Transforming a Print Collection

Brian Schoolar
Temple University, jbrian@temple.edu

Fred Rowland
Temple University, frowland@temple.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>

Part of the [Library and Information Science Commons](#)

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Brian Schoolar and Fred Rowland, "Transforming a Print Collection" (2013). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284315263>

Transforming a Print Collection

Brian Schoolar, Head, Acquisitions and Collection Development, Temple University Libraries

Fred Rowland, Reference Librarian, Temple University Libraries

We are all aware that the rise of e-books and demand-driven acquisitions (DDA) models are affecting print collections by moving funding away from print, not to mention the squeezing effect serials inflation continues to have on print acquisitions. This is as true at Temple University as elsewhere. However, at Temple we also have the extraordinary reality of a brand new library building on the horizon. Though slated to be a larger facility than what we currently have, it is also likely to house complementary support services such as the Writing Center and a Faculty Instructional Support center. Square footage for collaborative study spaces is a top priority. This will be a library, but our recently inaugurated president has made it clear it is not to be a \$190 million warehouse for books. We know some large portion of our existing collection will have to be housed in an on-site Automated Storage and Retrieval System (ASRS), leaving a much smaller footprint for the traditional open stacks, browseable collection. While all libraries face the prospect of transforming their print collection in some manner, at Temple we have a target date of Fall Semester 2017, only 4 short years away.

The Temple University Libraries is an Association of Research Libraries (ARL) Library, and Paley Library is its main library with a size of approximately 34,000 square feet housing a collection of approximately 2.1 million bound volumes, including 1.4 million monographs. Starting in 2008, we became increasingly active in acquiring e-books, purchasing with varying degrees of participation packages from Elsevier, Springer, Oxford University, and Project MUSE/UPCC and subscribing to many of the typical e-book packages like Safari, ACLS, and Books24x7. In 2011, we initiated a DDA program with MyiLibrary.

It is within this context that Fred Rowland and Brian Schoolar undertook two independent but complementary projects to look at our current

print monograph collection, with the hope of teasing out patterns of usage that might provide valuable insights for charting an intelligent path forward. Below, you will see two analytical lenses on a print collection. Brian Schoolar looked at the previous 2.5 years of books that circulated at least one time across the print collection in its entirety. Fred Rowland looked at three measures of circulation activity for books purchased in the last 10 years. The goal of our complementary studies is to provide relevant and actionable circulation data for consideration in a new open stacks collection of approximately 250,000 books. This requires a careful attention to copyright and acquisition dates and disciplinary categories in each of our studies.

In deciding how to allocate our print collection between an open stacks and an ASRS, the criteria for doing so is not entirely obvious because among the many options there is really no easy way to determine which ones would maximize circulation and discoverability. The Temple University Libraries could decide to include in an open stacks:

- Just the most recent materials;
- Just the highest circulating materials;
- Just a representative cross section of materials regardless of age or use; or
- Some variation of the above based on disciplinary patterns and needs.

Given that the ASRS is a recent innovation in academic libraries, we know little about user preferences with regards to it. It might turn out that many patrons prefer books to be in an ASRS because of ease of retrieval and checkout. In addition, any decisions on the optimal open stacks collection would have to be consistent with the logistical capabilities and requirements of the Access Services staff. On the whole, the policy would have to be understandable to patrons and the staff.

Figure 1. Percentage of Collection Circulated in Past 2.5 Years

In summer 2013, Brian Schoolar began receiving requests for collections data that required rapid decisions on the parameters of analysis—which copyright years and periods of circulation to focus on. Although a common-sense judgment pointed to using a publication date cutoff to determine which books were headed to the new open stacks—for instance, say, all books published after 2000, Brian Schoolar wanted to get some sense of the differences in disciplines to see how monograph-dependent disciplines in the humanities and several of the social sciences might differ from the sciences. If a particular discipline has relatively high recent circulation of older material, it might be best to try to keep more of this older material in open stacks. The reverse might also be sensible: if the publication dates of circulated items in a discipline falls off sharply after, say, five years, perhaps less of the recent material in this discipline should remain in the open stacks. Brian Schoolar chose to examine the most recent 2.5 years of circulation across the entire collection. He captured call numbers and publication dates and then grouped the publication dates by decade. This analysis differs from simply determining which items in the collection have ever circulated because one needs to get an idea of what is likely to circulate going forward. Therefore, recent use is an essential part of this analysis. Even if something had been checked out 25 times, if it had not circulated in 10 years there is really little need to keep it in an open stacks collection.

The scholarly communication characteristics of different disciplines are reflected in Figure 1,

Figure 2. Publication Date Profile of Items Checked in Past 2.5 Years

which shows what percentage of the collection in different categories circulated in the previous 2.5 years. For instance, 21.4% of items in the Arts and 15.0% of items in History circulated during this period. In descending order, the bar graph illustrates that less of the monographic collection was used as we move from the humanities to the social sciences to the sciences. Of course, it also needs to be understood that the size of the collections represented in Figure 1 vary quite markedly from one to another. Even though the Arts collection circulated at 21.4%, it is a much smaller collection than either the History or the Humanities (philosophy, religion, literature) collections. It is important to note, as well, that, over this period, most of the entire collection has not circulated at all.

Figure 2 shows the aggregate publication date profile of items checked out over the past 2.5 years. Of the books in the collection with publication dates of 2010 or later, 33.3% have circulated; publication dates between 2000–2009, 30.0%; and publication dates between 1990–1999, 17.9%. These are the top three categories, and the percentages gradually decrease by date down to the pre-1924 publication dates where just 3.2% have circulated. Not surprisingly, newer publications circulate at many times the rate of older publications, although the exact relationship will vary according to discipline.

The aggregate publication dates in Figure 2 are disaggregated into broad categories in Figure 3 to show some of the variability between the humanities, social sciences, and sciences. Despite

Figure 3. Publication Date Profile Broken Down into Broad Categories

Figure 4. Publication Date Profile for Language and Literature

Figure 5. Publication Profile of Psychology, Education, and Business and Economics

Figure 6. Publication Date Profile of Arts and Music Scores

this variability, it is important to note that each category shares the same curved slope of Figure 2. Arts, History, and Humanities are consistently above the average circulation rate, while STEM and the Social Sciences are (for the most part) below. The circulation rate for the Humanities category would be significantly higher if non-English language books were removed from the total since these books circulate at a very low rate.

Taking the Library of Congress (LC) subclass P, Figure 4 shows what the data look like as one drills down from broad categories to individual disciplines. The slopes of the curve, while resembling Figure 3, become increasingly erratic, and the difference between the top category—English, at over 50%—is 5 times that of the lowest—Germanic languages and literature.

Figure 5 is an example of individual disciplines within the Social Sciences. Across the whole range of publication dates, Psychology circulates more heavily than Education or Business and Economics. Of particular interest, Education and

Business and Economics both drop off to the low single digits in the decades before 1990, while Psychology circulation remains relative to the other two.

In contrast to Education and Business and Economics in Figure 5, Figure 6 shows that older materials for Arts and Music Scores circulate at a relatively high rate, even the pre-1924 materials. The high rate of use of Arts books in Figure 6 is consistent throughout our two studies.

What would be the effects of using 2000 as a cutoff publication date for including materials in the open stacks? Assume an ASRS and open stacks environment were currently in place and that everything published prior to 2000 was in the ASRS and everything post-2000 was in open stacks. Figure 7 shows the percentage of books that would be in each location with this arrangement. Education would have the highest percentage of circulating books in the open stacks at around 70%, while Classics would only have around 40%. This figure shows important disciplinary differences in the importance of

Figure 7. Hypothetical 2000 Cutoff for Open Stacks; Percentage of Books in Open Stacks

currency in various disciplines. The use of Physics, Art and Architecture, and Classics might all be impeded if 2000 were used as the sole cutoff date.

Schoolar's data highlight the fact that there are broad disciplinary (Humanities, Social Sciences, Sciences) differences in the use of materials by publication date and that specific LC classes differ from broader disciplinary patterns. One needs to find the right level of analysis in making decisions. Music Scores, the Arts, and Classics have most heavily used older materials and may benefit from having a greater selection of older materials in open stacks, while for areas like Business and Economics, Education, and Engineering, the more recently published items account for a high portion of recent circulation and older materials could likely go to ASRS with little loss.

While Schoolar focused on which books—across the entire Temple University collection—had circulated at least once over the previous 2.5 years, Fred Rowland's focus was on the circulation activity of books purchased in the previous 10 years (2003–2013), broken out by year of purchase. In contrast to Schoolar, Rowland looked at not only whether a book had circulated at least one time, but two additional measures as well. The books included in Rowland's data sets were limited to print books purchased by the library, so they excluded gift books. The data sets excluded e-books, music scores, DVDs, and journals. All languages were included in the data sets and circulation figures included course reserves. Rowland calculated three measures:

Aggregate Statistics: 2003-2013

274,692 Total Books
 151,809 Books circulated at least once
 441,762 Total checkouts
 55.27% Books circulated at least once
 1.61 Average checkout per book
 48% Top ten percent of books account this percent of total checkouts
 76% Top twenty-five percent of books account for this percent of total checkouts

Figure 8. Aggregate Statistics in Fred Rowland's Study

1. How much of the collection was used? Percentage of books circulated one or more times.
2. What was the overall usage of books? Ratio of total book checkouts to all books.
3. How even (or uneven) was the use of books? 10%/25 % of highest circulating books accounted for what percent of overall circulation?

Figure 8 shows the aggregate statistics of all ten data sets (from years 2003–2004 to 2012–2013). There is a very healthy 55.27% circulation rate and an average usage of 1.61 checkouts per book purchased during this time. When English language-only books are considered, these rates increase to 58.64% and 1.73, respectively. The concentration of checkouts is very high, with just 10% of books accounting for 48% of total checkouts and 25% accounting for 76% of usage.

When these aggregates were broken down by broad areas—Humanities (LC class: B–BD, BH–BX, M, N, P) Social Sciences (LC class: BF, G, H, J, K), Sciences (LC class: Q, R, S, T), and Other (LC class: A, C, U, V, Z) in Figure 9, they constituted 51%, 32%, 15%, and 2% of books purchased, respectively. Surprisingly, the percentage of books circulated at least one time and the total percentage of checkouts closely reflect these same percentages, as seen in Figures 10 and 11. This shows that, broadly speaking, the books purchased in the past 10 years have been used very evenly across the Humanities, Social Sciences, and Sciences.

Figure 9. Percent of Total Books Purchased Between 2003–2013

Figure 10. Percent of Books Circulating at Least Once Between 2003–2013

Figure 11. Percent of Total Book Checkouts

Figure 12. Total Book Checkouts Across Entire Collection, Between 2003–2013

Figure 12 shows the trend line for total annual book checkouts during the 10-year period across the entire collection. The books in Rowland's data accounted for 26.54% of these checkouts during this period. Although the number of annual checkouts rose substantially between 2003 and 2009, since then, annual checkouts have been dropping. This trend in Figure 12 roughly corresponds to the level of annual book purchases shown in Figure 13, with purchases peaking in 2007–2008 at 43,844 and falling to a low of 19,776 in 2012–2013.

We can look at the 10 academic years of book purchases as an artificial time series, with the 2012–2013 year, representing books that have been on the shelf one year, progressing down to year 2003–2004, representing books on the shelf for 10 years. Instead of looking at one “cohort” of

books over 10 years, we are looking at ten different cohorts examined at one point in time for their circulation data. The circulation data were collected in July 2013 for all ten data sets. Doing this, we can chart out our three usage measures across time. Figure 14 shows that 22.74% of books on the shelf up to one year had circulated at least once, while books purchased in academic year 2003–2004 (on the shelf 10 years) had circulated at a rate of 72.01%. The slope of the curve suggests that new books receive the greatest percentage of their usage in the first 3 years and then continue to increase in usage, albeit at a slower rate from years 3 to 10. However, these are ten different data sets and the circulation data associated with each year provide a snapshot in time. We have no way of knowing in what year the actual circulation was recorded for each cohort of books.

Figure 13. Book Purchases by Year, Between 2003–2013

Figure 14. Time Series of Books Circulating at Least Once

Figure 15. Time Series of Average Checkouts Per Book

Figure 16. Time Series of Concentration of Usage

Figure 15 provides the same time series using the measure of average checkouts per book. This ratio is calculated by dividing the total checkouts for a given cohort by the total number of books purchased. Between the first and fifth years, this ratio more than quadruples, and between years 5 and 10 it more than doubles. These results are similar to those displayed in Figure 14, with the greatest increase in the first half of the period and then continuing at a decreased rate up to year 10. Figure 16 shows how the concentration of usage changes with each increasing year. For books that have been on the shelf for 5 years, the top 10% and 25% account for 46% and 74% of total checkouts, respectively. At the 10-year mark, those percentages drop to 40% and 68%. This means that a relatively small percentage of books account for most of the total checkouts.

Figure 17 shows the three broad subject areas as a time series. As you drill down into the data, the trend lines become less smooth because these are ten separate and independent data sets. When all the data were aggregated (Figures 9, 10, 11, 14, 15,

and 16), the “law of large numbers” operates and the slopes are smooth. In contrast, at lower levels, as in Figure 17, it is not unusual for the lines to trend downwards at points. So for the Sciences, books purchased in 2004–2005 have circulated less than books purchased in 2005–2006, even though they have spent an additional year on the shelf. The Social Sciences have circulated at a higher rate than the Humanities and Sciences. However, the Humanities would be significantly raised if only English language titles were included in our data. In all areas—Humanities, Social Sciences, and Sciences—books purchased in 2003–2004 have circulated at a very robust rate of over 70%.

In Figure 18, we break down the Humanities and see greater dispersion in the data with Arts (Music and Fine Arts) circulating at a consistently higher rate than History and the combined Philosophy-Religion-Literature category. Figure 19 shows the Social Sciences broken down by LC class. It reveals a very consistent pattern of circulation except for LC class K, a very small part of the Temple University Libraries’ collections.

Figure 17. Time Series Broken into Broad Categories

Figure 18. Time Series of Three Subdivisions of Humanities

Figure 19. Time Series of Five LC Classes of Social Sciences

Figure 20. Time Series of Four LC Classes of STEM

Figure 20 shows the Sciences broken down by LC class, with very strong circulation in Medicine (LC class R), and weaker circulation in Science (LC class Q). Of the books purchased during the 2003–2004 academic year, 82.64% have circulated at least once for Medicine.

Rowland's study, focused on the annual purchases over the most recent 10 years, is an indicator of the effectiveness of the Temple University Libraries' purchasing program. Given the aggregate statistics, with 55.27% of books having circulated at least once and an average circulation of 1.61 checkouts per book, the Libraries clearly has an effective program. If one looks more closely at the cohort of books that has been on the shelf for 10 years (books purchased in 2003–2004), an even brighter picture emerges, with 72.81% of books having circulated and an average circulation of 2.89. Despite the move to e-books, it is clear that print books are still very popular with students and faculty.

Like Schoolar, Rowland found that there were significant disciplinary differences, especially when one drills down into the data. Rowland found

particularly strong usage in Music, Fine Arts, Medicine, and the LC class E (History of the Americas). Each were among the leaders in percentage of books having circulated at least once and average circulation per book. Additionally, each had low concentration of usage scores, meaning the use of books was spread out fairly evenly across each respective collection. As one would expect, Temple University purchased humanities titles in the greatest number followed by the social sciences and sciences. Surprisingly, the usage of these three categories was very evenly divided based on their numbers in the collection.

The total number of books purchased between 2003 and 2013 was 274,692, so most of these would fit into a new, reduced open stacks collection of 250,000 books. However, it might make sense to provide more open stacks time for those LC classes and subclasses whose older materials circulate more robustly and cut back on others where currency of publication puts only very recent books in demand. For instance, the LC classes E, M, N, and R might remain in the open stacks for 12–15 years and the LC classes Q, S, and T for 5–8 years.