

1-9-2018

The Secret Lives of Ebooks: A Paratextual Analysis Illuminates a Veil of Usage Statistics

Jonathan H. Harwell
Rollins College, jharwell@rollins.edu

Erin Gallagher
Reed College, gallaghere@reed.edu

Follow this and additional works at: http://scholarship.rollins.edu/as_facpub

 Part of the [Collection Development and Management Commons](#), [Scholarly Communication Commons](#), and the [Scholarly Publishing Commons](#)

Published In

Harwell, Jonathan H. and Gallagher, Erin, "The Secret Lives of Ebooks: A Paratextual Analysis Illuminates a Veil of Usage Statistics" (2018). *Faculty Publications*. 187.
http://scholarship.rollins.edu/as_facpub/187

This Article is brought to you for free and open access by Rollins Scholarship Online. It has been accepted for inclusion in Faculty Publications by an authorized administrator of Rollins Scholarship Online. For more information, please contact rwalton@rollins.edu.

This study is a grounded approach within the frame of library and information studies, in which the authors, two academic librarians, initially set out to use thick description to analyze paratextual elements of specific copies of ebooks as artifacts within an academic library collection, including access, discoverability, and administrative metadata such as usage statistics, pricing structure, and digital rights; and in light of the variation in usage statistics from different vendors, refocused the study on the problematic aspects of comparing that data. The ebooks have been selected on the basis of high usage by the library patrons of Rollins College, a small, independent liberal arts college in Florida with a graduate business school. The six titles represent the two most used books from each of three platforms. The approach is constructed from the data analysis. The initial aim of the study was to present and analyze various elements of paratext as a set of case studies of ebooks, allowing for the creation of a “biography” for each ebook. However, upon further research, we realized that it would be unwise to do so without addressing the complexities of comparing usage statistics of library ebooks across vendor platforms. Those complexities arose as the main point of the study, overshadowing the original intent and making paratext only a secondary consideration. Thus we include some descriptive notes, rather than an in-depth paratextual analysis.

The hardware and software associated with the study are different from those in most studies of ebook paratext, which focus on e-reader devices like the Kindle, Nook or Kobo, most of which are associated with e-reading in the retail market. Retail e-reading hardware can be and is used frequently in academic library settings, and many libraries purchase e-readers that patrons can use to download ebooks provided by the library. However, there are significant differences in ebook acquisition and usage between retail and academia. For example, Kindle is both a

device and a platform. In contrast, none of the content providers we discuss in this article maintain their own proprietary devices, only e-reading platforms. Purchasing, licensing, and pricing models differ greatly between academia and the retail world, with libraries demanding more flexible options to meet the particular preferences of their diverse user groups, and often paying more than retail customers for the same titles.

The high-use titles are drawn from usage statistics from our three major library ebook vendors: ebrary, EBSCO, and Springer. We focus on these three providers because their content represents the bulk of our ebook holdings, although academic libraries have a large universe of ebook providers among which to choose. These library vendors' interfaces, although they each have distinct designs, tools, and digital rights management, all use a standard browser (e.g., Chrome, Firefox, Safari, Internet Explorer, and the associated mobile applications) as the basic platform. The ebooks may be downloaded to the desktop or to a variety of e-reader devices; however the easiest way to access the text of ebooks (as well as articles or streaming audio/video) is online via the browser. The secondary platform is the technological format of the book as presented in the browsing interface. In terms of content, Rowberry notes that some ebooks are presented without front or back matter; some chapters have been renumbered; and the form of each page mimics that of the print format (10).

The service infrastructure of library ebooks from vendors such as ebrary, EBSCO, and Springer is complex and invisible to most readers other than library employees and vendor representatives. In terms of external integration, in our case the ebooks are discoverable by title, author, keyword, etc., in the Primo discovery interface developed by Ex Libris, owned by ProQuest (who also owns ebrary and is in the process of blending it with EBL to form the ProQuest Ebook Central interface). When a librarian acquires a book from ebrary or EBSCO,

for example, they might use an ordering platform provided by that company, or an aggregated platform provided by a distributor. In our case, we use the interface provided by the distributor GOBI Library Solutions (formerly known as YBP or Yankee Book Peddler, owned by EBSCO, a major competitor of ProQuest) and which also distributes books from ebrary and other competing vendors. GOBI displays the pricing and licensing options based on print and/or electronic format, and within the latter, the tiered costs of providing the ebook to single or multiple simultaneous users. These price differences and licensing options vary widely, and are the result of negotiations between the aggregator (ebrary or EBSCO) and the publisher. Meanwhile Springer sells some ebooks through ebrary and/or EBSCO, while their other ebooks are only sold to libraries in packages directly from Springer.

GOBI also indicates which ebooks are available to purchase through DDA (demand-driven acquisition, also known as PDA, or patron-driven acquisition). DDA enables librarians to add ebooks to their local collections without being invoiced unless those books are then used significantly (triggered) by the local patrons. Trigger levels vary across aggregators and are determined through negotiations with publishers. For example, EBSCO ebooks trigger purchase invoices when an individual title is viewed for ten minutes, or across ten pages (not including end matter), or when information from the title is copied, printed, or saved. ProQuest Ebook Central, the emerging successor to Ebrary and EBL, shortens the time trigger to five minutes and does not count page turns. These ebooks may be delivered automatically as a weekly batch of new titles that fit the library's predetermined criteria (similar to an approval plan), and additional books may be manually added to the weekly batch by a library employee (manual DDA). This stretches the library's budget further by adding content for potential readership without immediate costs, with many titles never triggered for invoicing by actual use. DDA is invisible

to the patron; the ebooks appear in the library's discovery or catalog interface alongside those that have already been purchased. The only difference is the invoicing workflow.

Springer ebook service infrastructure differs from ebrary and EBSCO. Springer publishes books and sells them individually in print format, but only sells ebooks to libraries in packages of multiple titles on their SpringerLink platform (and also sells PDF versions for individuals to download). This means that librarians may not purchase an individual Springer title for the collection unless it is included in an aggregate ebook platform such as ebrary or EBSCO. In one recent case at our library, a faculty member needed a specific psychology book published a decade ago by Springer. In order to acquire this one ebook, we had to purchase a package of 10-year-old psychology books at a cost that might be considered a bargain, had we needed all of them. Since some psychology books are considered as candidates for weeding from academic libraries after ten years, this is not the library's ideal service infrastructure. Because of the publisher's restrictive sales model, SpringerLink ebooks may not be integrated into a library's DDA plan for purchasing individual titles, and thus are perhaps underrepresented in the library collection.

Despite their restrictions on per-title purchasing, Springer is known for a liberal model of digital rights for ebooks licensed by libraries. Unlike other major ebook vendors, Springer enables interlibrary loan, and has collaborated with librarians on piloting the Occam's Reader project that was developed to facilitate the interlibrary loan of ebooks (Litsey and Ketner). They also provide DRM-free downloads in PDF format, and there are no limits on simultaneous users. It is common for individual publishers to offer less restrictive DRM (digital rights management) than aggregators, as they control the production, pricing, and delivery of the content. This also

provides an incentive for libraries to work directly with publishers when acquiring ebooks.

While advantageous from a DRM point of view, purchasing and maintaining access to ebooks on a variety of disparate publisher and aggregator platforms can lead to more complex workflows for the library and a thorny experience for users who must navigate multiple platforms with varying features.

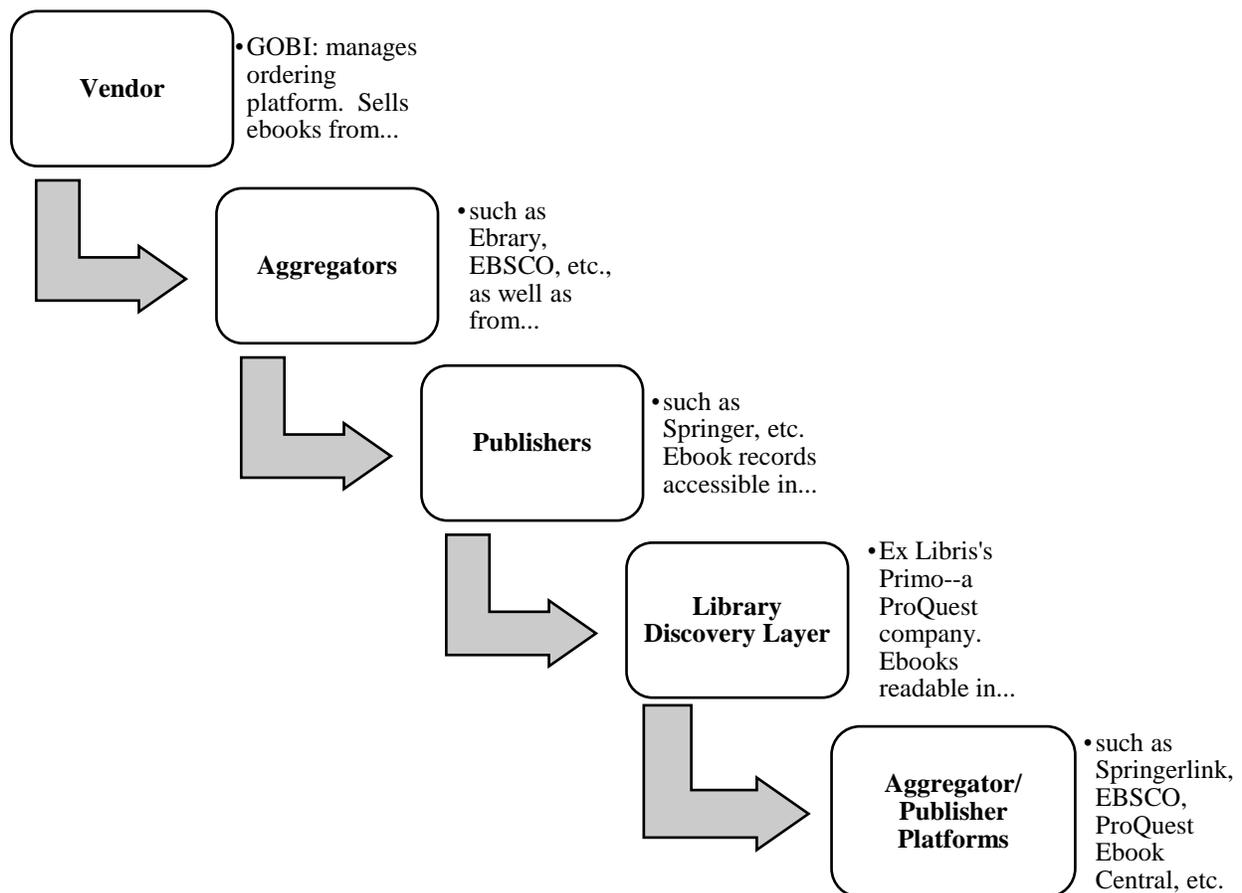


Figure 1. Service infrastructure

Literature Review

The study of the history of the book incorporates disparate frames of analysis, such as literary theory (e.g., Gerard Genette's seminal *Paratexts: Thresholds of Interpretation*, Cambridge

University Press, 1997), descriptive bibliography, media criticism, communication history, book arts, and ethnography of information. The book as artifact has evolved across distinct formats, including wax tablets, the manuscript codex, incunabula, letterpress, microforms, audiobooks, and ebooks.

The paratextual elements of books are those that describe its life and history beyond (though including) the printed page. Noting that “paratextual analysis of e-books designed for immersive reading is...virtually unexplored territory,” Birke and Christ describe the paratextual elements of ebooks as displayed on iPad and Kindle devices, but only in general terms rather than specific titles. McCracken does similar paratextual analysis of iPad and Kindle e-reading platforms. Northcutt Malone compares the Kindle interface to the paratext of the codex, in terms of orientation and readability. Rowberry provides a holistic framework for examining the paratext of the ebook, with two specific books as examples. He draws from platform studies to define a model with three layers: technology (hardware and software), text (format and content), and service infrastructure (internal tools and external integration) (3-4). Galey examines the theoretical underpinnings of the history of the book as artifact and how they apply to ebooks, with a specific ebook used as a case study. Smyth updates Genette’s concepts in *Paratexts* for application to electronic books, including audiobooks. In so doing, Smyth helpfully explains peritext (elements present in the physical construction of the book) and epitext (elements that expand beyond the physical space of the book) and demonstrates how they are manifest in the case of ebooks, including Amazon customer reviews.

Methodology

We gathered usage statistics from the complete calendar years of 2014 and 2015, seeking to identify the most used ebooks in that time frame. We have limited the study to three

platforms: ebrary, EBSCO eBooks, and SpringerLink. These three are the primary vendors of ebooks in our library collection that provide title-level usage statistics. Open access ebooks, as well as books from other vendor platforms, are beyond the scope of this study. Analyzing open access ebooks would be of interest; however generally speaking, their platforms do not yet report usage by library patrons, though some do provide aggregate usage statistics. As we are primarily interested in each platform’s most used ebooks by our library community, open access books are omitted from the current study. Our library no longer provides Kindle or Nook ebooks or devices due to low usage; those are omitted here as well.

COUNTER is a non-profit organization made up of librarians and content providers. They provide the Code of Practice, a set of industry standards that allow for consistency in reporting library usage statistics across vendor and publisher platforms. Vendors provide various combinations of BR (book report) and JR (journal report) data. For the purposes of describing the usage of books across three platforms, we focus on the standard usage reports that are common to all three: the COUNTER BR2 and BR3. BR2 contains the number of successful section (or chapter) requests by month and book title, whereas BR3 reports the number of turnaways by month and book title (books that our patrons are unsuccessful in accessing). Both are significant to our study, as there is the possibility that some ebooks we haven’t licensed (or that restrict access by multiple simultaneous users) are in higher demand than those we have.

COUNTER Book Reports	What is reported?	What does this show?
BR1	Number of successful title requests by month and title	Number of times users requested ebooks by title and month. Requests include viewing, downloading, emailing, and

		printing. Useful if the ebook is accessible as a single file.
BR2	Number of successful section requests by month and title	Number of times users requested sections of ebooks by section title and month. Sections usually mean chapters. Useful if the ebook is subdivided into individual chapter files.
BR3	Access denied to content items by month, title, and category	Also known as “Turnaways”. Number of times users attempted to access ebook content by title but were denied because the institution does not have access rights. Categories include simultaneous user limits and complete lack of access.
BR4	Access denied to content items by month, platform, and category	Same as BR3, but instead of reporting turnaways by Title, reports turnaways by publisher or vendor platform.
BR5	Total searches by month and title	Number of times users searched for ebooks within a vendor or publisher platform by title. Useful if searches can be counted at the title level, not at the platform level.

Figure 2. COUNTER book reports (Project Counter)

Titles analyzed in this study represent the two ebooks with the highest BR2 and BR3 usage statistics for our library from each of the three platforms. If an ebook is a multivolume work, even with unique ISBN’s per volume, this study treats it as a single ebook. For example, if volume one has 300 uses and volume two has 200, the book as a whole has 500. Some ebooks are published annually in a series, such as annual conference proceedings. In those cases, each annual installment is treated as a distinct ebook.

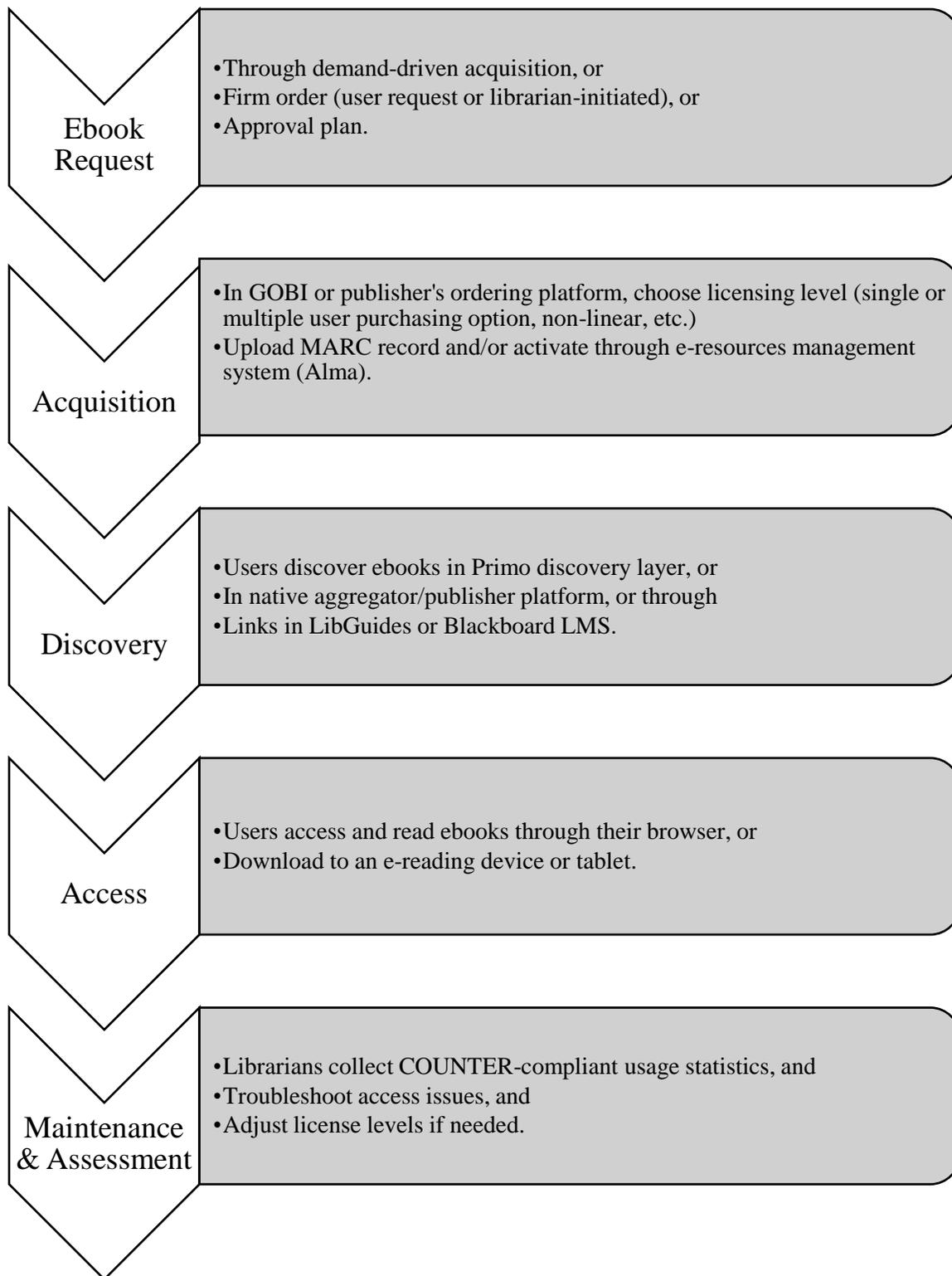


Figure 3. Ebook lifecycle chronology

Title Selection and Analysis

We will begin with ebrary. Our users have access to approximately 4,261 ebrary ebooks through our collection. The most used ebook across all three platforms, based on the common BR2 report, is a 2007 graphic memoir, *Fun Home: A Family Tragicomic*, by Alison Bechdel, 241 pages (seven titled chapters) published by Mariner Books (a paperback division of Houghton Mifflin Harcourt). This book about the lesbian cartoonist's childhood and family life was adopted as a required text for a Fall 2012 course at Rollins entitled "Bad Boys and Good Girls: Coming-of-Age in American Literature and Culture." The ebook was selected by a librarian for purchase from ebrary in March 2014, and accessed only three times during that year (all in August), before readership suddenly increased. *Fun Home* had been adapted into a Broadway musical that opened in October 2013. In June 2015 it swept the Tony Awards, winning five, including best musical. The book then became a local overnight hit on ebrary, from a steady zero usage through the preceding months to a total of 1005 (and 9 searches of the text) in June-November 2015. Since Fall 2015 it has been required reading for a Rollins course on the graphic novel. Remarkably, although this book is only available as a single-user title, it only had one turnaway, in October 2015.

Ebrary allows this book to be downloaded, only two pages to be printed, and no portion to be copied. *Fun Home* is listed at \$14.95 from ebrary for single users. It is available as a DDA title from ebrary; however, our library began DDA in December 2014, so that was not an option when we selected this book. In terms of cost per use for the years studied, we paid about a penny for each time a chapter was accessed, and the book is still being used as a required text.

The second highest usage was for Lin Noueihed and Alex Warren's co-authored, 371-page *Battle for the Arab Spring: Revolution, Counter-Revolution and the Making of a New Era*, published by Yale University Press in 2012. We purchased this book in February 2013 at the recommendation of a political science professor who had told us, "I personally prefer books in print format. However, I understand there may be cases where there are significant savings from electronic format." The book was purchased on the ebrary platform as a single-user title. Two years later, in February 2015, it was upgraded to a multiple (unlimited simultaneous) user version, which happened automatically when two patrons accessed it during the same time. The cost for a single-user license of this ebook is \$39.20, though we paid \$58.80 including the upgrade.

A scholarly history of the Arab Spring movement of 2011, this book was written by an editor from Reuters (Noueihed) and the director of Frontier, a privately owned research and advisory consulting firm (Warren). In Spring 2013 and Spring 2014, the book was used as a required text for a Rollins course, a seminar in international affairs, entitled "The Arab Democratic Spring of 2011." This course was in the Hamilton Holt School, a division of Rollins that primarily serves nontraditional students with evening and weekend classes. The professor was not the same one who had requested the book for the library collection, and was in a different department.

Ebrary allows patrons to download this book, to copy 55 pages, and/or to print 111 pages. The usage for this title during 2014-15 shows exactly 1000 instances of chapters being accessed (and two searches run within the book), and only during the months of January and February 2015. There was zero usage in all months before and after that. When we noticed the instant disappearance of usage for our second most used ebook, we investigated and found that this

ebook was not activated in the new unified resource management system to which we migrated in Spring 2015 (the migration was from a combination of Sirsi Symphony and ProQuest Summon, to Ex Libris Alma and Primo). We remedied the problem and in the process identified additional books with the same issue. The book had not been listed in our new Primo discovery system, although it remained available in the native ebrary interface. This instance suggests that the library's ebook was not used during the Spring 2014 section of the course that required the text; and it seems to indicate that no one went to the ebrary interface (which is included in our library's alphabetical list of databases) to locate the book following our system migration, instead relying on the library's discovery service until the ebook disappeared. The scenario also suggests that no one saved a permanent ebook link and used it later, and certainly no one notified us of the problem. While the ebook was still accessible, it was not as discoverable as it had been. It is plausible that this title's usage could have far exceeded that of *Fun Home* if it had not been affected by the problem in system migration. As it happened, however, the usage statistics are nearly equal for both books. Based on this measurement, the cost per chapter accessed during January-February 2015 was about six cents. This case illustrates how our research has informed and improved our practice of librarianship, by illuminating an access problem with some of our ebooks-- a problem we were then able to resolve.

Here we must reiterate that rather than describing the top six titles across all platforms, we are instead analyzing the top two titles (by usage) from each of the three platforms: ebrary, SpringerLink, and EBSCO. If we were to select the most used titles across the board, there would be no EBSCO books in this set; and as we will discuss, comparing usage across platforms is not a one-to-one analysis. As we began with ebrary, we turn next to SpringerLink. Our users have access to approximately 15,507 Springer ebooks in our collection.

The most used ebook on the SpringerLink platform is a book of refereed conference proceedings entitled *Computer Vision* from ECCV 2006, the 9th European Conference on Computer Vision in Graz, Austria, May 7-13, 2006, and published by Springer the same year, as part of a series entitled Lecture Notes in Computer Science which began in 1973. The three editors of this set of proceedings are Aleš Leonardis (University of Ljubljana, Slovenia), Horst Bischof, and Axel Pinz (both of Graz University of Technology, Austria). The overall topics are computer vision and pattern recognition systems.

There are four parts, or volumes, of the 2006 proceedings available via our library's discovery service. The full texts of all four parts are linked from a single entry; and are also listed separately as distinct items. A search for "Rollins" as a keyword finds no results, so there is no indication that a member of the Rollins community presented at that conference. Our library also provides access to the ECCV conference proceedings from 2002 and 2008, as well as some workshop proceedings associated with ECCV and ICCV between those years, also online via SpringerLink. The electronic versions of these titles are only sold to libraries in SpringerLink packages. Those are priced at \$119 each for the first three and \$109 for Part IV.

The most used volumes during 2014-15 were the second and third, with 245 chapter views in each. Part One had 192 chapter views, and Part Four had 184. All usage of the four-part set was during February 2015. No other ECCV or ICCV sets were used during those years. The total number of chapter views is 866.

The market for this publication is specialized and aimed at experts in such areas as image processing and computer vision, pattern recognition, computer graphics, and artificial intelligence (including robotics). These topics are listed on the publisher's webpage for the

book; in fact, all four parts or volumes have separate webpages, each offering a free two-page preview of a paper from the conference, as well as the option to “reserve an online book review copy.”

Next in high usage is another set of SpringerLink conference proceedings, also from the Lecture Notes in Computer Science series. This one is ICANNGA (International Conference on Adaptive and Natural Computing Algorithms) from 2007, published in the same year by Springer in two parts. The editors are Bartłomiej Beliczynski, Andrzej Dzielinski, Marcin Iwanowski, and Bernardete Ribeiro. The first three are affiliated with Warsaw University of Technology in Poland, and Ribeiro is with the University of Coimbra in Portugal. Computational intelligence is the overall topic. There is no mention of Rollins in the proceedings. Our library provides access to the 2009 and 2011 conferences as well, also via SpringerLink. The 2007 proceedings are listed for \$179 per volume in paperback from GOBI. The electronic versions are only sold to libraries in packages.

Part I of the 2007 set received 376 chapter views, and Part II had 336, for a total of 712. In addition, the 2009 proceedings (a single volume) had 63 chapter views; Part I of 2011 had 43; and Part II of 2011 had 41. All were during February 2015, the same month that the ECCV proceedings were used.

According to the publisher’s webpage for each volume, this publication is marketed to experts in artificial intelligence (including robotics), computation by abstract devices, algorithm analysis and problem complexity, programming techniques, software engineering, and image processing and computer vision.

Our library had acquired both Springer titles as part of the Computer Science Subject Collection. Based on the number of books in that collection, we paid 72 cents for each volume.

This totals \$2.88 for the four-part ECCV 2006, and \$1.44 for the two-volume ICANNGA 2007. Cost per use for the two years included in the study is less than one cent per chapter view, for both titles.

The remaining two titles to analyze are from the EBSCO platform. Our users have access to approximately 23,218 EBSCO ebooks in our collection. The usage statistics for EBSCO ebooks are much lower than those on the other platforms. We noticed that the report of book chapter views (COUNTER BR2) that we received from EBSCO lists an assortment of articles from databases, which are not also chapters contained in ebooks. We followed up with EBSCO and received a response that this is a “high priority service issue”. We also received a BR1 report with usage by book title. (BR1 measures successful book title requests, while BR2 represents successful book section requests.) So we are able to identify the most used books, but in a different way. The method of reporting might help to explain the low usage, since we have not received BR1 reports from ebrary or Springer.

According to the BR1 report, the most used title from EBSCO during 2014-15 was *Hannah's Child: A Theologian's Memoir* by Stanley Hauerwas, published in 2012 by Wm. B. Eerdmans Publishing Co. EBSCO is the only ebook platform listed for this book by GOBI. There is no DDA option. We purchased the unlimited version at \$40 in November 2013, following selection by a librarian. A cost per use calculation for 2014-15 indicates that we paid about 75 cents for each time the book was accessed.

This memoir of Hauerwas' life and development as a Christian theologian spans 307 pages. It was accessed 53 times, all in March 2015. We have no evidence of the book being used as a Rollins course assignment. However another book co-authored by Hauerwas and William H. Willimon, *Resident Aliens*, was required in 2008 for a Rollins course on religion and

politics in America. Along with clergy, the audience for the book includes the scholarly community in theological and religious studies, as well as a broader readership of religious memoirs.

The second most used book from EBSCO happens to be one that we had used in testing when we experienced problems with EBSCO's ebook platform in June 2015. All but one of the uses were during that month. The book is *Holy Feast and Holy Fast: The Religious Significance of Food to Medieval Women*, by Caroline Walker Bynum (University of California Press, 1987). Since most or all of the usage for this title is artificial and does not reflect actual readership, we will instead analyze the third most used book.

The next EBSCO book is *Chocolate: History, Culture, and Heritage*, a 975-page (1040 including front and back matter) scholarly collection of 57 chapters by 114 authors (none from Rollins), divided into eleven thematic parts, on the history of chocolate, edited by Louis Evan Grivetti and Howard-Yana Shapiro and published by Wiley in 2009. Both editors are affiliated with the Chocolate History Group created by the University of California, Davis, and Mars, Inc. Shapiro is the Global Director of Plant Science and External Research at Mars, "the largest chocolate company in the world," according to the publisher's webpage for the book.

We purchased the three-user version from EBSCO at \$167.93. A faculty member in chemistry had selected this book for the collection, to support a course on the science and culture of chocolate, an undergraduate course that is still being taught. We purchased the print version in March 2014. Having found it easier to provide readings to students with an ebook linked from Blackboard, the same faculty member asked us to add the electronic version as well. We purchased that one in July 2014.

EBSCO allows up to 60 pages to be printed or saved from this book; copying and pasting are not permitted. The ebook was accessed 39 times during 2014-15. We received it on July 29, 2014. There were two uses that August, six in September, two in November, ten in December; and in 2015, one in June, nine in July, four in September, and five in October. In this instance we can also report the circulation of the print copy, although print and electronic usage can not be compared as equal measures. We received it on April 18, 2014; it was checked out once to a faculty member on June 17, 2014, was repeatedly renewed, and remains checked out. This single use illustrates one reason for not treating print and electronic usage as equivalent. If we calculate cost per use for the print copy, it is simply the purchase price, \$111.95 for the single circulation. In comparison, the cost per use for the ebook (which cost substantially more than the print) during 2014-15 is only \$4.46.

[Insert Table 1.]

Discussion

This study uses six electronic books as a set of case studies. The intended approach of paratextual analysis arises from bibliographic research methods used in the study of the history of the book, in order to model an approach for library ebooks (following Rowberry). The purpose is not to generalize and predict high usage of the same titles in other libraries, but to examine the usage and histories of these books at a single liberal arts college. As we see, the resulting usage analysis has illuminated ways in which ebook usage statistics can be misleading if taken at face value and compared across vendors. This aspect is an unforeseen product of the research and points to the need for further study and action in the library ebook industry.

We have analyzed the two most used books in each of the three main ebook platforms used in this library: ebrary, EBSCO, and SpringerLink. We started with the standard usage reports available to us through vendor and publisher statistics gathering tools, and then followed up with each of the three ebook content providers in search of thicker usage data to augment COUNTER reports. We asked each aggregator/publisher for the following:

- Which specific chapters/sections/pages did users most access?
- How many pages were viewed on average?
- Did users highlight/annotate/save to folder? If so, on which pages?
- Which devices were used?
- What times of day/days of the week did most uses occur?
- How long (minutes, hours, etc.) did users spend in the book?

Even with standard reporting of usage statistics, identifying the most used books can be a complex question, and one in which not all “uses” are equal. All three vendors provide the ability to generate usage reports according to the COUNTER standard, but each delivers a different combination of these reports (BR1, BR2 and/or BR3, BR5) and all offer insight into the multifaceted and sometimes convoluted world of ebook usage reporting.

Ebrary

Ebrary presented us with a case in which short-term, situational access led to unexpected high use. As previously mentioned, our second most-used Ebrary title, *Battle for the Arab Spring: Revolution, Counter-Revolution and the Making of a New Era*, was not discoverable in our discovery layer for much of its lifespan. Our study spans the calendar years 2014 and 2015, and this title was only listed in our discovery layer for a limited portion of that time period. In this case, long-term, sustained discoverability did not correlate to high usage numbers.

Ebrary provides both COUNTER reports and three proprietary reports accessible through their administrative portal. These proprietary reports list usage elements that go beyond COUNTER standards, including:

- Title Report
 - Provides detailed usage for each title in our holdings, including views, copies, prints, etc. along with some document metadata.
- Category Summary Report
 - Summarizes the activity in our holdings grouped by subject, including number of user sessions, unique documents, pages viewed, pages printed, and pages copied.
- Site Activity Report
 - Provides a quick snapshot of the activity in our holdings, including number of user sessions, unique documents, pages viewed, pages printed, pages copied, and online turnaways.

We understand that the current reports, including proprietary, do not offer the additional data elements we would like to analyze, though the reports available through ProQuest's Ebook Central will offer more granularity. For example, reports through Ebook Central provide title-level usage down to the minute and second, free viewing time summary and detail, device-specific information, and average viewing time for a specific timeframe, among others.

EBSCO

In the case of our second-most used EBSCO ebook, *Holy Feast and Holy Fast: The Religious Significance of Food to Medieval Women*, we found that nearly all usage occurred during the one month our librarians had to troubleshoot downloading problems with that particular interface. This one represents high “usage” according to the statistical reports, but this

type of usage does not signify readership, nor would it tell a reliable story of usage to anyone not familiar with our library's history of troubleshooting technical issues.

EBSCO's Product Management team generously responded to our request for more granular data on our top-used titles:

1. How many pages were viewed on average?

Hannah's Child had 781 page turns (in 54 accesses) in March of 2015 and 17 page turns (in one access) on 3/22/16

Chocolate had 93 page turns on 3/23/16

Please note that page turn statistics only reflect page turns completed within the eBook Viewer. Any page turns completed after the user has downloaded the eBook are not trackable by our systems and will not be included in the page turn count. In this case, users did not download these books.

2. Did users highlight/annotate/save to folder? If so, on which pages?

Users only viewed the eBook online for these two books, they did not print/email/save/export/download/sample.

3. Which devices were used?

Device information was not available but users accessed using Windows OS.

4. What times of day/days of the week did most uses occur?

The only usage was 3/22 and 3/23, Tuesday and Wednesday. The March 2015 usage was from a different system so we don't have the days of the week logged.

5. How long (minutes, hours, etc.) did users spend in the book?

We cannot determine the exact value of this for these books, but I can tell you that the average session duration in March was between 30-40 minutes. This includes all time

spent browsing EBSCO and does not necessarily limit to browsing the eBook. We are unable to limit to time spent just browsing an eBook/eBooks.

A noteworthy outcome of this study was the discovery that EBSCO's COUNTER BR2 reports did not include any content from EBSCO eBooks, only from other EBSCOhost databases. It should be noted that BR2 reports are not required by the COUNTER Code of Practice if BR1 reports are supplied. Either report is acceptable, depending on whether books are delivered as single files or chapter files. Further, BR2 reports of "section requests" may have varying definitions, as Thomas, et al., explain: "eLibrary totals all successful page views, prints, and copies for its BR2 report" (2013), not just chapter views. The problem here is that a BR2 report is supplied by EBSCO, but the data is not measuring the usage of EBSCO eBooks, but other EBSCOhost databases that primarily consist of articles. Thus the report can be misleading.

Since we cannot compare BR2 reports from EBSCO to those from other vendors, and because "section requests" are defined differently by various vendors, our selection of the top two titles from each must not be interpreted as comparing equivalent usage data.

SpringerLink

Springer presented an intriguing mystery. Per the additional reports we received from our representatives, "a large number of titles were used a few hundred times." Since both of our top Springer titles were only used during February 2015, Springer responded that this was likely a product of a temporarily "hacked IP range", and thus the usage is not reliable as a representation of our library readership. We had no way of confirming this potential hack in our system, but we accepted this explanation as the most logical, considering that the level of usage was beyond what our user base could realistically achieve. If we omit those two months from our usage totals, the top two titles are *Essentials of Pain Management*, a 2011 textbook; and

Modelling, Computation and Optimization in Information Systems and Management Sciences, a set of conference proceedings from 2008. If we omit April as well, the top two titles are *Essentials of Pain Management* and *Sustainability Accounting and Reporting*, a book on environmental management accounting. These two perhaps accurately represent the top Springer ebook readership from our library during 2015.

In response to our request for thicker usage data, Springer informed us that “the features classically thought of as benefits of our content, being DRM-free, in this situation are playing against us”. Once an ebook title is downloaded (as a PDF or ePub version), there is no way to track granular usage elements. They are able to track users’ interactions with SpringerLink content through standard COUNTER BR reports, but “this is not representative of the complete hands-on use of an individual title’s use once it’s downloaded to a local device”. The ability to analyze additional data points would be ideal, but the lack of DRM and the ability to download Springer ebooks for offline reading are acceptable tradeoffs.

Conclusion

The initial aim of this study was to model the application of paratextual analysis, an established approach to the history of the book, to library ebooks. The study was to provide a model approach for descriptive bibliography of the paratext of selected ebooks from a particular library collection, with a similar but modernized methodology as that long applied to print books. As we have discussed, we instead found that it was necessary to refocus on a byproduct of this study, namely the realization that library ebook usage statistics are not calculated in the same way among vendors, even with industry standards (COUNTER) in place. This points to the need for further collaboration among librarians, publishers, and aggregators to ensure that the COUNTER standards are communicated and offered in a standard way. The purpose of this

study is not to draw conclusions about readership at our academic institution or at others, or to analyze the paratextual elements of ebooks in order to make recommendations to vendors and publishers (though there is much room for improvement). This is not a usability study or a platform comparison. The decision to analyze the top two most-used ebooks from our primary ebook content providers was about identifying a manageable set of texts, rather than drawing correlations between paratextual elements and high usage. Indeed, the latter would be impossible due to the unreliability of vendor-supplied usage statistics, which is a cautionary illustration for librarians. In fact, our top-used ebooks have little in common and present few shared patterns. We would have selected several more ebook examples and created a larger sample size if we had intended to unearth trends and commonalities.

Our analysis raises more questions than answers. Because one of the most-used books was not able to be found in our discovery layer for much of the time period of this study, we must question the correlation between access and use. Academic library professionals invest a lot of time, money, and human resources into maintaining access to content through our siloed discovery avenues; however, in this case, consistent library discovery was not crucial to experiencing a good return on investment. Also, “use” itself must be contextualized. Another much-used title made it into the top two due to usage by librarians troubleshooting platform issues. We, as the authors of this study, are aware of this phenomenon because we lived through it, but if we simply reported the numbers without context, readers could make untrue assumptions about the appeal of this ebook to our user community.

Many of the same complexities that surround print book usage arose in this study; e.g., the traditional components of determining print book usage (checkout, shelving, renewal), that fail to tell us much about the lives of these books when they are in use, are not present when

analyzing ebook usage. We can rely on COUNTER reports to determine which books are being “used” in the most basic manner. When combined with more detailed reports from our vendor partners, we get a slightly clearer picture of how users interacted with ebooks. However, the rich elements needed to portray the full lives and histories of these ebooks (often created when users download and subsequently annotate, highlight, and otherwise interact with PDF and/or ePub versions) are not trackable through current publisher or aggregator systems. We learn what we can through the available data, but raw usage tells us almost nothing about intent, context, and practice.

What began as a paratextual case study of top-used ebooks at a small, independent liberal arts college in Florida quickly morphed into an exercise in making sense of usage statistics. Even if our content providers were unable to provide the level of granular data we requested, the study is already paving the way for new possibilities. Librarians find ourselves interacting with the same sphere of vendor representatives, rarely getting “face time” with product developers or managers. As we sought to unearth the lives of this sample of ebooks, we found our vendors receptive to and intrigued by our suggestions, even going so far as to ask how we envision the formats of these reports. An unexpectedly positive byproduct of this study led to new opportunities for development partnerships with our vendors and publishers.

This study also illustrates the common phenomenon through which research can inform practice. When we realized that one of our top-used titles was not activated in our discovery service, we then worked to ensure our other ebook records were discoverable. In researching which subject collection one of our top-used ebooks was part of, we undertook a project to make sure the correct collections (and years of these collections) were activated in our electronic resources management system (ERMS).

Suggestions for Future Research and Action

As we have seen, ebook usage is not equal across vendor platforms. This presents limitations for statistical studies. Deeper analyses of ebook usage might benefit from ethnographic research to complement the quantitative data. As one reviewer suggested, interviews, research journaling, or usage tracking software could be used to describe the readership of each ebook. How do people discover it, why do they choose it, how do they use it, what do they say about it, and have they discussed it with anyone?

Research on the paratextual dimensions of ebooks, from marketing to reviews, fonts, and usage, is still relatively new territory. We can expand these studies to apply the concepts to other types of electronic resources, including streaming audio, streaming video, video games, and apps. We could similarly analyze websites, including search engines, social media, library websites, and publisher websites. Because this territory is mostly uncharted, we must ask how effectively we can apply the same frames of paratextual analysis of print books to their electronic counterparts, as well as other electronic media. Are we in need of new criteria when crafting descriptions of the lives and times of ebooks?

Our partners in the vendor world might also appreciate suggestions for where they can focus future development efforts in regards to usage reporting. Indeed, as we have found, there is action needed on the communication (the clarification, the interpretation, and the uniform application) of COUNTER standards for usage statistics for library ebooks. Librarians would be thrilled if there were a way to track use of ebooks once they have been downloaded to personal computers or devices. We could then begin to understand how our ebooks are being used (annotations, highlights, bookmarks, searches within books, etc.) instead of being content that they're viewed at all. Librarians must also play our part in this call to action, demonstrating to

content providers that access to rich, standardized usage data is as important as we say it is. The ability to retrieve and analyze usage reports that reflect modern information-seeking behavior, and that conform to a reasonable set of standards, does affect our purchasing decisions.

Librarians and content providers have successfully collaborated to include provisions for ILL and perpetual archiving of electronic materials in licensing agreements; is there a place in licensing for usage statistics delivery?

There is another potential avenue of study in terms of the readership of open access content. The primary difference there is that when usage statistics are available, they are given only in the aggregate for the universe of readers. This could be beneficial in a wider study of readership beyond a local library, although it hinders one from determining how open access information is used at a specific institution (a vital consideration when librarians are deliberating about which open access initiatives to support with funding). Though we heavily consider global benefit when deciding which open access and digitization projects to fund, the ability to assess local value will help to offset the financial uncertainties involved.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Birke D and Christ B (2013) Paratext and digitized narrative: Mapping the field. *Narrative* 21(1): 65-87.

Galey, A (2012) The enkindling reciter: E-books in the bibliographical imagination. *Book History* 15(1): 210-47.

Litsey, R and Ketner K (2015) Occam's Reader: The first library-developed ebook interlibrary loan system. *Collaborative Librarianship* 7(1): 13.

McCracken, E (2013) Expanding Genette's epitext/peritext model for transitional electronic literature: Centrifugal and centripetal vectors on Kindles and iPads. *Narrative* 21(1): 105-24.

Northcutt Malone, C (2015) Noses in books: Orientation, immersion, and paratext. *Interdisciplinary Science Reviews* 40(1): 17-28.

Project Counter (2017) Code of Practice Usage Reports. Available at:

<https://www.projectcounter.org/code-of-practice-sections/usage-reports/> (accessed 28 April 2017).

Rowberry, SP (2015) Ebookness. *Convergence: The International Journal of Research into New Media Technologies* 1-18. Web.

Smyth, P (2014) Ebooks and the digital paratext: Emerging trends in the interpretation of digital media. In: Desrochers N and Apollon D (eds) *Examining Paratextual Theory and Its Applications in Digital Culture*. Hershey, PA: Information Science Reference, pp.314-333.

Thomas, WJ, Racine H, and Shouse D (2013) Ebooks and efficiencies in acquisitions expenditures and workflows. *Against the Grain* 25(2): 14-18.