



ACADEMIC EBOOK USAGE IN  
LIBRARIES: A STUDY OF UNIVERSITY  
PRESSES AND AGGREGATORS



Alex Holzman and Frank Smith

[alex.alexpubsol@gmail.com](mailto:alex.alexpubsol@gmail.com)

[fsmith26@proton.me](mailto:fsmith26@proton.me)

Ebooks have been offered to libraries by aggregators for more than 25 years. In the past 10 years, and especially since 2020, the proportion of ebook revenue for university presses from library aggregators has increased significantly. Some presses report that library aggregator revenue represents one-third of all book revenue. But do presses understand their relationships with library aggregators and are they achieving their goals – and recognizing that these will likely be a mixture of revenue maximization and dissemination?

This project began with an invitation from De Gruyter eBound to examine data relating to recent ebook usage by academic libraries and interview university presses about their experiences managing library ebook activities. Eleven presses agreed to share 2023 usage data for their ebooks from three different aggregators and speak about their activities. We are not calling out specific data from any press nor identifying the presses or aggregators. Our purpose is to draw upon the usage data and conversations to describe the general current situation between presses and aggregators as it relates to managing ebooks portfolios in a way that aligns with a press's mission. We also aim to offer suggestions, most of which arose out of our interviews with presses, to improve press-aggregator relations in the future. Presses will have different goals in making books available through aggregators and there is no one correct way to manage ebook availability, although all presses do face the need to replace declining print revenue with increased digital revenue. Accordingly, we hope to demonstrate the value of being able to compare the performance of content with each aggregator in terms of revenue and the business models a press elects to participate in.

## **A brief history**

The first dedicated academic ebook offerings to libraries appeared in 1999: Netlibrary; the Gutenberg-e project managed by Columbia University Press and the Columbia University Libraries and funded by the Andrew W. Mellon Foundation; and the ebook project of the American Council of Learned Societies (ACLS). The Gutenberg-e project was an important early attempt to determine how scholarly ebooks might be offered to libraries, and it later merged with the ACLS project. Oxford University Press launched the first dedicated single-publisher ebook platform in 2004. The offerings began to achieve scale after 2010 with the launch of Cambridge Books Online, Books at JSTOR, De Gruyter, EBSCO, ProQuest, Project MUSE/UPCC, and others. Sales in the first decade of availability were modest, but they grew rapidly after 2010. The growth in sales after 2010 may have been a result of more offers, libraries' acceptance of ebooks, or some combination of the two.

## **Library sales**

Libraries are tasked with providing access to quality content and resources for their campus community, whether it be a research university, a community college, or anything in between. To do this, they maintain subscriptions to resources that make these content items searchable (the library catalog and discovery layer), cataloging and metadata services (i.e., OCLC Connexion), and interlibrary loan and document delivery services. In addition to books, the library also maintains subscriptions for journals and databases, which are responsible for a significant portion of the collections budget. And, of course, they also buy books, increasingly shifting to digital purchase rather than print.

This report will focus on eBook sales and purchasing. Ebook aggregators work closely with individual libraries and library consortia (the latter are often geographic, i.e., OhioLINK, and can include dozens of colleges and universities). A library will have relationships with more than one aggregator, though it may tend to purchase more from one in particular. The mission of the library is to provide comprehensive resources to the community to foster intellectual growth, advance knowledge, and support lifelong learning. They also seek to provide equitable access to diverse collections. To this end, libraries work with vendors and publishers to enhance their collections with quality content that meets the needs of their campus community while also working within the constraints of their budget.

The amount of content that can be offered to the library's patrons depends upon the deals they are able to negotiate and where they sit on the access versus ownership spectrum. Some libraries may prioritize book access models that offer access to a large amount of content, such as usage-based or subscription models, while others may prioritize purchase models, such as collection purchases or title-by-title selection. Both approaches have their benefits and drawbacks and depend on the library's collection strategy. Regardless of the approach, the commonality is the relationship between the aggregator and the library. This relationship is inherently dynamic.

Sales offerings from library aggregators include complex questions of how discounts (on the retail price set by the publisher for each title) are set and how libraries actually decide to buy books. Some sales models depend upon automated systems where purchases are made by a pre-arranged set of criteria worked out by the librarian and vendor representative. These titles will be funneled to the library, either to review for selection or to be part of a demand-driven pool, based

on previously set criteria. This is typically monitored and reviewed periodically to ensure the content being sent is relevant to the library's needs.

### **Sales models**

A non-exhaustive list of sales models would include:

- a) Title-by-title selection or firm orders (i.e., single-title purchase).
- b) A collection of books (by year, publisher, subject, etc.).
- c) Demand-driven acquisition (DDA), where an aggregator works with the library to identify criteria for books to be allowed into a pool. The library then makes this pool available to its patrons. Once titles meet a use threshold or “trigger”, the library will then be charged for the title. Any use prior to this threshold is free. Uses are additive – a trigger may be reached by a single user who reaches the threshold for purchase or by multiple patrons whose use together reaches that threshold.
- d) Evidence-based acquisition (EBA), where patrons of a library are granted unlimited access to a specified set of titles, agreed upon with the library, for a set period, i.e., 12 months or less. The library will make a deposit at the start of the access period to cover sales at the end. At the end of the access period, the library reviews the usage of titles, selecting those to convert to perpetual access. This selection is often, though not always, based on usage during the access period. At the end of the period, the library loses access to the unpurchased books unless they renew the EBA arrangement.
- e) A subscription model, where a large set of books can be accessed without limit during the subscription period. Access is lost when the subscription expires.

### **How presses manage their ebook offerings through aggregators**

Presses, often short-staffed and preoccupied with acquiring, publishing, and marketing their lists, may not have time to develop a sufficient understanding of how the sales models and procedures of library aggregators help achieve the dual and sometimes conflicting university press goals of maximum dissemination of content and revenue generation sufficient to continue operating.

Although presses receive widely varying kinds of reports about usage and sales achieved by library aggregators, these reports may not be sufficiently transparent for presses to understand how their books are sold and disseminated. It also may not be clear what revenue derives from an ebook sale and what revenue, depending on contractual arrangements, might be classified as subsidiary rights.

All aggregators offer unlimited-user models for books, whereby theoretically, all full-time employees of an institution may access a book simultaneously. Because these models offer different uses and conditions, publishers need to choose which model or combination of models offers the best fit for their business plan and mission. Publishers will have different reasons for putting books in unlimited-user models versus limited-user models. The latter, for example, might be chosen for new titles or for books the publisher thinks might be assigned for course reading and where they would like to protect the potential for print sales.

In addition, library aggregators collect usage data from their platforms that they may or may not share with presses or may not share in formats that presses find useful. This is despite the fact

that publishers of scholarly journals, which includes many university presses, have lots of experience in viewing and interpreting usage data. (Indeed, libraries expect to receive usage data from publishers and vendors who supply electronic content.) Standards for usage data for electronic materials – COUNTER – have been developed over the past 10 years by a consortium of librarians and publishers. Those standards, as they apply to libraries’ ebook usage, have been accepted by aggregators at the insistence of librarians. However, the ebook information aggregators provide to publishers is not yet sufficiently comprehensive, as evidenced by [COUNTER’s recent statement](#) urging improvements.

### **What is COUNTER?**

COUNTER is a community-led organization funded through the membership of libraries, publishers, aggregators, and technology providers. To quote from the COUNTER website:

“COUNTER exists to bring the knowledge community together to agree and adopt the global standard for measuring and reporting content usage through normalized metrics, and we do that by developing, maintaining, and extending the Code of Practice.”

Founded in 2003, COUNTER was originally designed to assist academic librarians in understanding the usage by their faculty and students of journals online. The core of COUNTER is a standard, or set of standards, that publishers agree to meet in providing usage statistics.

Publishers and aggregators are audited, usually when the COUNTER standards change and are updated, to ensure that they are compliant with the standards – in other words, so that librarians can feel certain that the usage statistics they see for a content item offered by publisher X will be

fully comparable to the usage statistics they see from publisher Y. It enables comparing apples and apples.

### **Publishers and aggregators and usage**

There are five major academic ebook aggregators: De Gruyter, EBSCO, JSTOR, Project MUSE, and ProQuest. Each tends to report results in different ways. We discovered in conversations with many of the participating presses that there is often a lack of knowledge of all the details that should be reported to them. This lack of knowledge is partly a consequence of inadequate staffing at smaller presses. Intensely busy staff will always have difficulty staying on top of all aspects of selling ebooks via aggregators, but the problem is greatly exacerbated by differences in the ways aggregators report data, right down to using different terms to describe the same things and which data is shared.

For example, downloads could mean different things for different platforms. Is the table of contents and front matter included in this number, or is it limited to the main content of the book? Are views and downloads duplicated if a user double-clicks in a single session? What does this mean if the press wanted to review usage of their portfolio compared with revenue to assess their participation in these various models?

Similarly, some aggregators share information about which library has made a given purchase or rental and some do not. It is our impression that most presses accept this, perhaps because they need the revenue an aggregator generates, and they do not wish to risk the aggregator relationship by demanding a standard of reporting. But it is our strong contention that publishers,



at minimum, co-own any data that the sale of their products generates. Aggregators should be transparent with all sales and usage data, ranging from discounts offered that differ from an announced discount schedule to where sales occur to the model under which they occurred. They should offer comprehensive reports outlining where downloads take place and at what institution. And they should offer these reports in a timely fashion – or create a system that allows a publisher to generate reports as needed and with a degree of ease (remembering the staffing issues).

As previously mentioned, 11 presses took part in this study, each supplying download data from three aggregators for 2023. We were able to distinguish downloads generated via open versus paid access for one vendor. The participating presses saw both the compilation of the raw data and the aggregated totals informing them where their sales took place on a country basis, how many downloads occurred in the top 10 nations, how many ISBNs they had on each aggregator's platform, and the subject areas that produced the top 80% of usage. Each participating press is also aware of the revenue it received from each vendor in 2023, but that information was not supplied. Hence, we can only work from the usage data without linking it to revenue.

Several things can be gleaned that are of interest to the overall scholarly community as well as to each individual press. One aggregator consistently outperforms the other two in distributing content over the greatest number of countries. But interestingly, that same aggregator usually had the fewest overall downloads among the three participating aggregators, so the very wide distribution was not deep. It is also generally true that publishers provided this aggregator with fewer ISBNs than they did the other two. The reasons for providing fewer ISBNs are not clear

but could include other aggregators emphasizing collections more than this one, assistance provided to presses in digitizing backlists, publisher preferences for some library acquisitions models over others (including open access), and even library and library patrons preferring other aggregators' platforms (though this would not explain why the aggregator with the widest geographical coverage has fewer downloads than the other aggregators, unless the downloads in high-volume countries simply swamped the number of downloads in low-usage countries). It is also possible that the aggregator had more ISBNs available. These did not show up in the reports due to not seeing usage during the reporting period, but this seems less likely.

We also can see that most presses do not put all their books on all platforms. Some have told us they do, but when we compare the number of titles that occur in the listings of all three aggregators, there are far fewer than the number for each individual aggregator. One press did not have a single title that was available on all three platforms. One caveat – only ISBNs that were used during the reporting period show up in our reports. Publishers may have made some titles available that received zero use, which could alter these figures.

Not surprisingly, total downloads skewed toward wealthier countries, especially English-speaking ones. Interestingly, while the US predictably led in total downloads, the most intense use came from the UK. This may be in part because of agreements some libraries appear to have to emphasize the platform of one of the participating aggregators almost exclusively or because of the availability of more titles because of the wider use of use-based models in Europe than the US for at least one aggregator. It may also reflect a different culture of university teaching and learning: UK university students in the pre-ebook era typically did not purchase course books

and generally relied upon the library for most assigned readings. It would be logical for this style of learning to carry over to ebooks. In any event, 80% of downloads routinely occurred among only five or six countries, with the US, the UK, Canada, and Germany very often among those five or six. Other frequent high users included Australia, Israel, and the Netherlands.

A variety of subject areas comprised 80% of usage among the individual presses, largely including social sciences and the humanities, history, literary studies, and other HSS subjects. Most of the participating presses publish little in the sciences, so the absence of usage in science is not surprising.

The lessons to be drawn from the data, while limited, indicate increasing acceptance of ebooks by academic libraries, usually at the expense of print books. Presses we interviewed said that ebook sales now comprise anywhere from around 10% of total book revenue to a high of 49% at one press. Anecdotally, ebook sales increased considerably during the pandemic and have recently leveled off or are showing a slower growth rate. But given academic library space issues and the acceptance of ebooks by librarians and their patrons, we suspect that library acquisition of scholarly monographs – relatively low circulation books by definition – will continue to shift toward ebooks and away from print.

Because few presses have their own platforms, we suspect that the reliance on aggregators for sales revenue models (and much open access material) will continue to grow. This makes it incumbent on presses to understand aggregators' goals, methods, and sales models, better than we sense many do now. We recognize that at smaller presses (groups 1 and 2 especially), it is

extremely difficult to devote increasing staff time to such activities. Hence, we suggest some areas where a press without a lot of e-publishing staff resources should focus.

### **Our recommendations**

Almost all university presses, not just the small ones, have trouble understanding and monitoring the elements involved in working with aggregators and understanding how library sales are (or are not) made. Properly doing so takes time, and staff at most presses are already overworked.

Almost every press told us that they do not feel they have adequate resources to manage their ebook sales. The challenges of understanding sales models and analyzing sales and usage reports were brought up by most presses. We therefore emphasize a few of the most important things we believe presses should track, as well as urging greater transparency and standardized reporting from all aggregators.

First, it is critical to monitor all the ways that aggregators generate revenue for presses. As we've seen, this involves single-title sales, collection sales, sales via DDA, sales via EBA, and short-term rentals. Some aggregators may have other models, but these are the main ones, though they do not always operate in the same way for each aggregator. Publishers must be aware of and make decisions about the discounts they are willing to offer for each type of sale and whether an aggregator is observing these discounts. If not, is the extra discount an aggregator grants to close a "big" deal worth it to the press? Does the aggregator discuss this prior to closing, discuss it afterward, or sometimes not really mention it? As the saying goes, follow the money.

Publishers set retail prices for their ebooks and aggregators offer discounts from those prices. But it is worth monitoring if aggregators actually start with this base price when negotiating a purchase or if they set what they feel is a true best market price, then apply discounts. We have heard anecdotally this may sometimes occur, but we cannot say with certainty that it does. It is worth keeping an eye on.

One press told us they have an annual audit of their ebook business where they examine all aspects of their relationships with aggregators, including terms, contracts, sales, and usage. They evaluate whether to stay in or move out of certain sales models and whether to shift titles out of models. Presses might consider this approach to managing their ebook business.

Not surprisingly, presses reported different attitudes about the nature of the returns from ebook aggregators. One press told us that while they understood the revenue per title from a subscription model might look low relative to other models, the gross revenue from participation in the model was satisfactory for the press's financial needs and they did not want to risk upsetting their relationship with that aggregator. Other presses reported that they wanted a greater understanding of the returns from all models but particularly from DDA and EBA.

Publishers should also be aware that putting books in DDA or EBA means some books may be used but never bought. One aggregator estimates that a maximum of 5% of books in DDA/EBA are ever purchased with the other 95% available without cost to libraries. For this reason, embargoing new titles from these models is common, but every press should periodically decide whether to embargo titles, how long the embargo period should last and whether the return using

these models exceeds the sales that would be achieved if the title were only available via purchase models. At the same time, putting a book in a DDA or EBA model may be the only opportunity to make sales to some libraries, as the funds that would go for these models may otherwise be diverted to non-ebook resources.

On the other end of the spectrum, long after any embargo period would make sense, several presses that we spoke to have either digitized their backlist (often out-of-print books) or collaborated with aggregators who digitized books and then enjoyed exclusive rights to sell access to the books for a defined period. Although the usage of these books was not as great as recent publications, it often lagged only slightly behind, provided some marginal revenue and helping the presses achieve their dissemination goals.

The monitoring and decision-making that presses need to undertake relies on data. Aggregators make the sales and accumulate the usage, and they need to provide complete and transparent data back to the publishers they represent. At present, they do not. For example, at least one aggregator does not divulge what libraries have purchased what books.

Other aggregators were reluctant to provide the download data for this study, and one small press had to withdraw from the study because an aggregator refused to compile it (though it did provide the data for other presses in our study – behavior we find as baffling as it is unacceptable).

The point here is not to criticize aggregators but to emphasize that presses, libraries, and aggregators must act as equal partners for the dissemination of scholarly monographs to continue. University presses must earn sufficient revenue to meet their costs and they face at best static support from their parent institutions for the foreseeable future. While open access advocates are experimenting with models that would provide sufficient revenue without requiring sales to libraries, we are a very long way from solving the problem that way. Sales revenue remains critical and if aggregators are to continue to provide services that generate sales, they must themselves earn enough to make the undertaking worth their while. But they must also provide publishers with complete and completely transparent data about those sales, making all the items we've listed above available to the presses who provide the new content they require every year.

To achieve the same transparency and fullness of data to presses that libraries already receive (though the information libraries require differs somewhat from the information presses need), aggregators and presses should work with a standards organization – COUNTER, which was described earlier, or the National Information Standards Organization, which helps develop voluntary standards for many aspects of scholarly publishing – to develop a set of standards in aggregator-publisher reporting that works for both sides. Besides transparency, or perhaps as a partial means to achieving it, efforts should be made to standardize terms, abbreviations, and categories across aggregator platforms. A partial list would include an industry-wide definition of a download, DDA, EBA, and others.

It is beyond the capabilities of any one press to pursue negotiating with aggregators and so we heartily recommend that the Association of University Presses (AUP) act on behalf of all its members to achieve these reporting standards. The AUP Executive Director and Board could determine how best to do this, activate any necessary committees – for example, a best practices committee – and undertake any other needed actions. However the AUP might proceed, its members need more and better data encapsulating ebook sales to library markets.

We close with this. University presses need to be able to disseminate monographic studies in the form of ebooks in a manner that yields sufficient revenue to survive. Open access does not currently provide such revenue and seems to us to be a long way from achieving that – so sales and rentals will continue. Aggregators play a vital role in ensuring libraries get the books and access they need. Libraries define what they need and provide it to their scholarly communities. All three legs of this stool are vital, and none can fulfill their missions without the others.

We have faith that the limited data we were able to compile provides a glimmer of hope – and that presses, aggregators, and libraries will work together to allow the scholarly ebook ecosystem to thrive.

*Alex Holzman and Frank Smith*



## About

*Alex Holzman* is director emeritus at Temple University Press and in semi-retirement heads his consulting firm, Alex Publishing Solutions when not watching Phillies games. He also worked in books and journals, and in editorial, marketing, and e-publishing capacities at Cambridge University Press, Ohio State University Press, and Charles Scribner's Son. In 2008-09 he served as president of the Association of American University Presses.

*Frank Smith* worked for Cambridge University Press from 1978 to 2011 as, successively, Editor for History; Editorial Director, North America; and Director of Digital Publishing, Global. From 2011 to 2019 he was Director of Books at JSTOR. He now spends most of his time watching birds and his lifelist stands at 3371.

This report was sponsored by De Gruyter eBound, a not-for-profit that supports the work of mission-driven academic publishers at a time when the scholarly system is undergoing fundamental change. You can read more at <https://www.ebound.org>.