

ALA American Library Association

APPLYING QUANTITATIVE METHODS TO E-BOOK COLLECTIONS

Melissa J. Goertzen

Library Technology Reports

Expert Guides to Library Systems and Services

MAY/JUNE 2017

Vol. 53 / No. 4

ISSN 0024-2586

Library Technology

R E P O R T S

Expert Guides to Library Systems and Services

Applying Quantitative Methods to E-book Collections

Melissa J. Goertzen



ALA TechSource
alatechsource.org

American Library Association

Library Technology R E P O R T S

ALA TechSource purchases fund advocacy, awareness, and accreditation programs for library professionals worldwide.

Volume 53, Number 4

Applying Quantitative Methods to E-book Collections

ISBN: 978-0-8389-5987-9

American Library Association

50 East Huron St.
Chicago, IL 60611-2795 USA
alatechsource.org
800-545-2433, ext. 4299
312-944-6780
312-280-5275 (fax)

Advertising Representative

Samantha Imburgia
simburgia@ala.org
312-280-3244

Editors

Patrick Hogan
phogan@ala.org
312-280-3240

Samantha Imburgia
simburgia@ala.org
312-280-3244

Copy Editor

Judith Lauber

Production

Tim Clifford and Samantha Imburgia

Cover Design

Alejandra Diaz

Library Technology Reports (ISSN 0024-2586) is published eight times a year (January, March, April, June, July, September, October, and December) by American Library Association, 50 E. Huron St., Chicago, IL 60611. It is managed by ALA TechSource, a unit of the publishing department of ALA. Periodical postage paid at Chicago, Illinois, and at additional mailing offices. POSTMASTER: Send address changes to *Library Technology Reports*, 50 E. Huron St., Chicago, IL 60611.

Trademarked names appear in the text of this journal. Rather than identify or insert a trademark symbol at the appearance of each name, the authors and the American Library Association state that the names are used for editorial purposes exclusively, to the ultimate benefit of the owners of the trademarks. There is absolutely no intention of infringement on the rights of the trademark owners.



Copyright © 2017
Melissa J. Goertzen
All Rights Reserved.

About the Author

Melissa J. Goertzen is the Collection Development Analysis & Support Librarian at Columbia University Libraries. She has ten years of experience working as a writer, project manager, and information professional at academic institutions across Canada and the United States. In 2016, she completed the E-book Program Development Study, an ambitious assessment project that documented the e-book landscape at Columbia University over the course of two years. The results provided a series of strategic best practices for collection development initiatives. She completed a double BA program in English and History at the University of Calgary, followed by a Master of Information Management (MLIS) at Dalhousie University. To learn more, please visit <http://melissagoertzen.wordpress.com>.

Acknowledgements

I thank the American Library Association for their support of this project. In particular, I would like to extend my gratitude to Samantha Imburgia and Patrick Hogan for helpful discussions and feedback on early drafts. Also, a special thank you to my wonderful collection development students, Karina Pantoja and Daisy Alarcon, who have provided invaluable assistance over the past four years, particularly in regard to conducting background research, collecting data, and editing manuscripts.

Abstract

Collection development activities that involve electronic content require knowledge of quantitative research methods. The ability to calculate cost per use, identify usage trends, and provide evidence for collection development decisions are essential skills in the digital age. Because of the dynamic nature of electronic resources, particularly e-books, it has been challenging to create standardized methods that support routine evaluations of collection materials. While the term quantitative analysis can seem daunting, I hope this issue of *Library Technology Reports* (vol. 53, no. 4), "Applying Quantitative Methods to E-book Collections," demonstrates that it is possible to hone quantitative skill sets and develop an evaluation framework for e-book collections based around readily available quantitative data sources regardless of the size of an individual library system or equipment budget.

Subscriptions

alatechsource.org/subscribe

Contents

Chapter 1—Analyzing E-book Collections in the Digital Age	5
The Concept behind This Issue of <i>Library Technology Reports</i>	5
Introduction to the E-book Program Development Study	6
Report Structure	6
Notes	7
Chapter 2—Establishing a Foundation	8
Brief Overview of the E-book Publishing Landscape	8
Brief Overview of Information Needs and Usage Trends	9
Local Information Needs at CUL	10
Suggested Resources	11
Notes	11
Chapter 3—Introduction to Quantitative Research and Data	12
What Is Quantitative Research?	12
Quantitative Research in Information Management Environments	13
Definitions of Local and External Data Sources	14
Indicators and Performance Measures That Support Quantitative Research	14
Challenges Presented by Quantitative Research	16
Designing a Quantitative Research Framework	16
Notes	17
Chapter 4—Putting It into Practice	19
Introduction to the E-book Collection and Quantitative Analysis at CUL	19
Developing a Baseline: E-book Cost Analysis Project	20
Applying Quantitative Methods at the Avery Architectural and Fine Arts Library	24
Applying Quantitative Methods to a Big Deal Package	26
Notes	27
Chapter 5—Translation of Quantitative Results to Collection Development Policies	28
Disseminating Quantitative Analysis Results to Stakeholders	29
Unresolved Areas	30
Notes	30
Chapter 6—Conclusion	31

Analyzing E-book Collections in the Digital Age

Information professionals work in an age of customized information consumption. The digital age opened the door to a virtual reality where knowledge is disseminated instantaneously around the globe and opportunities for collaboration and innovation are endless. In this era, individuals select from a wide variety of formats suited to unique learning styles and specific tasks. Convenience is the name of the game, and rapid technological advances in the consumer market shape expectations regarding how electronic content can be discovered, accessed, and used. Librarians play an important role in this landscape, particularly in regard to teaching, learning, and research environments; the ability to identify and acquire authoritative information links user communities to reliable resources that fulfill information needs. Over the past decade, a growing interest in electronic content has resulted in a greater emphasis on e-book collection development, which has required evaluations of new business models, license agreements, and collection assessment methods. The e-book market has yet to reach maturity and requires librarians to navigate a dynamic landscape that is often described as “the new wild west.”

Despite the exponential growth of available information sources, collection budgets often remain fixed and, in some cases, are reduced due to factors like the economy, demands on general institutional resources, or assumptions that digital information is free. In this environment, information needs surpass available resources, and librarians are required to justify purchases or requests for budget increases with quantitative evidence. Now more than ever, it is essential for librarians to demonstrate data-based collection development decisions or evaluate current holdings to

identify areas where resources can be shifted to support the teaching, learning, and research needs of a user community.

Collection development activities that involve electronic content require knowledge of subject areas and local needs, as well as skill sets related to quantitative research and analysis. The ability to calculate cost per use, identify usage trends, document how funds are allocated to acquire materials, and provide evidence for collection development decisions is essential. However, training opportunities in this area of library work are still in development and not always widely available. At the same time, the evolving nature of electronic resources, particularly e-books, provides challenges in regard to developing standardized methods of conducting quantitative analysis. When information managers share analysis methods currently used to evaluate electronic collections, opportunities for experimentation, feedback, and standardization become available.

The Concept behind This Issue of *Library Technology Reports*

In the summer of 2016, I was presented with an opportunity to write an issue of *Library Technology Reports* that discussed quantitative methods related to the analysis and evaluation of e-book collections. The initial concept was structured as a response to the article “The Evolving DDA Project at the Orbis Cascade Alliance” by Kathleen Carlisle Fountain.¹ Although the article focused on the results of a pilot consortial DDA project, a discussion surrounding the gaps in standardized approaches to e-book collection

analysis and general training opportunities spoke to many of the challenges that I encounter on a regular basis. As Fountain said:

I can only speak for myself, but I don't think that my experience is uncommon: I was not trained to purchase books after examining various sets of metrics. I bought books in anticipation of local needs that were based on plans, institutional research focus, the curriculum, and faculty requests. Those are two completely different skill sets, and I have learned on the job how to run analyses and use their results to make data-informed decisions. We as librarians need to continue to develop these skills that allow us to independently assess our success and adjust accordingly.²

Further in the article, she mentions that needs of librarians would be served by workable models that demonstrate how to analyze data, what it means, and how to act on it.³ This idea stayed with me and gradually evolved into the central thesis of this report.

For the past four years, I have worked in the Collection Development (CD) department at Columbia University Libraries (CUL) and regularly conduct analysis projects that assist with the overall management of the materials budget. Although CD oversees initiatives across a library system composed of twenty individual branches, the department is staffed by two professionals and one student worker. I work independently and rely exclusively on Microsoft Excel to complete most of my projects, including data collection and analysis. I mention this because I believe that many librarians who read this report may find themselves in similar situations. The term *quantitative analysis* can seem daunting, particularly when completing tasks alone or in small teams. I hope this report demonstrates that it is possible to develop quantitative skill sets and build evaluation frameworks for e-book collections based around readily available quantitative data sources, regardless of the size of an individual library or equipment budget.

Introduction to the E-book Program Development Study

Between the years 2013 and 2015, I conducted the E-book Program Development Study at CUL, a two-year assessment project aimed at gathering essential data to drive best practices related to e-book collection development strategies.⁴ The primary objective was to document the e-book landscape at Columbia University and understand how local challenges fit into the larger context of collection development and management within the academic community. A second objective was to create standardized methodologies that enabled librarians within the system to collect data and evaluate e-book holdings. The quantitative

data collected over the course of the study serves as a benchmark for the future evaluation of e-book holdings at CUL.

One of the successes of the study was the development of a cost-benefit methodology that has since been implemented at CUL. It is primarily used to evaluate the value of e-book subscriptions and packages. More recently, it has been applied to an evaluation of a big deal e-journal package. The work has involved documenting fund allocations across subject areas, calculating cost per use, examining title overlap, and identifying usage trends. This method has provided actionable results that were used by librarians at CUL to negotiate reductions in subscription fees and determine how to allocate funds to build collections. Because the study was based on local and external data that is readily available to the professional community, I believe that the methods described in the following chapters can be adopted and adapted by other libraries, regardless of size or budget.

Report Structure

As mentioned earlier, assessing the value of e-book collections begins with a foundational knowledge of patron information needs and publishing trends. Chapter 2 contains a brief overview of current trends in both areas followed by a discussion of local information needs documented through the E-book Program Development Study.

Chapter 3 introduces quantitative data and metrics; I discuss key characteristics and various types of research questions they can answer. Next, I list performance measures and indicators that can be used in information management environments to support conclusions and provide evidence for e-book collection development decisions. Finally, I provide a research framework that I rely on to plan and define my e-book analysis projects.

Chapter 4 provides three examples taken directly from my work at CUL. Each demonstrates how I have developed and applied a quantitative method to answer questions related to fund allocations, return on investment, usage trends, collection impact, and content distribution across subject headings. While my work was designed for the analysis of e-book collections, the same method has also proven successful in the evaluation of e-journal packages. This adaptation is described in the third example of this chapter.

Chapter 5 discusses how the results of quantitative research can translate into collection development policies and best practices. Evidence regarding budget allocations, usage trends, and return on investment can feed into established criteria for acquiring e-book materials and developing an ongoing strategy for collection development activities.

As the e-book landscape continues to evolve, I believe that opportunities to share analysis methods and techniques within the professional community will encourage ongoing experimentation and collaboration. The methods discussed in this report are not meant to be an end point but rather a springboard for future assessment projects and the development of standardized practices across the information profession.

Notes

1. Kathleen C. Fountain, "The Evolving DDA Project at the Orbis Cascade Alliance," *Against the Grain* 27, no. 5 (November 2015): 10–12.
2. Ibid., 12.
3. Ibid.
4. Melissa J. Goertzen, "E-book Program Development Study: Results and Recommendations," Columbia University Academic Commons, 2016, <https://doi.org/10.7916/D81Z44C3>.

Establishing a Foundation

Trends in the Publishing Industry and User Communities

As librarians navigate new realities of both information and financial landscapes, quantitative methods play a key role in analyzing the value of electronic content, particularly e-book collections, in order to demonstrate value and impact. To develop an effective framework for ongoing collection evaluation, it is important to establish a foundational knowledge of e-book publishing trends and local information needs that inform collection development decisions.

There is already a wealth of knowledge resources available to information professionals that discuss the above-mentioned topics in detail. Instead of focusing on them in depth, this chapter provides a brief summary of information that informed projects later discussed in this issue of *Library Technology Reports*. At the end of this chapter, I have also compiled a short list of resources that I used to fill personal knowledge gaps. This list is by no means exhaustive and does somewhat lean toward academic library environments, but I hope it may provide a starting point for others who are beginning their own investigations into the world of e-book collection development.

Brief Overview of the E-book Publishing Landscape

E-books are regarded as a disruptive technology and have required publishers, aggregators, and information professionals to completely rethink purchasing models, business relationships, and acquisition workflows. On all sides of the equation, content creators and providers search for models that offer quick, seamless, and affordable access to e-book content. This is a time of experimentation, and factors such

as publishing costs, intellectual property rights, and rapid advancements in technology mean that business models and workflows are often in flux.

One common but largely incorrect assumption is that electronic content costs less to produce than print materials. However, the “basic tasks involved in creating e-books are very similar to those of creating a print book: acquisition, financing, production, marketing, sales, and delivery of books.”¹ In regard to total publication costs, manufacturing and distribution expenses account for only 12 percent of a print book’s retail price.² While this percentage can be eliminated from an e-book’s price tag, the electronic format comes with three additional costs not associated with print counterparts: digitized preparation in multiple formats, quality assurance, and digital distribution through a range of retailers that have unique upload and management protocols.³

The e-book consumer market is based on a retail business model; customers pay for downloads of individual titles and content is stored on personal devices. This arrangement is similar to the way print titles are sold to consumers, including libraries.

The acquisition of e-books by libraries is based on online supply models, meaning that content is hosted on a third-party website and libraries pay for access. This model creates an ongoing business relationship between libraries, publishers, and aggregators. There is typically an exchange of annual fees (either maintenance or subscription) to maintain host platforms and ensure that content access is not interrupted. The terms of the business relationship are outlined by license agreements, which also determine how users discover, access, and use e-books.

In many cases, the shift in business practices creates challenges for both publishers and libraries. In

order to produce content and remain in business, publishers have to develop business models that are economically viable. Many fear that providing perpetual access to e-books through libraries can ultimately impact sales of digital titles, place increased pressures on bookstores, decrease royalties paid to authors, or provide opportunities for piracy.⁴ Libraries, on the other hand, state that e-book collections introduce texts to wider audiences and promote authors to patrons. Also, a central issue is the question of whether libraries own e-book titles or merely lease access as they would a subscription to an external database. Many questions regarding e-book ownership are wrapped into discussions of long-term access, preservation, and whether it is possible to transfer holdings to alternative platforms in the event that a publisher ceases to exist.

Currently, there exists a wide range of e-book business models that try to balance the diverse needs of content producers and libraries. Options include one book/one user, multiuser, unlimited simultaneous use, subscription, patron-driven acquisition, evidence-based acquisition, and short-term loan. Depending on the model, libraries can purchase titles in perpetuity or lease content for a specified period of time. Many models also incorporate digital rights management (DRM) that limit patrons' abilities to print or download sections of individual titles in order to prevent a negative impact on book sales or comply with publishing agreements. To learn more about various e-book business models, please see the list of suggested readings at the end of this chapter.

It is also important to note that in a highly experimental environment, it can take several fiscal years to determine if a business model is sustainable, which means it must be competitive and economically viable. One recent example is the widely popular patron-driven acquisition (PDA) model, which allows users to discover books that are not owned by the library and initiate their purchase for the collection. While the model provides exposure to a wide range of content, it is also highly unstable and can have dramatic impacts on bottom lines in the publishing industry as well as annual library budgets. Another concern is that PDA focuses all collection efforts on the short-term needs of a user community and does not consider "the long tail of need," which refers to robust collections that serve future users.⁵ Finally, there are concerns regarding the future of specialized scholarly monographs that cover narrow subject areas. Will small presses have the resources to publish these materials if they are not purchased by libraries at the time of publication?⁶ These are all challenges that impact the future of scholarly communication and the depth and breadth of subject-specific collections.

In this complex landscape, acquiring e-books for library collections begins with a consideration of the

source of content, whether that means striking a deal with the original content creator (i.e., the publisher) or with an aggregator who pulls work from multiple publishers into one platform. With many options available, librarians can locate business models that suit their needs if they familiarize themselves with the market and also with the information needs that exist in their local user community.

Brief Overview of Information Needs and Usage Trends

In the commercial e-book market, convenience is the name of the game. Customers access content anytime, anywhere, and from any device. Information consumption is highly customizable to individual needs or tasks at hand. The characteristics of commercial digital content inform the expectations and behaviors of information consumers, who also rely on libraries for access to materials.

A wide variety of studies and literature reviews by information professionals conclude that convenience is a critical factor in information-seeking choices, particularly in teaching, learning, and research environments.⁷ They discovered that the main benefits associated with e-books are twenty-four-hour access to materials, remote access, and the ability of multiple patrons to access one resource at the same time. Restrictions imposed by DRM, poor platform design, and limited file compatibility with popular devices served as deterrents.

In teaching, learning, and research environments, e-book use is most often task-oriented. Activities are linked to quick reference, in-text search functionalities, and citation checks.⁸ In other cases, e-books serve as a convenient means of previewing a text; students and faculty members peruse the e-version to gain a sense of the information, biases, or arguments presented in a scholarly monograph. If content is deemed useful, a print version is often requested for extended reading.⁹

In many cases, expressed preferences for print or electronic formats are linked to predicted levels of engagement with a text. For instance, e-books are often favored for discontinuous reading activities, which involve skimming, fact or citation checks, and searches. This observation is described as "use rather than read" behavior.¹⁰ Essentially, e-books are a tool that promotes power browsing. They allow users to preview texts or try out new authors in a convenient and low-risk fashion. Print formats, however, are typically favored for continuous reading activities. These involve a deep level of engagement with a text over an extended period of time. Associated activities can include learning new information, conducting research, or engaging with a plot.

Some researchers speculate that the relationship between format preference and reading behavior may account for the popularity of the PDF file format, particularly in academic environments. The format combines desired functionalities of electronic (search capability, convenience) and print (reading) into one convenient reading space. The format provides electronic renditions alongside the ability to produce high-quality prints, giving users freedom to choose the format most appropriate to their purposes.¹¹ Researchers suggest that when studying information-seeking behaviors, it is important to consider how users prefer to search for content and how they prefer to read content.¹²

Local Information Needs at CUL

Several months into the E-book Program Development Study,¹³ I realized that I was operating under the assumption that users prefer electronic content for research, teaching, and learning activities. Throughout society, we rely so heavily on the Internet, mobile technologies, and social media to gather and disseminate information that I assumed users would prefer e-books for all scholarly activities. However, as I started to collect usage statistics, examine discovery and access trends, and speak with faculty, students, and library staff, I realized that my initial impressions of content use were far too simplistic and did not tell the full story. As I dug deeper into data sets, I started to ask more and more questions about when, how, and why users gravitate towards certain formats (e.g., print, electronic, archival materials) to support scholarly activities and build knowledge around specific subject areas.

At first, I worried that the discovery of questions rather than solutions would result in a lengthy list of proposals for future studies. But the opposite turned out to be true. The inquiry process provided a focus and pulled everything I had observed into one overarching question: what is the intended use of e-book content? Users interact with information for a variety of reasons including course use, research pursuits, and general interest. These activities serve different functions within a community, rely on different forms or reading (e.g., continuous vs. discontinuous), and support different information needs.

When I consider intent of use from a collection development perspective, my mind automatically begins to separate activities into the categories of “current use” and “future use.”¹⁴ This seems appropriate as there are few business models or collection development strategies that address both requirements at once. I think this separation points to a general shift in the way academic libraries in the twenty-first century must approach collection development activities;

successful initiatives rely on a balance between “just in case” and “just in time” strategies. This balance allows information professionals to determine when it makes economic sense to invest resources in high-use materials for current users and when it is appropriate to purchase materials that may have low use but add to the long-term value and legacy of the collection.¹⁵ Again, having a strong understanding of information needs and how patrons intend to use e-book titles, packages, and subscriptions provides the insight required to make these calls.

My work on the E-book Program Development Study has convinced me that going forward, collection development decisions regarding business models, license negotiation, and title acquisition hinge on the libraries’ understanding of the intent of e-book use. Once we have successfully answered this question, we can enter into meaningful conversations with stakeholders, advocate for the needs of users during conversations with publishers and aggregators, and build collections that satisfy immediate and long-term information needs.

It is important to note that as the information landscape continues to evolve at a rapid pace, user intent will evolve as well. Below are several methods that were used in the E-book Program Development Study to document intent of use. They served as a starting point for the evaluation of e-book holdings, collection development policies, and general communication with the patron community. Please see chapter 3 for an in-depth discussion of indicators that inform collection development decisions and activities.

- Solicit direct feedback from patrons to monitor evolving information needs and the ways in which collection materials are used to support teaching, learning, and research activities. Based on feedback, develop a baseline to monitor usage trends that impact how collections are built (e.g., acquisition of print or electronic, focus on subscriptions vs. package deals that offer perpetual access) to fulfill “just in case” and “just in time” priorities.
- Develop strong channels of communication with subject specialists and liaison librarians to monitor changing information needs (e.g., shift in patron focus from one subject area to another) and document how collection materials are integrated into course reading lists, public programs and services, and so on.
- Consider whether e-book materials are most likely used to support continuous or discontinuous reading activities.
- Monitor usage trends across e-book subscriptions and packages. Evaluate new subscriptions over a period of three years to examine usage trends, return on investment, overlap, and quality of content.

Based on observations regarding intent of use within the patron community at CUL, I also compiled a list of criteria to inform a general evaluation framework that would standardize e-book evaluation methods across campus and establish a baseline of quantitative data:

- intent of use
- intended audience
- access at least equivalent to print
- sufficient access (e.g., multiple user access when a title is placed in course reserves)
- cost history
- range of subject coverage
- usage and availability of usage statistics
- return on investment
- content impact within the user community

Suggested Resources

Listed below are several resources I used to fill my own knowledge gaps regarding e-book business models, collection development practices, and resource management practices. This list is by no means exhaustive, and there is a wealth of information available to the professional community. However, I hope it can provide a solid starting point to information professionals who are developing e-book evaluation frameworks for the first time or would like to expand their knowledge on practices mentioned above.

- Clay, James. *Preparing for Effective Adoption and Use of Ebooks in Education*. JISC Observatory TechWatch Series, Report no. 4, final version. JISC Observatory, December 2012. http://opus.bath.ac.uk/34852/1/ebooks_in_education.pdf.
- Nelson, Jeremy. *Becoming a Lean Librarian: Lessons from the World of Technology Start-Ups*. Waltham, MA: Chandos Publishing, 2016.
- Polanka, Sue. *The No Shelf Required Guide to E-book Purchasing*. Chicago: ALA TechSource, 2011.
- Roberts, Stephen A. *Financial and Cost Management for Libraries and Information Management Services*. London: Bowker-Saur, 1998.
- Roncevic, Mirela. "E-book Platforms for Libraries." *Library Technology Reports* 49, no. 3 (April 2013).

Notes

1. Stanley M. Bensen and Sheila Nataraj Kirby, "E-books and Libraries: An Economic Perspective," American Library Association, September 2012,

- www.ala.org/transforminglibraries/e-books-and-libraries-economic-perspective.
2. Ibid.
 3. Ibid.
 4. Kathryn Zickuhr, Lee Rainie, Kristen Purcell, Mary Madden, and Joanna Berenner, *Libraries, Patrons, and E-books*, report (Washington, DC: Pew Research Center, June 22, 2012), <http://libraries.pewinternet.org/2012/06/22/part-1-an-introduction-to-the-issues-surrounding-libraries-and-e-books>.
 5. Kizer Walker, "Patron-Driven Acquisition in U.S. Academic Research Libraries: At the Tipping Point in 2011?" *Bibliothek: Forschung und Praxis* 36 (2012): 126–30.
 6. Ibid.
 7. Lynn Silipigni Connaway, Timothy J. Dickey, and Marie L. Radford, "If It Is Too Inconvenient I'm Not Going after It: Convenience as a Critical Factor in Information-Seeking Behaviors," *Library and Information Science Research* 33, no. 3 (2011): 179–90; Linda Ashcroft, "E-books in Libraries: An Overview of the Current Situation," *Library Management* 32, no. 6 (2011): 398–407; Amalia Beisler and Lisa Kurt, "E-book Workflow from Inquiry to Access: Facing the Challenges to Implementing E-book Access at the University of Nevada, Reno," *Collaborative Librarianship* 4, no. 3 (2012): 96–116.
 8. A. Noorhidawati and Forbes Gibb, "How Students Use E-books—Reading or Referring," *Malaysian Journal of Library and Information Science* 13, no. 2 (December 2008): 1–14; Selinda Adelle Berg, Kristin Hoffman, and Diane Dawson, "Not on the Same Page: Undergraduates' Information Retrieval in Electronic and Print Books," *Journal of Academic Librarianship* 36, no. 6 (November 2010): 518–25.
 9. Ibid.
 10. Jeff Staiger. "How E-books Are Used," *Reference and User Services Quarterly* 51, no. 4 (2012): 355–65.
 11. Xi Niu, Bradley M. Hemminger, Cory Lown, Stephanie Adams, Cecelia Brown, Allison Level, Merinda McLure, Audrey Powers, Michael R. Tennant, M. and Tara Cataldo, "National Study of Information Seeking Behavior of Academic Researchers in the United States," *Journal of the Association for Information Science and Technology* 61, no. 5 (May 2010): 869–90.
 12. Ibid.
 13. Melissa J. Goertzen, "E-book Program Development Study: Results and Recommendations," Columbia University Academic Commons, 2016, <https://doi.org/10.7916/D81Z44C3>.
 14. Greg Eow, Todd Gilman, Jill Jascha, Caitlyn Lam, Melanie Maksin, Scott Matheson, Colin McCaffrey, and Nathan Rupp, *The eBook Strategic Plan Task Force: Report of Findings and Recommendations* (New Haven, CT: Yale University Library, March 1, 2013), www.library.yale.edu/departments/collection-development/Yale-ebook-task-force-rpt.pdf.
 15. Ibid.

Introduction to Quantitative Research and Data

The foundation of any e-book analysis framework rests on knowledge of the general e-book landscape and the existing information needs of a local user community. From this starting point, quantitative methods, such as cost analysis, can provide evidence for collection development initiatives and demonstrate how they align with patrons' needs and the overarching goals of library administrators or funding agencies.

Essentially, “data stands in place of reality we wish to study. We cannot simply know a phenomenon, but we can attempt to capture it as data which represents the reality we have experienced . . . and are trying to explain.”¹ The data collected through quantitative investigations provides a baseline for future evaluation, evidence for when and how patrons make use of electronic collections, and promotes data-driven decisions throughout collection development departments. To get the most mileage out of the time and resources invested into quantitative investigations, it is essential to first understand what quantitative research is and what types of questions it can answer.

What Is Quantitative Research?

In the most basic terms, quantitative research methods are concerned with collecting and analyzing data that is structured and can be represented numerically.² One of the central goals is to build accurate and reliable measurements that allow for statistical analysis.

Because quantitative research focuses on data that can be measured, it is very effective at answering the “what” or “how” of a given situation. Questions are direct, quantifiable, and often contain phrases such as

what percentage? what proportion? to what extent? how many? how much?

Quantitative research allows librarians to learn more about the demographics of a population, measure how many patrons use a service or product, examine attitudes and behaviors, document trends, or explain what is known anecdotally. Measurements like frequencies (i.e., counts), percentages, proportions, and relationships provide means to quantify and provide evidence for the variables listed above.

Findings generated from quantitative research uncover behaviors and trends. However, it is important to note that they do not provide insight into *why* people think, feel, or act in certain ways. In other words, quantitative research highlights trends across data sets or study groups, but not the motivation behind observed behaviors. To fill in these knowledge gaps, qualitative studies like focus groups, interviews, or open-ended survey questions are effective.

Whenever I sit down to a new quantitative research project and begin to think about my goals and objectives, I like to keep a small cheat sheet on my desk to remind me of the trends quantitative data can uncover and the stories that I can tell with study conclusions. This serves as one quick strategy that keeps my thoughts focused and prevents scope creep as I discuss project plans with various stakeholders.

Quantitative Research Cheat Sheet

Six key characteristics of quantitative research:

- It deals with numbers to assess information.
- Data can be measured and quantified.
- It aims to be objective.
- Findings can be evaluated using statistical analysis.

- It represents complex problems through variables.
- Results can be summarized, compared, or generalized.

Quantitative findings can provide evidence or answers in the following areas:

- Demonstrate to what extent services and collection are used and accessed.
- Back up claims about use and impact.
- Provide evidence for how the budget is spent and whether adjustments should be made.
- Demonstrate return on investment when presenting budget figures.
- Inform decisions regarding packages and subscriptions that are or are not worth pursuing.
- Demonstrate evidence for trends and prove or discount what is known anecdotally.
- Provide a method to make information accessible to audiences.
- Provide evidence of success and highlight areas where unmet information needs exist.

Main advantages of quantitative research:

- Findings can be generalized to a specific population.
- Data sets are large, and findings are representative of a population.
- Documentation regarding the research framework and methods can be shared and replicated.
- Standardized approaches permit the study to be replicated over time.

Main limitations of quantitative research:

- Data does not provide evidence for *why* populations think, feel, or act in certain ways.
- Specific demographic groups, particularly vulnerable or disadvantaged groups, may be difficult to reach.
- Studies can be time consuming and require data collection over long periods of time.³

Quantitative Research in Information Management Environments

In the current information landscape, a wealth of quantitative data sources is available to librarians. One of the challenges surrounding quantitative research in the information management profession is “how to make sense of all these data sources and use them in a way that supports effective decision-making.”⁴

Most libraries pay for and receive materials through multiple routes. As a result, a quantitative

research framework for e-book collections often consist of two central components: an examination of resource allocations and expenditures from funds, endowments, or gifts; and an examination of titles received through firm orders, subscriptions, packages, and large aggregated databases.⁵ In many cases, examining funds and titles according to subject areas adds an extra layer of knowledge that can provide evidence for teaching, learning, or research activities in a specific field or justify requests for budget increases.⁶

Many of the quantitative research projects that I have conducted over the past four years are in direct response to an inquiry from library administrators. In most cases, I have been asked to provide evidence for collection development activities that support expressed information needs, justify expenditures, or project annual increases in preparation for a new fiscal year. Study results are often expected to describe or weigh several courses of action in the short and long term. Essentially, my work is categorized into three basic concepts related to library management:

- Distinguish between recurrent and capital expenditure and projects, and between past, present, and future states.
- Accommodate priorities and determine how resources are spread across collections.
- Indicate the ways of allocating resources at input, monitor performance, and assess performance at output.⁷

To assist in my prep work for a quantitative research project, I put together a file of background information about my library system and local user community to ensure that the project supports institutional goals and aligns with the general direction of programs and services on campus. Below are seven categories of information that I have on file at all times:

- the institutional identity of the library
- the stakeholder groups to be served
- collection resources
- financial resources
- library personnel
- facilities and equipment
- the various programs and services related to the quantitative investigation⁸

Typically, I take a day or two at the beginning of each fiscal year to update this information and ensure that it accurately reflects the landscape of collections and services available at CUL. From this starting point, it is simple to look at new project descriptions and think about the data required to support high-level decisions regarding the allocation of resources,

to assess the effectiveness of collections and services, or to measure the value and impact of collections.

A wealth of local and external data sources is available to librarians, and each one can be used to tell a story about collection size, value, and impact. All that is required is an understanding of what the data measures and how different sources can be combined to tell a story about a user community.

Definitions of Local and External Data Sources

The remaining sections of this issue of *Library Technology Reports* discuss how I use quantitative data, what evidence I have uncovered to support e-book collection decisions, and how I apply quantitative findings in practical library settings. For the purposes of these discussions, I will use the following terminology:

Bibliographic record: A library catalog record that represents a specific title or resource.

Catalog clickthroughs: Counts of patron use of the catalog to access electronic full texts.

Citation analysis: Measurement of the impact of an article based on the number of times it has been cited.

Consortia reports: Consolidated usage reports for consortia. Often used to view usage linked to each individual consortia member.

COUNTER (Counting Online Usage of Networked Electronic Resources): An international initiative to improve the reliability of online usage statistics by providing a Code of Practice that standardizes the collection of usage data. It works to ensure vendor usage data is credible and comparable.

Cost data: Factual information concerning the cost of library materials, annual budget allocations, and general acquisitions budget.

FTE (full-time equivalent): The number of full-time faculty and students working or studying at a specific institution.

IP (Internet Protocol) address: A numerical label usually assigned to a library router or firewall that provides access to a private network (e.g., school or library network).

Link resolver statistics: Information regarding the pathways users take to access electronic resources.

Overlap data: Measurement of the degree of duplication across a collection.

Publication analysis: Measurement of impact by counting the research output of an author. Metrics include the number of peer-reviewed articles, coauthor collaborations, publication patterns, and extent of interdisciplinary research.

Title lists: Lists of e-book titles available in subscriptions, databases, or packages. These lists are generated and maintained by vendors and publishers.

Turnaway statistics: The number of patrons denied access to a specific title.

Vendor use data: Electronic use statistics provided by vendors.

Indicators and Performance Measures That Support Quantitative Research

I regularly use several indicators and performance measures to analyze e-book collections. Local and external data sources (listed in the section above) inform these investigations and provide the necessary “ingredients” to conduct cost analysis, examine return on investment, or measure the value of e-book collections to the community at CUL. Below is a breakdown of how I classify data and relate it to different indicators.⁹

Input Cost Measures

Data source: Cost data pulled from Voyager reports (or your institution’s ILS system).

In general, cost data demonstrates how funds are allocated across a budget. Analysis can identify areas where additional resources are required, monitor cost changes over time, and flag collection areas where funds can be pulled (e.g., overbudgeted funds, subject areas that no longer support the curriculum, etc.) and “reinvested” in the collection to support current information needs.

Each of the investigations described in the following chapter began with a review of cost data. I relied on a basic knowledge of how e-book acquisition budgets are distributed across subject areas or pooled to purchase interdisciplinary materials. Essentially, these investigations involved the identification of fund codes linked to subject areas, expenditures across set date ranges (e.g., calendar years, fiscal years, academic years), and bulk versus long-tail purchases.

Tip: When working with cost data and examining input cost measures, I have found it helpful to categorize data by fund type. E-book collections at CUL are often built with general income (GI) funds, endowments, and gifts. Policies and procedures regarding how funds can be transferred and what materials can be purchased impact how resources are allocated to build e-book collections. Before beginning a cost analysis project at your institution, it may be helpful to review the policies in place and determine how they relate to overarching institutional goals and collection priorities.

Collection Output Measures

Data sources: Cost data, title lists, overlap data, bibliographic records (particularly subject headings).

Collection output measures are related to the quantity and quality of output. Examples include the number of e-book titles included in a subscription or package deal acquired by a library, the number of e-book records acquired over a given period of time, the number of publishers and unique subject areas represented in an e-book collection, the currency of information (e.g., publication year), and the percentage of title overlap, or duplication, within a collection.

At this stage in my cost analysis projects, it is often necessary to combine data to create a snapshot of how funds flow in and out of subject areas to acquire research and teaching materials. For example, many of our large e-book packages are interdisciplinary. By pulling cost data, I can determine how the total cost was split across subject divisions based on fund code counts. Then, I break title lists apart by subject to determine what percentage of total content relates to each library division. By comparing the cost breakdown and title list breakdown, it is possible to determine what percentage of total content each library division receives and if it is on par with the division's financial contribution.

Effectiveness Measures and Indicators

Data sources: Cost data, title lists, COUNTER reports, vendor reports, consortia reports, resolver statistics, turnaway statistics, Google Analytics.

Examining input and output measures is an effective way of determining how budgets are allocated and the quantity and quality of materials available to patrons. To develop a quantitative baseline for the general value of e-book collections, measures like rate of use, cost per use, and turnaway rates can be very effective.

Again, this form of analysis relies on data from multiple sources. The ability to combine cost data, title lists, and COUNTER data (or vendor data) has yielded actionable results at my library. For instance, I combine data from these three sources to measure the value of databases. By pulling cost data covering three fiscal years and matching title lists against COUNTER reports, I have been able to examine trends in annual increase rates, examine overlap between subscriptions in the same subject area, and calculate cost per use to determine what percentage of the user community makes use of subscriptions.

Finally, by looking at turnaway statistics (also found in COUNTER data), it is possible to determine if sufficient access is provided to users. For instance, I look at turnaway statistics to evaluate if e-books listed on course reading lists provide sufficient access to a class of students over a semester. In cases where access is limited to a single user, I may look at the budget to find areas where funds can be shifted to

purchase simultaneous usage instead.

Together, the data sets mentioned above provide evidence for how funds are invested, if they are invested in materials that are heavily used by patrons, and if access models are suited to the needs of the local user community.

In some cases, particularly when dealing with foreign language materials, I have encountered challenges because COUNTER data is not provided, and in some cases, it is difficult to obtain vendor reports as well. In the absence of usage data, I have experimented with link resolver statistics to determine what information they provide about user activities and the value of e-book materials.

Link resolver statistics provide information about the pathways users take to access electronic resources.¹⁰ Resolver statistics show that a patron made a “request” via the link resolver and started the process of trying to view a full text. If the patron successfully accesses the full text, this is counted as a “clickthrough.”

It is important to note that link resolver statistics and usage statistics (like COUNTER) are not comparable because they measure different activities. Link resolvers measure attempts to connect while usage data measures usage activity. However, comparing sets of link resolver statistics against each other may provide insight into which resources patrons attempt to access most frequently. This can provide a ballpark idea of resource value in cases where usage statistics are not available.

Domain Measures

Data sources: FTE (full-time equivalent), IP address, demographic information.

Domain measures relate to the user community served by a library. They include total population, demographic information, attributes (e.g., undergraduate level, graduate level), and information needs.

In my work, domain measures impact subscription or package costs because campus-wide access is often priced according to FTE. Due to the size of CUL's student body, access to essential collections can become extremely expensive and fall outside of the budget range. When this occurs, examining patron access by IP address has opened the door to negotiation, particularly when dealing with content that is discipline-specific. For instance, when negotiating subscription prices for science materials, IP data provided evidence that usage is concentrated at the library router located in the Science and Engineering Library. This allowed science selectors to negotiate pricing models based around the FTE of natural science programs as opposed to the campus community as a whole.

Cost-Effectiveness Indicators

Data sources: COUNTER reports, vendor reports, turnaway statistics, citation analysis, publication analysis.

Cost-effectiveness indicators are related to measures like cost per use and ultimately examine the general return on investment. They evaluate the financial resources invested in a product and determine if the investment brings added value to the existing collection.

In my work, I often combine cost data with usage data to calculate cost per use and also capture usage trends spanning at least three calendar years. The results provide a benchmark regarding whether the financial investment in the product is equivalent to its general “demand” within the user community. A recent project with colleagues at the science and medical science libraries has examined how to use citation and publication data to determine general impact of electronic resources.

Challenges Presented by Quantitative Research

One of the challenges surrounding quantitative research in library environments is a lack of standardization across data sets, particularly vendor reports. The general situation has improved in recent years due to widespread compliance with the COUNTER Code of Practice, but there is still work to be done. It is difficult to interpret the meaning of vendor usage data that is still not COUNTER-compliant because clear definitions of *use* do not exist. This can create significant roadblocks when running quantitative projects that examine multiple e-book collections to get a sense of comparative value.

Also, usage data is generated outside of libraries by publishers or aggregators and vendors. Factors like turnover, company mergers, or password changes result in significant time lags between when usage statistics are generated and when libraries receive them. Also, some vendors pull down usage statistics after a period of months. In most cases, librarians need statistics captured over two or three years to meet reporting requirements, and data dating back this far can be difficult to obtain. Finally, annual usage statistics are provided according to calendar year. However, librarians look at usage by fiscal year and academic year as well. In many cases, this means that multiple usage reports have to be stitched together in order to capture the appropriate timeframe for reporting purposes. This process is labor intensive and takes a considerable amount of time to complete.

These challenges emphasize an ongoing need to build positive working relationships with publishers, aggregators, and vendors to discuss challenges and

develop solutions that benefit all stakeholders. It is important to note that libraries have valuable information that is not available to content providers, namely how e-books are discovered and used. Strong relationships allow for the transparent exchange of information between all parties, which ultimately benefits patrons by providing a seamless e-book experience.

Designing a Quantitative Research Framework

As mentioned earlier in this chapter, data stands in place of a reality we wish to study, quantify, and explain. In order to prevent scope creep and pull together bodies of data that add value to local work environments, it is essential to begin any quantitative research project with a set of clearly defined objectives, a strong understanding of the stakeholder group or audience, and knowledge of local information needs. These bits of information serve as markers to measure progress and ensure the project stays on track.

It is tempting to dive straight into a project and investigate if anecdotal information or assumptions are correct, but time spent developing a project outline is never wasted. The development of a successful plan requires “a clear idea of what it is to be achieved among the stakeholders. Clearly articulated objectives are the engine that drives the assessment process. This is one of the most difficult but most rewarding stages of the assessment process.”¹¹ Creating a roadmap for research projects can save countless hours down the line and ensures the correct quantitative method is selected. The plan also provides focus when the analysis phase of a project begins. Keep in mind that the data set you end up working with will be large; approaching it with stated goals and objectives saves significant amounts of time, which is especially important when working under a tight deadline!

Below is a checklist that I use at the beginning of any research project. It is based on recommendations made by Bakkalbasi, Sundre, and Fulcher.¹²

1. Project goals, objectives, and desired results
While goals and objectives are closely related, they are not the same. Project goals should state exactly what you hope to learn or demonstrate through your research. Objectives state what you will assess or measure in order to achieve your overarching project goal.
Example of a project goal:
 - Gain insight into how local patrons use the library to support teaching and learning needs.Example of project objectives:
 - To learn what activities local patrons engage in when using library facilities.

- To assess the degree to which patrons make use of subscribed e-book content.
 - Consider how results may support improvement of collection development initiatives or lead to evaluation of existing workflows, policies, and procedures.
2. List of key stakeholders
 - What questions and/or evidence are required by stakeholders?
 - What information do stakeholders require to make decisions?
 - How will results support the improvement of collection development initiatives?
 - How will results be made accessible to stakeholders?
 - Are the results intended for internal use, or will they be shared with the professional community?
 - Will findings be used to support grant or funding applications?
 3. Project timeline
 - Is there a stated project deadline? What methods or resources will allow you to collect data, conduct analysis, and provide findings within the stated timeframe?
 - Does the project coincide with other activities that may require your attention (e.g., fiscal year, subscription renewal period)?
 - Are there colleagues at the library who may be able to provide assistance given the timeline of the project?
 4. Confidentiality
 - What data collected through the study cannot be shared with external stakeholders (e.g., cost data, FOIP compliance, etc.)?
 - Are there any permissions required before study results can be disseminated to external stakeholders?
 - Is clearance required to collect data from a user community?
 5. Data collection process
 - What data sources are most valued and meaningful to your library?
 - What data sources will allow results to be applied at your library?
 - What data collection methods will be most effective?
 - What data collection methods will provide valid and reliable results?
 - Are there parameters such as specific fiscal years, calendar years, or academic years that you are required to report on?
 6. Data analysis¹³
 - How will data be summarized and described?
 - What features of the data set are most relevant to project objectives and goals?

- What are the relationships between different data sets?
7. Presentation of results
 - How is data evaluated?
 - How is data interpreted into meaningful results and conclusions?
 - What are the recommendations for action or improvements?
 - How will findings be communicated to stakeholders?

The data sets collected through quantitative methods are large and can easily be examined from a variety of perspectives. As the project develops, mentally frame emerging trends into a story that can be shared with stakeholders. This process determines how results will ultimately be applied to collection development initiatives. Background knowledge of the local patron community and institutional goals serves as a compass; use it to shape results that bring value to your library or the greater professional community.

From my experience, each quantitative project that I work on allows me to expand my skill sets and understand how I can structure my daily activities to support overarching institutional goals. During many projects, I have encountered unexpected challenges or had to improvise when quantitative methods did not yield expected results (e.g., low survey response rates). However, each challenge equipped me to take on larger projects, better understand how our budget is structured, or build stronger relationships with patrons and colleagues.

One skill that has been invaluable to my work is the ability to develop a quantitative research plan. I hope that by sharing this structure, along with performance measures and data sources that I use, readers have a behind-the-scenes view of my process and all of the moving parts that I work with to conduct e-book collection analysis. And of course, now to the fun part! It is time to get down to the nitty-gritty and demonstrate how I conduct analysis to inform budget decisions and collection development activities at CUL.

Notes

1. Bob Matthews and Liz Ross, *Research Methods: A Practical Guide for the Social Sciences* (Harlow, UK: Pearson Education, 2010), 45.
2. *Ibid.*, 465.
3. Based on information provided by Stephen A. Roberts, *Financial and Cost Management for Libraries and Information Management Services* (London: Bowker-Saur, 1998), 140–41.
4. Darby Orcutt, *Library Data: Empowering Practice and Persuasion* (Santa Barbara, CA: Libraries Unlimited, 2009), 106.

5. Northwestern University Libraries, "DataBank: How to Interpret Your Data: Financial Support," LibGuide, last updated December 8 2015, <http://libguides.northwestern.edu/c.php?g=115065&p=748741>.
6. Ibid.
7. Roberts, *Financial and Cost Management*, 132.
8. Ibid.
9. For further information regarding indicators and performance measures, please see Roberts, *Financial and Cost Management*, 140–41.
10. Orcutt, *Library Data*, 107.
11. Nisa Bakkalbasi, Donna Sundre, and Kenton Fulcher, "Assessing Assessment: A Framework to Evaluate Assessment Practices and Progress for Library Collections and Services," in *Proceedings of the 2012 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment, October 29–31, 2012, Charlottesville, VA*, ed. Steve Hiller, Martha Kyrillidou, Angela Pappalardo, Jim Self, and Amy Yeager (Washington, DC: Association of Research Libraries, 2013), 538-545.
12. Ibid.
13. Based on information provided by Matthews and Ross, *Research Methods*, 345.

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

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

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%

use after purchase, the average cost of an e-book, and cost per use⁴ (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/Undgraduate	General	\$7,500.00	0.75%
2300	Chinese Studies	World History	\$6,555.37	0.65%
2465	General/Undgraduate	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	Environmental 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

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

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 website⁶ (table 4.7).

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.

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.

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

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.

Translation of Quantitative Results to Collection Development Policies

The results of a quantitative analysis project often bring about a greater understanding of how budgets are allocated to build e-book collections, general usage trends, and impact. This knowledge can be applied to the improvement of library services or the creation of policies surrounding collection development priorities and activities.

At Columbia University Libraries (CUL), results from projects like the E-book Program Development Study are used to inform best practices that guide decisions related to the selection and acquisition of e-book materials. Study results, combined with knowledge of the local user community, suggest that e-book collections are particularly supportive of “just in case” collection development activities; patrons often turn to e-book materials to support current teaching and learning activities.

To expand on this idea, Collection Development recommends that subject specialists on campus consider at the time of purchase whether materials will be used for continuous (e.g., reading for extended periods of time, conducting in-depth research, exploring subjects in depth) or discontinuous (e.g., reference, citation confirmation, searching for keywords, skimming chapters) reading. The answer to this question will be a strong indicator regarding which format is most appropriate. Based on study results, it is recommended that print serve continuous reading needs and electronic serve discontinuous reading needs. The results were consistent across the major disciplines observed during this study (i.e., humanities, social science, science, and fine arts).

Over the coming years, there will be greater pressure placed on available financial resources as information is made available through a growing variety of formats and services. It is not expected that the

budget will grow alongside demand for these services. As a result, Collection Development is establishing an assessment framework that promotes continuous evaluation of collection materials. This is particularly relevant to new and current subscriptions. It is recommended on campus that subscriptions be evaluated every five years to monitor general usage trends, duplication, annual costs, and quality of content. As research interests shift and courses develop (particularly due to the growth of online programs), usage statistics will be systematically collected to ensure that the collection budget is invested in resources that support the growth and development of research, teaching, and learning activities.

Defining a Future Path: Collection Development Goals

The overarching theme that has emerged in my research over the past several years is intent of e-book use. The findings suggest that collection development activities, ranging from the selection of business models and formats to the negotiation of license agreements and the development of preservation strategies, hinge on one central question: what is the intent of use?

In reality, no e-book solution simultaneously meets both the “current use” and “future use” requirements.¹ In some cases, it makes economic sense for libraries to purchase titles in electronic formats without consideration for long-term access. In other cases, it is appropriate to purchase materials for preservation purposes despite levels of current user demand.² Today, it seems that the success of collection development initiatives relies on a balance between “just

in case” and “just in time” strategies. As a result, it is imperative for librarians to have a comprehensive understanding of how patrons intend to use specific e-book packages, subscriptions, and even titles. With this knowledge, librarians gain the insight required to determine when it makes economic sense to invest resources in high-use materials for current users and when it is appropriate to purchase materials that may have low use but add to the long-term value and legacy of the collection.

Based on the results of quantitative investigations, the following goals and recommendations support activities related to resource allocation and collection management activities at CUL:

1. Allocate the materials budget and perform selection in a systematic manner that maximizes coverage, minimizes gaps, and avoids unnecessary duplication.
2. Anticipate and respond to users’ needs based on evidence surrounding usage trends.
3. Consider the complete life cycle of e-books at the time of purchase.
4. Continually evaluate collections and monitor research trends on campus.

To achieve these goals, I am experimenting with methods to map intent of e-book use to collection depth indicators. Evidence from quantitative analysis projects has consistently suggested that e-books support discontinuous reading behaviors (e.g., power searches, citation verification) as well as teaching and learning activities. At this time, the following recommendations have been made to align selection and acquisition activities with collection depth indicators:

1. **Basic Collection: E-books Recommended**
Supports lower-division undergraduate research; includes the core of the discipline or subdiscipline as it relates to the curriculum. This level describes materials that serve to introduce and define subjects, including selected databases, fundamental materials, introductory works, historical surveys, and reference works. ILL is expected to augment the collection.
2. **Extensive Collection: E-books Recommended**
Supports graduate course work; information is adequate to maintain knowledge of a subject required at less than research intensity. Examples of content include primary and critical resources, reference resources, specialized databases, and bibliographical resources. ILL is expected to augment the collection.
3. **Research Collection: Print Recommended**
Supports research leading to a doctorate, faculty research, or independent study. It includes resources supporting the framework for the

methodology and implementation of original doctoral research. ILL is expected to augment the collection.

Criteria for Acquiring E-book Materials

Based on the body of evidence generated through quantitative analysis projects, below are the recommended criteria at CUL that guide decisions to purchase e-book packages and titles:

1. Intent of use
 - a. Research: Print format recommended
 - b. Teaching and learning: Electronic format recommended
2. Intended audience
 - a. Basic collection: Electronic format recommended
 - b. Extensive collection: Electronic format recommended
 - c. Research collection: Print format recommended
3. Access at least equivalent to print
 - a. Format preference: PDF (supports both continuous and discontinuous reading behaviors)
4. Sufficient access (e.g., multiple user access for course reserves)
5. Cost history/cost per use
6. Availability of usage statistics for collection evaluation purposes
7. Relevance of the content to the full collection (e.g., fill existing gaps, align with collection priorities)
8. Degree of overlap
9. Courses and programs supported
10. Price and price comparison with competitors

Disseminating Quantitative Analysis Results to Stakeholders

Because of the many stakeholder groups that make use of e-book collections, any analysis project requires strong communication channels with the individuals who will implement or be impacted by change. Based on my experiences over the past several years, I put together a workflow that I use to disseminate quantitative analysis results and provide opportunities for feedback:

1. Alert Collection Development, division directors, and subject specialists when an offer, subscription, or package is flagged for evaluation.
2. Present project goals and objectives to relevant subject selectors for feedback.
3. Work with subject selectors to coordinate feedback from the user community regarding a specific subscription, package, or offer.

4. Compile results related to goals and objectives into a brief report that can be distributed to internal stakeholders.
 - a. Present report to relevant subject selectors for feedback.
 - b. Present report to relevant division directors.
 - c. Present report to the Director of Collection Development.
 5. After internal stakeholders have provided feedback, post results in the Academic Common and distribute link to external stakeholders for review.
 6. Present findings to publisher, vendor, or aggregator.
2. Increasingly, e-books are made available as PDF files that are sold by publishers or authors. How will libraries acquire electronic texts that are not made available through a host platform? How will this material be made widely available to patrons?
 3. How will open-access initiatives change the way e-books are acquired, discovered, and accessed in the future? What types of financial commitments and collection development activities will individual libraries take on?

Despite these questions, I feel optimistic that continued discussions with stakeholders, information managers, and publishers and vendors will lead to solutions that provide evidence and allow all parties involved to better understand how e-books serve patron communities.

Unresolved Areas

It is important to note that e-books are still an evolving format; libraries and content providers are grappling with the development of sustainable acquisition and discovery models and will continue to do so for the foreseeable future. There are currently several challenges that are under investigation at CUL, and further experimentation will be required before standardized assessment models are developed:

1. How will librarians acquire e-books that they currently do not have the technical capacity or legal rights to do so?

Notes

1. Greg Eow, Todd Gilman, Jill Jascha, Caitlyn Lam, Melanie Maksin, Scott Matheson, Colin McCaffrey, and Nathan Rupp, *The eBook Strategic Plan Task Force: Report of Findings and Recommendations* (New Haven, CT: Yale University Library, March 1, 2013), www.library.yale.edu/departments/collection-development/Yale-ebook-task-force-rpt.pdf.
2. Ibid.

Conclusion

Over the past decade, a growing interest in electronic content has resulted in a greater emphasis on e-book collection development, which has required evaluations of new business models, license agreements, and collection assessment methods. In this digital environment, information needs surpass available resources, and librarians are required to justify purchases or requests for budget increases with quantitative evidence. Now more than ever, it is essential for librarians to demonstrate data-based collection development decisions or evaluate current holdings to identify areas where resources can be shifted to support the teaching, learning, and research needs of a user community.

The evolving nature of electronic resources, particularly in regard to e-books, provides challenges in developing standardized methods of conducting quantitative analysis. The abilities to calculate cost per use, identify usage trends, document how funds are allocated to acquire materials, and provide evidence for collection development decisions are essential components to developing e-book collections that address “just in case” and “just in time” information needs. However, training opportunities in this area of library work are still in development and not always widely available. By sharing methods currently used within the professional community, opportunities for experimentation, feedback, and standardization become available.

This issue of *Library Technology Reports* demonstrates the steps that I took to develop quantitative analysis skill sets and an evaluation framework for e-book collections at Columbia University Libraries.

The methods that I developed and used are based around local and external data sources that are widely available to all libraries. All the examples of quantitative research discussed in this report were conducted using Microsoft Excel, and projects were completed through independent work or collaboration with small teams. Based on the data sources and resources that supported my work, I am confident that the methodologies described can be implemented at most libraries, regardless of size or staff or equipment budget.

The reality that the e-book landscape is constantly evolving was factored into decisions regarding the overarching analysis framework guiding described studies. Research designs were created so that they can be replicated regardless of how e-books evolve in the coming years. Because the designs are flexible and adaptive in nature, they promote continued assessment, evaluation, and strategic planning as a regular component of e-book programs.

The methods discussed in this report are not meant to be an end point but rather a springboard for future analysis activities and, ultimately, development of standardized practices across the information profession. I am excited to find out what the future holds for e-book collection development practices and to learn more about the quantitative analysis methods that are developed and implemented at libraries across the country.

Thank you for taking this journey through quantitative analysis methods with me. I look forward to growing with and learning from colleagues across the information profession in the years to come.

Library Technology

R E P O R T S

Upcoming Issues

July 53:5	3-D Printers for Libraries, 2017 Edition by Jason Griffey
August/ September 53:6	Open Source Platforms by Marshall Breeding
October 53:7	Open Virtual Libraries by Mirela Roncevic

Subscribe

alatechsource.org/subscribe

Purchase single copies in the ALA Store

alastore.ala.org



alatechsource.org

ALA TechSource, a unit of the publishing department of the American Library Association