E-Books in Academic Libraries

Ward, Suzanne M, Freeman, Robert S, Nixon, Judith M

Published by Purdue University Press

Ward, Suzanne M, et al.
E-Books in Academic Libraries: Stepping up to the Challenge.
Purdue University Press, 2015.
Project MUSE.muse.jhu.edu/book/43208.

3 | Production, Marketing, and Legal Challenges: The University Press Perspective on E-Books in Libraries

Tony Sanfilippo

ABSTRACT

A university press's mission is to disseminate scholarship, but the challenge is to fulfill that mission by issuing quality books at low cost, but with high impact. This paper explores topics such as the workflow for print books and e-books, the many options for including e-books on aggregator platforms, the challenges involved in digitizing backlist titles, a variety of legal issues, the reasons for pricing differences between print books and e-books, and placing titles where scholars and nonscholars alike will discover them.

INTRODUCTION

In preparation for writing this chapter, I asked my colleagues on the Association of American University Presses (AAUP) general listserv which university press first published an e-book and when that occurred. It would seem a simple enough question with a straightforward answer, but it wasn't. Nine different presses claimed to have published the first e-book, eventually causing many of them to dig into their archives to determine the actual release dates of their candidates. But the one factor that most of their books shared was that the customers for those first e-books probably weren't libraries. In fact, since all but two were published as floppy disks or CDs packaged in a sleeve attached to the back cover of the physical book, it's quite likely that most libraries that purchased the book actually removed those e-books and discarded them with the dust jackets.

From a university press's perspective, it's not a surprise that the first e-books were primarily add-ons to print books, or that the target audience wasn't actually the library market. Setting aside for now a couple of outliers, most university presses started experimenting with e-books in the mid to late 1990s, and almost always worked with outside partners, such as Voyager Expanded Books and Eastgate Systems, to create those early e-books. Although most print production at that time had moved to using digital tools, those tools were still producing files that were specific to print production. Complicating and frustrating the development of an e-book workflow in the 1990s was the lack of e-book standards and devices that could display e-books. Early software tools for print book creation, namely PageMaker and Quark, simply mimicked the page layout and composition work done by hand before digital tools were available. These processes were specific to fixed text and were only efficient at automating the parts of book production related to print products; features like automated header placement and page number placement, line and figure spacing, and note insertion were all features of those tools. Even today, the most widely used tool in page layout, Adobe's InDesign, contains only rudimentary tools that pertain strictly to e-book design, flowable text, and complex e-book file creation.

One of the likely reasons for the lag in university press e-book production is because of the realities of the current book marketplace. E-books are not the primary market for most books sold in the United States (Packer, 2014), nor are they for university presses, and they still aren't the primary market for most books sold to libraries. It also is difficult to say with certainty which format, print or digital, patrons prefer. Although use of electronic content seems to be growing rapidly, often outpacing circulation of the print versions, surveys of students and faculty seem to show a preference for print (Sacco, 2014). Defining use of digital material also can be difficult. Comparisons of digital access to print circulation are problematic because they are not measuring the same thing, and they do not include noncirculating/in-library print use, which, unlike digital access, is very difficult to measure. The initial triple-digit growth of the e-book market after the introduction of the Kindle has also slowed significantly, and the majority of that growth was and is concentrated in genre fiction, such as science fiction and mysteries, rather than in the humanities scholarship that university presses more typically publish. Even within the broader world of

scholarship, compared to university presses, the for-profit STEM publishers seem to be reaching a larger proportion of their audience with their digital publications rather than with the print versions of the same content. The most likely reasons for this are the greater need in the STEM disciplines for fast delivery. The journal article typically is the preferred venue for scholarly communication in these fields, and the market and platforms for digital journals are more mature than those for digital books. According to AAUP's annual sales statistics, e-books make up less than 10% of sales for most university presses, with only one press's sales reaching 21%, and that was only for one year (American Association of University Presses, 2014). It also is worth noting that for that particular press, the majority of those books were sold on Amazon, which makes it highly unlikely that libraries used them. As of October 2013, Michael Zeoli (2013) of YBP, the largest U.S. academic library wholesaler, noted in a presentation at the annual Charleston Conference that e-books still only accounted for about 20% of the units his company sold. It also is worth recognizing that although the proportion varies from press to press, based on the composition of their lists, library sales do not typically make up the majority of monograph sales, although it is difficult to say this with great certainty as one can't be sure where books end up after being sold through certain wholesale distributors or online retailers (Esposito, 2014).

One of the reasons that market demand and file production workflows are important to understand when exploring how university presses allocate resources is that, so far, the expense of file creation for e-books remains an investment rather than a recoverable cost of a good sold. There's been a long running misunderstanding of the economics of book production, and specifically file production. Consumers who complain about the cost of e-books frequently point to the lack of a physical product at the end of the production line as justification for why e-books should have a significantly lower price when compared to the print price. But that rationalization often ignores how low the typical unit cost is on a print product. For a 300-page, 6 x 9-inch, all-text monograph, the paperback unit cost, including printing and binding, is about $5, or even less if the quantities printed are in the thousands rather than in the hundreds. The expense of the book is not so much in producing the physical object with its printing and binding, but rather in the book's editing, design, and marketing, and those expenses do

not decrease when dealing with e-book editions. In fact, those expenses actually increase because very different files need to be created, and publishers incur very different marketing and distribution costs. If we include higher-end features in an e-book, such as robust tagging or embedded animation, those would add even more expenses to an electronic edition that will probably generate only a fifth of the demand compared with a typical university press title produced in print.

E-BOOK COSTS OUTSIDE THE PRINT WORKFLOW

So what are the added costs for the e-book workflow? The most obvious one is the creation of the digital files. There are three basic file types needed for submission to the largest e-book platforms a university press would want to use. These types would include a web PDF, with embedded fonts, downsampled image files, chapter and section bookmarking, and all print artifacts removed, such as crop marks. Most library-facing platforms could use that type of file. Next is the MOBI file format, a proprietary format used only by Amazon for the Kindle platform. And finally the EPUB file format, which is sold directly to consumers by some platforms and can be converted to PDF or MOBI for use by others. EPUB also typically is the most useful file type for a press to use for its archive as it offers the most flexibility in subsequent file conversion and modification. If a print book's page composition is outsourced, most commercial compositors also can create all three of those file formats for an additional cost of a few hundred dollars.

The next cost incurred is file submission. Although there is no cost for the actual submission to a given platform, there is a cost for the labor necessary to prepare and submit the files. This varies based on the number of platforms a press works with and whether or not any additional file manipulation is required before the file is submitted. Typically, modification is limited to changing the name of the file to meet the platform's specifications for naming conventions, but it can sometimes include removing third-party content to which the press does not have the digital rights. Although some platforms do not allow submissions without all of the content included in the print book, other platforms allow publishers to remove third-party content. File modification also might include changing references to the ISBN on the CIP page of a book or the removal of a barcode referencing a print edition from the back cover of the book's jacket file.

Beyond the labor involved in file modification and submission, there also is a significant amount of labor involved in the submission of the associated metadata for an e-book. Each platform has unique requirements for metadata submission, so each title requires a separate metadata submission, typically submitted in the form of an Excel spreadsheet. If the platform offers multiple sales and distribution models, metadata pertaining to each of the models also are required. These would include elements such as pricing, use restrictions or the lack of them, and regional restrictions or the lack of them.

To realize the greatest potential from e-book sales, the most important platforms a university press would want to be on would include the following: Amazon's Kindle, Apple's iBook, Google's Play, Barnes & Noble's Nook or Yuzu (Nook is for the retail consumer market, Yuzu is for the textbook market, and a publisher may only submit a title to one of the two Barnes & Noble platforms), ProQuest, EBSCO, JSTOR, and Project MUSE. These eight platforms make up about 90% of the market for a typical university press's e-books. My latest count found over 60 different e-book platforms taking submissions worldwide, but since 60 different file and metadata submissions per title typically are not economically feasible, limiting submissions to those eight would cover most of the audiences a university press would want to reach. If, however, a press wants to expand beyond those eight, it might consider a third-party digital asset distributor (DAD). A DAD handles file submission, including specific metadata submissions and submission using the multiple naming conventions. That choice, however, comes with a price. DADs charge by the number of files stored and/or distributed or by taking a cut of each e-book sold (or both), depending on the DAD and the agreement negotiated. For a press with over 500 e-book titles, that cost could easily pass $10,000 a year. If a press submits titles only to those eight core platforms and publishes about 50 e-book titles a year, the file and metadata submission could easily take up to 10% of a staff member's time, so using a DAD could pay for itself by both freeing up that staff member and increasing the market reach of a press's titles.

LEGACY TITLES, OR THE BACKLIST, AS E-BOOKS

Publishers generally divide the list of their publications into two categories: the frontlist, books published in the last year or two, and the backlist, books published prior to that. The reason for doing this is again economic. The

frontlist represents the majority of costs a publisher is likely to incur in a given year, and the backlist represents costs already incurred. This makes the biggest difference when looking at cost versus revenue. In a given fiscal year, a typical university press publisher is likely to have about 10% of its frontlist titles earning back their costs. This leaves the other 90% of the annual title output in the red. The backlist, on the other hand, typically has a minimal cost after that first year, limited primarily to warehousing overhead and royalties, so the revenue it produces is essential to make up for the new titles published that had greater costs than revenue. In any given fiscal year, a mature publisher with a substantial backlist can expect half of all revenue to come from the frontlist and half from the backlist. New titles, of course, typically sell more copies immediately upon publication than the average title on the whole list, but the exponentially larger number of titles in the backlist, selling fewer copies of each title, can match or even surpass the amount of revenue the frontlist produces. For a midsize university press, this means that the thousand or so titles in the backlist are as important for their sustainability as the 20 to 50 new titles it will publish in a given year.

So it may seem odd that publishers do not always offer their full title list as e-books if the revenue potential seems equal to that of their new offerings, but there are very good reasons for this. Again, cost is one of the primary reasons. Digitizing the backlist is expensive, typically hundreds of dollars per title, depending on the complexity and length of the book, but cost isn't the only reason. The other reason is legal obstacles, among them third-party rights. Whenever a permission was sought and granted, the permission to use that material typically came with restrictions. University presses are in a unique position among book publishers in that as educational mission-based organizations, permissions almost always are granted and frequently without a permission fee, but the permissions frequently come with restrictions on the number of iterations, or on formats that did not or would not cover digital use. So a book with a photograph might have a cost-free permission for use in a university press book, but if the permission noted that it was for the hardcover edition of the title, that photograph would need to be repermissioned for use in a digitized version. For some titles that sell well year after year, it might make sense to spend the time to find the photograph's copyright owner and get it repermissioned for digital use, but for a backlist of a thousand titles, there isn't a simple way to do this on a large scale.

There is also the issue of the copyright infringement liability clause in author contracts and how the practice that developed around that clause has created a third-party permission documentation problem. A very typical infringement liability clause in a university press author's contract will put the onus of ensuring everything in a manuscript is either the work of the author or, where fair use might not apply, that the author has secured permission to use any third-party content. Typically it is the author who has secured the permission, not the publisher. This situation creates a problem if a publisher wishes to digitize an older title; the publisher needs to know who owns the third-party material and what permission parameters were first granted. Unless the publisher asked the author for a copy of all of that documentation when the book was initially published, and then kept those copies of that documentation, staff would need to start the permission process for a digital version from scratch. If the book is heavily illustrated, or if it includes poetry or song lyrics, the resources needed to do all that work would likely far exceed any revenue a digitized version of the book might be expected to earn, and thus the book simply doesn't get digitized.

Another possible legal obstacle is the author contract itself. Do older contracts with authors that include no specific wording about e-books still allow a publisher to release an e-book? Some author contracts might include language permitting the publisher the right to publish a manuscript "in all forms," and many publishers consider that wording sufficient to assume that it is permissible to issue the title as an e-book; however, further down the contract where royalty payments are enumerated, there would not be an e-book royalty listed, that is, no guidance on how the author should be paid. Many contracts might have an "all other uses" royalty clause that typically refers to subrights such as translations, serial rights, or film versions, but those often default to 50%. In the case of an e-book, if the publisher must incur the cost of repermissioning and digitizing, and then return half of all e-book proceeds to the author, does it still make sense to bring out that e-book version when it will compete with the print version on which the publisher is more likely to only be paying a 10% royalty and of which there are still likely to be plenty of print copies?

The approach to these challenges during my years at the Pennsylvania State University Press were twofold. First, we sent a letter to the authors of

our backlist titles explaining these challenges and asking those who could afford it to sign an addendum to their original contract waiving royalties on e-books. The second thing we did was use our books database to identify titles that were very likely to have no third-party content. We also built a web portal in that database that used the Google Books Project API to allow us to inspect the CIP page of each title, where permissions are frequently mentioned, and sample a few pages looking for third-party content use. By filtering out illustrated books and examining the CIP page to sample each book's content for possible third-party material, we were able to identify low-hanging fruit and select a couple hundred titles where the risk of unauthorized use of third-party content seemed to have been minimized. In cases where we had author addenda, but the books didn't fall under the filtering criteria, we looked at each book to assess the level of difficulty that repermissioning would entail, and we assessed the market demand of the print version. Again, those titles that seemed to have enough revenue potential to cover the cost of repermissioning and digitizing were included in the digitizing effort. To my knowledge, no plans have been made for the rest of the backlist where there was significant repermissioning needed or where royalties would need to be paid at 50%. As of this writing, fewer than one-third of the titles in Pennsylvania State University Press's active backlist have been, or are scheduled to be, digitized.

THE FRONTLIST AS E-BOOKS

With new titles, the cost of digitizing doesn't exist since the press creates the books using a digital workflow. For most university presses, digital files of some kind exist for books published after 2000, so, other than nominal file conversion costs, most of those titles can be added to e-book platforms. It's useful to note, however, that although post-2000 contracts often expressly include e-book rights and a sustainable royalty, the third-party rights issue continues to impede the inclusion of some books. Some rightsholders have been reluctant to allow the use of their work, especially illustrations, in digital form. They have also sometimes set parameters that simply wouldn't be feasible in a digital context, like limiting the number of "views" or even iterations, which cannot always be measured. Other third-party rightsholders might impose a time limit, like five years, after which the image must be removed. Many rightsholders see the ease of duplication in a digital

medium as a threat to the control of their content and thus charge a premium for digital rights. These kinds of restrictions may sometimes mean that only a print edition will be published, or that a particular title can't be included on a particular platform, because that platform's model might not limit the use of the content sufficiently. An example of this would be Project MUSE, where book chapters are allowed infinite downloads, which would conflict with both an iteration limit and a time limit.

For the most part, publishers would prefer to include as many titles as possible on as many platforms as possible, so the default tends to be inclusion, unless a legal issue prevents it. But it also is becoming evident that certain models are becoming rather problematic for publishers, so inclusion on all platforms may not be in a book's or a publisher's best interest. Demand-driven (or patron-driven) acquisitions and the typically accompanying short-term loan option, in which rentals or purchases only occur when certain use thresholds are reached, is one example. Although there may not yet be enough evidence to conclude that this model always will produce significantly lower revenues for university presses, one thing is immediately clear: this model is guaranteed to delay the majority of a title's revenue until one year after publication. Frontlist e-book titles put in platforms using this model are less likely to earn the same amount of revenue that their print counterparts do or—perhaps the past tense is more appropriate—did. Not only is the revenue deferred but, with its growing popularity among academic libraries, this model is also significantly cannibalizing print sales. This leads to inevitable price increases and then complaints from librarians about those higher prices (Stearns & Unsworth, 2014).

Another problematic model for publishers are those of Project MUSE and JSTOR, whose platforms allow a library to purchase a title at or close to the single copy price, thus impacting the textbook market for a title. Books that have textbook potential have had up to ten times more downloads at institutions where the books are used in courses. The only recourse that a publisher might have to this situation is to raise the price well beyond that of a single copy. If a title will be downloaded ten times more than all other titles published in a given year, is a tenfold increase in price even enough considering the course is likely to be taught year after year? Not only does a model allowing this practice hurt frontlist textbook revenue, but it will continue to hurt that title's revenue even in the backlist.

CONCLUSION

Raising prices on books isn't something that university presses want to do. The word *revenue* has come up often enough in this chapter that it is worth addressing why this is an overarching theme in how university presses think about libraries and e-books. Most university presses have the core mission of knowledge dissemination; cost recovery is an ancillary goal, so one might think that putting a book in a library where it can get ten times the average use at the same price as any other book would mark the pinnacle of success for that mission. In reality, the parent institution may not support its press's mission to that extent. Instead, as universities embrace the trend to be more like businesses, they look at their university presses as places with low costs and high impact. They tend to see university presses less as educational instruments and more as revenue centers, not unlike how many universities see their online course programs. In thinking about scholarly communication in the digital age, it would seem that the cost of that dissemination should be going down, trending toward zero, and the system should be embracing open access; but that's not happening. There are a couple of reasons for that, beyond administrators who are not currently interested in replacing a revenue center with a cost center. One has to do with the nature of what university presses do. As noted earlier, it is not the production of the physical book that creates the bulk of the cost, it is the need for very smart people to edit, design, and market the book. Although technology has aided significantly in reducing the costs associated with book production, algorithms are not yet sophisticated enough to do what editors or designers do.

The other reason that a shift to an entirely open program may not be the best idea is something that is not easy to see or admit, and that is that markets can often be significantly more efficient at informing people about a piece of scholarship, and about getting it to them quickly, than the open web. I am astonished at the number of scholars I know who prefer to look up a book on Amazon than on either Google Scholar or their own institution's online public access catalog. They also can have that book delivered right to their office in a couple of days or, in some places, on the same day. Although open access publications do a great service for scholars and scholarship, it is also important that they be in places where scholars look, and that means being in markets. The other benefit of engaging in markets is

to reach nonscholars, or those who might never know about scholars' work were it not included in Amazon's catalog, and, unfortunately, to be in that catalog, the content needs a price.

Getting back to those first university press e-books, knowing which two were not bundled as disks packaged with physical books enlightens where we have been and where we may be going. The very first university press e-book actually turns out to have been two books, both from Oxford University Press: *The Oxford Dictionary of Quotations* and *The Oxford Shakespeare: The Complete Works*. Both of these came as freeware bundled into every NeXT computer that went on the market in October 1988; Steve Jobs negotiated a 74 cents-per-copy royalty (Isaacson, 2011). But the other early e-book that was not bundled with a print copy was the University Press of Virginia's *Afro-American Sources in Virginia: A Guide to Manuscripts* ("First book," 1994). The most notable fact about that e-book might just be its price—it was free. It also was produced in a collaboration between University of Virginia's press and its library. Perhaps the place where university press e-books are going is the exact same place where they first started.

REFERENCES

Association of American University Presses. (2014). *Digital book publishing in the AAUP community: Survey report: Spring 2014*. Retrieved from http://www.aaupnet.org/images/stories/data/2014digitalsurveyreport.pdf

Esposito, J. (2014, March 24). A survey of university presses. *The Scholarly Kitchen*. Retrieved from http://scholarlykitchen.sspnet.org/2014/03/24/a-survey-of-university-presses/

First book to be published on Internet by a university press will focus on Virginia's extensive African-American history records. (1994, January 25). Retrieved from http://www.upress.virginia.edu/plunkett/mfp-release.html

Isaacson, W. (2011). NeXT: Prometheus unbound. In *Steve Jobs* (pp. 224–225). New York, NY: Simon & Schuster.

Packer, G. (2014, February 17). Cheap words: Amazon is good for customers. But is it good for books? *New Yorker*. Retrieved from http://www.newyorker.com/magazine/2014/02/17/cheap-words

Sacco, K. L. (2014). Results of the 2013-2014 student & faculty ebook survey. Retrieved from http://www.fredonia.edu/library/about/EBook%20survey%20results%2020140730.pdf

Stearns, S., & Unsworth, J. (2014, May 27). Ebook pricing hikes amount to price-gouging. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/blogs/letters/ebook-pricing-hikes-amount-to-price-gouging/

Zeoli, M. (2013, November 8). Searching for sustainability: Strategies and choice in the ebook supply chain. Paper presented at the Charleston Conference, Charleston, SC.