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

Tricia L. Clayton and Skye C. Hardesty, "Book Usage Is Rollin' Down: Multifaceted Assessment of Monograph Collection Performance to Optimize Purchase Decisions" (2016). *Proceedings of the Charleston Library Conference.*
http://dx.doi.org/10.5703/1288284316443

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

Tricia L. Clayton, Collections Services Librarian, Georgia State University

Skye C. Hardesty, Department Head, Collection Development, Georgia State University

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.

https://doi.org/10.5703/1288284316441

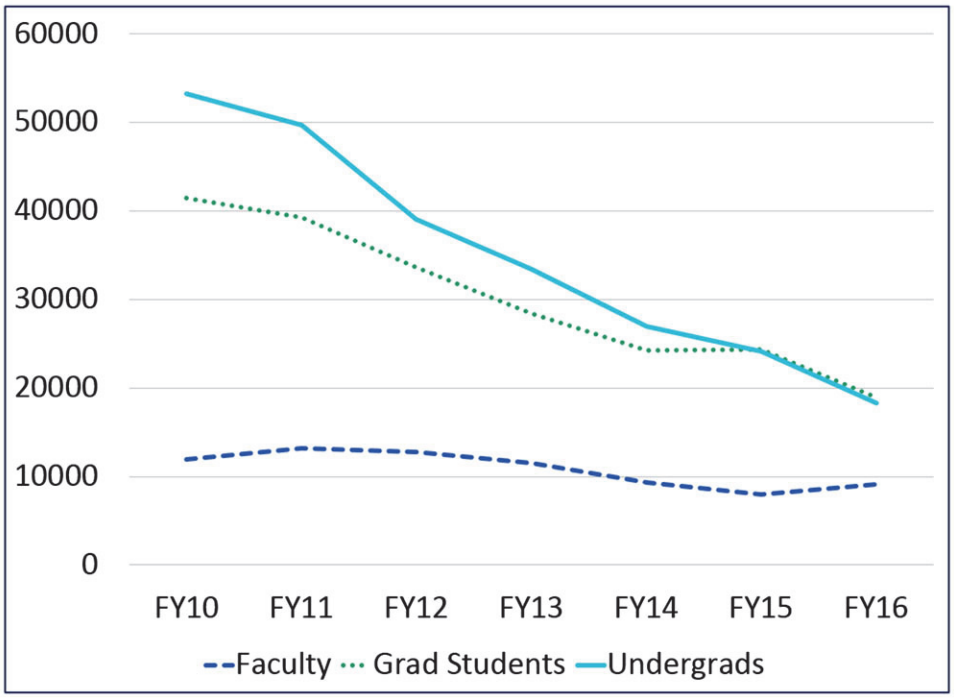

Figure 1. Circulation by patron group FY10-16.

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.

		Approval Ebook				Approval Print				FirmSlip Ebook				FirmSlip Print				All 4 Purchase Categories		
LC Class	LC Classification Description	Not Used	Used	Total Bought	% Used	Not Used	Used	Total Bought	% Used	Not Used	Used	Total Bought	% Used	Not Used	Used	Total Bought	% Used	Total Bought	Total Used	% Used
Q		**353**	**31**	**384**	**8.07%**	**966**	**291**	**1257**	**23.15%**	**141**	**90**	**231**	**38.96%**	**504**	**304**	**808**	**37.62%**	**2680**	**716**	**26.72%**
Q	Science (General)	21	2	23	8.70%	113	23	136	16.91%	8	4	12	33.33%	39	33	72	45.83%	243	62	25.51%
QA	Science. Mathematics	147	11	158	6.96%	375	110	485	22.68%	36	22	58	37.93%	143	77	220	35.00%	921	220	23.89%
QB	Science. Astronomy	9	2	11	18.18%	67	8	75	10.67%	4	3	7	42.86%	23	19	42	45.24%	135	32	23.70%
QC	Science. Physics	49	2	51	3.92%	135	21	156	13.46%	22	8	30	26.67%	99	37	136	27.21%	373	68	18.23%
QD	Science. Chemistry	2		2	0.00%	36	3	39	7.69%	22	11	33	33.33%	41	28	69	40.58%	143	42	29.37%
QE	Science. Geology	29	1	30	3.33%	51	10	61	16.39%					4	2	6	33.33%	97	13	13.40%
QH	Science. Natural histor	28	6	34	17.65%	76	34	110	30.91%	18	16	34	47.06%	66	41	107	38.32%	285	97	34.04%
QK	Science. Botany					2		2	0.00%	2		2	0.00%	9	3	12	25.00%	16	3	18.75%
QL	Science. Zoology	1		1	0.00%	5	5	10	50.00%	9	6	15	40.00%	26	13	39	33.33%	65	24	36.92%
QM	Science. Human anato									4	2	6	33.33%	2	4	6	66.67%	12	6	50.00%
QP	Science. Physiology	67	7	74	9.46%	104	77	181	42.54%	12	12	24	50.00%	44	40	84	47.62%	363	136	37.47%
QR	Science. Microbiology					2		2	0.00%	4	6	10	60.00%	8	7	15	46.67%	27	13	48.15%
Grand Total (All LC Classes)		**4210**	**951**	**5161**	**18.43%**	**16385**	**8079**	**24464**	**33.02%**	**535**	**433**	**968**	**44.73%**	**6847**	**4506**	**11353**	**39.69%**	**41946**	**13969**	**33.30%**

Figure 2. Summary analysis of purchase categories by LC class, FY12-14.

LC Class	Subclass	LC Classification Description	FY10 size	Circs FY10	FY10 % circ	FY11 size	Circs FY11	FY11 % circ	FY12 size	Circs FY12	FY12 % circ	FY13 size	Circs FY13	FY13 % circ	FY14 size	Circs FY14	FY14 % circ	% Change FY10-14
D	DD	World History. Germ	2827	277	9.80%	2861	189	6.61%	2913	251	8.62%	2940	205	6.97%	2954	129	4.37%	-55.43%
D	DE	World History. Greco	291	23	7.90%	297	25	8.42%	304	31	10.20%	307	32	10.42%	313	22	7.03%	-11.07%
D	DF	World History. Greec	1580	161	10.19%	1610	229	14.22%	1633	219	13.41%	1650	109	6.61%	1657	105	6.34%	-37.81%
D	DG	World History. Italy -	2458	343	13.95%	2502	390	15.59%	2541	277	10.90%	2571	292	11.36%	2592	194	7.48%	-46.36%
D	DH	World History. Low C	144	6	4.17%	144	1	0.69%	145	3	2.07%	146	4	2.74%	146	5	3.42%	-17.81%
D	DJ	World History. Neth	128	4	3.13%	128	4	3.13%	130	6	4.62%	130	6	4.62%	131	2	1.53%	-51.15%
D	DJK	World History. Easte	195	20	10.26%	199	13	6.53%	205	9	4.39%	209	15	7.18%	212	10	4.72%	-54.01%
D	DK	World History. Russi	4142	274	6.62%	4193	202	4.82%	4256	295	6.93%	4296	211	4.91%	4323	194	4.49%	-32.16%
D	DL	World History. North	325	38	11.69%	325	37	11.38%	325	18	5.54%	327	16	4.89%	328	9	2.74%	-76.53%
D	DP	World History. Spain	1158	182	15.72%	1173	191	16.28%	1198	110	9.18%	1216	108	8.88%	1227	102	8.31%	-47.11%
D	DQ	World History. Switz	94	4	4.26%	94	7	7.45%	94	2	2.13%	95	3	3.16%	95		0.00%	-100.00%
D	DR	World History. Balka	1196	166	13.88%	1233	221	17.92%	1266	160	12.64%	1284	107	8.33%	1306	96	7.35%	-47.04%
D	DS	World History. Asia.	18740	2734	14.59%	19200	2214	11.53%	19641	1752	8.92%	19932	1787	8.97%	20163	1312	6.51%	-55.40%
D	DT	World History. Africa	7005	1519	21.68%	7166	1356	18.92%	7320	1135	15.51%	7419	1096	14.77%	7513	726	9.66%	-55.44%
D	DU	World History. Ocea	1165	47	4.03%	1169	55	4.70%	1175	82	6.98%	1185	33	2.78%	1194	44	3.69%	-8.66%
D	DX	World History. Roma	69	24	34.78%	69	13	18.84%	71	3	4.23%	73	1	1.37%	75	14	18.67%	-46.33%
E	E	History America (11-	33396	6460	19.34%	33994	6348	18.67%	34605	5488	15.86%	35068	4509	12.86%	35402	3838	10.84%	-43.95%
F	F	United States local h	19533	2920	14.95%	19799	2574	13.00%	20038	2572	12.84%	20251	2178	10.76%	20364	1653	8.12%	-45.70%
G	G	Geography (General	4519	267	5.91%	4608	269	5.84%	4691	360	7.67%	4767	248	5.20%	4788	178	3.72%	-37.08%
G	GA	Mathematical geogr	643	71	11.04%	651	35	5.38%	662	27	4.08%	673	19	2.82%	677	20	2.95%	-73.25%
G	GB	Physical geography	1838	93	5.06%	1856	103	5.55%	1874	78	4.16%	1891	64	3.38%	1894	37	1.95%	-61.39%
G	GC	Oceanography	832	26	3.13%	858	54	6.29%	871	21	2.41%	872	19	2.18%	875	20	2.29%	-26.86%

Figure 3. Analysis of collection size and circulations by LC class, FY10-14.

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:

- The e-book approval plan had very low use.
  By the end of FY14, the e-book approval plan had 18.43% use, and by the end of FY15, it still only had 18% use.

- The books acquired by the library in the last few years had decent use.
  Overall, 33% of titles acquired by the library from FY12-14 were used; including data for FY15 moved usage up to 35%, which was better than the group anticipated.

- 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 was 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.

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

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

Atima, Zhuang, H., Vedvyas, I., & Dole, R. (2013, December). *OpenRefine tutorial v1.5.* Retrieved from http://casci.umd.edu/wp-content/uploads/2013/12/OpenRefine-tutorial-v1.5.pdf

[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.