

The Creation, Management, and Export of Digital Books

By Jack Illingworth, Publishing Consultant
jack_illingworth@canbook.org

Context

This paper is both a backgrounder on the growing business of assembling and monetizing digital book content, and a summary of developments in the marketplace since the 2007 publication of Guylaine Beaudry et. al.'s *The Stake of Publishing*. Since it has been commissioned to brief both the AECB board and secretariat on these issues, it is necessarily broad. At the request of the AECB, this paper moves thematically through the major areas of concern in digital publishing. In the interest of readability, I have chosen to lead with the matters that I believe to be of most interest to the AECB: business models, export sales opportunities, and issues of concern to the Canadian-owned publishing sector. Technical discussion of formats, platforms, and technologies follows, and the paper concludes with a short list of international conferences worth attending and resources for monitoring current developments.

The impetus for this discussion is the intense interest in digitization that has taken root in the Canadian publishing community. This interest was recently heightened by the 2008 sale of almost \$12 million in e-books by 30 publishers to a consortium of academic libraries. In the United States, wholesale sales of e-books for retail to consumers continues to grow dramatically. In Q3 2008, 15 of the major trade publishers sold \$13.9 million in e-books through this channel alone.¹ Growth tracked through the International Digital Publishing Foundation's statistics program tends to be on the order of 55%-70% per year. While \$13.8 million isn't much in the context of the U.S. market, the consistent growth in this sector means that it must be taken seriously. In the past 18 months, many major book trade players, including Amazon, Sony, Indigo, and Ingram, have all made significant moves to consolidate their positions in this emergent market. L'ANEL and the ACP have both undertaken collective projects to help their members move their businesses into this new territory.

¹ International Digital Publishing Foundation statistics.
http://www.idpf.org/doc_library/industrystats.htm

Business Models

There are, at present, only two major business models that have found widespread acceptance among publishers and succeeded in creating significant revenue from e-books: library sales/subscriptions, and retail to readers.

The bulk of Canadian e-book activity has been in the area of library sales, particularly to the academic market. (The domestic library market for e-books published by AECB-client publishers is probably in the neighbourhood of \$4.5 million/year). There are many players in this area, particularly internationally, and some will be discussed in the sales opportunities section. All of these players follow more-or-less the same model: they assemble libraries of content, usually in PDF format, and license that content to libraries. Libraries may then offer the content to their clients through on-site terminals, through authenticated catalogue sessions, and sometimes even through digital downloads. Typically, libraries may choose between subscribing to content or purchasing it for perpetual access, and between single-user and multi-user access. Libraries do not usually select digital titles on an à la carte basis. Instead, they typically purchase aggregated collections of hundreds or thousands of books at any given time.

Occasionally, a similar model is used for direct-to-consumer sales; the Safari subscription platform for technical computer e-books is the notable example.

E-book sales to consumers usually replicate the print retail model in a digital context, selling complete books at a lower price than the print SRP, often using DRM to tie the consumer to a particular store or a particular reading device. Sony, Amazon, Powell's, and most of the major mobile phone e-book retailers use this model.

Other models certainly exist, though none are as dominant in terms of revenue for publishers. Google Book Search pays publisher partners a portion of advertising revenue generated by click-throughs from their books. Wowio has experimented with placing advertisements within e-books, with various degrees of success. DailyLit offers subscriptions to e-books that are delivered daily on a by-the-chapter basis. Bookriff, which has yet to publicly launch, will allow users to purchase pieces of book-content for assembly into a customized POD book. Indigo appears ready to launch a platform that will sell e-books by the chapter. All bear watching, but for the moment the traditional models of library loan and by-the-book retail are strongest.

Export Sales Opportunities

Digitization is beginning to muddy the distinction between an export sale and a domestic sale in an unprecedented way. The overwhelming majority of resellers that are entering the domestic market for digital book content are foreign firms. The few Canadian firms that are making a serious investment in direct digital sales are addressing an international market. For English-Canadian firms, the U.S. market is a direct pathway to certain Canadian sales opportunities. International sales are, of course, worth investigating in their own right.

The largest immediate sources of export sales revenue are the library resellers. The dominant English-language businesses – Ingram's MyILibrary, EBSCO, Follett Digital, Ebrary, Overdrive, Alexander Street Press, and their peers – are all signing up content on a non-exclusive basis, and have proven willing to work with virtually any partner able to put enough content on the table to catch their attention.

Ingram has publicly indicated that it is looking to take advantage of Government-funded content acquisition programs in Europe, the Middle East, Asia, and Latin America. In particular, the governments of Hong Kong, Taiwan, and Japan are subsidizing the acquisition of English-language digital book content for their library systems. Spanish, German, and Italian are noted as the non-English languages that are showing the most growth so far as e-book sales are concerned, but the content that is available internationally is overwhelmingly English.

Aggregators that serve the Francophonie are undoubtedly growing, and much of their business will be out of sight of the New York-dominated English-language e-publishing community.

The biggest challenges in working with these aggregators are:

- Amassing enough content to attract their attention. Most are only willing to negotiate for large collections of books at the present time.
- Negotiating for acceptable terms. Most basic terms offered by aggregators are not very favourable to publishers. However, they are flexible, and publishers who negotiate collectively can secure better deals.
- Proactively identifying and acting on specific market opportunities, especially outside of the Anglosphere. The major aggregators are not always transparent with their publisher partners in terms of disclosing the sales opportunities that they are addressing, or the particular titles that they choose to offer. Opportunities that may be of particular interest to Canadian publishers (such as offering collections of Canadian e-books to universities that host Canadian Studies programs) are undoubtedly being missed.

The export retail market for e-books is also a noteworthy opportunity, although the small size of each sales transaction means that only books or publishers that already have a significant international audience are likely to find the sales to be worthy of the price of admission. Hard statistics on the volume of e-book retail sales through the major channels are unavailable; however, Amazon claims that, in cases in which a book is available in both print and Kindle editions, the sales through the Kindle average 12% of Amazon's sales of the equivalent print edition. Since the Kindle is generally believed to be the most successful e-book reading device in the world (except for, possibly, the iPhone, which has a huge installed base but few in-copyright books available), it can be inferred that only bestselling titles are recouping publishers' investment in accessing this market.

Finally, the settlement of Google's lawsuit with the Association of American Publishers and Authors Guild should be viewed in part as a sales opportunity. This settlement, which will create a non-profit book rights registry that will then allow Google to become a serious player in the library licensing market, could wholly transform the way that the U.S. e-book market functions. Since most Canadian publishers control U.S. rights to their books, they are technically parties to the suit, and their books will be included in the product offerings that Google will bring to market if and when the suit is settled.

Challenges and Concerns Specific to Canada

The chief challenge facing Canadian publishers in this environment is implicit throughout this paper: scale. While digital publishing is often touted as a force that will create a level playing-field for book publishers, it has in fact done the opposite. The real benefit of doing business digitally is realized by those companies that can amass licenses to colossal libraries of content and make it available to millions of users at a very low cost. Independent Canadian firms, who have relatively small lists and

frequently publish to modest audiences, must learn to act collectively in order to secure the terms of trade and access to markets that are available to multinational-owned firms. This will require ongoing cooperation between firms, and industry-wide collaboration that spans publishing associations, policymakers, and private businesses. The necessity for collective action may be ameliorated somewhat once the digital market has grown – a twentyfold expansion in the market for digital content would allow publishers to access markets with a reasonable return – but its benefits are not even negated in that context.

Even if Canadian content can be licensed on a scale that can generate a real return on investment, the future of copyright, and of Canada's relationship to international copyright law, will determine the continued viability of Canadian publishers. Since Canada has not ratified the WIPO treaties, Canadian publishers are not granted certain reciprocal enforcement privileges in other territories. While the enforceability of copyright in an online context is always a difficult proposition, the lack of reciprocal provisions makes the issue even more dire.

Digital Services and Suppliers

The digital services most frequently retained by publishers are conversion (from print to digital, or from one digital format to another) and digital asset management and distribution. At the present, both activities are generally outsourced to foreign firms, although Canadian middlemen are often involved. Conversion services are not well-documented at the moment, but the ACP's conversion project will lead to at least some analysis of the process of retaining these firms.

In order to fully engage with export markets for digital content, most publishers will require a digital asset distributor (DAD). These services act as warehousing and distribution services for digital content. They are not inexpensive, and many of the tasks that they automate (file storage and transfer) are seemingly trivial. The main value they bring to publishers is their ability to manage metadata (the descriptive information that allows files and transactions associated with them to be tracked) and, in some cases, to facilitate content sales to readers or publication online for marketing purposes, as documented in the Beaudry report.

L'ANEL has entered the development phase of a DAD system for its members, with a contribution from the Quebec government and an expected grant from BPIDP. The ACP is investigating securing group purchases of DAD services from third-party suppliers for its members, and is entering pilot-phase discussions with several suppliers.

The chief concerns that are motivating industry-wide action in this area are access to markets, cost, management, and control. A DAD's ability to automate file management and transfer makes it much less time-consuming to send content to resellers and other partners, thereby extending a publishers' market reach and lowering the cost of doing business with those partners. If the digital assets of many publishers can be combined in these transactions, the incremental cost of each sale gets lower, once again making it easier for both publishers and their partners to do business. Collective purchases further lower costs to publishers by securing bulk rates for services, and by amortizing the often large start-up fees over several firms. ANEL's approach has the unique advantage of ensuring that publishers and their associations retain full control over their files, and that the file repository is located in a domestic jurisdiction (therefore freed of some concerns about territorial rights, and likely exempt from requests for user information under the U.S. PATRIOT Act and similar laws).

Reading Software and Hardware

Digital content continues to be read on three classes of devices: personal computers, mobile phones, and dedicated e-book readers. The personal computer remains the most important of the three, since the majority of e-book monetization continues to happen through library resellers (who primarily deliver content to PC desktops) or online publication on the open web.

Mobile phones are the second most significant platform, at least for the moment. However nightmarish reading on a tiny iPhone or Blackberry screen might seem, it is becoming popular. The Stanza e-book software package for the iPhone has reportedly been downloaded over 500,000 times. However, the majority of content available for these devices either originates in the public domain or from unauthorized sources. Publishers in the west have been slow to believe that an opportunity exists on these platforms. However, the mobile phone has been the dominant platform in Japan for many years, and may well prove to be a good means of delivering content to younger readers.

E-paper based reading devices are the scarcest e-book platform in the wild, but their high cost of ownership means that only serious readers are likely to buy them. While an iPhone owner may download Stanza and never buy a single e-book, the same cannot be said of these devices. Amazon.com's Kindle is the dominant device in the U.S. market, but its dependence on prepaid deals with national cellular phone networks has made it difficult for Amazon to launch it in other territories. In English Canada, the U.K., and France, the Sony Reader is the premier, relatively low-priced reading platform. Sony reports that they have difficulty keeping up with demand, but does not release firm sales data.

The salient difference between the two devices, so far as publishers are concerned, is their openness to third-party retail. Amazon is the only company that can sell content for the Kindle. Sony, while it maintains its own content stores for the Reader, has also opened the device to competition from other retailers. Publishers who wish to access the Kindle must accept Amazon's terms, while any firm that doesn't wish to do business with Sony may still open its own store for the Reader.

E-book formats

The electronic storage and transmission of books is usually a two or three stage process. Publishers typically produce a book in one format (a production format), then convert it to another before delivering it to a printer or an electronic reseller (a business-to-business, or B2B, format). Most electronic resellers then either "stream" the book to their customers through an online platform, or convert the e-book before delivering it to the customer as a standalone file (a business-to-consumer, or B2C, format).

Production Formats

Most trade books are still authored in traditional desktop publishing programs like Quark and Adobe Indesign. In recent years, publishers have started to use these programs in parallel with XML-based content management systems that allow the content of books to be quickly and efficiently repurposed through database products, custom books, online syndication, and so on. Adopting an XML based workflow is not a straightforward proposition and may be extremely difficult for some trade publishers.

B2B Formats

The most important B2B formats are PDF and EPUB. PDF is a page-description format: it details exactly how each page of a book should look, and those pages will be displayed the same way on all

devices. It remains the industry standard for delivering page layouts to printers, and is not threatened in that capacity in any way. It is also still the best way of delivering complex layouts to any platform, and is favoured by academic library resellers (like Ebrary, MyILibrary, and EBSCO) for use in their subscription platforms. For that reason, it is probably the most economically important digital book format. All publishers create PDFs at some stage of their production process.

Because it is a fixed description of a page, PDF is not a good format for reflowable text.² A reflowable book changes its formatting automatically to suit the device that it is displayed on, as well as the tastes of its users. This is the way text is usually displayed on handheld readers and mobile phones, but reading reflowable files on a computer is also much more natural than using a plain PDF. The EPUB standard released by the IDPF in 2007 is increasingly the preferred choice for managing reflowable files, and many major international publishers have announced that they will only offer EPUB files of their books to their trading partners.

For the most part, resellers will take EPUB files, and some, such as Amazon.com, prefer them to any other file format.

EPUB creation remains a difficult question. Most publishers outsource this process for now, but Adobe has built EPUB-generation capability into Indesign CS3 and CS4. This has been reported to be problematic in CS3 but may be more effective in CS4, which was only released in October of 2008.

B2C Formats

If one includes the multitude of e-book retailers and devices that exist in the market, there are too many B2C formats already in play to address here. However, most of these formats are proprietary and are used by single retailers (or small groups of retailers) to deliver content to their customers. As a result, publishers need not concern themselves with the minutiae, except for one overarching issue: some retailers, such as Amazon.com, appear to be investing heavily in proprietary B2C formats in an effort to create customer lock-in and therefore develop market dominance. Such market dominance then allows a retailer to dictate prices to suppliers. Apple's iTunes achieved just this: Apple combined a popular device with an easy to use, well-stocked store backed up with proprietary DRM. In doing so it established \$0.99 as the de facto price for a song purchased online.

The formats most commonly used by publishers to sell e-books directly to customers are PDF, EPUB, and Mobipocket. Everything said about PDF above applies here – it is easy to use, good for complicated layouts, and inadequate for portable reading. PDF can be protected with DRM, but most options for this file format are very weak.

EPUB is problematic as a B2C format. While, as an open standard, implementation should be easy, the handful of devices that read it (newer models of the Sony Reader, some iPhone reading applications, the Adobe Digital Editions software application) do not follow the full specification. This means that an e-book may pass all validation tests as an error-free EPUB but still fail to display. Publishers engaged in converting to EPUB need to be aware of these issues. The IDPF is developing best practices and real-world practices guidelines for the format, but they have not yet been completed.

² Adobe's PDF MARS format, which combines PDF with XML, is reflowable, but is not yet widely adopted. For most purposes EPUB is a preferred option.

Adding DRM to an EPUB is possible, (especially by using Adobe Content Server 4, which works with Sony Readers and Digital Editions) but there is no DRM standard associated with the format at the moment. This means that the potential for customer confusion is high. To my knowledge Adobe's DRM for EPUB has not yet been cracked, but it is surely only a matter of time.

Mobipocket is the most dominant of a variety of proprietary formats (others include Microsoft's .Lit and Palm) designed for delivering e-books to handheld devices. All of these formats may be used by publishers and booksellers to sell direct, but they have been eclipsed by EPUB – particularly because none have been able to establish a foothold on the iPhone.

Digital Rights Management (DRM)

The issue of DRM is closely tied to questions of retail strategy and file formats, and is discussed in context at intervals throughout this paper. Some brief generalities about DRM may be useful as well.

DRM is a catch-all term for any technical measure that limits access to or use of a digital file. Sometimes, “social DRM” – the application of digital watermarks to a file to track its unauthorized circulation – is also discussed. However, most DRMs are designed to enforce limits to the situations in which a file may be accessed, or the ways in which they are used. Their stated purpose is usually to prevent the illegal redistribution of copyright material. However, they are also often used to create digital business models that simulate print-based conditions, such as virtual library loans of digital content.

Only those publishers who sell files directly to the public need to engage with the technicalities of DRM. Publishers who work through third-party retailers currently have little or no say in determining whether DRM will be applied to their books before distribution. Typically, the retailer makes this decision; publishers can either sell through that retailer on its own terms or not.

As noted in the “B2C Formats” subsection, retailers who sell files to individual consumers often use proprietary DRMs to create consumer lock-in. In these cases, the uses that the DRM limits are not always in the interests of the publisher: letting a retailer develop a captive customer base that cannot choose from other suppliers is inevitably bad for suppliers. Other business models, such as those used by library licensing companies, use DRM to enforce limits on end-users' use – i.e. they make sure that the content can't be printed, or that it expires once the user's virtual loan period ends.

In general, consumers will eventually resist DRM that is applied to content that they buy to own, mainly because it makes it difficult for them to move to new computers or new reading devices. DRM that is applied to subscription-based services, however, tends to be less of an issue.

DRM can be expensive to implement and its effectiveness at curtailing unauthorized distribution is debatable. Once a DRM becomes popular among content producers, someone inevitably finds a way to crack it: even the encryption on Blu-Ray discs, which was one of the most robust DRM technologies in the world, has been broken. Once a DRM has been cracked, it becomes easy for moderately sophisticated users to unencrypt content. It only takes the online posting one copy of an unencrypted e-book to seed that book throughout the world. However, studies that have been conducted to date, including an experiment with DRM-free mp3 audiobooks published by Hachette U.S.A. and distributed by emusic.com, suggest that consumers who purchase DRM-free content seldom, if ever, post that content to peer-to-peer networks.

Furthermore, many consumers dislike the ways in which DRM can curtail formerly legitimate uses of content. The music industry, which pioneered using DRM in an online context, has now all but abandoned it for this reason. The only major holdout in selling DRM-encrypted music is Apple, and this is now the decision of some of the major recording labels, who appear to be attempting to reduce iTunes' market dominance by only allowing its competitors to sell unencrypted files.

E-commerce and online stores

Since the publication of Guylaine Beaudry's report *The Challenges of Digital Publishing*, there have been relatively few transformative developments in the online retail of print books. The changes that have happened have been questions of marketing, acquisitions, digital books, and long-term strategy. Amazon.com has been the most active retailer in this area: it has purchased the secondhand bookselling platform Abebooks.com and the book-driven social networking site Shelfari. In purchasing Abebooks.com it acquired a minority stake in Librarything, a competing book-based social networking platform. Amazon's core digital offering, the Kindle, is discussed elsewhere in this paper. All of these measures serve to consolidate Amazon as the dominant online retailer in the English-speaking world, and they give it a strong technology base which could be leveraged to improve its standing in other markets. Indigo, too, has rolled out a book-based social networking site, which they report draws a significant international user base.

Print-on-demand (POD)

It is important to distinguish true POD, in which print runs may be as small as a single copy, from the short-run digital printing that is sometimes used for class set reprints, advance reading copies, and so on. While there is plenty of short-run digital printing capacity in Canada, and most publishers have integrated it into their publishing or marketing programs, there is still very little domestic POD capacity. Many Canadian publishers are therefore looking to POD to extend their export reach or make their export programs more efficient, as it offers a cost-effective means of keeping small volumes of particular titles available in territories where demand does not justify shipping or warehousing traditional stock. It also allows publishers to respond to spikes in demand more quickly than the traditional book supply chain permits.

The North American leader in this area is the Lightning Source division of the U.S. wholesaler Ingram. Lightning Source is closely integrated with Ingram's wholesale business, meaning that books that are held on a print-to-order basis in Lightning Source's system are visible to booksellers as in-stock at Ingram's wholesale operation. This has the obvious benefit of making these books perpetually in-stock in the one of the largest U.S. trade wholesalers. Amazon.com has a similar relationship with its POD division, Booksurge.

Ingram has announced its intention to expand to the UK and Asia, and some European companies, such as CPI Publishing Solutions, are working to offer similar platforms in Europe. The benefit to publishers of participating in a territorially or globally integrated POD platform is obvious: more territories may be served more efficiently, with less time and money lost to shipping and warehousing. However, for large volumes of books, traditional printing and supply chain management are still more cost effective.