"Ebooks: The Next Big Deal?"

John Unsworth
Presented at
eBooks and the Academic Library Market: The Charlotte Initiative Open Conference
March 9, 2017

First of all, thank you, Ivy and others, for inviting me to join you at this meeting (more briefly than I would like) and to share some thoughts about Ebooks. As will become clear in my talk, I am fully supportive of the Principles of the Charlotte Initiative, namely

- Provision of irrevocable perpetual access and archival rights.
- Allowance for unlimited simultaneous users.
- Freedom from any Digital Rights Management (DRM), including (but not limited to) use of proprietary formats, restricted access to content, or time-limited access terms.

I'm also looking forward to the promised outcomes of the project, especially alignment with publishers whose goals are commensurate with library priorities, and the identification of platforms and services for ebook preservation. If I could add anything to your list, it would be to develop agreement among librarians as to a core set of requisite elements in ebook licenses, including explicit definitions of key terms ("purchase," "perpetual access" etc.), and preferred pricing models.

Ebooks and ebook aggregators are looming large these days on the research library horizon, but many unanswered questions remain. Among those questions:

- What does rational pricing look like?
- What information needs to be shared from library to aggregator to publisher, and vice-versa?
- What happens when titles disappear from aggregations?
- Are we headed down the road to something that will be better than our present state, or are we looking at the next Big Deal?

I'll address these and other issues by telling you the back-story behind an angry letter about eBook publishing that I wrote one day in May of 2014, a letter that got published in the Chronicle of Higher Education, and prompted some useful exchanges, in spite of its intemperate tone.

Most of the little I know about eBooks, I learned not at the University of Virginia, but in my previous post as University Librarian and CIO at Brandeis University, where the library is a member of the Boston Library Consortium, and where I was, in the spring of 2014, incoming president of that consortium. At that point, nine BLC libraries-- Boston College, Boston University, Brandeis University, Northeastern University, Tufts University, University of Connecticut, University of Massachusetts at Boston, University of Massachusetts at Dartmouth, and University of New Hampshire—had agreed to participate in the second phase of a pilot program with the Ebrary and YBP, which expanded on a 2012 initial trial in which some of these same libraries pooled about $80,000 for patron-driven acquisition of ebooks. In 2014, we were looking at consortial licensing of ebooks from commercial publishers such as Wiley and Taylor & Francis, as well as from a number of university presses. The libraries, publishers, and aggregator involved in this pilot were experimenting with short-term loans, collective ownership, different pricing models, and terms of use—just about everything. It was pretty clear, in other words, that none of us really knew how this was supposed to work, and… that's what a pilot is all about. As my friend Stephen Downie says: "if we knew what we were doing, it wouldn't be called research."

As it turns out, 2014 was something of a turning point, even a watershed moment, in the evolution of ebooks in academic libraries. As Mark Zeolis notes in his September 2016 blog post on No Shelf Required ("Academic Libraries Are Shrinking, While Content Is Growing. How Did We Get Here?" http://www.noshelfrequired.com/academic-libraries-are-shrinking-while-content-is-growing-how-did-we-get-here/),

> the ebook acquisition trend for one [unnamed] research library over 6 years [shows that] including print, the library was acquiring roughly 1,200 fewer print books and ebooks each year while expending significant funds on Short-Term Loans. While meeting short-term demand, a gulf was opening in resources available for the medium- and long-term.

Michael's on a panel here later today, and will probably be discussing some of these same topics and trends, so I hope this isn't a spoiler, but the graph that accompanies this example is striking: it shows ebooks ordered and ebooks on approval heading down a steep slope from 2014 to 2016, ebooks purchased on PDA creeping up slowly from the bottom of the Y axis, and short-term loans, which had been a negligible factor in 2013, rocketing up the Y axis in 2014 and 2015. In an accompanying graph that represents ebook business in these categories (DDA, STL, etc.) for one university press, Michael notes

> In 2014, their ebook orders had begun to decline, notwithstanding the steady production of new titles. DDA sales were not compensating for overall revenue loss, including sharp print sales decline. STL was the fastest growing category of use, however, the value of that use was inadequate.

The way these trends played out in the BLC pilot was both alarming and instructive.

In March of 2014, at a meeting of the BLC Board, executive director Susan Stearns delivered an upbeat assessment of the consortial ebook pilot to that point. Among her headlines:

- Over 10,000 ebook titles are available to any user of the 9 libraries = **immediate access** for resource sharing
- Short-term loan usage for many titles results in **less expensive resource sharing** – average of $12.29/loan vs. $30+ for ILL
- Each of the 9 participating libraries now owns the 154 **e-books purchased collaboratively** with a total value of $16,152 per library
- The average cost per library is between $6,085 and $8,780 depending on whether or not short-term loan costs are included = **clear financial savings**
- Usage of the purchased titles has been consistent across all of the libraries = **value for each participant**
- Usage across multiple libraries increases over time = **increased value over time**

In other good news, she reported:

- Of the 154 titles purchased through end Feb, 133 or 86% were used by 2 or more of the libraries
- 99 or 64% were used by 3 or more [up from 61% only a month ago]
- 52 or 43% were used by 4 or more [up from 35% only a month ago]
- Usage numbers are rising as the program continues over time

From the perspective of cost-effectiveness, this was going swimmingly:

- Based on an individual library's contribution of $6,085.44 to the purchase of the 154 titles [1/9 of $54,769] and a list price of $16,152.33 for these titles, the savings per library is $10,066.89 with a consortial savings of $90,602
- Total savings across the libraries ranges from a low of $7,919 to a high of $11,777
- If the cost of the short-term loans is added to the library's cost, the savings per library is $7,371.53 with a consortial savings of $66,343.77

However, the view was evidently rather different from the other side of the aggregator. On May 12, Susan received a communication from ProQuest/Ebrary, saying that short-term loan costs that had been 10-15% of list price range were being increased (by some but not all of the publishers represented in our pilot) to 25-40% of list price. We believed that we had negotiated terms at the beginning of the pilot that would hold for the year, at least: we did not expect a sharp price increase in the area of greatest activity, short-term loans, at a point in our fiscal year when most of our collections funds were already spent or committed.

High dudgeon ensued. Susan and I sent a letter to the Chronicle of Higher Education written on behalf of the BLC and endorsed by its management board, which the Chronicle ran with the title "Ebook Pricing Hikes Amount to Price-Gouging." In the letter, we said:

> The BLC recognizes that the scholarly ebook market is a developing one and that publishers need to ensure that they have a sustainable revenue stream as they invest in evolving digital technologies. However, this move looks like an experiment in predatory pricing, designed to make the most of rising demand, but without justification in terms of either production cost or use value. Academic libraries and the universities they serve have already seen the results of this kind of experiment, in the pricing of scientific journals, which sky-rocketed as publishers transitioned from print to electronic delivery.
> http://www.chronicle.com/blogs/letters/ebook-pricing-hikes-amount-to-price-gouging/

There’s actually a pretty lively and informative exchange of views in the comments that follow that letter online, in which an open-access advocate declares that libraries have lost the war and should just give researchers money to buy things for themselves, the Copyright Clearance Center advertised its “Get it Now” program, and a publisher observes that

> STL was sold to publishers as an alternative to ILL and to make additional content not previously available to patrons (owing to financial constraints) discoverable and potentially available in an instant. It was not understood as a tool to be used *instead of* purchase. There's nothing wrong with rental, but the financial model must be sustainable for publishers as well as libraries. The expected use and actual use of PDA/STL do not match.

Susan and I also had some follow-up conversations with publishers, including one of particular interest with a Wiley representative, in which we learned that STLs which didn’t trigger a sale until after the window of time in which their accounting rules considered a title needed to be sold would effectively prevent them from ever recording a sale, even if one was eventually triggered. In other words, they were writing off half-sold books. In other responses to our letter, and to other push-back, some publishers delayed their price increases until the next budget year, others reduced the amount of the increases—and some sailed blithely onward.

In general, the increases came from commercial publishers, but the university press publisher who responded in comments to our Chronicle letter argued that this was just a question of response-time, and that university presses would eventually also act in accordance with market rationality. On the other side of the ledger, Rick Anderson, from the Marriott Library at Utah, wondered aloud on the LibLicense email list, whether

> What we[1]re seeing here … is not just the relatively elastic nature of demand for scholarly books … but also a formerly irrational system--one where books were sold in numbers that had nothing to do with the amount of demand for them--gradually becoming more and more rational as sales start to come more and more into line with demand. That[1]s what PDA/DDA does--it starts to expose what has, up until recently, been largely hidden by the library[1]s traditional just-in-case collection-building practices: the actual amount of

> reader/researcher demand for scholarly books. And the results would be pretty terrifying to me if I were a publisher.

Stepping back from the drama of that moment in May 2014, it seems to me that while we were right to object to unilateral price increases in the middle of a fiscal year, levied without discussion or negotiation, we (on the library side) didn't fully understand what was driving that behavior, and part of the problem was, in fact, the opacity introduced by the presence of an aggregator. What information sharing might have produced a more constructive response to evolving circumstances?

Our understanding from publishers was that they didn't get very granular information from the aggregator about what users were doing with their materials---what chapters they looked at, for how long, what sequence of ebooks they examined in what order, what material is getting annotated, etc.: all of these are facets of usage that could lead publishers to more nuanced interpretation of short-term loans, demand-driven acquisition, etc. It's not that this information doesn't exist: Bob Nardini's 2015 ProQuest Blog entry "A Day in the Life of An Ebook" makes it clear that aggregators' logs do contain much of this information, and probably more. But publishers didn't seem to be getting much of it (http://www.proquest.com/blog/pqblog/2015/A-Day-in-the-Life-of-an-Ebook.html)

Libraries would benefit from more granular usage information as well, especially because (according to Michael Levine-Clark's 2015 study of usage patterns in EBL and ebrary, published in UKSG's journal Insights) there are significant local variations in those patterns. It is, by definition, difficult to generalize about local variations, but we can assume that any library would be more interested in what **their** users are doing with ebooks than in what users in general do with ebooks. Levine-Clark's study also makes it clear that use patterns vary by subject matter. He writes:

> Readers of books in the LC Class T (Technology) look at more pages per session than anyone else except for readers of F (History of the Americas). The other history LC Classes are also high on the list of page views per session. T is right at the top of downloads per session, while the History classes are at the very bottom. Meanwhile,

> readers spend more time per session in the History classes than any others, but less time in Technology books than all but two. Users of History books look at a lot of pages, they spend a lot of time in the book, and they rarely download. Users of Technology books look at a lot of pages and download a lot but spend very little time in the book. http://doi.org/10.1629/uksg.240

This sort of information would be invaluable in negotiating licenses, in designing support services around ebooks, in forecasting their role in instruction, etc.

In May of 2014, we would probably have been better off if we had been dealing directly with the publishers—and subsequently in some cases that is what we chose to do. Dealing directly with the publisher would also have spared us, at Brandeis, the ire of the faculty member who spent considerable time using the aggregator's platform to annotate an ebook of interest in his research, only to find one day that the title had been withdrawn from the aggregation and his annotations had evaporated. There's some hope on this front in the recent announcement by the W3C of a formal recommendation on based the open standards already implemented by Dan Whaley's team in a shared web annotation toolkit called hypothes.is, but for those annotations to persist, commercial ebooks would need to have a permanent address.

That part of the problem can be made easier by choosing a different aggregator, one that aggregates DRM-free ebooks. By 2016, Susan was reporting to the BLC Board that

> The BLC's JSTOR DDA program, which includes 10 of the member libraries, is in its second year and going strong. Usage of the close to 40,000 titles is high and the libraries have purchased over 600 titles at an average cost of just over $30.00. Unfortunately, the evidence-based pilot program with Taylor and Francis, which includes 5 of the member libraries, has not proven successful for a number of reasons and the libraries are in the process of winding down the program this calendar year.

But communication was not the whole story of that day in May, nor are market forces.

The University of Virginia is a relatively modest consumer of ebooks, having only about half a million titles represented in our catalogue in 2016 (which is on the low side within ASERL, our regional library group). What we can see, even with a modest collection, is that most of the use goes to a small number of the titles in these packages. This is quite reminiscent of the usage patterns across the bundles that characterize the Big Deal in commercially published journals, and I expect that we will see bundling strategies gain in popularity with ebook publishers, as a way of amortizing risk across their lists. As Susan and I said in our letter to the Chronicle, we have seen this before, and we should not stand for it again. But why did we stand for it with ejournals? And are we in any materially different position with ebooks?

Brandon Butler, the University of Virginia Library's Director of Information Policy, recently brought to my attention an essay in the February 2017 issue of College and Research Libraries by John Wenzler, Dean of Libraries at California State University, East Bay (and University of Rochester History Ph.D.). The title of this essay is "Scholarly Communication and the Dilemma of Collective Action: Why Academic Journals Cost Too Much." Wenzel's answer to that question relies on Mancur Olson's 1965 book "The Logic of Collective Action: Public Goods and the Theory of Groups," which argues that

> even if all of the individuals in a large group are rational and self-interested, and would gain if, as a group, they acted to achieve their common interest or objective, they will still not voluntarily act to achieve that common or group interest.

The dilemma of collective action, Wenzel explains, is more commonly known as the Free Rider problem:

> Because everyone benefits from the creation of public good regardless of whether or not they help to produce it, individuals are tempted to become "free riders" who exploit the efforts of others. Even those who have no intention of free-riding for themselves are reluctant to invest in a public good because they worry that their effort will be wasted if too many other people chose to ride for free. In small communities, it often is possible to build bonds of reciprocal trust that allow groups to achieve collective ends, but it is more challenging for larger and more distributed groups to do so.

With respect to ejournals, the implication of this theory is that a small community (in this case, publishers) may be able more effectively to promote their interests than a larger but more diffuse community (libraries). In the end, after examining several possible responses to this dilemma, Wenzel recommends that libraries either negotiate in one large bloc, as a union (or what I would call a universal buyers' club), or buy back the commercial platforms (an idea I actually proposed to an AAU provost when Springer was up for sale a couple of years ago), or that they work together to build open-access journal publishing platforms (like Ubiquity, perhaps?) that would allow them to compete with publishers. I think a mix of all three is probably necessary, plus (I would say) the support and promotion of aggregators like JSTOR and new business models for open access like Knowledge Unlatched that are born out of the library community. But more fundamentally, what's needed is a university strategy with respect to risk. In fact, I think what's needed is the strategic acquisition of risk by universities, to push back against commercial publishing's strategic management of risk through bundle pricing.

Universities (and their general counsels) are generally risk averse, especially when it comes to the risk of being sued and paying judgements. In situations where there are counterposed risks—for example, when universities make copyrighted content accessible for those with documented disabilities—they will generally choose whatever seems at the moment to be the lesser of two risks, and that calculation may change as case law evolves. It's harder for universities to take risks in situations where there is no counterweight, though: that's why I still think it is so remarkable that the University of Michigan chose to take a risk on the HathiTrust project. Granted, they are a public university with the mantle of a state's sovereign immunity, and granted, Google had the deep pockets in the original situation: nonetheless, getting sued must have seemed a near certainty, and winning such a suit must have seemed like a long shot. Perhaps it's no accident that an economist interested in the tragedy of the commons, Paul Courant, played a pivotal role in that action, but even so, it is difficult for me to come up with a better—or even another—example of the strategic acquisition of risk by a university. I say strategic, because the particular risk was acquired in the long-term interest of the university's mission and constituency—indeed, not just in the interest of **that** university's mission and constituency, but in the interest of **all** universities' mission, to educate and to do research, and all universities' constituencies—students, faculty, and the future public good.

I mentioned Knowledge Unlatched earlier: they have demonstrated that libraries can direct their purchasing power in ways that reduce the risk of university presses and achieve open access for monographs, in short order. I mentioned JSTOR: their experiments with ebooks demonstrate that there are alternatives to DRM that prevents experimentation with the format (for example, shared annotations like those of Hypothes.is), and that interferes with sharing and preservation. I also mentioned Ubiquity: their pricing model makes it clear that properly edited, professionally produced, peer-reviewed content can be produced for a fraction of what we now pay for it.

Why have we been de-funding our university presses and steadily increasing our dependence on commercial entities for the publishing of our research? Because we have preferred to outsource risk. Publishing is a risky business, in which it's a truism that 20% (or less) of what you publish pays for the other 80% (if you're very lucky), and you never really know which 20% that will be. From a less granular and more university-level perspective, it is often not clear to presidents and provosts why their budgets should be paying for an activity that, almost by rule, will benefit the faculty of other institutions more than their own. University press publishing must often look to them like an absurd exercise in altruism. And considered locally, it is. Considered collectively, though, university-based acquisition of the risk involved in publishing our own work would be a very strategic move, because it would insulate us from further exposure to price increases that we do not control and that are every day presenting us with less and less palatable options. Do we drop information resources on which our research and teaching depend? Or do we reduce library staff (and therefore, library services)?

In universities with Responsibility Centered Management financial models, the choices are even more stark. Want to keep that chemistry journal? You'll have to cut that faculty line in the chemistry department. Don't worry, though: commercial scholarly publishing will find a way to produce and market chemistry research. The logical conclusion to the outsourcing of risk, in other words, is the outsourcing of research itself. At that point, the university will become what some open-access pessimists would say the university library already is—an organization without content of its own, that exists simply to deliver commercially produced and remotely held information to a local audience. Avoiding that future seems worth considerable risk—

beyond the local or individualized risk that now characterize our responses to this unsustainable situation. We need to bump that risk up the food chain: presidents, provosts, and university counsels need to invest in the strategic acquisition of risk. Major universities need to agree on which investments are strategic, and they need to accept that the very nature of this risk is that it will never be evenly distributed across institutions, faculties, or disciplines. The Charlotte Initiative will provide some valuable guidance on which risks are strategic, in the emerging ebook market, so …. thank you.