



Challenges for The Publisher in an Uncertain Market

Opening Observations

First, it is always advisable to know the perspective of your observer. In this case, the observations come from a long term player in the industry. More specifically, this writer has been in production management for STM/E and professional publishing for decades. While there can be limitations to the views of an industry participant, there is the advantage of first-hand experience and even a personal sense of urgency about the challenges now facing publishers.

Those challenges are multi-fold. Of course, there is the unsettled nature of the current market. Rapidly changing technology is fueling the drive into as yet-fully-defined digital world. And there are important questions about how a publisher should respond.

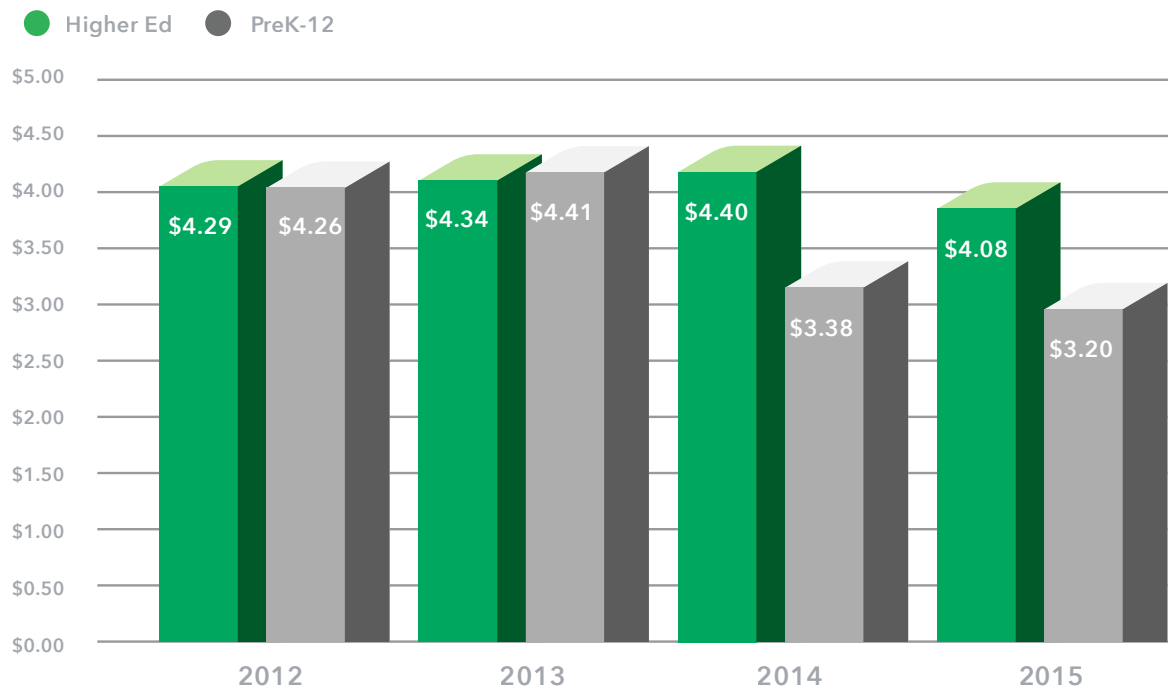
I believe that it is the very nature of unsettledness that helps inform a way forward for us publishers. There are critical adjustments and preparations that one needs to make in order to be flexible and ready to respond rapidly to changing demands. The focus, not surprisingly, is on content.

The digitization of content has been an obvious goal for many years, and it takes on greater significance with the economic pressure on print output. As publishers, we are now discovering that we must make some critical and not-so-obvious decisions about the creation, structure, execution, and user experience for digital products. When dealing with massive collections of archival files, a single wrong move could be costly in terms of both time and money. There are some common mistakes that can be avoided or corrected now, during the critical stages of a digital strategy initiative.

Chart 1

U.S. Textbook Revenues 2012- 2015

Higher Ed and PreK-12 Textbook Rev (\$billions)



Source: Association of American Publishers' Stat Shot

The Uncertain Market

"What the Blazes Happened to Print Sales?"

The demise of print has been famously and erroneously predicted for years. In the early 1990's, the CEO of a major professional publisher announced that "print is dead." To his credit, he publicly recanted that statement several years later. Despite the incredible advances in digital technology and new opportunities for selling e-product, print sales have remained the bread and butter of most STM and professional publishers.

Print, however, is not the solid foundation it used to be. In recent years, most sectors have seen an overall decline in print sales. There has been a definite downturn for Pre-K through 12 textbook revenue. From 2013 to 2014, revenues dropped an astounding 23.3%, and then another 5.3% drop in 2015. Higher education textbook performance has also dropped, showing a 7.3% slump between 2014 and 2015 (See chart 1)

Chart 2

Trade Units sold Print vs. eBooks in U.S

Print vs eBook Unit Sales (millions)

● Print ● eBook



Source: Neilson Book Scan

For Trade titles, according to the Association of American Publishers, demand for print actually increased by 5% in 2015. It is important, however, to take the long view here. A particularly telling picture is what has happened with print and e-book unit sales over the past 6 years (see Chart 2). Despite the recent modest increases, print unit sales have declined significantly since 2009. While we are likely to see some variations over time, the overall trend does not seem likely to reverse.

Is it possible that we are now experiencing the tremors that foretell a major tectonic shift in the industry? Well, we do know that electronic product has been growing in many sectors (with a recent exception of Trade), and we have most certainly been told for years that it is the future. But I believe most publishers would agree that future has yet to define itself.

The promise of revenue from electronic product remains strong but somewhat elusive. No one can predict what new form of product will provide the greatest potential. We are in something of a limbo right now, in which new e-products and market opportunities have yet to come clearly into focus.

Publishing strategy meetings can be frustrating affairs. There are many intriguing notions of what will be the next big thing, but the truth of the matter is that no one knows for certain. However, there is one strategy that is both obvious and critical—be prepared for all eventualities.



Preparing for Multiple Eventualities

"Jack Be Nimble; Jack Be Quick"

The key to being prepared to move in whatever direction the market dictates is the publisher's own content archives. Ensuring safe storage and easy accessibility to content is an obvious first step. In the past, publishers have been known to be all too casual about the preservation of files. In many cases, we publishers paid almost no attention to archival storage, instead depending on third party typesetters and printers to maintain our intellectual property. Even now, some smaller and mid-sized publishers are guilty of such dangerous practices.

I have witnessed any number of acquisitions in which the acquired company is unable to produce a complete set of archives. Part of the acquisition process becomes a somewhat frantic search among the files of third party vendors. In one case, a vendor who served as the sole source of a publisher's archives had just shut down an old office. As a result, all the old files were destroyed just weeks before the purchasing company requested them.

Larger publishers have known for some years now the importance of maintaining an archival system. In some cases, unfortunately, their files are a relative hodge-podge, with no automated or systematized method of retrieval. The larger the collection of old titles, the more problematic this becomes. Finding the correct format of the correct title is one challenge; ensuring that the file is still in existence is another. Many major publishers have tended to adopt some sort of product-data repository (PDR) or content management system (CMS). Whether internally developed or vendor supported, such systems become the foundation for all existing and future uses of content.

In any case, the publisher needs a complete and fully searchable system for finished product files. Such a system is essential in responding to an emerging and evolving market. Whether creating the 2.0 version of an existing e-product, developing a new online site, or merely converting to a format required by a promising aggregator, the publisher needs to be quick. Such projects can no longer take months or even years to complete. Those publishers able to respond most rapidly to market changes will have the greatest potential for survival.

Overlooking the Need for Consistency in Vendor-Produced Files

“Trust But Verify”

Not only does a publisher need easy access to files, but those files need to be in a consistent, ready-to-use condition. A publisher must be diligent about what is being archived. Publishers often believe that they are maintaining a single version of their content and are surprised and frustrated to find they are not. A familiar mistake derives from the common practice of using multiple third party vendors for the production of front list titles. It is not unusual to ask vendors to provide final files in the various formats needed for each e-product title. There is nothing wrong with this practice except that the publisher needs to be aware that every vendor will develop each format to its particular set of specs. As a result, a publisher using seven different vendors is likely to end up with seven separate interpretations of the same digital format. So, instead of one consistent ePub3, one ends up with seven different ePub3 versions.

In an example from math publishing, consider that the standard software used by vendors for interpreting equations is MathType. But MathType exists in multiple versions, and each version has differences in the way it interprets and displays equations. There are presently at least three versions in use, and in all likelihood the publisher using multiple vendors will discover all three types of math in its files.

Similarly, consider the situation with fonts. Unless all of a publisher’s vendors have been certified to carry all the fonts used by a publisher (good luck with that), there is the strong possibility of an unwanted substitution for any irregular font. Font collections cost money, and vendors are reluctant to purchase a new font to handle one particular job. It is therefore not uncommon for vendors to pick up that stray Greek symbol or foreign phrase as an image. Although that symbol or phrase will display fine in the print version of the product, all semantic meaning has been lost in the publisher’s files.

The differences will essentially corrupt the publisher’s archives and may remain transparent until it’s time to re-purpose content. Any automated conversion program depends upon a consistent file structure. Without such consistency, the program will fail. At that point those seven different versions of one format become one huge headache. Oh, and as for those Greek symbols or foreign phrases, embedded as images, in the digital version, those characters will not scale. When the end user changes font size, the image stays the same.

Some publishers may argue that this problem does not apply to them because they provide their own detailed specifications for each format to every vendor. That may be true, but it underestimates just how easily inconsistencies can be introduced. Content needs to be treated as a most valuable asset. No CFO would argue that since everyone has been provided financial regulations, there is never a need for an audit.

A number of publishers have now built in a review-and-correct (Quality Assurance) step that occurs prior to archiving files. This QA process may be an internal exercise or a task assigned to a third party vendor. In any case, the objective is to identify any discrepancies from the publisher’s own specs and then to return the file to the original vendor for correction. So, when dealing with multiple vendors, best practice is to take Ronald Reagan’s famous advice, “Trust but verify.” That simple measure for ensuring quality/structure consistency is key to a publisher’s readiness for any eventuality.

Overlooking the Need to Make DTDs and Schemas Specific to the Publisher's Content

"A Model Is Not a Spec."

Industry standard DTDs and schemas are good models for storing a publisher's content. They allow consistency across businesses for storing files on multiple platforms. However, choosing a DTD or schema for content is just the beginning of the process.

Equally important is the specification created for a publisher's content. The publisher's spec ensures that content will be retrieved and archived with a consistent structure and at an appropriate level of granularity. For instance, BITS is the standard DTD for books. Choosing BITS for book titles makes sense since it is being embraced across the book publishing world. However, it is important to keep in mind that by the very nature of its structure, BITS is extremely generic. By itself, the DTD will not address specific nuances of a publisher's marketable output. No two publishers of philosophy titles will organize and present content in the same fashion. Each publisher will want the spec to address the tags required to capture and identify all aspects of its content at a granular level. And that will require a spec created for its own brand of BITS. A DTD without a built-in publisher's spec puts that publisher in a vulnerable position, dependent upon how vendors might determine to tag content.

If the publisher does not have someone within the organization who can complete a spec for their particular content, then it would be prudent to seek outside help. There are consultants who can assist. It is a mistake to adopt a DTD and charge ahead converting content without first addressing the specifications required to preserve the qualities of that content. Such a move could have costly consequences in the near future.

Choosing a System

"Do We Need a Honda or Do We Need a Tank?"

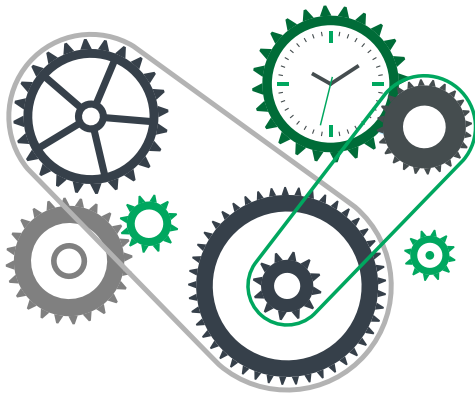
At this point the discussion circles back to a topic raised earlier in this paper: the need for a dynamic system to manage content. The question then comes up as to what kind of system best suits the publisher's needs.

Basically, there is the need for an interactive system that warehouses finished goods along with their associated metadata. A repository containing both content and its associated metadata will be called, for the purposes of this paper, a product data repository, or PDR. Note: the term "repository" is used here with some hesitation; there are unfortunate connotations associated with it. Still to be found in Webster's is a definition of repository describing "a burial vault or tomb," and it is just that notion publishers need to avoid. A good way to think of a PDR is not as some static storage area but as an active digital warehouse for managing content and data that are vital to the business. Words have power.

Typically, a PDR will provide an overall structure for storage, as well as features for search and retrieval, utilizing associated metadata. Content stored in a PDR will normally be in the form of finished goods and not be accessed by any dynamic delivery system, like a web server.

A PDR is often a simple relational dbase, providing sophisticated search and retrieval while remaining user friendly. It can be internally developed or can easily be adapted from a third party system to the publisher's particular needs. Simple scripts can execute the publisher's own business rules by evaluating the metadata. Workflow management can be easily introduced to the PDR architecture.

The economic advantage here is that such a system is relatively inexpensive and quick to employ. The PDR is an assembly of content. It is not possible to semantically search across an entire library of digital content in a PDR. One can, however, prepare content at a very granular level for possible future use. With digital object identifiers (DOI's) stored as metadata, a publisher can tag individual chapters or even objects within those chapters for a time when they might be accessed dynamically.



Perhaps the most important question for publishers to ask is will a PDR meet their most immediate business needs. Often, those business needs are tied to storing, tracking, accessing, and retrieving finished goods content quickly and reliably. Such modest needs could be metaphorically represented as what mode of transportation is needed to drive from New York to Boston. If the publisher is not planning to go to battle along the way, then perhaps a Honda is the best bet. If it serves the publisher's most urgent needs, then a PDR may be the cheapest, quickest means to an end.

On the other hand, while a content management system (CMS) can replicate the functionality of a PDR, its power lies in the structured assembly of content, allowing the publisher to perform a semantic search across an entire library and dynamically retrieve and re-purpose discrete elements. It is key to interactivity between user and content. Imagine that an STM publisher has customers with an urgent commercial need to search out specific data tables from within a collection of technical titles and then re-display those tables as bar graphs. Now imagine that the customer's requirements are such that he needs to update those bar graphs on his website every month. Going back to the original data tables, he makes the necessary updates and sees them automatically reflected in the bar graphs. The software and interactivity required for this is absolutely dependent on a CMS repository.

It is probably true that most large publishers are going to need a CMS eventually. What is little recognized is that publishers rarely need such a powerful tool when managing content at the book, article or even chapter level. The market is tempting us with possibilities, but there are presently not that many real life commercial demands requiring CMS functionality. Much of what we publish today conforms to the book paradigm, with the eBook being little more than a digital facsimile. The PDR fits most present day commercial needs. However, as web-based content delivery platforms take hold (such as for education and scientific research), we are seeing a steady migration toward workflows with digital content taking precedence over print. And with the advent of digital-first authoring environments (Inkling, Habitat, Metrodigi's Chaucer), the CMS becomes inevitable.

A CMS is not going to be an out-of-the-box solution. A publisher's particular business rules and workflow management needs will require months of development. In order to function as intended, the CMS will require consistent tagging across all content. A CMS is not inherently user-friendly. If everyday users are going to easily retrieve and validate downloads from a CMS, they will either need to be familiar with query language (unlikely) or the system is going to need an interpretive layer to accommodate them in a more user-friendly mode.

Going back to the comparison of the PDR to a Honda, the CMS is something of a tank. Systems are incredibly dense, powerful, often based on Alfresco software or adapted from pre-existing systems, such as R-Suite, based on MarkLogic. The dynamic interactive re-purposing of content at a granular level is the metaphorical equivalent of going into battle—it requires a tank. It is complicated, expensive, and time-consuming to build, but it is the right vehicle for the job. It is not, however, the right vehicle for driving from New York to Boston.

A PDR is easily ported to a CMS when the time comes. The two system fit hand in glove. For most publisher's immediate needs, the answer may be to start with a PDR and let the market dictate when a CMS has become necessary.



Summary and Conclusions

“Where Are We, and Where Do We Go From Here?”

Publishers are facing a number of potential pitfalls going forward. Some of these are related to an uncertain market and evolving technology, but some can be of our own making. It is a challenge to reconcile our own business strategies with fast-changing and complicated technologies. And there is always the chance of making costly mistakes.

The market for publishers appears to be in a shifting world between print and electronic. While the fortunes of print continue to decline, the promise of electronic product revenue has yet to be fully realized. There is confidence that the digitization of content is going to eventually recoup lost print revenues, but there is not yet any clear vision of what the market is going to demand going forward.

The best strategy is to be prepared for any eventuality. Most significantly, that means ensuring that archival content is consistent, well structured, and easily retrievable and that new titles are properly vetted for quality and consistency before being archived. It also means not letting content structure be defined by full service suppliers and being diligent about creating specs for the DTD or schema.

Such diligence and forbearance is often lost on senior publishing executives. And a certain amount of C-level humility is a healthy thing. It takes a bold CEO to say to his tech team, “I don’t really get it. Explain it to me in layman’s terms, and let me repeat it back in my own words. Then, please, correct me when I am wrong.”

Here in 2016 there is a good deal more certainty about the digital market for content than there was a decade ago, when Amazon launched Kindle. Still, a number of markets are far from mature. What is certain is that the bold executive decision isn’t about choosing one format over another. It’s not about predicting what markets will do. It’s about being prepared to execute and deliver content to whatever market presents itself.

Print is not dead; it’s just not the forgiving market it once was. Digital is also here to stay; it’s just not quite certain what form (or format) it will take. Publishers need to put their houses in order to serve both markets. For no matter what, the economic future of our business will be decided by both.

Brought to you by CodeMantra.