Planning and Compiling the Data**

To best answer our research questions, three project deliverables were needed. First, we needed a comprehensive master spreadsheet of all books acquired by the library during FY12-15 and all DDA records added to the collection. This spreadsheet also needed to include acquisitions, bibliographic, and usage data. Second, we wanted a summary analysis of a subset of the master spreadsheet that examined usage of print and e-book approval titles, print and e-book firm order titles, and total combined usage by Library of Congress Classification System (LC) class. Third, we wanted an analysis of overall collection performance across LC classes since FY10.

To determine what we were buying and how it was performing, we collected data from the Voyager ILS, GOBI, and the e-book vendors ebrary, EBSCO, and E-book Library (EBL). We collected reports on what was bought in all formats for the past three years (including DDA acquisitions), collection size, the number of items added annually, e-book usage (COUNTER and non-COUNTER data), circulation of recently purchased titles, and annual circulation by LC class. The circulation data did not include renewals.

The most challenging aspect of synthesizing all the data for this project was creating the master spreadsheet of recently purchased titles and DDA acquisitions. Primarily this was due to the sheer number and variety of reports that needed to be combined, but it was also a challenge because the data was combined twice. Because the project ultimately extended longer than originally projected, the project deliverables were produced for the period covering FY12-14, then again for FY12-15.

The following strategy was used to construct the master spreadsheet:

1. Combine approval and firm order title lists from Voyager.
2. Remove e-books to deal with separately.
3. Combine acquisitions data with circulation and bibliographic data for print titles.
4. Collate all the e-book reports into one document, integrating purchases and DDA titles together with their usage.
5. Reunite the print and electronic titles back together.

For a variety of reasons, the most difficult task in this process was collating all the e-book reports together. E-book data varied widely, and fields and formatting were inconsistent across vendors and systems. Even the COUNTER data was not unified. For example, EBL provided BR1 (title requests) reports, while ebrary provided BR2 (section requests) reports. DDA titles that were never triggered were not included on the COUNTER reports. Fund code and cost information came from the integrated library system (ILS), but unique identifiers varied across systems. For example, ebrary vendor reports used a DocID, while Voyager acquisitions data contained a BibID. Titles, call numbers, and ISBNs were unreliable matching points; thus, finding reliable match points between all the reports was a challenge.

To keep track of the process of merging many files into one document, several project management techniques proved effective. These basically boiled down to thorough documentation and active file and folder management. Our documentation included a data dictionary for the final spreadsheet, an annotated list of e-book files detailing the contents of the various types of reports collected, and a match point map. This map charted the files we were working with in conjunction with the fields they contained so that the best match points could be identified. Our file management strategy included using versioning and naming conventions in file names, moving “integrated” files into an archive folder, and making backups of critical files.

**Using OpenRefine**

The open source tool OpenRefine (http://openrefine.org) proved very useful for helping to combine the various reports and create the master spreadsheet of purchased titles. OpenRefine is an excellent tool for exploring,
cleaning, and transforming messy data. Platform independent, it runs locally in your browser. Documentation and help can be found on the OpenRefine wiki on GitHub: https://github.com/OpenRefine/OpenRefine/wiki.

OpenRefine is only one of many tools useful for exploring and manipulating data. For this project, both Excel and OpenRefine were used extensively, as they both have different strengths. OpenRefine is more visual and interactive. You can see the impact of your actions before you execute a transformation, and there are powerful undo features. Editing happens in batch, meaning that editing usually happens one column at a time, across many rows of matching criteria. In Excel, editing happens one cell at a time, and formulas must be copied to other cells. On the other hand, Excel is often a better choice if you need to enter a lot of data, perform calculations, or make pivot tables (Atima, Zhuang, Vedvyas, & Dole, 2013).

The basic principle in OpenRefine is that you use facets and filters to select the rows you want to work with. Then you choose a column on which to operate. Using options in the column’s drop-down menu, you perform mass edits to either transform the data in that column or create a new column based on the data in that column. These can be simple, built-in transformations (to number/text/date format, to titlecase, trim trailing whitespace, etc.) or more complex manipulations based on the regular expression language used by OpenRefine (Google Refine Expression Language or GREL).

The following examples describe some practical tasks that OpenRefine was used to complete in order to create our master spreadsheet title list.¹

- Execute multiple simultaneous search and replace commands, for example, to clean up extra characters (brackets, dashes, parenthesis, and punctuation marks) in the publication date field.

- Isolate the ebrary DocID from the URL in the catalog record in Voyager, thus providing a match point to additional ebrary data in the vendor reports.

- Bring columns of data from one report into another using a common match point (the “cell.cross” function). This procedure was used extensively throughout the project.

- Clean up messy publisher data by clustering. Clustering is a feature in OpenRefine where various algorithms are used to merge text that looks related into the same text string. Thus, it greatly speeds up the process of cleaning up variations such as “Academic Press,” “Academic Press Inc.,” “Academic Press, Inc.,” and “Academic Press, an imprint of Elsevier.”

First Project Deliverable—Master Spreadsheet of Recently Purchased Titles

All this cleaning and merging of reports in OpenRefine resulted in the project’s first major working document. This final title list of over 100,000 titles allowed examination of purchases and DDA at the title level over a four-year period (FY12-15). This document included a lot of fields to allow flexibility with how the data could be filtered and examined, both immediately and in the future. The included fields were: E-book DocID, BibID, Pub Date, Purchase Category, Fiscal Period, Title, Fund Name, Cost, Publisher, Imprint, Normalized Call Number, Display Call Number, LC Class, LC Subclass, Old Circulation Counts (FY12-14), New Circulation Counts (FY12-15), Old Section Requests (FY12-14), New Section Requests (FY12-15), Used? (Y/N), E-book Platform, ISBN print, ISBN electronic, and # of Triggers (EBSCO).

The purchase category field was not one that already existed in the reports we were merging, but one we added in as various reports were incorporated. Since one of our questions was what are we buying and how, it was important to have this information available, both for filtering this document and in creating the summary analysis.

We included two versions of the call number. While the normalized call number enabled accurate sorting, the prefix in the display call number, if present, indicated if a title lived in something other than the general circulating collection (e.g., reference). Similarly, we included columns for the first letter of the LC class but also the LC subclass. This enabled more filtering choices for the user, but it also made it easy to summarize the data in pivot tables at both LC class levels of description.
The document incorporated several columns that represented usage. For print titles, we had circulation counts for both data collection periods. For e-books, the section requests columns also sometimes represented title requests and/or triggers, depending on what data was available. Since comparing use across formats is tricky (how does a book circulation really relate to a section or title request?), we opted to include a column that measured usage as a binary metric. If something was used regardless of frequency during the data collection period, it was assigned a value of yes; otherwise, it received a value of no. This also made it easy to filter the document based on use regardless of format.

**Data Analysis**

The second project deliverable was a summary analysis of a subset of the titles from the master spreadsheet. This document compared the performance of four purchase categories (print approval, electronic approval, print firm order, and electronic firm order) across LC classes and with each other. DDA and media titles were excluded, since those were going to be reviewed later. To create this document, a pivot table was created that examined the number of titles used by LC class and subclass, broken down by format and purchase type. The pivot table data was then copied to a clean spreadsheet to allow additional columns to be added and allow more flexibility with formatting. For each LC class, the percentage use for each purchase category and across all categories was calculated. A column was also added with the LC class descriptions.

An excerpt from this analysis is shown in Figure 2, where the Q class, as well as the overall totals for all LC classes, is visible. For each of the purchase categories examined, the document shows the total titles purchased, the number of titles used and not used, and the percentage used. Summary columns on the right show the total titles bought across all four categories, the total used, and the percentage used.

This document is beneficial because it reveals areas of strength and weakness, both with the approval plan and with firm order decisions. We can also see how much or how little is being selected and whether format is an issue. We have the ability to look at the entire class, what was bought in it, and compare across acquisition method and format. We can also compare across multiple LC classes.

The next component of our analysis examined the collection more broadly. The first two project deliverables focused on recent purchases; we also wanted to look at the performance of the collection as a whole across LC classes. Going back to FY10, we looked at circulation in relation to collection size. For this analysis, collection size equaled the number of items (not titles) in the circulating collection. Data for this analysis came from two types of Voyager reports: Items added by fiscal year and annual circulation. Both were broken down by LC class, so it was fairly straightforward to combine these reports in Excel.

![Figure 2. Summary analysis of purchase categories by LC class, FY12-14.](image-url)
Figure 3 shows the final analysis document examining circulation in relation to collection size for FY10-14. For each LC class and for each fiscal year, we included the collection size, number of items added, number of circulations, and the percentage circulated. The final column on the right shows the percentage change in circulation between FY10-14 and is conditionally formatted by color scale. Red (dark shading) represents the lowest values, and the yellows are midrange values. This excerpt showing much of the D-G LC range reveals that the middle range for our circulation change was in the −40s. Overall, examining these trends in circulation and collection size in conjunction with the document analyzing the performance of recent purchases provided us with a holistic view of collection usage.

Results and Future Directions

When the group examined the documents, the compiled data revealed some interesting information about the collection:

- There were big decreases in circulation within the humanities LC classes. Between FY10 and FY15, PN had a decrease in circulation of 45%. PR had a decrease in circulation of 50.56%, with only 6% of the PRs circulating in FY15. Similar trends were seen in D, E and F.

- Overall, titles chosen by selectors circulated more than approval titles. At the end of FY15, titles chosen by selectors circulated 44%, compared to 36% circulation of the approval plan titles. Because there was a noticeable gap in the usage between approval titles and selected titles, we knew we needed to make adjustments to the approval plan, but we also wanted to know more about those titles chosen by selectors.

Most selectors did not track or categorize their purchases, outside of keeping an eye on their budgets, so all the group had was anecdotal evidence and a lot of assumptions about why librarian selected titles fared better. In general, the group assumed that there were enough faculty requests and course reserve requests to account for much of the usage, but we wanted to know more. In order for the group to learn more about the reasons subject librarians used to make firm order purchase decisions, we asked selectors to code their selections to record the decisions underlying their FY16 purchases.
The Materials Budget Review Group developed a list of codes reflecting what we thought would be the major reasons why a librarian would purchase a book or media item. We wanted the list to be manageable rather than exhaustive; thus, we instructed the librarians to choose only one code per item.

Most of the selectors participated in the project, but some selectors were inconsistent in when they put codes in and when they did not. The library’s selector for psychology did not participate, which explains why 15% of the books and media purchased in FY16 did not receive a selection code. Most of the top reasons for purchasing an item make sense; however, the group was surprised by how few faculty requests there were. Also, it was curious that there were more items purchased in areas of undergraduate coursework/research than graduate coursework and research given our circulation by patron group. We have not yet tracked usage of these items, but that is something we are planning to do in the future.

The data also showed us where there was a demonstrated need for book content that we were missing. For example, circulation of RT had decreased significantly over the last five years, but it was clear from the data that we were not adding many new titles or allocating a lot of money toward building that collection. However, we learned that the nursing program was growing and DDA usage was fairly high. We decided that we needed to increase DDA for nursing but also look at purchasing e-book subscription packages and other ways to get more nursing book content for our patrons.

<table>
<thead>
<tr>
<th>Selection Code</th>
<th>Titles Assigned</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>In scope of collection development policy for the subject area</td>
<td>874</td>
<td>27.29%</td>
</tr>
<tr>
<td>None</td>
<td>502</td>
<td>15.67%</td>
</tr>
<tr>
<td>Topic/subject is in area of faculty research/publication</td>
<td>341</td>
<td>10.65%</td>
</tr>
<tr>
<td>Think it will get used based on my subject expertise</td>
<td>325</td>
<td>10.15%</td>
</tr>
<tr>
<td>Received positive review(s)</td>
<td>264</td>
<td>8.24%</td>
</tr>
<tr>
<td>Faculty request</td>
<td>234</td>
<td>7.31%</td>
</tr>
<tr>
<td>Topic/subject is in area of undergraduate coursework/research</td>
<td>168</td>
<td>5.24%</td>
</tr>
<tr>
<td>To fill a gap in the collection</td>
<td>116</td>
<td>3.62%</td>
</tr>
<tr>
<td>Topic/subject is in area of graduate coursework/research</td>
<td>104</td>
<td>3.25%</td>
</tr>
<tr>
<td>To replace a lost or damaged item</td>
<td>96</td>
<td>3.00%</td>
</tr>
<tr>
<td>Reflects trends in the discipline</td>
<td>81</td>
<td>2.53%</td>
</tr>
<tr>
<td>Graduate student request</td>
<td>43</td>
<td>1.34%</td>
</tr>
<tr>
<td>Course reserves</td>
<td>33</td>
<td>1.03%</td>
</tr>
<tr>
<td>To replace a previous edition</td>
<td>19</td>
<td>0.59%</td>
</tr>
<tr>
<td>Undergraduate student request</td>
<td>3</td>
<td>0.09%</td>
</tr>
</tbody>
</table>

Figure 4. Results of the selector coding project.
While we were unable to use the data to rework the materials budget in the way the group originally intended, we did make some strategic decisions to help the collection. The data showed that we needed to discontinue the e-book approval plan. We also needed to make some significant changes to the print approval plan and increase content in certain areas. There are multiple possibilities for employing this type of collection data analysis. Projects we hope to implement in the future include improving firm order allocations and DDA evaluation.

References


1 Due to length considerations, it was impractical to include full explanations and accompanying visuals of OpenRefine processes in these proceedings. However, the presentation slides and accompanying notes, which are posted on the Charleston Conference website, do provide this information.