Looking Forward by Looking Back: Books at the End of the Book

Darby Orcutt

North Carolina State University, darby_orcutt@ncsu.edu

Genya O'Gara

North Carolina State University, gogara@gmu.edu

Follow this and additional works at: <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>.

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>.

Darby Orcutt and Genya O'Gara, "Looking Forward by Looking Back: Books at the End of the Book" (2010). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284314831>

LOOKING FORWARD BY LOOKING BACK: BOOKS AT THE END OF THE BOOK

Darby Orcutt (darby_orcutt@ncsu.edu) - Assistant Head Collection Management, North Carolina State University Libraries

Genya O'Gara (genya_ogara@ncsu.edu) - Librarian for the Student Leadership Initiative Special Collections Research Center, North Carolina State University Libraries

Last year, the NCSU Libraries, as other academic research libraries, recognized two potentially historic changes underway in library collections: the immediate threat of greatly diminished financial resources in a troubled economy, and the increasingly rapid transition of users and marketplace towards digital formats. With dual goals of fiscal responsibility (through cuts, selection efficiencies and more strongly aligning acquisition with use) and predicting trends in monographic use (print and electronic) in the increasing hybrid/digital environment, we conducted an ambitious collection-wide monographic use study. Recognizing this key moment in our collection's print to digital shift, we established a print monographic use data set that is comprehensive, covers a decade of use, offers a baseline for future comparisons, and allows us to represent data in new and useful ways; for example, we color-coded the copious data to facilitate selectors' rapid apprehension of large amounts of information, allowing easy visual comprehension of trends in use.

Serving immediate needs, the use study allowed NCSU Libraries selectors to quickly assess collecting patterns with regard to use metrics, and much more easily make decisions, especially regarding approval plan coverage of particular call number ranges. We are currently using the study to answer questions, including: How may print use predict e-use? How do use patterns differ by discipline? How may historic use predict future use? Furthermore, the study informs our decision-making regarding our developing "purchase on demand" ebook program. Finally, supporting longer-term needs, this study serves as a baseline for future study of how monographic use may change in the NCSU Libraries as a result of the digital shift.

BACKGROUND

In the library community, we often remind ourselves to "never let a good crisis go to waste." In that spirit, when faced last year with cuts to the NCSU Libraries' materials budget, we dusted off long-held plans to conduct a study of monographic use patterns across all subject areas. Our immediate desire was for data to support our everyday selection decisions. For individual subject selectors, more readily accessible and digestible use data would certainly aid decision making with regard to title-level, firm-ordered selections. In addition, this data would assist in our thorough review of our book approval plan, approval materials constituting a budgetary line item that had exhibited considerable upward "creep" in recent fiscal years.

Apart from our immediate needs, however, we recognized tremendous further value in conducting a study of this sort, in a time of transition. The NCSU Libraries, as many other academic libraries, is in the midst of at least two potentially historic changes. Firstly, our immediate threat of greatly diminished fiscal resources, especially with respect to our expanding collecting scope (now including e-formats, data sets, and new media), may very well *not* be a

temporary phenomenon. This year's continuing state budget shortfalls tend to bear this prediction out. The "new normal" in academic libraries, as in our parent institutions, may be that we need to do more with less. Secondly, the rapid shift of much of our user base and of the marketplace towards digital formats means that the complexion of our collections must change. While the electronic revolution has hit the book sector well after transforming most academic journals, ebook formats availability, delivery, and pricing models seem to be stabilizing only in the last few years. The NCSU Libraries has been purchasing many more ebooks in recent years, and as more publishers move towards issuing ebook versions under more attractive and reasonable perpetual access terms, we anticipate exponential growth in electronic versus print acquisitions.

Our resulting monograph collection already looks much different than the previous print-only one, and the greater availability of online texts already seems (on the surface) to be one factor negatively impacting the level of use of much of the print collection. In embarking on our use study of print monographs, we recognized immediately that this would be our last opportunity to capture and analyze use data from a largely non-hybrid monographic world. In other words, this study was the last chance to examine monographic use prior to the introduction of the uncontrolled variable of ubiquitous ebook alternatives. Could historic use of print books predict ebook use? How might electronic availability change use patterns within different disciplines? How might historic use data inform decision-making with regard to new models, including purchase on demand? These were just a few of the questions we hoped to fruitfully investigate using this baseline use data as a starting point.

DATA

First, we offer a little background on the data we examined for this study. NCSU Libraries currently uses SirsiDynix (Sirsi) as our ILS and the data we analyzed was pulled from there as well as information culled from our approval plan with Baker and Taylor's Yankee Book Peddler (YBP) Library Services. Although this study was conducted at the beginning of 2009 we chose to examine the 10-year period ranging from 1997 through 2007. We did not look at 2008 for several reasons: there was a jump in e-book ordering at the end of 2007 and beginning of 2008 which had the potential to skew our use data; our fiscal year begins on July 1st so we would have had an incomplete data set; in choosing 2007 as our end date we could be assured that all monographs we examined had been in the collection for at least 2 years, and so had had opportunity to circulate; and lastly, the data before 1997 was inconsistent, due to previous database migrations and changes in technical services processes.

We extracted information¹ about each individual circulating monograph added during this time period and created an Access database to house this data. Obviously, this provided us with a very specific snapshot, rather than a more holistic view, of use of the circulating collection. However, since one of the purposes of this study was to adjust our current approval plan,

¹ Information included the date the book was added to the catalog, the unique number associated with each individual item in the catalog, the date the item was last checked out, whether or not the item was a firm order, an approval order, or a gift, the Library of Congress (LC) call number (which corresponds to the field "shelving key"), the total charges, the price (this included both "total" price and "unit price"), title, series information, edition, author, publish year, publisher, binding (cloth or paper), current status (checked out, removed), and which library location held the title.

examining use prior to 10 years ago would have made our study a hypothetical exercise since the approval plan had changed over that time period. Additionally, the data we could consider prior to 1997 had more and more inconsistencies with each preceding year we might have wished to examine. It is also important to note that we pulled data by unique item id rather than by title id—this prevented us from losing data on additional copies of older books added in recent years. Because selectors were asked to examine particular titles in this fashion, they were also encouraged, when it was unclear what total use might be at the title level, to examine other circulation data within Sirsi associated with specific titles.

This data had some issues, the most difficult of which was that not all years included the same information. This was due to a variety of factors, including migration from one database to another, type of acquisitions data tracked from year to year, approval plan changes, and human error. As an example, in 2002 the Libraries migrated from the Data Research Associates (DRA) system to Sirsi. The DRA system did not collect either an order type (which includes whether an item came as a firm, approval or gift) or pricing information as Sirsi did. Additionally the Libraries modified the approval plan in 2004 – so even once order type was tracked it couldn't necessarily be used as an indicator of current practice. Another difficulty was that although we have total circulation numbers and last date circulated, there is no way within the current Sirsi system to track other dates the materials may have circulated. For example, if an item has circulated 17 times, and the last circulation was May 7th, 2010, we have no way of knowing if those circulations happened within the same year (e.g. heavy use within a single course) or if they were spread out over an entire decade.

ORGANIZATION

In order to organize this data into a manageable form for selectors to review, we queried the database by LC call number range as mapped to the Libraries' approval and slip plan with Baker and Taylor's YBP Library Services (YBP) into subject specific blocks or fund codes.² After querying the data against the designated subject fund ranges in our approval plan we created a spreadsheet for each subject. Because we applied a fund (subject) name to these items strictly by call number range, spreadsheets could include items that were firm ordered on another fund code, but classified within this range. This was useful in that there was no duplication of items across the subject spreadsheets; however it did mean that selectors needed to be vigilant in looking out for interdisciplinary items that might not have been classified in their area or were a better fit for other areas.

Each spreadsheet included two tabs. Tab 1 included all monographic items added between fiscal years 1997 and 2007 and tab 2 included only YBP book approval items from the most recent 2 ½ years. By breaking out the last 2 ½ years (December 2005-June 2007) onto a separate spreadsheet we were able to see order type (firm, approval, gift) as mapped to our current approval plan and compare this to overall circulation trends.

² Our approval plan with YBP is broken down by LC number and designated to a specific selector based on subject code applied to each area. For example, any books within the LC call number range GE 1-9999 would be designated "Environmental Science" or fund code ENV5. Some of these come on approval to the library, and others titles are selected on a case-by-case basis from notifications (virtual "slips").

It is also important to note that we kept an “exceptions” spreadsheet of items that resisted mapping to a subject fund. These items amounted to less than 1% of all items and the list was made available to all of the selectors.

We then color-coded the spreadsheets in the following way:

- RED denoted an item that averaged more than 2 circulations per year over its life in the collection.
- YELLOW denoted an item that **averaged** more than 1 (and up to 2) circulations per year.
- WHITE denoted items not fitting a color category.³
- LIGHT BLUE denoted an item that circulated only 1 time ever.
- DARK BLUE denoted an item that had never circulated.

(See example below)

1	lib	Fund	Shelvingkey	title	Publisher	pubyear	datecreated	totalcha	lastcharged	YearlyAv	curloc	UnitPric	TotalAmtPaid
77	DHHILL	MATH	QA 000001 .A647 VOL 000	The Boltzmann equation a	New York : Springer-Verlag, c1988.	1988	20040413	2	20060402	0.4	STACKS	84.95	78.92
78	DHHILL	MATH	QA 000001 .A647 VOL 000	Linear integral equations /	New York : Springer, c1999	1999	19990624	22	20070511	2.2	STACKS		0
79	DHHILL	MATH	QA 000001 .A647 VOL 000	The energy method, stable	New York : Springer, c2004.	2004	20040130	0	0	0	STACKS		
80	DHHILL	MATH	QA 000001 .A647 VOL 000	The geometry of Minkows	New York : Springer-Verlag, c1992.	1992	20040715	0	0	0	STACKS		
81	DHHILL	MATH	QA 000001 .A647 VOL 000	Inverse acoustic and elc	New York : Springer, 1997.	1997	19981120	35	20081205	3.5	STACKS		0
82	DHHILL	MATH	QA 000001 .A647 VOL 000	Analysis and simulation of	New York : Springer, c2000.	2000	20000308	3	20030517	0.33333	STACKS		0
83	DHHILL	MATH	QA 000001 .A647 VOL 000	Dynamics : numerical expl	New York : Springer, c1998.	1998	19980415	2	20031201	0.18182	STACKS		0
84	DHHILL	MATH	QA 000001 .A647 VOL 000	Elements of applied bifurc	New York : Springer, c1998.	1998	19981106	14	20080825	1.4	STACKS		0
85	DHHILL	MATH	QA 000001 .A647 VOL 000	Elements of applied bifurc	New York : Springer-Verlag, c2004.	2004	20041021	12	20090526	3	CHECKEDOUT		
86	DHHILL	MATH	QA 000001 .A647 VOL 000	Optimization : algorithms a	New York : Springer-Verlag, c1997.	1997	19971030	21	20051213	1.90909	MISSING		0
87	DHHILL	MATH	QA 000001 .A647 VOL 000	Topological methods in hy	New York : Springer, 1998.	1998	19980623	22	20090623	2	STACKS		0
88	DHHILL	MATH	QA 000001 .A647 VOL 000	Weakly connected neural	New York : Springer, c1997.	1997	19971031	4	20080428	0.36364	STACKS		0
89	DHHILL	MATH	QA 000001 .A647 VOL 000	Inverse problems for paria	New York : Springer, c1998.	1998	19980316	5	20001113	0.45455	STACKS		0
90	DHHILL	MATH	QA 000001 .A647 VOL 000	Invariant manifolds and fib	New York : Springer, 1997.	1997	19980120	0	0	0	STACKS		0
91	DHHILL	MATH	QA 000001 .A647 VOL 000	Analysis of spherical symm	New York : Springer, c1998.	1998	19980317	0	0	0	STACKS		0
92	DHHILL	MATH	QA 000001 .A647 VOL 000	Robust control theory in H	New York : Springer, c1998.	1998	19980320	3	20030508	0.27273	STACKS		0
93	DHHILL	MATH	QA 000001 .A647 VOL 000	Introduction to the therm	New York : Springer, c1998.	1998	19980513	22	20030515	2	STACKS		0

The threshold of more than one circulation per year was selected based on the general rule of thumb that, especially for items available within our local consortium, the Triangle Research Libraries Network (TRLN), borrowing through ILL up to once per year would cost less than acquiring, storing, and preserving a copy of our own.

The color-coding scheme took advantage of the physiology of the human visual apparatus. Studies have proven that red-blue contrast is perhaps the easiest way for people to quickly recognize differences and patterns. Indeed, the color-coded spreadsheets made aspects of the

³ White items were those that averaged less than one circulation per year but had circulated more than once in their lifetime.

data quick to comprehend, to the extent that selectors could scroll through a spreadsheet faster than any of its text could be read, yet still easily pick out areas of lower or higher use, i.e. areas demanding more focused attention.

REVIEW GUIDELINES

Selectors were able to adjust their portions of the approval plan in whatever way made most sense, to them, drawing on their individual subject expertise. This meant there were a variety of ways they could approach this review. Looking at the spreadsheets from a subject-blind, sheer use standpoint, red and yellow items were high enough use that we might want to make sure to have them in the local collection (approval), and dark and light blue items, arguably, might not be required, as they are probably most cost-effectively borrowed when (or if) needed.

Selectors for each fund were able to sort the sheets in a variety of ways, but we recommended that everyone start sorting by call number (shelving key). The color-coding allowed them to spot any obvious “hot” (red and yellow) or “not” (dark blue and blue) call number ranges. Each selector examined their subject areas against our YBP approval plan to look for areas that might be obvious candidates for revisions. For example, in our first pass at Food Science, we noted a range of call numbers that was currently marked as send on approval for which the majority of the books were dark blue (i.e. never circulated). We were able to then move these from automatic sends to slips. This was done with an understanding that, while we might still want particular books regardless of anticipated circulation, we were unlikely to want to automatically receive a whole range of call numbers that had never circulated. We could then look at price of the areas we had moved from a particular year to estimate approximate savings in unpurchased approval titles for the following year. Good practice dictated that we wanted our approval plan to catch what we definitely wanted without catching lots of titles that we didn’t. Of course it was not always obvious; sometimes areas would have lots of blue and red, at which point a selector would have to determine what distinguished these items from one another. Was a certain imprint circulating? Was it a particular sub-topic that made certain books relevant and others not? There are a multitude of reasons like titles may have circulated or not – it was up to the selectors at this point to determine, where they could, why this was.

The dataset itself allowed selectors to investigate use patterns across disciplines as well. For example, in many humanities titles call number ranges, circulation often increased the longer they were part of the collection. The opposite trend could be seen in some science call number ranges, where particular titles might circulate heavily the first few years available and then rarely after 5 years.

FUNDS (THE MONEY)

At the end of the study each selector needed to determine where they would take 20% and 40% out of their funds; this could come out of either their firm order money or their approval money. Each selector was asked, based on total pricing from the last few years, to estimate how much of a cut they were able to take from different portions of their funds, i.e. approval or firm order. The amounts had to be saved, but could come disproportionately from either firm or approval based on their review findings.

As selectors altered the approval plan by moving different ranges from approval to slip or vice versa, they estimated costs based on the average pricing in that call number range for the previous two years. Predicting both use and approval savings is an imprecise art, but the hope was that the figures would balance out across the disciplines and be close enough for planning purposes. So, for example, if a call number range was moved from approval to slip, the approval estimate was decreased (assuming there were items being received in that range).

The net amount "saved" was subtracted from the approval plan estimate for that fund, but then added to the estimated savings and/or firm allocation for that fund. For example, if a particular area were changed from approval to slip because no books in that range were being used or are desired, then the selector would shift the full approval savings to the total estimated savings for that fund. Of course, when moving an area from approval to slip, a selector would likely still purchase some of those titles, just not all, in which case they would anticipate what percent they thought they might select. For example, if they anticipated selecting, say 50% of the titles (cost-wise), then they would move half of the estimated approval savings into the savings column and the other half into the firm. One selector found that there were no obvious low use areas in either firm or approval books within one of her subject areas – she chose to take the cuts in another subject, where the monographic circulation was not as heavy. By taking the cuts proportionally and leaving the review decisions up to individual subject specialists' discretion, we were able to create a more targeted approval plan and gain a deeper understanding of the way materials in different subject areas were being utilized.

It is important to note too that removing an area from our approval plan did not mean that we were not interested in any books in this area, just that we did not wish to receive them automatically. All of these decisions had to be carefully balanced by individual selectors. Unless by the profile criteria they were able to easily identify likely high-use materials and exclude low-use items, they wouldn't want to place a heavier emphasis on book approval. They had to keep in mind the extra work of selecting from slips, and not simply move wholesale to an item-by-item selection, an unsustainable practice that would not produce long-term cost-savings for the library, when librarians' time is factored in.

MANAGEMENT BENEFITS

From a management standpoint, this approach had significant benefits. For one, its holistic aspects forced selectors to see more of the forest than the individual trees of routine title selection. It is easy in any institution to begin to think separately about approval books and firm ordered titles, although they combine to form a single collection for our users. Examining usage patterns across these acquisition methods prompted selectors to think in terms not of changing or tweaking methods of acquisition, but of essentially rethinking them altogether in their subject areas.

In many areas, unexpected trends and/or high usage data suggested audiences beyond those of a particular disciplinary group. For selectors, this encouraged broader thinking about our users and collections. In a truly excellent research library, it is important for selectors to have a vision for themselves as selectors for *subjects*, not for specific user populations (e.g., departments).

This project aided our selectors in seeing approval plans as useful and adaptable tools. Often, book approval plans are not very well managed in practice. On the one extreme, they may be ignored; so long as they are quietly and effortlessly (for the selector) bringing in enough relevant materials (and nothing overtly out of scope), then they do not cry out for attention. On the other end of the spectrum, some selectors and institutions micromanage book approval plans through meticulous title-by-title reviews and excessive returns, effectively losing many of the advantages such a plan can provide. This project helped NCSU Libraries selectors clearly see the utility of a well-honed book approval plan, especially when considering the work of selection in terms of quantities of decisions. Ideally, to our thinking, a book approval plan should help minimize the number of positive decisions that a selector must make. In a perfect world, everything that we would select would come automatically; in the real world, a healthy book approval plan brings a preponderance of desired material and saves more in dollars, time, and effort than the cost of any undesired material.

The lens of historical usage data often indicated clear places for lines to be drawn. Because selectors had to balance their budgets across the lines of approval versus firm, each had to weigh the estimated expense of book approval in a range versus the opportunity cost of having to examine notifications – and of "mis-selection." If anything, our study showed that *selected* materials were very often not *used* materials. In subjects where selectors intended decisions primarily based on heavier use, they could examine whether this use might even be predictable; if not, one might of necessity need to accept the chaff with the wheat (strong incentive for pursuing patron-driven acquisition in these areas).

PRELIMINARY RESULTS

Each selector, using this tool, was able to identify both 20% and 40% cuts in their budgets across subject areas. Luckily, in 2009 we only ended up having to carry out the first tier of cuts, at 20%. Some selectors found that the approval plan was where it needed to be while others moved large call number ranges from slip to approval or approval to slip. As mentioned earlier, when comparing traditional science subject areas to humanities subject areas, the humanities circulation tended to increase the longer they were part of the collection. And although this trend was not seen in some specific science call number ranges, it was observable in the overall collection; the proportion of items that had not circulated after approximately two years in the collection was reversed after five years. Further, a majority of items (over 80%) added in FY 1997/1998 (after 12 years in the collection), had at least one checkout.

A portion of these have to be discounted due to questionable data⁴ but even when we remove those titles that leaves us with some serious questions about how long titles should be available to patrons in purchase on demand programs, in which many libraries (including ours) are participating. Since November 2009, the NCSU Libraries has provided access to an ebrary patron driven collection. If users "open" a book in this collection a certain number of times, the NCSU Libraries automatically purchases the title. As of October 5th, 2010, the title breakdown into very broad categories includes 148 titles in Engineering, Computer Science and the

⁴ For a short period of time the Metadata and Cataloging department was registering a checkout in the Sirsi system when they cataloged the book, therefore one checkout does not necessarily reflect an actual circulation depending on the year the book was added.

Physical Sciences, 463 titles in Humanities and Social Sciences, and 130 titles in Agriculture and Life Sciences.⁵ We do not plan to remove titles from our catalog, although many schools do. For example, the University of Kansas drops titles from their catalog if they aren't requested in the first 6 months.⁶ Our study indicates that, across disciplines, this approach will leave users without access to needed titles, if print use turns out to be indicative of e-book use.

Can monographic use patterns predict e-use patterns? We do not yet have the comprehensive data sets that would allow us to answer this question definitively one way or another. Additionally, because of the wide variety of ways to count use of an ebook, analysis is not straightforward. Should we examine and compare page views, time viewed, clicks, and/or something else? How do these uses compare to a traditional monographic circulation? One approach we have taken is to review call number range page views in comparison with other ebooks in a subject area: for instance, in examining the QA call ranges for the subject area math, with use being counted as pages viewed, many of the higher-use and lower-use call number ranges from the monographic use study appeared to line up with how patrons have been using our ebrary titles since 2005.

FUTURE STUDIES

We anticipate further explorations of this data. We also encourage other institutions to similarly capture this data while they still can, and hope that we can share data and study trends across institutional contexts as well.

Future use of this data set includes a comprehensive look at the data in terms of use via publisher; continued use of this data to select efficiently in "hot" areas; use as a baseline for future monographic studies to examine the change in use patterns with the growing number of e-titles; and finally, as a baseline from which to continue to examine whether or how print use may predict e-use.

⁵ PDA customer report – NCSU Libraries, 20101005.

⁶ Gold, Sarah. "Facing the Facts: University Presses in the Digital Age." *Publisher's Weekly Industry News*. <http://www.publishersweekly.com/pw/by-topic/industry-news/trade-shows-events/article/43600-facing-the-facts-university-presses-in-the-digital-age.html> (accessed September 15, 2010).