Biz of Acq — Toward Data-Informed Collection Decisions: 4+ Years of PDA Insights at Winthrop

by **Antje Mays** (Professor / Acquisitions & Description Librarian, Ida Jane Dacus Library, Winthrop University, 824 Oakland Avenue, Rock Hill, SC 29733; Phone: 803-323-2274; Fax: 803-323-2215) <maysa@winthrop.edu>

Column Editor: **Michelle Flinchbaugh** (Acquisitions and Digital Scholarship Services Librarian, Albin O. Kuhn Library & Gallery, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250; Phone: 410-455-6754; Fax: 410-455-1598) <flinchba@umbc.edu>

Introduction

Since the October 2011 inception of **Winthrop University's** eBook patron-driven acquisitions (PDA) program, the eBook program has matured and generated usage and expenditure data to a sufficient degree to spot trends between print and eBook preferences by discipline. Usage and expenditures for eBooks and hardcopy materials were analyzed through February 18, 2016. The first in a series of three articles, this article shares broad print and eBook usage and expenditure findings from year 2011/12 through 2014/15 gleaned from our old integrated library system (ILS) and offers insights for data-informed collection decisions. The second article (December 2016-January 2017) will show expenditures and usage trends in more depth by discipline for print and eBooks from year 2011/12 through 2014/15, as well as preliminary usage data gleaned from our new ILS through February 18, 2016. The third article (February 2017) will dig deeper into discipline-specific eBook expenditures and usage trends by examining usage and expenditure patterns by eBook collection type within each discipline.

Outline of the Study

What was measured? This study examined eBook and print usage and expenditure data since the October 2011 inception of **Winthrop's** PDA program. Expenditure data include PDA purchases, short-term-loan (STL) payments, eBook firm orders, and print book purchases broken out into 30 academic disciplines and professional fields. Because actual financial amounts could not be published, the study uses indexed values as a compromise for documenting trends and proportionality across formats and disciplines. Usage data include actual eBook usage broken out by perpetually owned titles, the PDA discovery pool, and the academic eBook subscription collection, in addition to print circulation. All data were examined in summary and broken out by 30 disciplines. Because July 2015 marked **Winthrop's** go-live with a new ILS that measures circulation differently from the previous system, historical comparisons run from the operating years of 2011/12 through 2014/15. Preliminary comparisons for data since July 1, 2015 were conducted through February 18, 2016.

Gleaning the Data: eBook data were gleaned from eBook-aggregator usage reports, broken out by type of use: our library's eBook usage types include usage of titles in our PDA discovery pool, of perpetually owned titles, and titles in a large scholarly eBook subscription database. Hardcopy circulation totals and expenditures for print and eBooks were gleaned from our ILS statistics and analytics modules. All raw data were exported to Excel spreadsheets, then normalized with Access, and analyzed with Excel.

Standardizing Data from Multiple Sources: eBook aggregators' subject mapping, ILS circulation data, and ILS financial activity by fund code rarely, if ever, match up seamlessly. Standardizing statistics on circulation transactions and expenditures in the ILS and vendor-sourced eBook usage data is prone to challenges in the quest for equivalent measurements across all pertinent data sources: eBook providers' usage reports summarize usage data to one set of subject groupings; the ILS statistical summaries of circulation transactions are grouped by LC classification groupings, while expenditure data by fund code may introduce yet another subject-clustering scheme based on how specific areas within disciples are clustered in a given university's academic units. To ensure that usage comparisons are meaningful, eBook and print usage data must be categorically equivalent and comparable. Our initial ILS provided circulation transaction totals, while eBook usage reports include unique titles used, pageviews, and total usage. Total eBook usage was used for direct comparison with total eBook usage.

Moreover, our ILS change introduced the need for different usage measures for comparison: in July of 2015, our library switched to a new ILS. Unlike the old system's statistics showing total circulation transactions, the new system's analytics provide the number of titles which have circulated broken out by subject clusters, but not the total number of circulation transactions. As the library went live with the new system on July 1, the analysis extracted the number of hardcopy items which have circulated at least once between July 1 and February 18, 2016. To match this hardcopy circulation measurement, comparison matches these data to the unique-titles-used data from the eBook reports.

Database Design to Automate Data Standardization: Manually normalizing every instance of vendor and ILS data to uniform measurement categories would be prohibitively time-consuming. As a solution, a relational database using Access was designed to automate this otherwise time-consuming process. First, a master table was created to associate the ILS circulation data's call-number clustering with corresponding fund codes and broader disciplines. A similar master table was created to associate the vendor's subject labeling of eBook usage with corresponding fund codes and broader disciplines. A third master table was created to associate fund codes specific to academic units with broader disciplines. Next, tables were created to ingest the raw data from ILS financials, ILS hardcopy circulation, and eBook reports by type of usage for each fiscal year during the measurement period. Database queries include mappings of the raw transaction and expenditure data's subjects to the master table's standardized discipline groupings, cross-referencing queries to catch any transaction data unmatched to disciplines in the master table, as well as queries totaling all transactions by major discipline for each year's financial data and usage type. By automating these data-standardizing steps, considerable time is saved and manual mapping errors are avoided. Lastly, database queries designed to combine usage and expenditure totals were used to create new comprehensive Excel spreadsheets with normalized data which formed the basis for analysis.

Analyzing the Data: After standardizing and summarizing the usage and expenditure data to the broad print and eBook categories reported here and to the disciplines examined in the later phases of this analysis, the trends and proportionality of usage and expenditures were analyzed and graphed using Excel.

Broad Usage Findings: eBook and Print, 2011-2015

This starter comparison set out with the broadest of measures: cumulative all-time **eBook usage** was analyzed from our eBook purchasing's 2011 inception through June 30, 2015, along with cumulative circulation through June 30, 2015, the final day with our old ILS. That day, our eBook collections had 189,583 total titles (used 42,943 times in total, 18,083 unique titles were used, and 648,821 pages were viewed). Of these eBooks, three subsets were examined in this study: 31,067 titles in the PDA pool (total usage: 9,287, with 3,690 unique titles used and 138,870 pageviews), 212 perpetually owned titles (total usage: 3,808, with 184 unique titles used and 92,090 pageviews), and a scholarly eBook subscription database of 132,132 titles (total usage: 29,199, with 13,123 unique titles used and 431,897 pageviews). **Hardcopy usage:** of a total of 448,366 volumes, the library had 258,629 circulating items in the collection which had circulated a total of 1,461,814 times. No additional usage measures could be gleaned from the ILS beyond total hardcopy circulation. **Aggregator-provided eBook variables** include total usage, unique titles used, and pageviews — these multiple variables help further illuminate the density, concentration, and depth of eBook usage. This section focuses on eBooks (perpetually owned,

PDA pool, and eBook subscription titles) and hardcopy in general; discipline-specific examination of these measures will be covered in the next two follow-up articles.
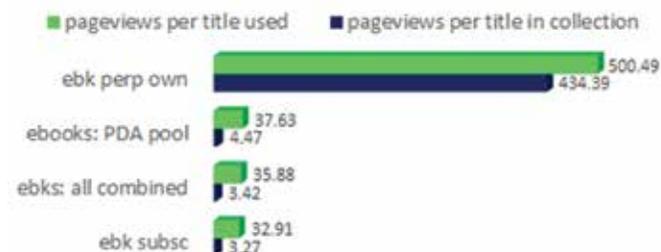
Density of use for the broad eBook collections was measured through **overall usage per held title by collection type:** the small perpetually-owned eBook collection (assembled through automatic PDA purchases and occasional efirm orders) was used most actively, followed by print: cumulative all-time usage data through June 30, 2015 show nearly 18 uses per title in the perpetually owned eBook collection, followed by nearly 6 uses per hardcopy item. These two far outdistance the 0.3 use per title in the PDA pool and the 0.22 use per title in the subscription eBook collection.

usage per title held, by collection type, through 6/30/2015

|	
ebk perp own	17.96
cumul hdcpy circ	5.65
ebooks: PDA pool	0.30
ebks: all combined	0.23
ebk subsc	0.22

Concentration of use was measured by **cumulative pageviews per eBook title held**. **Depth of use** was measured by **cumulative pageviews per title used**. The perpetually owned eBooks' concentration of title use and extent of collection use far outpace those of the other eBook collections: first, the over 500 pageviews per used title reveal the deepest use of all the eBook collections. Second, the over 434 pageviews per title *held* is only 13% below the 500 pageviews per title *used*. In contrast, the other collections' large gaps between pageviews per title used and pageviews per title held reflect that relatively small portions of the PDA discovery pool and subscription collection are used in depth.

Concentration of ebook collection use 2011-15

■ pageviews per title used ■ pageviews per title in collection

|	
ebk perp own	500.49 / 434.39
ebooks: PDA pool	37.63 / 4.47
ebks: all combined	35.88 / 3.42
ebk subsc	32.91 / 3.27

Year-to-year changes in pageviews per title used show year-to-year usage trends for each year. The second-year dip in pageviews per title used for the perpetually owned collection and the PDA pool resulted from the first full year of PDA purchases in 2012/13 and the early years' disproportionately large growth of the PDA pool.

Pageviews per title used, year-to-year 2011-2015
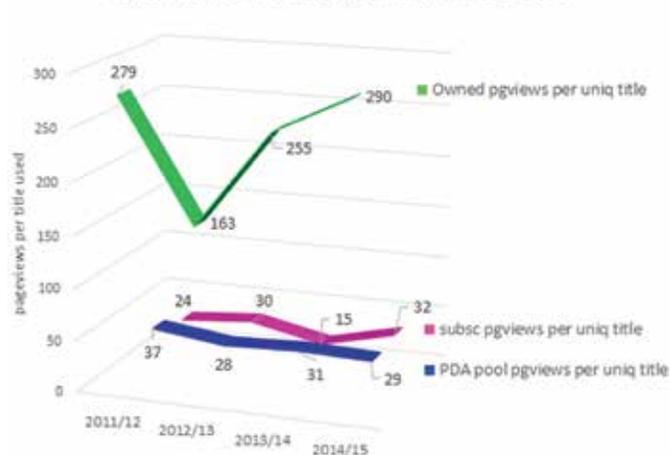
Year-to-year changes in eBook usage show that the perpetually owned collection was the only consistent gainer in usage, while usage of the larger PDA discovery pool and subscription collection fluctuated and actually declined in some years. Moreover, **year-to-year hardcopy circulation volume** was virtually twice that of eBook usage. In fact, after the 2011/12-to-2012/13 dip from 75% to 56% of all usage, hardcopy climbed back to 70% of all usage, despite the ease of online eBook access.


YTY hardcopy & ebook usage, with % of year totals

■ perpetually owned ■ PDA pool ■ ebk subsc hardcopy
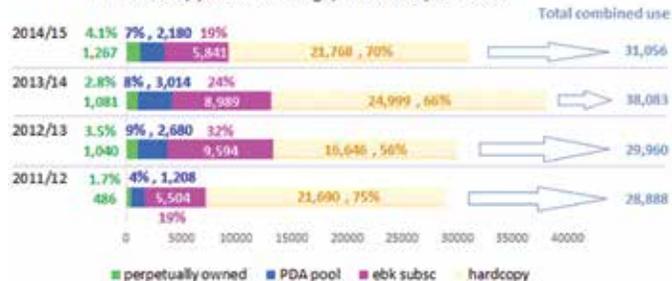
As shown in the next chart, hardcopy usage constituted 67% of total 4-year usage between 2011/12 and 2014/15.

4-YR USAGE

perpetually owned, 3,874, 3%
PDA pool, 9,082, 7%
ebk subsc, 29,928, 23%
hardcopy, 85,103, 67%
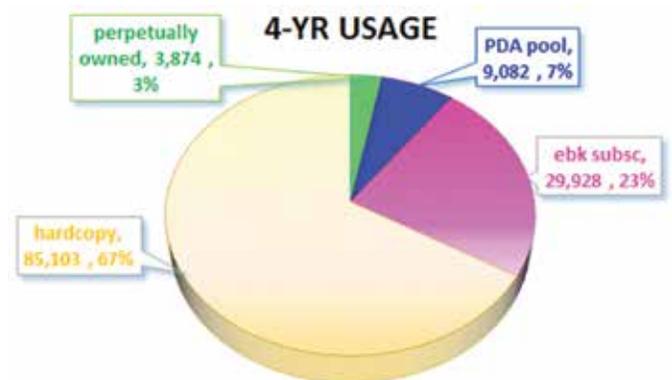
Year-to-year percentage changes in usage were mostly in tandem across collection types and diverged the most between the eBook collection types. The two largest year-to-year percentage jumps occurred in the PDA pool usage (+122%) and subscription eBooks (+74%) from 2011/12 to 2012/13, followed by print (up 50% between 2012/13 and 2013/14). The two biggest usage drops occurred in subscription eBooks (-35%) and the PDA pool (-28%) from 2013/14 to 2014/15. Only usage rates of the perpetually owned titles rose every year.


YTY usage # and % changes, 2011-2015
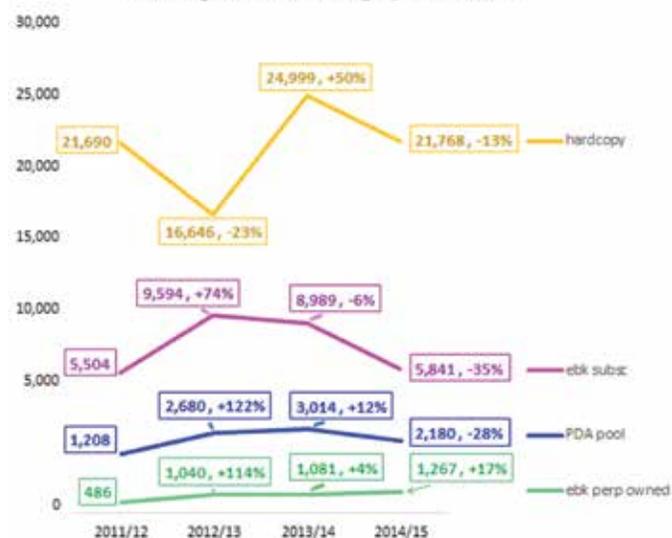
Cumulative 2011-2015 percentage changes in usage: cumulatively, all collections' usage rose. The next chart shows the cumulative percentage changes in usage from 2011 to 2015: of eBooks, the perpetually owned collection's usage rose the most (up by nearly 161%); the subscription collection's use rose the least (+6.12%). Hardcopy use rose by 0.36%.

% change in ebook and hardcopy usage: cumulative and YTY
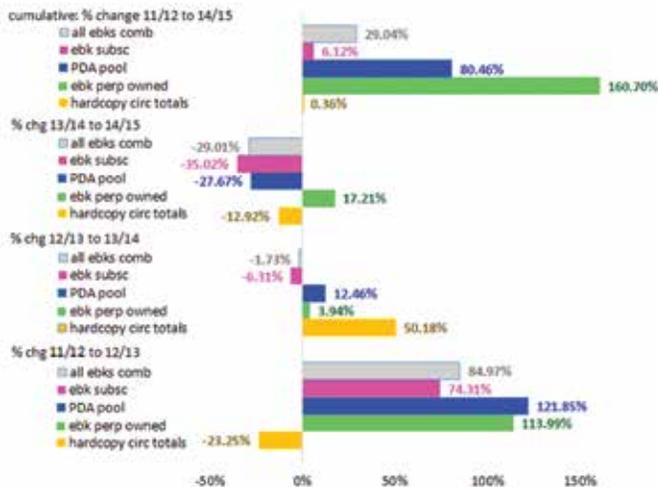
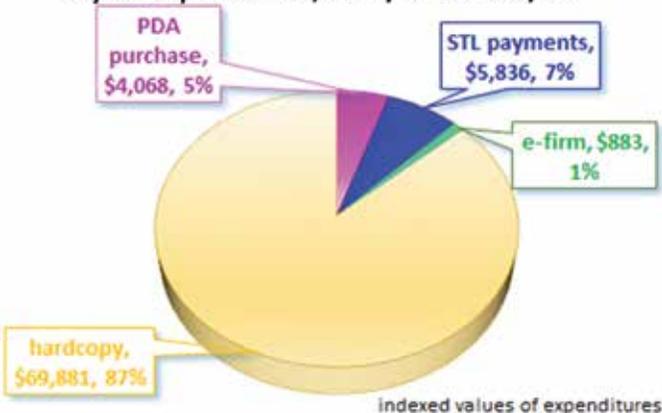### Broad Expenditure Findings: eBook and Print, 2011-2015

Print and eBook expenditures are shown as indexed values as a compromise because actual costs could not be published. The indexed values used throughout this study preserve the proportionality of the actual eBook and print expenditures. Despite year-to-year fluctuations, the overall trend shows that hardcopy went from 96% of total expenditures during year 2011/12 (when the eBook PDA plan was begun) to 82% in year 2014/15.

YTY hardcopy & ebook expenditures, with $ of year totals

Of the 4-year cumulative print and eBook purchases between 2011/12 and 2014/15, hardcopy expenditures constituted 87%, while usage-driven eBook purchases comprised 13%.


4-year expenditures, 2011/12 to 2014/15

PDA purchase, $4,068, 5%
STL payments, $5,836, 7%
e-firm, $883, 1%
hardcopy, $69,881, 87%
indexed values of expenditures

## Overall Trends: Cumulative Expenditure Changes from 2011 to 2015

While 2011-2015 print purchases declined and fluctuated dramatically year-to-year (primarily owing to budget caps), eBook expenditures grew during the same period. The chart below shows year-to-year expenditure fluctuations.

YTY expenditures $ and % changes, 2011-2015
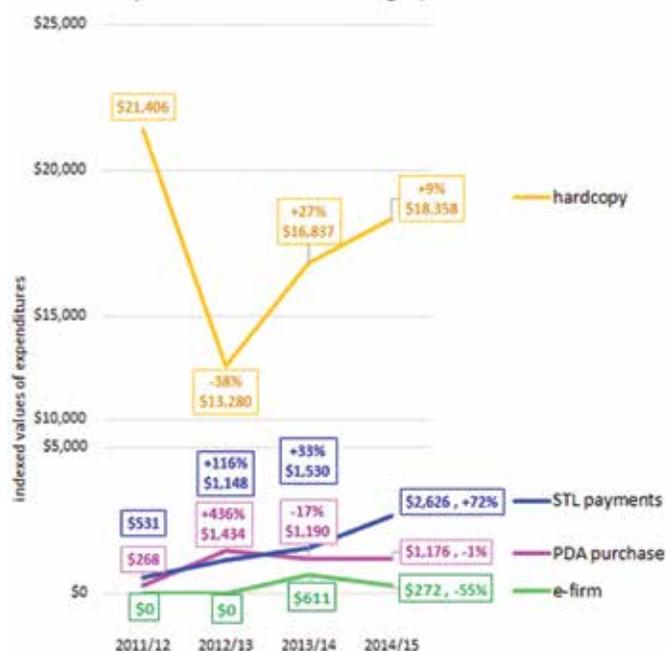
The following chart shows an at-a-glance overview of year-to-year as well as cumulative 2011-2015 percentage changes in expenditures. For all eBook subsets, percent rise of expenditures outpaced usage increases; for print, purchasing fluctuated more gently than did usage between 2011 and 2015.

% change: ebook & hardcopy expenditures, cumulative and YTY
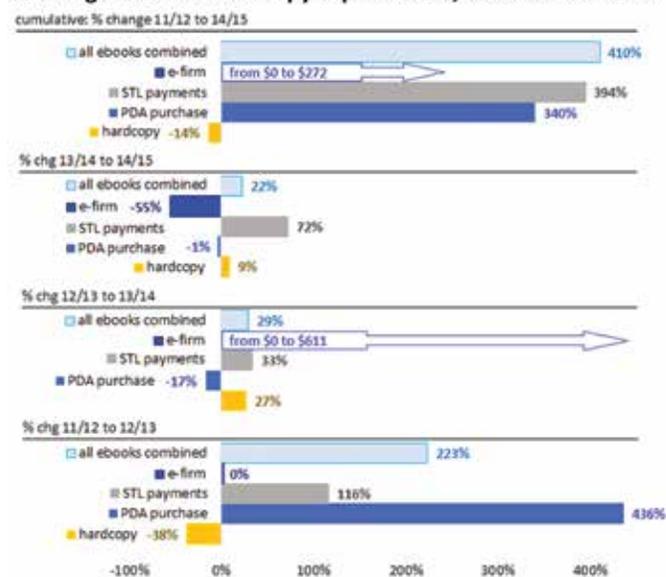
Implications For Collection Decisions

Conclusions: eBook usage and purchasing have risen dramatically in the four years since their 2011 beginnings from zero. Despite the ease of accessing eBooks and despite capped print budgets, usage of print cumulatively edged up between 2011 and 2015. After rising rapidly in the first two years, eBook usage began to level off and print regained

higher proportions of overall usage. These preliminary findings suggest that eBooks, despite their convenience, are complementing rather than replacing print. Among eBook collection types, the perpetually owned titles are used most heavily (18 uses per held title), greatly outdistancing the PDA pool (0.3 use per held title) and the subscription-based eBook collection (0.22). This pattern suggests continued high demand for titles owned through automatic purchase or efirm order. Moreover, the PDA pool is generating a healthy mix of short-term loans across the breadth of the collection, in some cases leading to automatic PDA purchase — factors pointing to an effective PDA profile.

Caveats: These broad findings alone are not sufficient for collection decisions responsive to the needs of a diverse mix of study programs. While overall trends have shown rapid rise of eBook use and steady usage of print books, nuance-sensitive decisions require insights into usage and expenditure patterns by discipline — the subject of next issue's article.