Chapter 4

# Putting It into Practice

## *Quantitative Methods at CUL*

I came to the Collection Development department at Columbia University Libraries (CUL) in May 2013. Originally, I was hired to complete the E-book Program Development Study, an ambitious two-year assessment project aimed at gathering quantitative and qualitative data to drive the development of best practices related to e-book programs and services.[1] It was during this time that I immersed myself in quantitative methods and developed a strong research framework that has formed the basis of my work over the past several years.

Essentially, I spent two years creating a baseline of data that has been used to conduct more granular collection evaluations in the years following the completion of the e-book study. For the purposes of this chapter, I have pulled together three examples of projects that I completed both independently and through collaborations with small teams to demonstrate how local and external data sources were used to document usage trends, calculate cost per use, and establish a method to examine collection impact.

The first example is taken from the E-book Program Development Study and demonstrates how I analyzed the e-book collection from a "bird's-eye view." Over the course of eight months, I analyzed ninety-six e-book subscriptions and thirty-five e-book packages to document collection development trends across all major disciplines represented on campus. Much of this work involved a large cost analysis project that resulted in a $50,000 savings for the 2015 fiscal year. The project also provided a methodology and baseline that I, as well as colleagues across campus, have continued to use and build upon. The method consistently yields actionable results and is highly adaptable to the unique information needs and budget considerations that exist at individual libraries throughout the system.

The second example discusses how I applied the same baseline and method to an analysis of e-book holdings at the Avery Architectural and Fine Arts Library at Columbia University. The goal was to examine budget allocations, subject coverage, and cost per use. The project findings are currently being used by subject specialists at the Avery to plan for e-book spending and collection development in fiscal year 2018.

The final example demonstrates how the cost analysis methods developed for e-book collections can be applied to the evaluation of e-journal packages and combined with citation and publication data to examine collection impact. I highlight this particular example because I think it demonstrates how knowledge of local information needs, establishment of a baseline of quantitative data, and experimentation with flexible and adaptable methods can expand quantitative analysis work across format boundaries to provide standardized collection evaluation strategies within a library system.

***Note**: Many of the results in this chapter were calculated using cost data. For the purposes of this report, the numbers have been recalculated to maintain confidentiality, but they reflect trends discovered during each quantitative project.

### Introduction to the E-book Collection and Quantitative Analysis at CUL

Columbia University in New York City is a research institution that brings together a community of approximately 4,300 faculty members and 26,000 students; three-quarters of the total student body is

enrolled in graduate and professional programs. CUL supports this research community with a collection that contains more than 12 million volumes, 160,000 journals and serials, and extensive electronic resources, audiovisual materials, and archives. Services and collections are organized across nineteen libraries that are run by a team of 450 professional and support staff. Our challenge is to provide the most comprehensive coverage of collection areas in formats that are suitable to the diverse information needs of our user community.

The Collection Development department oversees the development, management, and storage of collections in a variety of formats. The overarching goal of all departmental activities is to support the research and learning environment at Columbia University by providing access to information that supports scholarly activities. Collection Development also serves the research community by participating in cooperative efforts to collect, access, and preserve knowledge at the local, national, and international levels.

E-book collection development initiatives date back to the 1990s, beginning with experiments such as the Online Books Project, the Virtual Reading Room, and pilot programs with netLibrary. In 2004, CUL began purchasing e-books in an experimental capacity. Due to positive reception by faculty and students, CUL expanded e-book collections to support research, teaching, and learning activities across campus. Over the last four years, the e-book collection has increased exponentially in size; CUL now provides access to over two million e-book titles, and expenditures comprise more than 25 percent of the total book budget.

The concept for the E-book Program Development Study was developed in response to the growth of the e-book collection. The primary objective of the study was to document the e-book landscape at Columbia University (a) internally, (b) within the context of the academic community, and (c) within the context of the e-book publishing industry. A second objective was to create a sustainable analysis framework that enabled librarians to collect data and evaluate e-book holdings in a standardized fashion. The research design was guided by the following four principles outlined in the CUL/IS Strategic Plan 2010–2013:

1. User-focused design;
2. Data-driven decision making;
3. Continuous assessment of results;
4. Flexible and adaptive response to user needs.[2]

The framework of the study was structured around the e-book life cycle. Primary areas of focus included selection and acquisition, discovery, access and use, preservation, and evaluation. The following framework guided the E-book Program Development Study:

1. Develop a set of recommendations that address challenges related to selection and acquisition, discovery, access, and preservation at CUL.
2. Examine how study findings provide opportunities to evaluate and strengthen collaborations with partner institutions and content providers (e.g., vendors, publishers, aggregators).
3. Establish an evaluation framework that facilitates regular assessment and planning so that collection development strategies can be updated and revised as the e-book landscape evolves.

The reality that the e-book landscape is constantly evolving was factored into decisions regarding the overarching assessment framework guiding this study. The research design was created so that it can be replicated regardless of how e-books evolve in the coming years. Because the design is flexible and adaptive in nature, it promotes continued assessment, evaluation, and strategic planning as a regular component of e-book programs. The success of the project involved its ability to explore these areas in an innovative and methodical fashion.

## Developing a Baseline: E-book Cost Analysis Project

One of the initial goals of the E-book Program Development Study was to create an inventory of all e-book holdings at CUL.[3] It was soon apparent that due to the volume and complexity of the data set, this project was outside the scope of the study.

At this point, I began looking at available data sets and decided to experiment with a new methodology that would compile a snapshot of e-book collection development activities (i.e., subscriptions, packages, and firm orders) made throughout the 2013 fiscal year (FY2013). The goal was to document how funds were allocated to purchase e-book content and determine if acquired content is of value to the user community.

I determined that e-books are most often purchased on the EO (e-book subscriptions) or EB (purchased content including e-book packages and firm orders) fund codes. As the primary objective was to develop a quantitative analysis method that would promote the ongoing analysis of the e-book collection, a small subset of total purchases was selected for further investigation. For this study, data collection was limited to titles, packages, or subscriptions that had fund activity during the 2013 fiscal year.

To collect financial data for all e-book purchases, I ran a Voyager query (CUL's integrated library system) for all library funds ending in EO or EB. After running the cumulative query, I created a base list for each of the following categories: subscriptions (EO) and package purchases (EB packages).

**Table 4.1.** Example of budget breakdown of the EB fund (e-book packages and firm orders) in FY2013.

| EB Fund (E-book Purchases) | # of Titles | Amount |
|---|---|---|
| Entire EB Budget | 736 | $1,100,342.03 |
| EB Fund (Packages) | 35 | $1,002,031.98 |
| Bulk (~68%) of Package Budget | 6 | $739,833.62 |
| Tail (~32%) of Package Budget | 29 | $262,198.36 |
| EB Fund (Firm Orders) | 701 | $98,310.05 |
| Bulk (~70%) of Purchase Budget | 195 | $57,248.45 |
| Tail (~30%) of Purchase Budget | 506 | $41,061.60 |

**Table 4.2.** Example of statistical analysis of the EB fund (e-book packages) in FY2013.

| EB Fund (Packages) | |
|---|---|
| Total EB Cost | $1,100,342.03 |
| Package Cost | $1,002,031.98 |
| % of Total EB Cost | 91.07% |
| Number of Packages | 35 |
| Average Cost | $28,629.49 |
| Median | $8,033.53 |
| High | $218,891.59 |
| Low | $1,200.00 |

**Table 4.3.** Example of the cost breakdown of the top six e-book packages (EB funds) in FY2013.

| Top Titles for EB Packages | Price | % of Total Cost |
|---|---|---|
| Package A | $218,891.59 | 21.84% |
| Package B | $216,888.00 | 21.64% |
| Package C | $114,950.00 | 11.47% |
| Package D | $66,907.63 | 6.68% |
| Package E | $62,515.00 | 6.24% |
| Package F | $59,681.40 | 5.96% |

Next, spending for each of the three categories was totaled, and calculations were made to identify the top 70 percent (bulk) and bottom 30 percent (tail) of purchases within each budget (see table 4.1). Finally, I calculated the total, average, median, high, and low costs of each category (see table 4.2).

I decided to dig deeper into the data set and determine if the value CUL receives from e-book collections is closely aligned with the associated costs. To accomplish this task, I pulled usage data into the analysis to determine if CUL's funds are invested in heavily used e-book resources. Again, given the scope of this project, I decided to limit assessment to the top six e-book subscriptions and packages, ranked by cost (table 4.3). Then, I pulled the corresponding title lists and BR2 COUNTER reports from vendor or publisher websites. Finally, I filtered the full data set to remove titles published prior to 2013.

At this point, I encountered an unexpected challenge; in several cases, multiple collections from the same vendor are purchased as separate items on the EO or EB fund codes. However, there is no apparent way to filter COUNTER reports by collection. I considered analyzing the data by vendor or publisher instead of by collection, but decided that this method would skew results because of the discrepancies in cost, size, and use. Instead, I filtered the data for a second time by matching the 2013 title lists with COUNTER report data using Excel.

Next, I adapted a cost analysis framework used by the University of Western Australia to analyze DDA e-book models and calculated the number of titles loaned, number of loans, percentage of titles without use after purchase, the average cost of an e-book, and cost per use[4] (see tables 4.4 and 4.5).

## E-book Subscription Cost Analysis

After analyzing the cost and usage data of the top e-book subscriptions (see table 4.4), it was determined that the cost per use of Subscription D was high ($9.17 per use) compared to Subscription A ($0.21 per use), Subscription B ($0.18 per use), and Subscription C ($0.73 per use).[5]

The results were presented to the Selectors' Group at CUL. The consensus was to conduct a second analysis of Subscription D based on the following criteria: evaluation of content, overlap analysis, and interface analysis. The results indicated that Subscription D contained a large number of outdated technical manuals (96 percent published before 2011), a high number of titles available through other platforms, and incomplete multivolume sets.

When the decision was made to cancel Subscription D based on cost analysis results, CUL had the opportunity to speak with the vice president and a team of sales representatives on three separate occasions. CUL received a reduced quote, and again, I conducted a cost analysis using the method described above to examine its value. Again, I determined that the quoted cost was not reflective of the value of content provided and the offer was declined.

At that time, the sales team requested feedback from CUL regarding how to improve the platform and were provided with study findings. After a series of negotiations, Subscription D was renewed for one year at an 80 percent discount, and CUL received an annual savings of $51,000 on this subscription.

**Table 4.4.** Results of the cost and usage analysis for e-book subscriptions based on BR2 COUNTER data.

| EO Fund Analysis | Subscription A | Subscription B | Subscription C | Subscription D |
|---|---|---|---|---|
| 2013 Cost | $54,788.40 | $21,280.00 | $16,071.33 | $57,005.85 |
| No. of titles | 89,529 | 6,631 | 116 | 14,339 |
| No. of sections viewed | 33,834 | 2,185 | 85 | 1,649 |
| No. of loans | 2,581,195 | 119,725 | 21,899 | 6,219 |
| % of titles without use after purchase | 62.21% | 67.05% | 26.72% | 88.50% |
| Average cost of e-book | $0.61 | $3.21 | $138.55 | $3.98 |
| Cost per use | $0.21 | $0.18 | $0.73 | $9.17 |

**Table 4.5.** Results of the cost and usage analysis for e-book packages based on BR2 COUNTER data.

| EB Fund Cost Analysis | Package A | Package B | Package C | Package D | Package E | Package F |
|---|---|---|---|---|---|---|
| 2013 Cost | $59,681.40 | $62,515.00 | $113,377.81 | $216,888.00 | $114,950.00 | $66,907.63 |
| No. of titles | 704 | 446 | 408 | 5,988 | 1,553 | 1,945 |
| No. of titles loaned | 153 | 79 | 111 | 4,509 | 294 | 686 |
| No. of loans | 2,937 | 486 | 2,143 | 119,085 | 8,620 | 11,360 |
| % of titles without use after purchase | 78.27% | 82.29% | 72.79% | 24.70% | 81.07% | 64.73% |
| Average cost of e-book | $84.77 | $140.17 | $277.88 | $36.22 | $74.02 | $34.40 |
| Cost per use | $20.32 | $123.63 | $52.91 | $1.82 | $13.34 | $5.89 |

**Table 4.6.** FY2011 to FY2014 e-book package assessment based on cost data (Voyager) and usage data (BR2 COUNTER).

| Package F | 2011 | 2012 | 2013 | 2014 | Total |
|---|---|---|---|---|---|
| Cost | $69,300.00 | | | | |
| No. of titles | 1,091 | | | | |
| No. of titles loaned | 365 | 824 | 566 | 203 | 942* |
| No. of loans | 5,822 | 28,855 | 11,430 | 2,463 | 48,570 |
| % of titles without use each year | 723 (66.27%) | 264 (24.20%) | 522 (47.85%) | 885 (81.11%) | Average: 503 (46.10%)** |
| % of titles without use after purchase | | | | | 149 (13.66%) |
| Average cost of e-book (cost/titles) | $63.52 | | | | |
| Cost per use (cost/use) | $11.90 | $1.99 | $1.50 | $1.43 | |

* Number of titles from the frontlist that have circulated at least once.
**The average was calculated based on the 2011–2013 data sets.

After the negotiation was completed, the cost analysis methodology described in this chapter was adapted by a colleague in the Science and Engineering Library Division (SEL) to assess a series of eleven subscription packages. Again, findings pointed to high cost per use and title overlap rates. After viewing the results, the SEL division decided to cancel the two packages with the lowest usage rates and negotiated a flat price increase (on an annual basis) for the remaining nine packages. In total, this assessment resulted in a savings of $10,000 on an annual basis. The findings indicate that this cost analysis methodology can be adapted and applied across the library system at CUL to yield results.

## E-book Frontlist Cost Analysis Project

At CUL, a large percentage of resources are directed toward frontlists. When I analyzed the top five packages ranked by cost (see table 4.5), the cost per use appeared high (averaging at $36). A closer examination of the data revealed that many frontlist titles are not available to CUL users until the end of the year (largely due to publication dates). It seemed that evaluating the cost per use of 2013 frontlist titles based on 2013 COUNTER usage reports did not accurately reflect their value.

In order to develop a method to evaluate the cost per use of e-book frontlists, the 2011 Package F frontlist was selected for evaluation because of the fact that it provided access to a rich data set (e.g., title lists, usage statistics) spanning several years. I located the Package F title list and matched it against BR2 COUNTER reports ranging in date from January 2011 to April 2014. Then, I experimented with a method to observe how usage and cost change over time (see table 4.6).

I expected that the number of loans would also increase over time, but the results indicate a different trend. Between 2011 and 2012, loans increased by more than 80 percent. In the following year, the number of loans dropped by more than 50 percent. After

**Table 4.7.** Analysis of EB fund allocations (Voyager data) and subject coverage across all e-book packages purchased in FY2013.

| Fund Family | Subject Area | LC Category | Amount | % of Total Spending |
|---|---|---|---|---|
| 2242 | Geology | Science | $218,891.59 | 21.84% |
| 2160 | Western European History/ German Literature | World History | $216,888.00 | 21.64% |
| 2054 | Statistics | Social Science | $144,644.05 | 14.44% |
| 2239 | Biology | Science | $71,814.86 | 7.17% |
| 2850 | Statistics | Social Science | $71,708.96 | 7.16% |
| 2700 | Engineering | Science | $70,590.00 | 7.04% |
| 2235 | Psychology | Phil/Psych/Rel | $66,907.63 | 6.68% |
| 2560 | Iberian Studies | World History | $30,000.00 | 2.99% |
| 2464 | Science | Science | $16,498.22 | 1.65% |
| 6491 | Engineering | General | $16,495.00 | 1.65% |
| 2115 | British/Irish Studies | World History | $13,404.17 | 1.34% |
| 6496 | Engineering | General | $8,033.53 | 0.80% |
| 2301 | Tibetan Studies | World History | $8,000.00 | 0.80% |
| 2029 | General/Undegraduate | General | $7,500.00 | 0.75% |
| 2300 | Chinese Studies | World History | $6,555.37 | 0.65% |
| 2465 | General/Undegraduate | General | $6,180.00 | 0.62% |
| 2133 | French Studies | Language & Lit | $4,760.00 | 0.48% |
| 2012 | Philosophy | Phil/Psych/Rel | $4,556.50 | 0.45% |
| 2107 | East Asian Studies | World History | $3,368.48 | 0.34% |
| 2111 | Envrionmental Studies | Geo/Anthro/Rec | $2,025.00 | 0.20% |
| 2750 | Journalism | Language & Lit | $2,025.00 | 0.20% |

considering collection content and usage trends, my thought is that high-use titles were included in course reading lists. During this analysis, an attempt was made to identify all titles that were included in course reserves over the past three years. However, the time involved to extract this data is not conducive to the timeframe for the study. The topic has been flagged for a future examination.

## Cost Breakdown according to LC Class

While a significant focus of the E-book Cost Analysis Project focused on the value of collection materials to the user community, I also wanted to examine the distribution of content across subject areas and determine how funds were allocated to build subject collections.

Again, I looked at subscriptions (ninety-six total) and packages (thirty-five total) to determine how spending occurs. Because many of these materials are interdisciplinary, costs are split between library divisions based on the percentage of content that relates to subject areas. To begin, I collected cost data from Voyager to determine which fund codes were included on subscription and package invoices. Then, I determined what percentage of the total cost was charged to each fund code.

After collecting cost data, I collected title lists from publisher and vendor websites and filtered lists according to subject area. In cases where Library of Congress (LC) Class headings were not included with title lists, I matched broad subject headings against LC Class headings available on the Library of Congress website[6] (table 4.7).

**Table 4.8.** Analysis of subject coverage based on LC Class across all e-book packages purchased on EB funds in FY2013.

| LC Category | % of all packages purchased | # of packages |
|---|---|---|
| General | 17.14% | 6 |
| Phil/Psych/Rel | 5.71% | 2 |
| World History | 17.14% | 6 |
| Social Science | 31.43% | 11 |
| Geo/Anthro/Rec | 2.86% | 1 |
| Science | 20.00% | 7 |
| Language & Lit | 5.71% | 2 |

Finally, I calculated the total expenditures related to each LC Class to determine the percentage of subject coverage and counted the number of packages (total of thirty-five) and subscriptions (total of ninety-six) related to each LC Class (table 4.8).

After I had experimented with the above-mentioned methodologies to measure cost, usage, and distribution of content across subject areas, I was eager to find an opportunity to determine if the same methods could be adapted to the needs of an individual library or subject collection. Luckily, I did not have to wait for long as an opportunity to conduct an analysis project

at the Avery Architectural and Fine Arts Library presented itself in FY2017.

## Applying Quantitative Methods at the Avery Architectural and Fine Arts Library

### Brief Background to the E-book Collection at the Avery Library

For the past several years, librarians at the Avery Architectural and Fine Arts Library have considered how to build an e-book collection that meets the information needs of a group of faculty and students heavily engaged in research activities. Much of this work involves investigations surrounding art history, painting, sculpture, and architectural history. The patron community working in this area requires access to high-resolution images that allow them to examine and study fine details in art and architecture.

The Avery Library traditionally focused collection development activities on print materials because they met the needs of the patron community. In recent years, the collection has outgrown stack space, and subject specialists face challenges regarding where and how to store print materials. At the same time, requests for electronic resources that support teaching and learning needs are growing. Many patrons have long commutes to and from campus and have expressed an interest in electronic resources that can be accessed remotely from the subway.

Subject specialists at the Avery Library are interested in purchasing a greater number of e-book subscriptions or packages, but have found it difficult to locate materials that contain high-resolution images. Many times, they purchased e-books only to discover that images were either missing or published at low resolutions. Feedback from patrons suggests that these e-books are not suitable to their needs because access to quality images is imperative to their work.

In the summer of 2015, I had a chance to speak with faculty and students at Columbia University who conduct research in areas related to the fine arts. In addition to missing images, they mentioned that the main factor preventing e-book use is a general lack of convenience. When asked how they define the term *convenience*, students provided the following three criteria:

1. Availability (e.g., remote access, not having to "wait in line");
2. Accessibility (e.g., sync devices for access at home, office, and commute);
3. Usability (e.g., search annotations, print chapters, download chapters).

When all three criteria are present, students are likely to use e-books to support scholarly activities. When they are not, participants search for alternative formats, such as print, that provide access to the images that they require.

### Quantitative Analysis of the Avery Library E-book Collection

When I started work on a quantitative analysis of e-book holdings at the Avery Library in 2016, subject specialists received several offers from publishers regarding academic e-book collections targeted towards the fine arts and architecture. Previews of sample titles and discussions with publishers indicated that the e-book packages provided all the features, namely high-resolution images and a DRM-free environment, that our users desire. However, we felt that further investigation was required before investing funds in these e-book packages. We wanted a larger body of evidence to either support or contradict whether patrons make use of e-books related to the Avery Library's collection.

To begin the analysis, I collected a random sample of titles listed in the offers we received from publishers and compared them against current e-book collection holdings listed in Voyager. The sample suggested a high degree of overlap (85 percent) with an existing subscription, which coincidentally had been part of the cost analysis conducted during the E-book Program Development Study (see previous section). Using my baseline of data, I already knew that the cost per use for this subscription was under $1.00, which is quite reasonable. However, because the subscription is interdisciplinary, I wanted to pull it apart and run a cost analysis on the titles related specifically to the Avery Library. I wanted to find out how the cost per use compared to my baseline and if patrons working in the areas of fine arts and architecture valued the subscription content.

First, I worked with subject specialists to create a list of all LC Classes that relate to the Avery Library's collection. Then, I pulled a master title list from the vendor's website and filtered it according to our list of LC Classes. Now, I had a base title list (14,802 total) to work with and uncover general usage trends.

Next, I went back to the vendor's website to collect BR2 COUNTER usage statistics. To look at general trends, I pulled a sample that covered three full fiscal years. Then, I copied my base title list and usage statistics into the same Excel spreadsheet and ran a MATCH formula to identify Avery titles with 1+ use.

By comparing this data to the baseline established in the previous section, I determined that the number of uses related to Avery collections was on par with general usage across campus; over a three-year period, 36 percent of titles related to the Avery Library

**Table 4.9.** Results of the cost and usage analysis based on cost data (Voyager) and usage statistics (BR2 COUNTER data).

| Avery Title Count and Usage Statistics | 2013 | 2014 | 2015 | Total |
|---|---|---|---|---|
| Total Ebrary titles related to Avery collections | | | | 14,802 |
| # of Avery titles with 1+ use | 2,506 | 3,315 | 3,458 | 5,328 |
| % of Avery titles with 1+ use | 16.93% | 22.40% | 23.36% | 36.00% |
| Total number of loans* | 276,041 | 301,740 | 378,898 | 956,679 |
| # of Avery titles without use | 12,296 | 11,487 | 11,344 | 9,474 |
| % of Avery titles without use | 83.07% | 77.60% | 76.64% | 64.00% |

*Loan = One successful chapter use

received at least one use. Cost per use at the Avery Library averaged at $0.25 across three years; the campus average was $0.21 (table 4.9).

To drill down one step further, I also used the BR2 COUNTER usage report to conduct an analysis according to publisher. I filtered the data alphabetically by publisher, counted the number of Avery Library titles in each publisher group, and counted the number of unique chapter uses associated with each title (table 4.10).

The results suggested that patrons at the Avery Library make heavy use of content from university presses. This information provided subject specialists with insight into the type of e-book content that is of value to patrons. This is information that can be used to evaluate future title lists that are included with e-book offers.

**Table 4.10.** Subset of unique e-book chapter uses broken down according to publisher.

| Publisher | # of Avery titles with 1+ use | # of Chapter Uses |
|---|---|---|
| Aarhus University Press | 5 | 141 |
| ABC-CLIO | 6 | 294 |
| Academic Studies Press | 3 | 1,209 |
| Academica Press, LLC | 1 | 25 |
| Addicus Books | 1 | 2 |
| AEI Press | 1 | 1 |
| Aeon Books | 2 | 14 |
| Africa Institute of South Africa | 3 | 133 |
| Agate Publishing, Inc. | 1 | 64 |
| Algora Publishing | 3 | 226 |
| Al-Maktoum Institute Academic Press | 1 | 423 |
| AltaMira Press | 9 | 1,942 |
| American University in Cairo Press | 5 | 218 |
| Amsterdam University Press | 45 | 3,886 |
| Ashgate Publishing Ltd | 119 | 10,553 |
| ASP | 5 | 256 |
| Australian Academic Press | 2 | 14 |

Next, to gauge the level of e-book collection development activities at the Avery Library, I wanted to count the number of bibliographic records related to the same list of LC Classes that had been used to analyze cost per use (see above analysis). Then, I wanted to compare the results against all e-book holdings at CUL.

To conduct this analysis, I collected data from two sources: Voyager and the library catalog. First, I isolated all bibliographic records related to the full e-book collection at CUL. Then, I broke the list apart by LC Class and categorized records according to type (e.g., databases, packages, subscriptions, and firm orders). I conducted counts of records across the categories I had created and decided to conduct a final and more granular count of specific subject areas within each category. Table 4.11 is a snapshot of results from the database category.

The results suggest that holdings related to the Avery Library account for an average of 23 percent of all e-book content available at CUL; its holdings account for 26 percent of databases, 40 percent of e-book packages, and 4 percent of subscriptions respectively.

Finally, to understand collection development activities from a financial perspective, I conducted one final analysis to determine how the Avery Library spread funds across print and electronic collections. Again, I pulled cost data from Voyager spanning across three fiscal years to determine expenditure trends (table 4.12).

Please note that confidential cost data has been removed from published results. However, the distribution of funds represented in the charts below provides a snapshot of spending activities.

The results indicated that there is still a high level of financial commitment to building print collections. However, based on new insight regarding the use and value of e-book collections at the Avery Library, subject specialists are using the results of this analysis to examine areas where funds can be shifted to e-book acquisitions. One area that has been flagged for further investigation is print subscriptions that are over five years old. However, there are still investigations to be completed in this area and results are not yet available. Regardless, subject specialists now have a baseline and methodology that can be used in future evaluations of collection holdings.

**Table 4.11.** E-book content purchased through databases. Comparison of database holdings at the Avery Library against all database holdings at CUL.

| Databases & Ongoing Collections (Purchased Primarily on EO & G funds) | # of Bib Records |
|---|---|
| # of databases listed in CLIO | 1,559 |
| # of databases related to Avery collection | 404 |
| % of total CLIO database records related to Avery collection | 26.00% |
| ***# of Avery Databases by Subject Heading*** | |
| Arts, Architecture and Applied Arts | 94 |
| History and Archaeology | 41 |
| General (e.g. Ebrary) | 269 |
| ***Total*** | ***404*** |

**Table 4.12.** Distribution of fund expenditures across print and electronic formats.

| Summary: FY16 Fund Expenditures (Print and E-book Formats) | | | | |
|---|---|---|---|---|
| ***Format*** | ***Fund Type*** | ***% of Total Avery GI*** | ***% of Total Avery Endowment*** | ***% of Total Avery Budget*** |
| Print Books | GI Funds | 29.00% | - | 26.21% |
| Print Books | Endowments | - | 95.00% | 9.23% |
| Print Total | | | | 35.00% |
| Electronic Ongoing | GI Funds | 10.00% | - | 10.05% |
| Electronic Ongoing | Endowments | - | 1.00% | 0.62% |
| E-book Ongoing | GI Funds | 0.07% | - | 0.07% |
| Elec. Subscriptions Total | | | | 9.22% |
| E-book Firm Order/Packages | GI Funds | 1.98% | - | 1.79% |
| E-book Firm Order/Packages Total | | | | 2.00% |

## Applying Quantitative Methods to a Big Deal Package

In July 2016, I was asked to join a working group that evaluated e-journal costs, particularly in relation to "big deal" packages. The group was composed of four colleagues working at the Science and Engineering Library (SEL) and the Health Science Library (HSL), as both campuses jointly acquire electronic materials that support scientific and medical research across campus.[7] Current business models present challenges to both SEL and HSL because they tie up a large percentage of annual collection budgets. The working group was tasked with developing a set of metrics to inform decisions relating to budget allocations, equitable contribution, and acquisition models that support future collection development initiatives.

While this project relates to big deal e-journal packages, all working group participants use similar cost analysis frameworks to evaluate collections of electronic resources, including e-books. Due to the flexibility of cost analysis models, like the one described at the beginning of this chapter, it was possible to combine our strategies and incorporate an additional citation and publication analysis to the method. The project demonstrated that it is possible to standardize collection analysis methods and utilize strategies that were originally developed to examine e-book business models and apply them to e-journal packages as well. Building upon previous assessment work completed at SEL and HSL, the Working Group on Journal Costs decided to examine three key indicators of journal value: cost per use, publication, and citation.

I will not spend a great deal of time on how cost per use was calculated, as the method was discussed in detail earlier in this chapter. However, I will say that like previous investigations, the big deal analysis started with cost data covering the previous three fiscal years, the number of titles available through the package, and the percentage of titles used at least one time. The results of the analysis indicated that cost per use increased each year due to annual price increases and an overall decrease in the number of uses.

One point of interest is that this analysis brought in IP range data, to examine *where* users access journal content. As a brief aside, examining IP range allows librarians to determine the location (e.g., campus location) from which a resource is used or downloaded. Essentially, the data indicates which server within the library system is used to access content. Results indicated that usage was split fifty-fifty between SEL and HSL locations. This metric was taken into consideration as librarians determined how best to calculate an equitable cost split between the two library locations.

Regarding citation analysis, the committee examined data from the Web of Science to determine which journals authors affiliated with Columbia cite in their research. Ten or more citations in the past three years

mark a journal as having collection value. The analysis indicated that 10 percent of titles (189 out of 1,800) met criteria for retention. The citation analysis also provided evidence for collection impact within the local patron community. Based on the results, we identified titles within the journal package that support research projects at Columbia University. Again, the idea of collection impact provided additional insight into the value of the big deal package for library users.

Publication analysis was conducted in a similar fashion; one peer-reviewed research article published in the previous year, or two in the past three years, mark a journal as having collection value. Data was also collected from Web of Science. The results indicated that 25 percent of titles (453 out of 1,800) met criteria for retention. Again, as mentioned above, the results of the analysis also provided evidence for collection impact. Based on the results, we identified titles within the journal package that disseminate research findings generated by our user community. This data also fed into considerations regarding the overall value of package content and a final cost split.

Based on the findings of the analysis, a recommendation was made to stay within the big deal package to provide the same level of content coverage at a rate that was in line with the library budget. While findings are still under review at the writing of this issue of *Library Technology Reports*, a general discussion of the methodology demonstrates how it is possible to combine flexible quantitative methods, regardless of format, to assess the value of content to user communities and measure collection impact.

While the three examples described in this chapter utilize quantitative methods to support a variety of study objectives, I hope they demonstrate how it is possible to develop a collection analysis strategy around an understanding of local information needs, resources, and readily available local and external data sources. There is still a great deal of experimentation that can be done regarding quantitative analysis and collection development activities, but discussions of strategies that have provided actionable results at individual libraries may bring us closer to a standardized method that can be applied across the entire professional community.

## Notes

1. Melissa J. Goertzen, "E-book Program Development Study: Results and Recommendations," Columbia University Academic Commons, 2016, https://doi.org/10.7916/D81Z44C3.
2. Columbia University Libraries, "Strategic Plan, 2010–2013," 2010, https://culis.columbia.edu/content/dam/staffweb/admin/strategicplanning/2010-2013/CULIS_Strategic_Plan_2010-20131.pdf (university login required).
3. Portions of the E-book Cost Analysis Project, specifically the methodology and results of the e-book subscription analysis, were presented at the 2014 Charleston Conference on November 6, 2014. The full results are reported in the conference proceedings: Melissa Goertzen and Krystie Klahn, "The Buck Stops Here: Assessing E-book Subscriptions at Columbia University," in *The Importance of Being Earnest: Charleston Conference Proceedings, 2014*, ed. Beth R. Bernhardt, Leah H. Hinds, and Katina P. Strauch (West Lafayette, IN: Purdue University Press, 2016), 156-161.
4. Tony Davies and Michelle Morgan, "E-books Down Under," in *Too Much Is Not Enough: Charleston Conference Proceedings 2013*, ed. Beth R. Bernhardt, Leah H. Hinds, and Katina P. Strauch (West Lafayette, IN: Purdue University Press, 2014), 167-175.
5. The full results of this project were presented at the 2014 Charleston Conference on November 6, 2014: Goertzen and Klahn, "The Buck Stops Here."
6. "Library of Congress Classification Outline," accessed December 30, 2016, https://www.loc.gov/catdir/cpso/lcco.
7. Members of the Working Group on Journal Costs: Amanda Bielskas, Melissa Goertzen, Michael Koehn, Jeremy Serrano, Krystie Wilfong.