Ebrary on the Radar

Julie Linden, Angela Sidman, and Sarah Tudesco

Introduction

Yale University Library has a long history with ebrary, subscribing to its collection of academic ebooks since the days when this product was known as "Ebrarian." Long since rebranded as Academic Complete, and now owned by ProQuest, over the ensuing thirteen years the package has grown to include well over 100,000 titles. Regular review of Yale's usage data demonstrated that Academic Complete was well used and delivered a great deal of content for a low cost per use. Balancing this, however, was a well-established perception among librarians that users found the interface difficult to navigate and that the subscription content was unstable. In the spring of 2014 three librarians and an intern turned their attention to this product.[1] Attracted by the long record of usage history, the group took a deep dive into ebrary data, determined to separate facts from feelings and explore the role of this foundational ebook package within Yale's larger collection. The study was limited to the subscription collection. Though Yale owns additional content on the ebrary platform, any reference to "ebrary" in this paper refers to Academic Complete.

Data Collection Overview

The ready availability of robust data was a significant part of this project's appeal. Within the ebrary administrative interface, customers have access to MARC records and usage reports that track their entire history with the product. For Yale, this meant thirteen years' worth of bibliographic and usage data, covering the period from 2002 to 2014. This combination of MARC records and usage reports allowed us to conduct a multidimensional study of Academic Complete, examining patterns of use by subject and publisher.

Ebrary provides two types of usage reports: site activity reports and COUNTER reports. The site activity reports include information on the number of user sessions, pages viewed, pages printed, pages copied, and turnaways. One advantage of the site activity reports is that they have been available since we began subscribing in 2002. The available COUNTER reports include Book Report 2 (number of successful section requests), Book Report 3 (turnaways), Book Report 5 (total searches), and Platform Report 1 (total searches, result clicks, and record views). Though our preference is to work with COUNTER data, the standardized reports do not cover the first several years of Yale's subscription to Academic Complete. As COUNTER data is available only back to January 2009, we used ebrary's site activity reports for our deep dive.

The MARC record files available on the system provided a history of titles added and deleted over the years, showing both the overall growth of the collection and fluctuations in content at a granular level. The MARC record data also allowed us to analyze the collection by publisher and subject.

Finally, package cost information was gathered from our local records to calculate cost per use over time.

Ebrary Collection Growth, Cost Per Use, and Summary Usage Patterns

Yale University Library has subscribed to ebrary since 2002, as noted in the *New York Times* early that year, "In the next few months, library patrons at Yale's and Stanford's university libraries, as well as some libraries in Northern California, may be surprised to find that

Julie Linden is Assistant Director of Collection Development, Yale University Library; Angela Sidman is Electronic Resources Librarian, Yale University Library; Sarah Tudesco is Assessment Librarian, Yale University Library

their electronic catalogs offer links to digital books."[2] The article describes that early iteration of the service and also serves as a reminder of just how novel ebooks in libraries were a decade ago.

Since its inception, Academic Complete has been a subscription collection with no perpetual access rights to the subscribed content. Titles are added and removed according to agreements between publishers and the vendor. There have been far more titles added than deleted over time. Title counts show that the collection grew from 10,793 titles in 2004 to 121,970 titles in February 2015. Usage has also grown significantly. There were 4,145 user sessions in 2003 and 60,998 user sessions in 2014. A user session is the number of times a title is opened and the user performs at least one action (copy, print, page turn, or download). We calculated the cost per use based on Yale's subscription costs and user sessions. The results confirmed the perception that ebrary is indeed a low cost-per-use product. The cost per use has fluctuated over time; since 2004, it averages out to approximately $0.55 per user session. As a point of comparison, Yale's 2014 cost per use of a well-used large journal package was $1.10 per full text download.

Library of Congress / Subject Analysis

Ebrary provides BISAC subject headings in its usage reports, but we wanted the option to compare the content in ebrary with the content in our general print collection using Library of Congress Subject Headings. To this end, we extracted LC call numbers from the MARC records.

Analysis of the MARC records showed that the current ebrary Academic Complete package is evenly distributed between Humanities, Social Sciences, and Sciences content. LC class H (Social Sciences) is the largest category, with over 24,000 titles, followed by P (Language and Literature), Q (Science), and B (Philosophy, Religion, Psychology).

Usage Frequency

With such a large, diverse collection, we were interested in how much of the collection gets used. With the extensive usage data available, we were able to determine if a title has ever been touched by a user. There are titles that have been in the collection for more than twelve years that have not been accessed once by Yale users. To calculate usage frequency, we divided the number of titles that have at least one re-

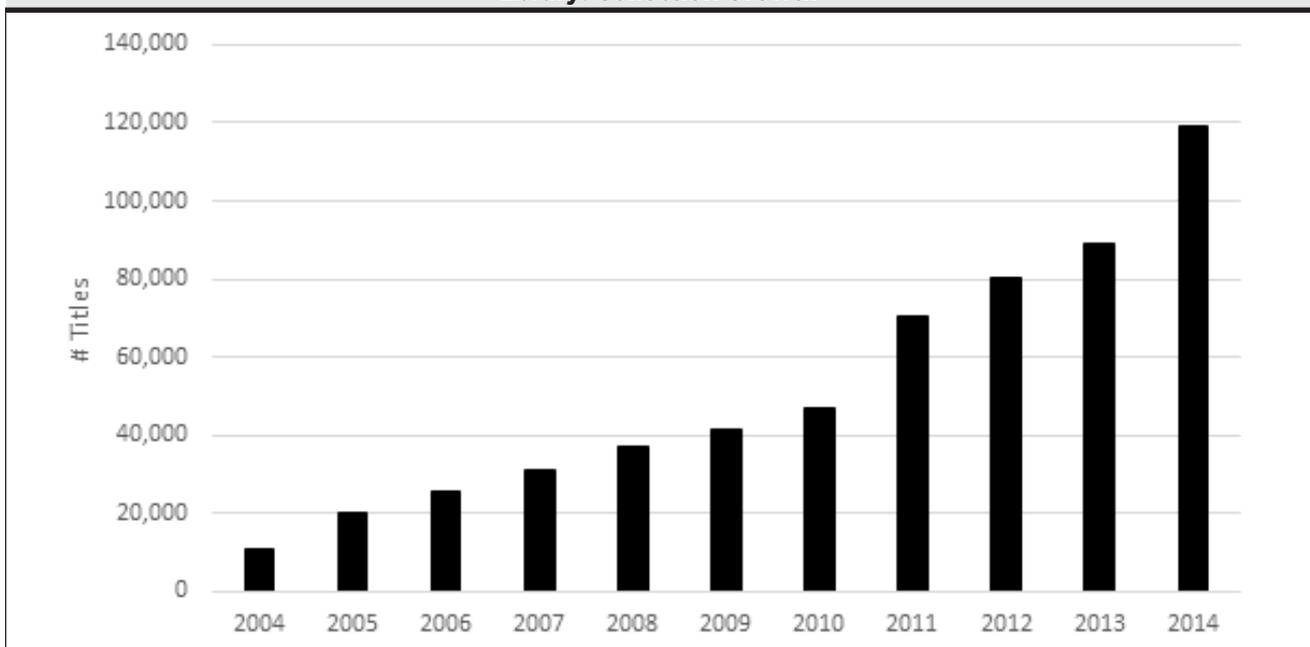**FIGURE 1**
Ebrary: Collection Growth
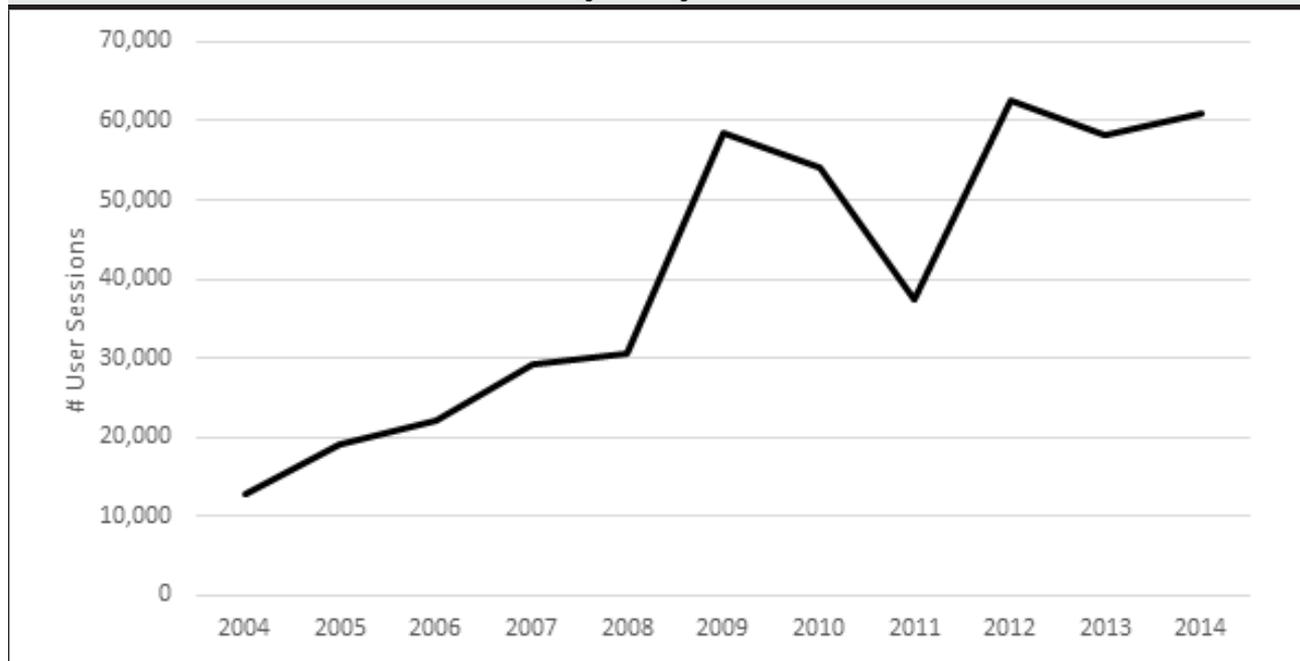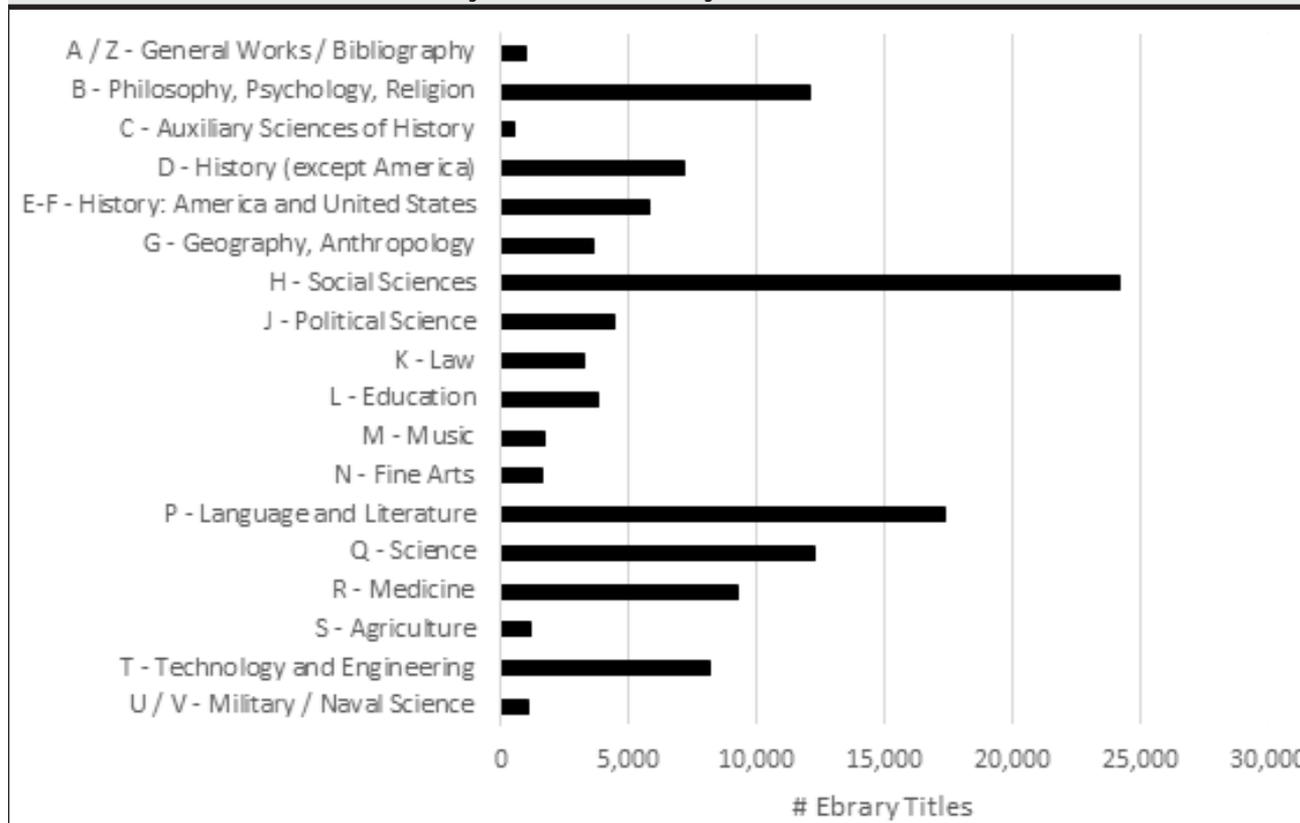
FIGURE 2
Yale Library: Ebrary User Sessions

FIGURE 3
Ebrary: Titles Available by LC Classification

corded use by the total number of titles in the collection.

Usage Frequency = Titles with Usage / Total Titles

The overall usage rate for the Academic Complete collection is 33%. Over the lifetime of Yale's subscription 38,000 titles have been accessed at least once and over 80,000 titles had not been accessed at the time we ran the data.

The usage frequency varies from subject to subject. The five Library of Congress subjects with the highest frequency of use do not correspond to the largest subjects by title count available in Academic Complete. Medicine, Music, and Fine Arts are the top three subjects by Usage Frequency, even though the number of titles in those subject areas is small compared to Social Science and Literature.

Building on the premise that usage statistics and MARC data could combine to reveal new understandings of how the collection was being used, we explored a new metric, Performance.

Performance

Michael Levine-Clark introduced a performance metric in his 2014 presentation "Diving into Ebook Usage: Accessing the Swell of Information."[3] We adapted this methodology to examine Yale's usage.

Performance = % of Usage - % of Collection

In applying this metric, our goal was to look at how various categories performed in relation to their place in the larger collection. We applied the metric to the LC subject classification. The table below sorts Performance from highest to lowest. R (Medicine)

FIGURE 4
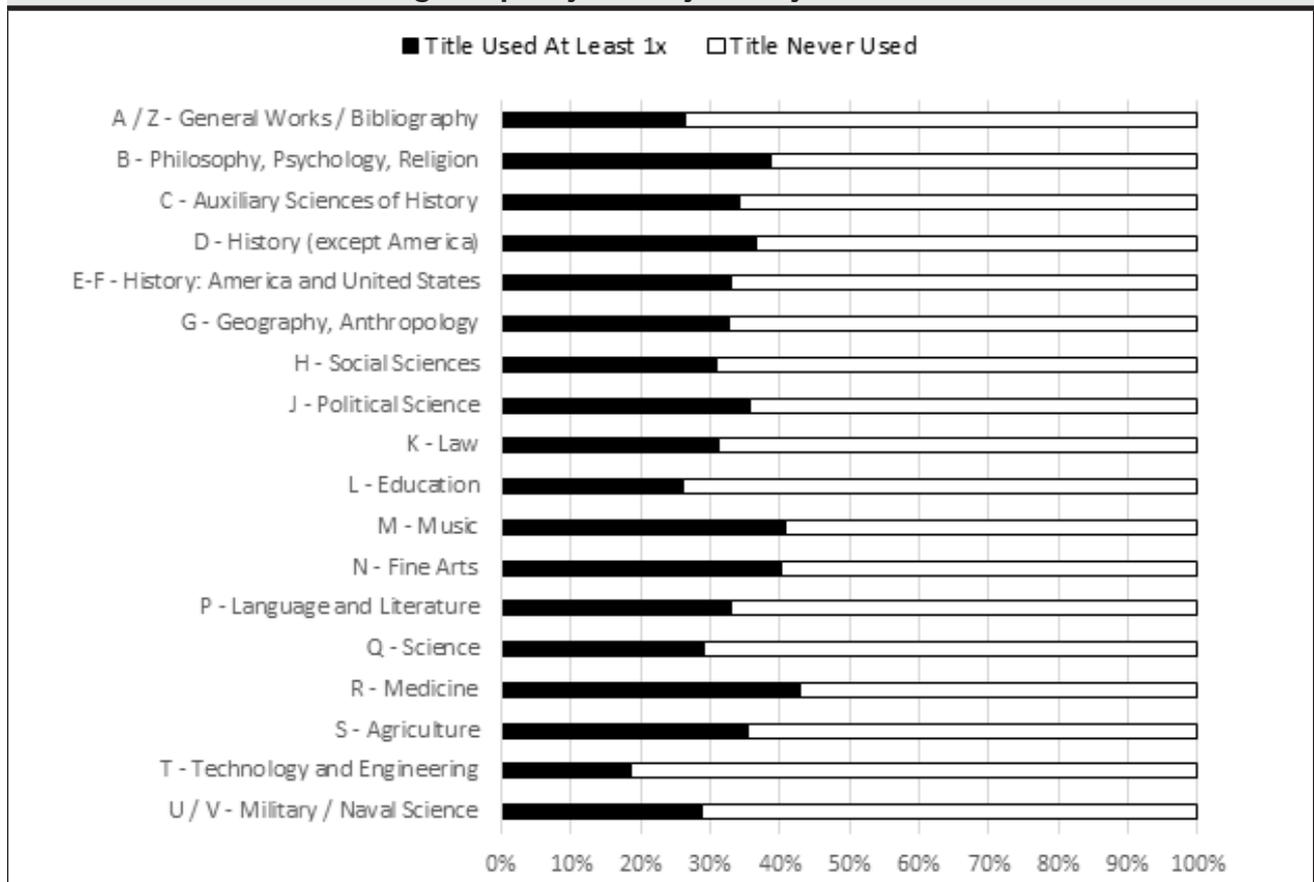**Yale Usage Frequency of Ebrary Titles by LC Classification**

TABLE 1
Ebrary Performance Ranking by LC Classification

LC Class	Usage	Collection	Performance
R - Medicine	10.18%	7.91%	2.27
B - Philosophy, Psychology, Religion	12.09%	9.84%	2.25
D - History (except America)	6.82%	5.80%	1.02
J - Political Science	4.08%	3.64%	0.45
M - Music	1.76%	1.36%	0.39
N - Fine Arts	1.65%	1.32%	0.34
E-F - History: America and United States	5.02%	4.80%	0.22
S - Agriculture	1.00%	0.95%	0.05
P - Language and Literature	14.83%	14.78%	0.05
C - Auxiliary Sciences of History	0.47%	0.43%	0
A - General Works	0.09%	0.09%	0
G - Geography, Anthropology	3.01%	3.01%	0
U - Military Science	0.74%	0.76%	-0.03
V - Naval Science	0.07%	0.12%	-0.05
Z - Bibliography	0.59%	0.69%	-0.11
K - Law	2.63%	2.80%	-0.18
L - Education	2.52%	3.21%	-0.69
Q - Science	9.00%	10.06%	-1.05
H - Social Sciences	19.21%	20.63%	-1.42
T - Technology and Engineering	3.94%	7.36%	-3.42

and B (Philosophy, Psychology, Religion) demonstrate a high performance rating, and H (Social Sciences), which is the largest subject in the package, performs towards the bottom of the rankings.

An examination of performance allows for a more nuanced understanding of the value of the content within this package. A common perception of ebrary is that it is a "general" package, so the revelation that Medicine is the highest performing subject in the package may lead to a new understanding about which students and faculty use the content, how and to whom we promote the package, and what constituencies are affected by changes to package content. An even more granular assessment of performance is possible. For example, if we teased out Psychology (BF) from Philosophy and Religion and added the Psychology figures to the Social Sciences figures, the Social Sciences performance ranking could change significantly.

Publisher Analysis

The methodologies used to examine the subject makeup of Academic Complete can also be applied to the more than 700 publishers currently represented in the package. Title counts as of January 2015 show that John Wiley & Sons (more than 7,000 titles), Oxford University Press (more than 6,000 titles), and Taylor & Francis (more than 4,000 titles) are the best represented publishers within the package.

Usage frequency of the top twenty publishers shows wide variation. Deriving meaning from these numbers requires an understanding of the publishers' ebrary content within a broader ecosystem. Three key data points to bring to the analysis are:

- Extent of title overlap between Academic Complete and other sources (such as purchases on ebrary or other platforms, other subscription packages, and/or print holdings)

FIGURE 5
Ebrary: Top 20 Publishers
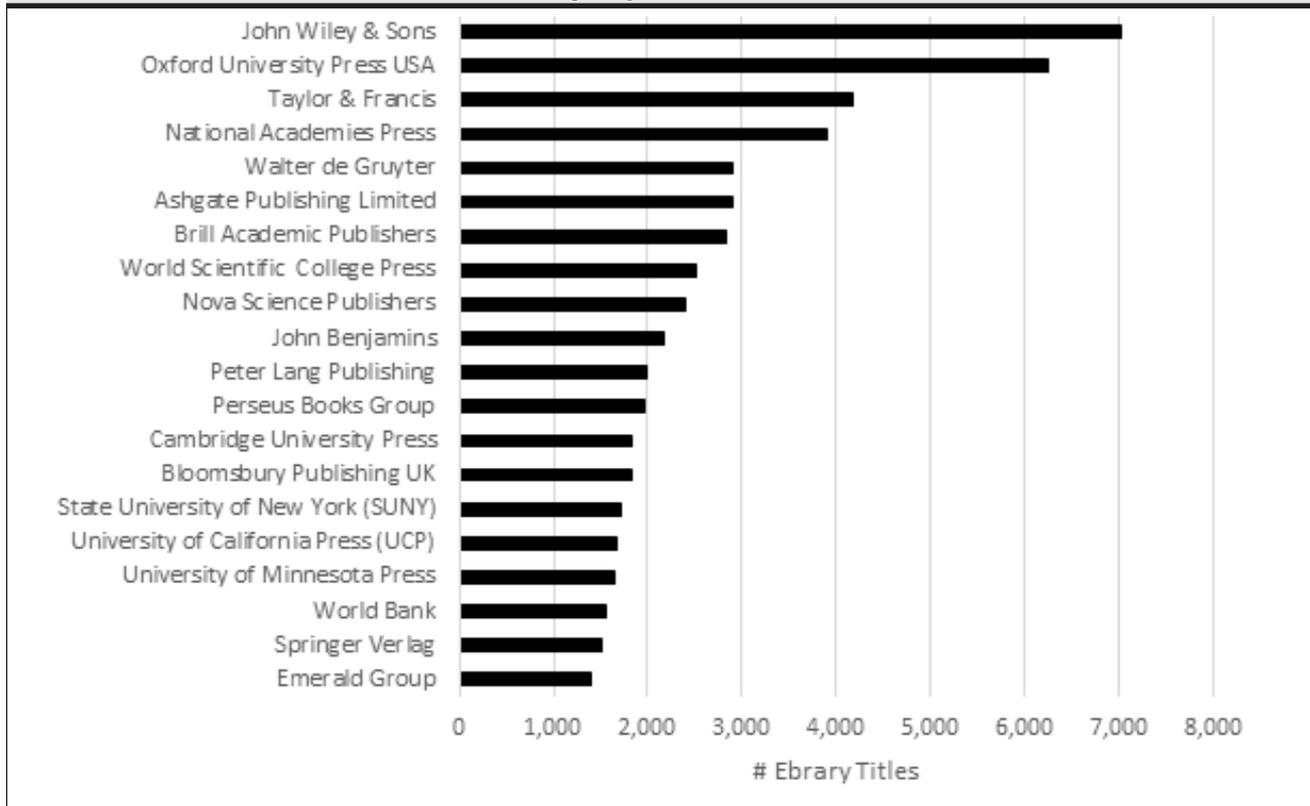
- Discoverability (not only in the online catalog and other discovery systems, but also inclusion on course reserves lists and syllabi)
- Usage trends over time

For example, at Yale we might ask questions like:

- Is the high usage of Cambridge titles influenced by the fact that Yale has not thus far acquired a Cambridge ebook package? Here we would want to look at title-level use of the Cambridge titles within Academic Complete as well as print circulation of those titles in our collection.
- Why have more than 50% of Springer titles within Academic Complete been used although Yale has purchased the Springer ebook package for several years? Was the Academic Complete usage primarily of Springer archive titles, which Yale purchased in 2012, and if so, did that Academic Complete usage

drop off once those titles were available on the Springer platform?

- Does the low use of International Monetary Fund titles mean that users are not interested in those books, or that they are accessing them from another source?

Ranking publishers by usage frequency surfaces publishers that are not in the top twenty by title count, such as Electric Book Co., Palgrave Macmillan (UK), and Johns Hopkins University Press. These may be publishers to investigate at a title level, to ask questions such as: are only a few titles being used heavily, or is usage distributed across hundreds of titles? Is use fairly consistent over time, or are there spikes that may be due to course reading?

The publisher data not only serves as a starting point for deeper dives into each publisher's unique set of circumstances, but also allows for assessing the entire Academic Complete package from yet another angle. The performance formula, which brings to-

FIGURE 6
Yale Usage Frequency of the Top 20 Ebrary Publishers
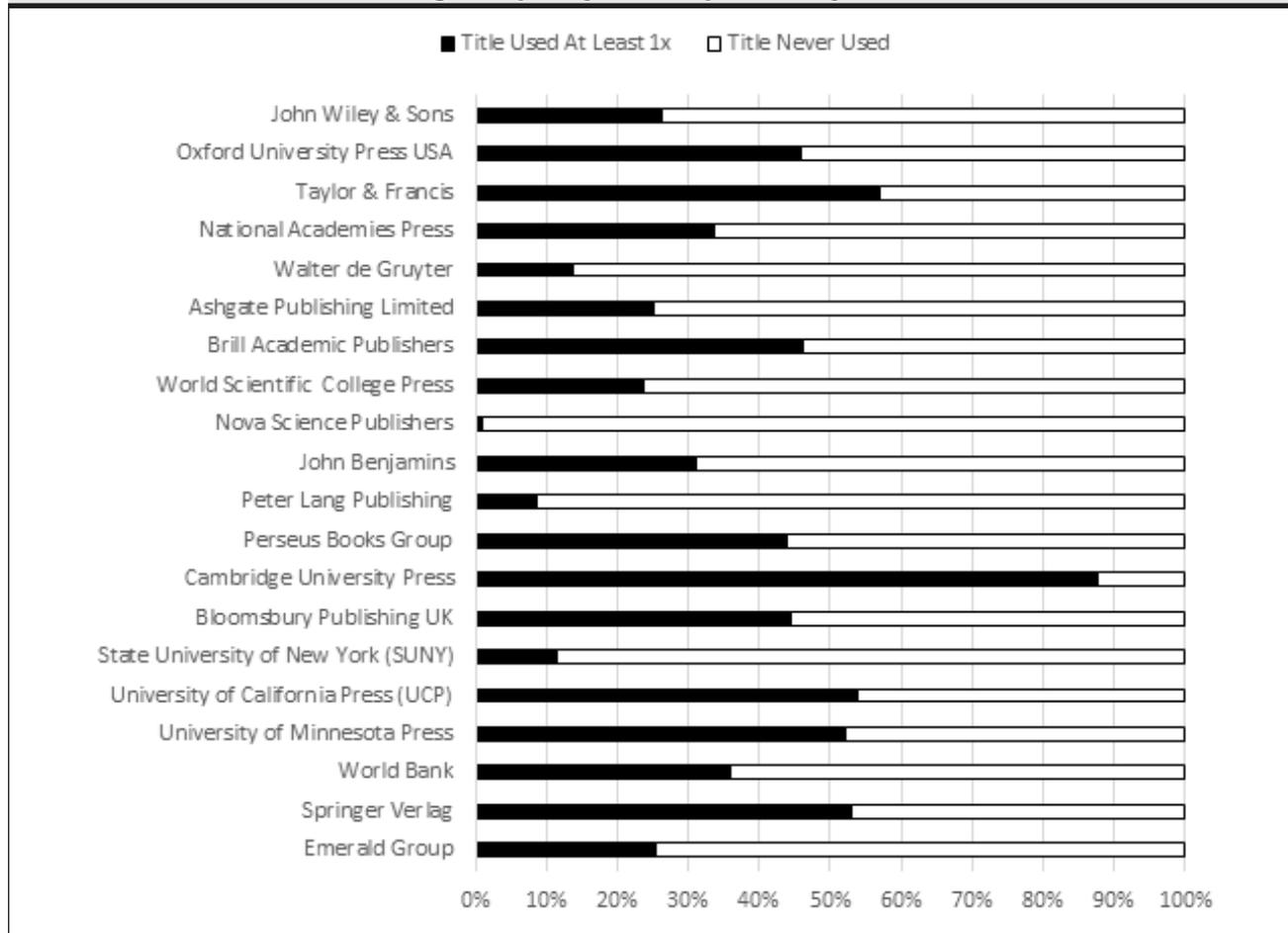
gether the publisher's prevalence in the collection with its usage, brings up interesting questions of value. In this context, qualitative judgments come into play. At Yale, we might ask what is the value of a package that includes the very high performing Cambridge and the underperforming Wiley—both publishers that are well-regarded by Yale's librarians and users alike. What is the value of the Cambridge and Wiley titles *within* this package?

The answers to these questions, while time-consuming to pursue, can help inform a library's decisions about future investments in ebooks.

Collection Stability

In April 2013, a Yale faculty member sent two librarians an e-mail with the subject line "Urgent book access is-

sue." She had just discovered that an ebrary book she'd assigned to her class was no longer accessible, although it had been available when she created her spring semester syllabus. One of the librarians later noted that this was "the second time that this same professor has run into this problem this year alone." This scenario has replayed itself enough times that librarians have formed a perception of Academic Complete as an unreliable package in which content might be here today, gone tomorrow.

Because ebrary provides MARC record files of titles deleted from the package, it is possible to move beyond anecdotes to data analysis. Yale has access to deletion files from March 2005 to the present via the ebrary administrative site.

Since March 2005, over 17,000 titles have been deleted from the Academic Complete package. The first

FIGURE 7
Ebrary Publishers with the Highest Usage Frequency
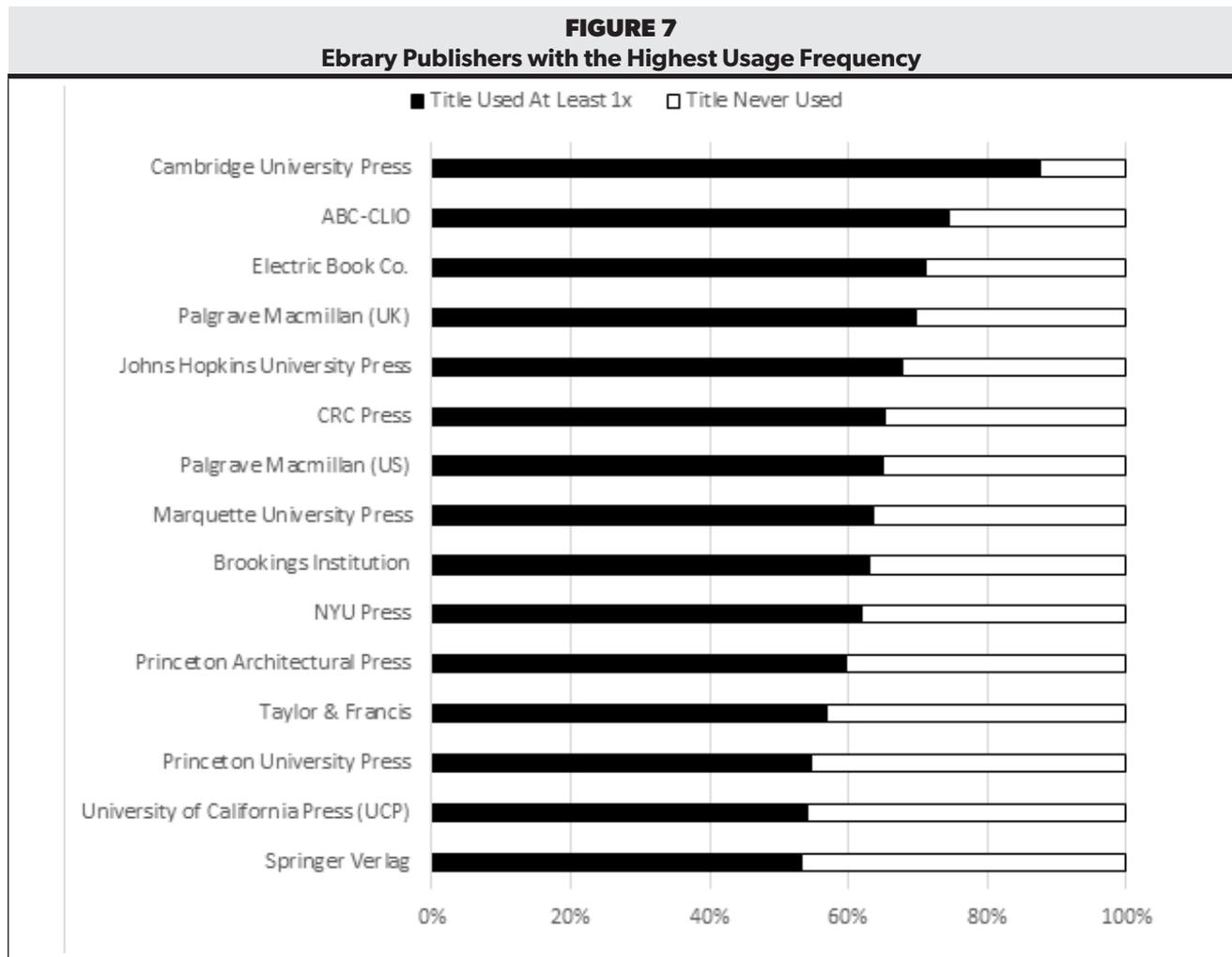
seven years of deletion data show a biennial pattern of far more deletions in odd-numbered years than in even-numbered years. Titles are typically removed when a publisher loses the electronic rights or when publishers change title offerings (for example, a publisher may change from a subscription to a perpetual-access model).

The 2011 spike in removals seen in the chart above was primarily due to a single publisher. Over 1,900 McGraw-Hill titles were removed from the package in a single monthly deletion cycle.

In recent years, the pattern of removals has evened out to nearly 3,000 titles removed in both 2013 and 2014. In 2014, ProQuest changed the deletion schedule from monthly to twice yearly (June and December), which should cause fewer mid-semester disruptions to libraries and users.

Studying the usage of dropped titles helps us understand the impact of their deletion. Analysis reveals that 82% of the titles dropped from Academic Complete were never used at Yale. Not only is this something of a relief—the majority of deleted titles were not causing major havoc—but this percentage also compares favorably to the 67% of never-used active titles in Academic Complete as a whole. In other words, the deleted titles, as a set, had less use (18%) when they were in the package than the set of active titles in the package (of which 33% have been used).

Pushing the data further, we found that among the small number of deleted titles that had seen usage were some of the most-used ebooks in the entire collection. Analyzing the heavily-used deleted titles

TABLE 2
Performance Ranking by Publisher

Publisher	Usage	Collection	Performance
Cambridge University Press	7.29%	1.39%	5.90
Oxford University Press	9.22%	4.62%	4.60
Taylor & Francis	6.99%	3.10%	3.89
Perseus Books Group	1.95%	1.45%	0.50
Brill Academic Publisher	2.55%	2.10%	0.45
John Benjamins	0.76%	1.60%	-0.84
World Scientific College Press	0.61%	1.86%	-1.24
National Academies Press	1.57%	2.89%	-1.33
Peter Lang Publishing	0.11%	1.48%	-1.37
Ashgate publishing Limited	0.45%	2.14%	-1.69
Walter de Gruyter	0.38%	2.15%	-1.78
Nova Science Publishers	0.01%	1.79%	-1.78
John Wiley & Sons	2.62%	5.20%	-2.58

added yet another dimension to our understanding of how the package is used, how deleted content potentially impacts users, and how perceptions are formed and reinforced.

We confirmed that some of these titles were used in course reserves (*Development as Freedom, Worlds on Fire*) and suspect that other titles may have been cited in courses outside the established course reserve system. One title, *Getting Things Done: The Art of Stress-Free Productivity* by David Allen, is a *New York Times* bestseller and extremely popular business title. Having a data-driven understanding of the stability of the content and the impact of deleted titles allows us to assess the value of ebrary—as content and as a service—within the larger context of its collections.

Conclusion

Yale's examination of ebrary's Academic Complete and its role within the larger library collections answered many of our initial questions and raised several new ones.

Our original questions were aimed at understanding ebrary's place in relation to Yale's larger collecting practices:
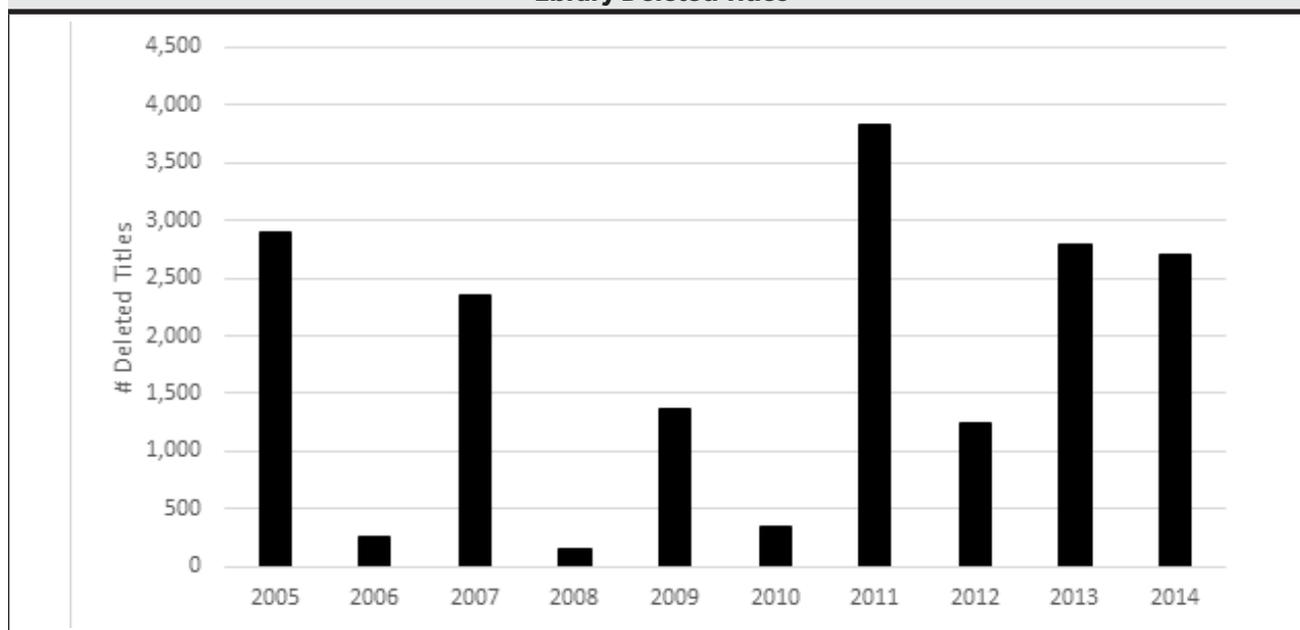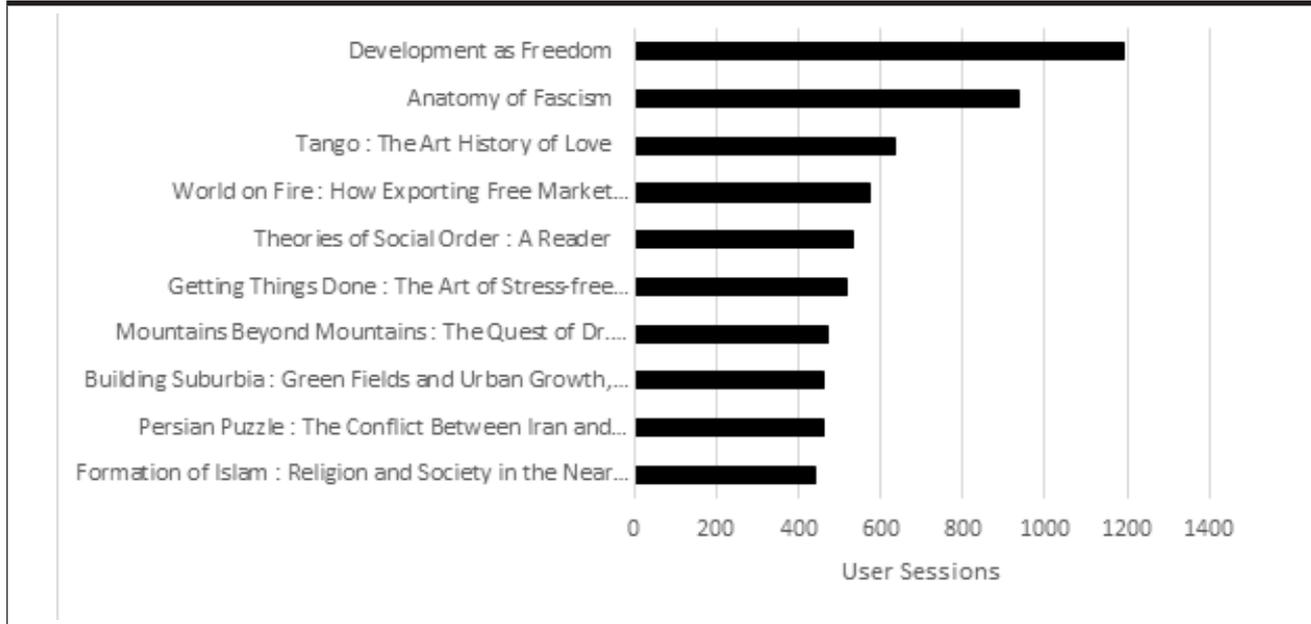
FIGURE 8
Ebrary Deleted Titles

FIGURE 9
Yale: Tope 10 Deleted Ebrary Titles by Usage

- Although we knew that ebrary was an overall low cost-per-use resource, what could we learn about the composition of titles and usage *within* the package? What could the site activity reports and bibliographic data reveal that COUNTER reports could not?
- Do the usage patterns suggest that for certain subjects or publishers we might want to invest in purchasing ebook collections?
- How does our perceived sense of collection stability compare to the actual impact of titles deleted from Academic Complete?

Though we have many questions left to answer, this initial round of work succeeded in its primary goal of separating facts about ebrary from feelings. We learned that despite poor librarian perceptions of the interface, our users were finding and using titles from top academic presses. We also gained a more nuanced view of collection stability. A low percentage of deleted titles had ever been used, meaning impact on users was minimal. However, among those deleted titles were some top performers that ranked among the most-used ebooks in the collection. The loss of

these few high-use titles had an outsize impact, creating and then reinforcing the idea that ebrary content as a whole was unstable.

This exercise lined up with Yale's larger interest in creating a culture of assessment, leading us to ask:

- What data sources and methodologies are available to us when assessing usage and content?
- How do we set parameters for analysis so that we balance a deep dive into data with a contextual assessment of a broader set of resources?
- How can we most effectively communicate findings to stakeholders?
- How do we turn our findings into priorities, action items, or guidelines for future decision making?

Through the process of asking questions, our analysis of data to inform decision-making becomes more sophisticated, multidimensional, and nuanced. A deep dive into usage data has sparked productive conversations about ebooks, ebook models, and ebook support and will inform our future decisions about products, services, and collections.

Notes

1. Mélanie Roche, an intern from ENSSIB, contributed greatly to the gathering and computation of Yale's historic ebrary data.
2. Guernsey, Lisa. "In Lean Times, E-Books Find a Friend: Libraries." *The New York Times*, February 21, 2002, sec. Technology / Circuits. http://www.nytimes.com/2002/02/21/technology/circuits/21BOOK.html.
3. Levine-Clark, Michael. "Diving into eBook Usage: Assessing the Swell of Information." presented at the Electronic Resources & Libraries, Austin, TX, March 17, 2014. http://www.slideshare.net/MichaelLevineClark/levineclark-michael-diving-into-ebook-usage-assessing-the-swell-of-information-electronic-resources-libraries-austin-march-17-2014.