Analyze This: Usage and Your Collection

Gracemary Smulewitz
Rutgers University, smulewi@rutgers.edu

Follow this and additional works at: http://docs.lib.purdue.edu/atg
Part of the Library and Information Science Commons

Recommended Citation
Smulewitz, Gracemary (2012) "Analyze This: Usage and Your Collection," Against the Grain: Vol. 24: Iss. 6, Article 42.
DOI: http://dx.doi.org/10.7771/2380-176X.6256

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Rutgers University is a large institution with 28 schools and colleges, including New Jersey’s only pharmacy school. It offers more than 100 undergraduate majors, as well as more than 180 graduate programs. Supporting these programs across three campuses are 26 libraries, with a collection that encompasses 84,000 electronic subscriptions and more than 500 databases. Rutgers has two major research libraries — one in science and one in humanities — six specialized libraries, and two undergraduate libraries. The John Cotton Dana Library, in Newark, is home to the Institute for Jazz Studies, which is the largest jazz archive in the world. And the Paul Robeson Library, in Camden, has a broad liberal arts collection. Our collection development serves a very large, diverse community.

To date, our usage analysis has been a progressive process, and we have found practical ways to employ our usage and performance data. As we progress, we are continuously revisiting the following questions and working toward new developments to answer them:

What data are we collecting, and why? How do we interpret the value of the data we collect? How can we enhance our data?

Initially, the subscription analysis project began with developing and implementing a basic tool that has become an essential component of collection and budget analysis. A script was created to identify subscription orders for print and electronic material that were encumbered and had not been paid at the end of the fiscal year approached (the encumbrances were still committed, but nothing had been received). The intent was to implement a clean-up project to release encumbered funds, where appropriate, so that money could be used for other purchases.

We began collecting statistics to make sensible decisions about resources. We compiled these statistics into a report and named the report “Encumbered and Not Paid.” It provides a list of unpaid subscriptions; many, after evaluation, were classified as poor performance subscriptions. Some of our findings included: duplicate subscriptions; ceased publications with open orders; orders that had been created, yet for which the full acquisition process had not been completed; and a myriad of other subscription irregularities — there were many in all areas.

Codes were created to group similar problems. We took action on each category: canceling duplications, closing poorly-performing subscriptions, and reimstating many that had lapsed. When looking at both print and electronic subscriptions for a single title, we incorporated statistics into the analysis. The first report listed more than 3,000 titles. Currently, the number has been reduced to approximately one thousand, but the number fluctuates because of the dynamic nature of serials. The report is processed every year and repeated as the fiscal year nears to a close. The current report is compared to previous years to further analyze performance. This work has created an investigatory culture in our department and has enabled us to look at all of this data from different perspectives.

To survive a very large budget cut, we decided to enrich our analysis by collecting extensive statistics to inform decisions about resources. Our goals were threefold: to develop a tool to comprehensively analyze packages; to develop a mechanism for print usage analysis; and to provide more information for selectors that would enable them to compare content. (They had to be engaged in the process. Up to this point, due to the complexity of the big packages, our selectors felt they were somewhat removed from the collection development decision-making.)

At the time, we were renewing a very large package, but we felt that we didn’t fully understand the structure and all contents of the package. In general, packages had been invoiced as single-line items and accompanied by title lists. The goal was to break down the package by title and apply a fund code with a subject identifier to each title. A purchase order was added in our ILS to accompany the bibliographic record for each title, and funds were encumbered. Although a single payment was sent to the provider, in the ILS each title had a payment posted using a load from a spreadsheet.

This process was carried out for all packages and group purchases. Subsequently, the titles, the package name, the funds, and the payments were captured and entered into a spreadsheet. Usage statistics were added, and cost-per-use was calculated. Staff downloaded statistics manually — going to each site and pulling data to better understand the process. The comprehensive tool allowed comparison by title. The spreadsheet could be filtered by any of the elements so that a selector had access to all titles, his or her department, all packages with usage and cost-per-use. Cost and use of each title from year to year, across all packages, could be analyzed. This was helpful in cancellations and in making decisions for the swapping models that many packages offer.

Our next objective was to design a tool to capture print usage so that print and online usage for the same title could be compared. Active title and subscription information was extracted from the ILS serial control records. The serial control records had a controlled vocabulary in specified fields so that terms and funds, location, and other information about each subscription could be extracted with a report. Very little free text was used in setting up the controls. For example, current loose issues shelved by title in our reading rooms had entries of “RR” in the first line labeled “shelving location.” We were able to capture the reading room issues in a report using a script.

Included in the report were: the title number for easy access to the bibliographic record; ISSN; the print holdings statement; all URLs (if the title was also available online); the purchase order and fund code; and the owning library. Our Access Staff alerted us to a feature in our ILS called “Marked as Used.” This term means items with barcodes could...
be read as they were being re-shelved, and each reading would be counted as an in-house circulation. We don’t barcode our current periodicals, so to record use we used duplicate barcodes. A bar-coded item was added to each bibliographic record for each print title. The item was shadowed so that it was not visible to the public. For the same title, a duplicate barcode was added to the report. Signs were placed throughout the Reading Room asking users not to re-shelve journals. Before re-shelving journals, the staff matched the title of the issue with the entry in the report and recorded the barcode to mark it as a use. This process prompted the “Mark as Used” function and recorded an in-house circulation for that item. A report ran monthly to capture the collective in-house use for each title. The original barcode contained the URLs of the periodical titles that also had electronic access. The data was filtered by the providers so that usage statistics could be added.

The comparison supported the cancellation of print where electronic use was overwhelmingly greater, and the source for access was stable. We also changed formats. We moved to online if it was economical and the source for access was stable. In one particular collection, the print usage was zero. The comparison between print and electronic was shocking. We understoodprint usage was low, but seeing the numbers made the contrast more striking.

Based on the data from our project, 864 print titles were cancelled where electronic was acceptable, and 472 titles were changed to online-only. The remaining periodicals after this cancellation were the 1,000 subscribed titles.

Currently, we are participating in a beta project for EBSCO’s usage tool, EBSCONET Usage Consolidation. We are pleased that the SUSHI protocol is part of the process and is easy to configure. The protocol is incredibly helpful. The information is automatically downloaded, and usage can be tied to titles that are serviced by EBSCO. This offers an easy cost-per-use calculation. Having acquisition and usage information available seamlessly is ideal. We’re happy to be participating in this project.

Our future plans include comparing unfilled ILL requests to subscribed electronic and print holdings to determine why the request was not filled. We are also considering gathering usage data at the volume level and even the article level. Also, we are working on adding Impact Factor and Eigenfactor to our current reports. It is complicated work but very worthwhile for collection development.

Interested in sharing your experience with usage statistics? Contact Rossi Morris <rmorris@ebsco.com>.

Curating Collective Collections—Shared Print Collections Reaching Maturity

This month we’ll take a look at the evolving “big picture” of shared print collections for journals and for books.

Lizanne Payne provides an update on two mature journal archiving programs: WEST and the United Kingdom Research Reserve. And I commend to your attention a must-read report for librarians working to develop coherent models for shaping and sharing supra-institutional collections of printed books: Print Management at “Mega-scale”: A Regional Perspective on Print Book Collections in North America by Brian Lavoie, Constance Malpas, and J. D. Shipengrover (OCLC Research, 2012).

The following is just a teaser to prompt the reader to Google the report and enjoy a glimpse into a possible future, and to speculate on how your library will participate in the emerging collective collections.

Follow on the brilliant Cloud-sourcing Research Collections: Managing Print in the Mass-digitized Library Environment (OCLC Research, 2011), in the “Mega-scale” report Constance Malpas and colleagues at OCLC Research’s paint a picture of a possible “mega-regional” framework for regional shared print collections. While the “Cloud-sourcing” report focused on the implications of mass-digitization for preservation of a national shared print collection and on models like JSTOR (the Stock model), WEST (a Distributed model), and CIC (a Hub model), the “Mega-scale” report begins to describe the characteristics of a Flow model, as exemplified by BorrowDirect. The Flow model addresses the reality (amply demonstrated in the fascinating analysis of WorldCat data included in the report) that the North American Book collection is highly diffused by positing the formation of a virtual collection that integrates discovery and delivery across thousands of local collections.

The concept of “mega-regions” is derived from satellite imagery capturing night-time clusters of light around the globe that transcend political boundaries and map concentrations of population and economic activity. This results in 12 mega-regions in U.S. and Canada, such as BOS-WASH, CHI-PITTS, TOR-BUFF-CHESTER, CHAR-LANTA, and SO-CAL. Borrowed from a paper by Richard Florida (etc.), OCLC Research used this geographical framework to analyze WorldCat data for U.S. and Canada using zip and postal codes to develop fascinating data on the North American book collection (comprised of “distinct imprints or editions of books in printed form”).

OCLC Research used this geographical framework to analyze WorldCat data for U.S. and Canada using zip and postal codes to develop fascinating data on the North American book collection (comprised of “distinct imprints or editions of books in printed form”). National and regional collection metrics on the 45.7 million print books (889.5 M holdings) in N. America include: degree of uniqueness and overlap within and across regions, extent of holdings in academic libraries and in public libraries, measures of rareness and scarcity, global diversity of holdings, extent of “regional flavor”/uniqueness, analysis of pairings of large regions, and examination of the aggregate holdings of “extra-regional” libraries that fall outside the 12 mega-regions.

Based on these data, the authors examine implications for shared print models including: the need for supra-institutional coordination; why and how scale and models of cooperation will likely vary among regions; possible pairing of regions and thoughts about the challenges of address preservation of extra-regional collections; the key role of the HathiTrust digital collection in shaping large, multi-regional collection management; and the financial implications of a tendency of smaller institutions to view stewardship of print legacy collections as the responsibility of a small number of research-intensive institutions.

The picture painted in this report is of a rich, highly diffused, and asymmetrically-distributed national collection that will be challenging, but not impossible, to shape at mega-regional scale. While existing models such as CIC, ASERL, Orbis-Cascade, MINITEX, NLM, COPPUL, and WEST have much to commend them, this report suggests that existing cooperative infrastructure may not be equal to the task of shared print management at mega-regional scale. In addition, the authors point out that to round out the “supply-side” picture presented in the “Mega-regional” report, there is a need for a more complete characterization of the “demand side” through large-scale analysis and projection of inter-library lending data. This is a more exciting view of potential paths forward in evolving shared collection management for print books. Check it out!