Commerce in Digital Publishing: Trends

November 2007

VISTA

a division of
Publishing Technology

Part 3 of MANAGING AND EXPLOITING PUBLISHING ASSETS
A series of research reports commissioned and published by
VISTA, a division of Publishing Technology

COMMERCE IN DIGITAL PUBLISHING: TRENDS

This study of Commerce in Digital Publishing represents the third report in a recent series from VISTA, a division of Publishing Technology. The subject areas explored by these papers include:

- *Digital Asset Management,*
- *Print on Demand*, and now
- *Commerce in Digital Publishing.*

This new report is based on a synthesis of surveys and structured conversations with executives at a dozen companies whose products include trade, professional, higher education, and scientific, technical, and medical publishing. The report focuses particularly on trade e-books, and journals and books in digital form, with some reflection on digital higher education materials. Although digital audio and video are of increasing importance, this study deals only with products that were, are, or might once have been, books and journals. The report makes no claim to be a statistical snapshot of the publishing industry as a whole, but rather is an informed overview based on these conversations and survey responses as enriched by other white papers, reports, conferences and the like.

In order to provide further insights about digital initiatives in trade publishing, a case study delving more deeply into major publisher HarperCollins' approach to digital commerce in trade publishing has been added as an appendix.

Finally, the three studies together present a comprehensive examination of the issues underlying the digital marketplace and indicate the approach being taken to the digital future by publishers of all kinds.

Sincere thanks for their continuing support go to George Lossius, Morayéa Pindziak, Charlie Rapple and other members of VISTA and Publishing Technology, without whose insight and contributions this work could not have been carried out.

prepared for VISTA, a division of Publishing Technology
by James Lichtenberg
November 2007

Contents

EXECUTIVE OVERVIEW

- The predominantly "large" and "traditional" publishers with whom we spoke have considerable experience in selling text-based digital content, some for more than a decade.

- Nonetheless, the evolving digital marketplace still feels "new" to all the players. "It is hard to function in this new, fascinating world, where people expect content for free."

- All segments of publishing are increasingly united by the single most powerful force in the digital marketplace: search and discovery (Google *et al*).

- A wide divergence in terms of file formats, business models, and delivery avenues, as well as the powerful, unpredictable impact of search, has created an atmosphere of deep uncertainty in the industry.

- The granularity of discovery (a chapter, an article, a page, a drawing) that search makes possible threatens to dissolve the commodity forms on which publishers' business models are based.

- There is remarkable confidence that digital sales will increase over the next three years, even though millennial predictions for the digital future were considerably overstated (particularly in trade publishing).

- Full interoperability across devices and content format is seen as being the most important driver for growing the marketplace for digital content.

- A new focus on the ultimate consumer, as opposed to retail or distribution partners, characterizes many responses.

- Conversion of all books to digital format is essential, as the use of search engines and online stores continues to grow, in order to drive sales of *physical* books.

- Digital products represent a business that is "riskier" than that with which publishers are normally comfortable.

- In order to be successful, publishers must become truly digital companies, including work-flows, infrastructure, and business models.

THE MANY BUSINESSES OF PUBLISHING

As it has evolved over the last century, book and journal publishing comprises at least half a dozen *different* businesses. For 600 years, a single production technology, the printing press, masked such differences. These separate businesses, however, are distinct in many ways:

- content,
- customer uses,
- supply chain interactions,
- business models, and
- marketplace dynamics.

In the future, the common content form (a digital file) and delivery mechanism (online) may blur the distinctions between these separate functions. However, in these relatively early days, digital technology has served to further disengage them. As this report demonstrates, the landscape and operational considerations of trade books in digital form is proving to be markedly different from the business dynamics of traditional print, which also holds true for all the different types of publishing.

Consequently, publishers now face a set of complex challenges. The objective of this white paper is to understand the strands of this complexity, to inform publishers as they attempt to transition successfully into the digital world.

The dozen companies with whom we conversed define themselves as “large” and “traditional.” All of them are currently publishing content in digital form. Almost all have been doing so for at least five years – and a third of the companies for more than 10 years.

An important question that emerged during research for this white paper is: since the sale of digital content is not new, why does it *feel* that way even to the publishers with a number of years of experience?

The simplest answer is "uncertainty", of which the nature and consequences are examined in this paper. It is further demonstrated that much of the uncertainty results from the speed with which the industry evolves. As another media CEO has been quoted as saying, "We are in a time of maximum paranoia, and minimum clarity." Yet despite all the uncertainties, when asked whether sales of digital content will increase, fully 100% of those with whom we spoke were confident about growth in the digital future.

Do you anticipate over the next three years that your company's sale of digital content will increase?

	Response Percent
no, it will decrease	0.0%
minor increases	0.0%
moderate increase	45.5%
significant increase	**54.6%**

Nonetheless, the scope and extent of digital commerce remains very unbalanced across publishing's several business lines. Karen Hunter of the STM giant Reed Elsevier is quoted as saying (*Trends 2007* p. 107), "In the scientific and technical professions, the migration to digital is almost completed." (It should be noted, however, that the large markets of health and other professions are still in the process of migration.) By contrast the sale of trade e-books remains at $25 million annually, a fraction of the millennial estimates by Arthur Andersen that, by 2005, digital sales would represent greater than 10% of total trade sales – or $2+ billion. And there is even less progress in higher education, as will be discussed below.

'SEARCH' AND ITS IMPACT ON BUSINESS MODELS

The unfulfilled Arthur Andersen prophecy epitomizes the pre-millennial expectation that the digital future would unfold incrementally and seamlessly, ploughing a straightforward furlong that would reflect and extend existing publishing trends and practices.

However, the digital future was moved onto a profoundly different track from that which was confidently expected during the high enthusiasm of 1999–2000. The catalyst for the change of direction was the explosion of search technologies offered by behemoths such as Google, Amazon and Microsoft. The spur to publishers' gradual and then acute anxiety in relation to their own digital capabilities was Google's announcement, in December 2004, of its agreement to digitize *all* the holdings of six major libraries, to which several more have since been added.

This announcement caused much perturbation, and not just from the perspective of the business issues tied up in copyright. Search was then the only unifier of digital publishing, enabling any kind of digital content, with the proper metadata, to be digitally discoverable. Yet this heralded problems: notably, that it challenged the invulnerability of traditional products. The granularity of discovery (a chapter, an article, a page, a drawing) introduced by more comprehensive search technologies threatened to dissolve the commodity forms on which a wide swathe of publishers' business models were – and in many cases, still are – based. Publishing-specific technology companies responded with 'information commerce systems' which have become the instrument by which publishers turned the potential threat of increased discoverability into an opportunity, to be optimized through granularizing the sale and delivery of online content.

However, embracing the commercial opportunities of this digital de-structuring of published content does pose a major challenge to publishers in terms of content forms and business models. STM publishing is an $8 billion enterprise *(Trends 2007*), with a primary B2B business model based on journal subscriptions to libraries, as well as academic, scientific, technical, medical and business institutions. A significant percentage of this business is in digital form. But, as one respondent put it, "What happens when the customers and ultimate users realize they don't necessarily need the whole journal and start requesting just the article(s) they require? Will subscriptions, as a business model, still be viable?" And, one must then ask, could the resulting business model of 'pay per article' support the current infrastructure of STM journal publishing? As Harvard Business School professor Clayton Christensen (*The Innovator's Dilemma*) has noted in his studies of "disruptive" innovation, the real disruption power of new technology is not the new device itself, but rather the disruption of a company's legacy business model. Companies that can't adapt quickly to this kind of disruption usually don't survive.

Defining viable business models in digital publishing is therefore among the most important challenges facing publishing over the next decade. Curiously, although digital commerce in fiction and narrative non-fiction e-books remains stuck at a very low commercial velocity, those works that involve immersive reading and are carefully structured may ultimately be the *least* vulnerable to the radical de-structuring into component parts that digital search and discovery makes possible.

Business model concerns are based on even more fundamental questions about digital works. Conferences and blogs have been buzzing with relevant exchanges, and high on the list of concerns are some basic issues, e.g. what constitutes a digital work and what do you "own" if you purchase one. As Evan Schnittman, Oxford University Press' Vice President of Business Development mused in a recent blog posting:

> "Consider this - a consumer has found a book (or journal article) and the full content is available instantly, for a price. But what price? Should a publisher establish the prices? Should Google [in the case of content that it has digitized] establish a specific pricing policy? What should the relationship be between print pricing and 100% online access pricing? ... These questions are neither academic nor is there consensus on the answers ... Pricing will surely be one of the most complicated issues for publishers as very few of us have experience with *business models that offer content access, not ownership.*"

Other thoughtful commentators such as Peter J. Dougherty, Director, Princeton University Press, have wondered what books will even comprise as the evolution of digital works becomes more robust.

Trade publishing has, to date, been spared these issues; in the late 1990s, major retailers were panicked by the possibility that digital versions might underprice the printed version and cannibalize their sales. They therefore pushed back so decisively that publishers abandoned experimentation in gauging market response to bundling print and digital products or different pricing models. Higher education publishers, in contrast, have long produced high quality electronic products, but have largely side-stepped the issue of the price of digital materials as these are offered as marketing inducements and are provided mostly for free once a professor adopts an edition of a printed textbook.

HIGHER EDUCATION: MUCH SMOKE, LITTLE FIRE

In the mid-1990s when the possibility of digital textbooks appeared on the horizon, there was great excitement among college publishers. Digital technology could lead to always up-to-date materials, all kinds of audio-visual teaching elements, and perhaps the end of the used text book market. The arrival of the University of Phoenix and its totally online courses that required purely digital materials, and major users of digital work such as the University of Maryland's University College, were seen as hopeful signs.

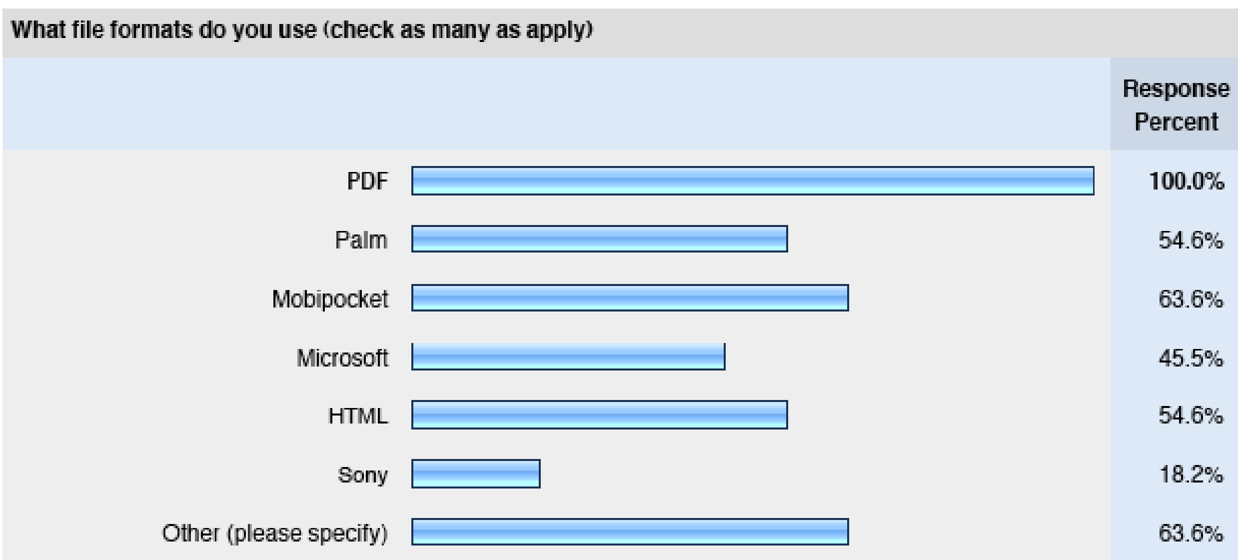
In addition, even as the market place of some two dozen publishers contracted, through acquisition, to three large players and three smaller ones, publishers continued their focus on creating excellent digital materials. As noted above, these were mostly given away free as ancillaries to encourage professors to adopt a physical book. However, as an executive at one of the larger houses observed, 96% of that which college publishers produce is still traditional textbook materials. Of the 4% that is digital, fully a quarter is produced just for the University of Phoenix. This particular company continues to create digital products, having published over 500 e-textbooks in 2006 as well as pursuing other digital pilots. To be sure, all major textbooks have websites and digital ancillaries, but there is what this executive called "a fly in the ointment." By and large, professors aren't interested in using digital textbooks. Thus, for digital content to be significant in the college marketplace, it will require “a profound change in way professors teach”. And the world of professoriate moves slowly.

RENEWED ENERGY IN DIGITAL TRADE BOOKS

By contrast with the slow-moving progress toward digital materials in higher education, the trade e-book marketplace has shown signs of renewed energy and focus in the past year. These include:

- the new Sony Reader
- Amazon’s acquisition of the French e-book company, Mobipocket, and its ensuing announcement of an e-book reader
- the standardized e-book file formats recently developed by the International Digital Publishing Forum.

However, this energy is flowing against a multi-year tide of complexity in file formats and reader systems that has effectively chilled early customer interest. While PDF is used universally by publishers, Adobe (unlike Apple with iTunes and the iPod) has not rushed in to lead the market by creating a successful reading device and a customer-friendly business model.



As if this chart were not diverse enough, the publishers who responded “other” also pointed to an equally large array of other formats including XML, HTML, SGML, ETI, and WMA. In terms of the use of file format, one publisher candidly declared: “As long as the DRM (digital rights management for security) is good, we go into it.”

A similarly broad spectrum exists with regard to business models used for digital commerce, although in this case, diversity may work to the advantage of customers by allowing choices as to how to access content. Fortunately, a number of companies are bringing technological solutions into the marketplace that make it relatively easy for publishers to create a whole range of new digital products, whether simply elements of an original longer product or highly customized aggregations that serve a specific customer's individual needs. Going forward, publishers' ability to create new business models as needed to fit these new product categories could give them significant competitive advantage as the market for digital content grows and evolves in richness and complexity.

Which business models do you use to sell content? Please estimate, even if only very roughly, the percentage of tot each category. If it is zero, please enter 0%

	Response Average	Response Total
Subscription	**48.25**	**386**
Individual titles (ebooks)	25.40	254
Chunks	17.57	123
Bundled (Print plus digital)	27.00	162
Time based access	28.33	85
Pick and mix (credit-based)	3.67	11

To capitalize on this diversity, however, publishers will need to work back up the chain to the contract with the author. This will need to be similarly flexible in order to permit delivering content to the customer across a spectrum of different format, delivery and price options, as well as incorporate rights for onward delivery of data held within the article (such as individual images), as may become possible going forward.

Who are your major retail outlets; and approximaly what perecentage of your digital sales are they responsible for

	Response Percent
eReader.com	27.3%
Sony	18.2%
Fictionwise	9.1%
Others (please specify)	**90.9%**

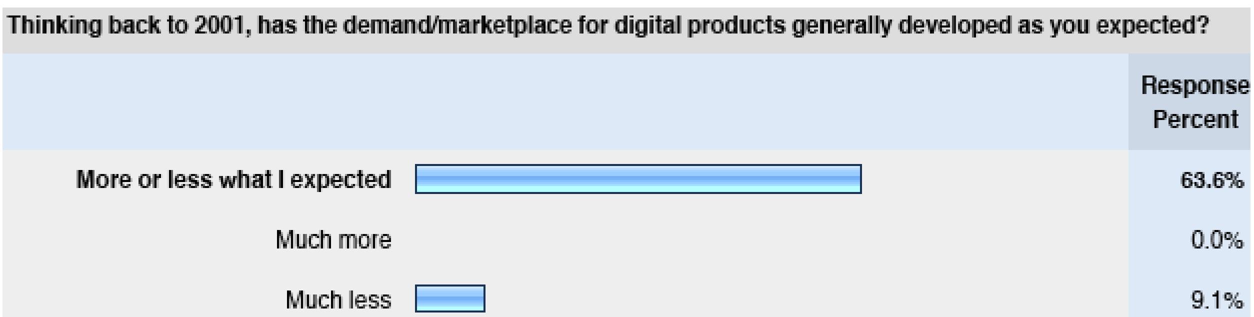
Multiplicity also applies to the retail outlets used by publishers to sell their content. Along with their own website, used by about 75% of those we spoke with, quite a range of outlets were cited beyond the several options listed in the survey.

The "other" category was not only the largest in terms of percent, but the one that garnered the most individual comments.

- “Amazon”
- “e-books.com”
- “do not use any major retail outlets”
- “Microsoft, Palm, Audible, Sony, although digital sales only totals 1% of the company revenue”
- “We are banking on Amazon and Google... In the handheld market, we do the most business with Handango”
- “MBS Direct, Follett, Nebraska”
- “ProQuest, Books 24X7, Lightning Source, Overdrive”

FACTORS DRIVING DIGITAL COMMERCE

As noted above, the atmosphere at the turn of the millennium was quite different to today. In hindsight, respondents say that they expected pretty much what has happened in terms of the digital marketplace.



Some of the specific comments give a more nuanced picture, such as: “B2B electronic content licensing has been a very good business for a long time. But *e-books,* simple electronic versions of a print book, *have been much less than expected or hoped for.* That distinction is very important.”

At the turn of the millennium, the expectations for trade e-book sales were such that a gap of this sort was not anticipated. Technology companies and industry experts were confident that the e-book market would grow quickly and robustly. Unfortunately, as indicated above, trade e-book sales were hobbled right out of the starting gate by a combination of:

- immature technology in terms of readers and software,
- competing file formats from players like Microsoft, Adobe and Palm (each of which seemed to believe it could corner the market),
- lack of consumer familiarity (not to mention confusion) as to use and value of books in electronic form, and
- a frustrating lack of interoperability across platforms for those who had the courage to try this new format.

Scientific, technical and medical publishers, on the other hand, had worked through many of these issues for their products published in the early 1990s on a CD-ROM platform, so that moving to an online screen version, often from SGML to HTML, was not a technological hurdle. Moreover, their professional audiences and library customers were experienced with digital content and able to handle the change without much difficulty. In fact, they welcome it, as can be seen by the wholesale move to digital in this marketplace.

Quite interesting are the comments about what is driving further growth of the market, particularly in trade publishing, as can be seen by the aggregate responses, in the chart below, to the question*:*

"If you anticipate change in your approach to bringing digital content to market, what is the major driver for this change?"

	Response Percent
Better DRM	0.0%
Standards	9.1%
Consumer interest	36.4%
Improved internal digital processes/applications	9.1%
Outsourcing	0.0%
Others (please specify)	**45.5%**

Consumer interest is, reasonably enough, seen as an important driver by about a third of the respondents, though once again, "other" is the largest category and includes some interesting comments:

- "better graphical user interfaces"
- "devices will narrow down; we expect to explore selling chunks to websites ... I don't know, but there are millions of cell phones [which] might work for certain kinds of content."
- "migration from books and journal to the web for information needs."

However, when asked directly about the importance of a "holy grail" device – "the right device at the right price", one that gives the reading experience of a traditional book – surprisingly, twice as many respondents answered that this was "*unimportant*" as said "important". Whether the newly-released iPhone, in some future iteration that supports text from books and journals, or Amazon's e-book, Kindle, will change both perceptions and marketplace dynamics remains to be seen. Certainly, as has been observed by the press, Sony's eReader is not yet a game changer.

Respondents' expression of the importance of "interoperability" (indicated by the following chart), can also be seen in the evolution of the cell phone market. The hype before the release of Apple’s iPhone notwithstanding, consumers in fact prefer a range of devices as long as content can be used on any or all of them. In sum, the key to marketplace success with trade e-books would appear to be not one device, but many, as long as there is “interoperability” across them.

Interoperability across platforms	Response Percent
Important	**90.9%**
Somewhat Important	9.1%

Essentially, respondents agreed with the importance of a system that works more like the original DVD marketplace: any disk, with any content, on any player. (It is interesting, now that movies and other rich video content can be downloaded with greater and greater ease, that content providers have again fallen to squabbling about the technology of the next generation of discs – which could well be the last.)

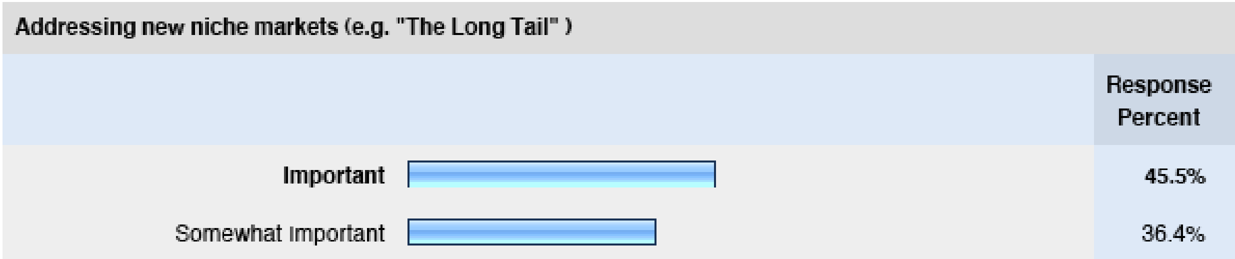
Similarly, most respondents felt that satisfying latent customer demand is also a major driver in growing the market.

Satisfying customer demand that would otherwise go unsatisfied	Response Percent
Important	**63.6%**
Somewhat Important	18.2%

The primacy of consumer interest as a driver to growth was echoed in the importance that publishers gave to the idea of “greater familiarity and comfort with digital content among consumers”.

Greater familiarity/comfort with digital content among users	Response Percent
Important	**72.7%**
Somewhat Important	18.2%

A corollary to this issue is the growing perception of new markets and opportunities which were originally described, by Chris Anderson of *Wired*, as the Long Tail. Once again, the focus is on satisfying a much broader range of niches and customer demand.



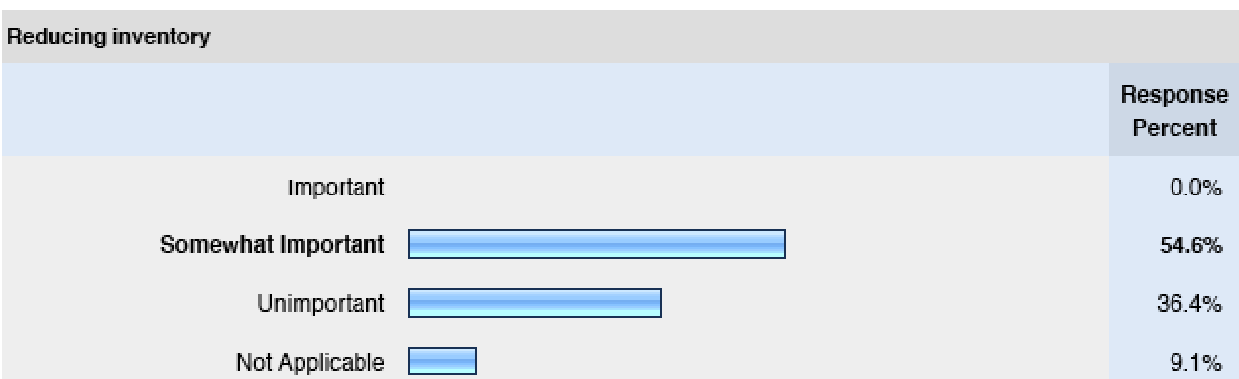
This is the consequence of a digital world in which customers are able to search, discover and acquire books from among some three to eight million titles (depending on the service), as opposed to being limited to perhaps 100,000 that a large bookstore chain can handle. One estimate reads as follows:

Product Type	*Online retail*	*Bricks and mortar retail*
Books	3,000,000	40,000 – 100,000
CDs	250,000	5,000 – 15,000
DVDs	18,000	500 – 1,500

This reality highlights a point made in the HarperCollins case study (see appendix to this paper), that publishers *must* have their products – or at least, metadata about their products – in digital form if for no other reason than to make them accessible to the search engines that will drive increasing revenue via the sale of what remain *physical* products.

OTHER CONSIDERATIONS

Interestingly, the questions of reducing inventory, unit costs of production and distribution, insuring that books don’t go out of print, and keeping rights from reverting to authors were not by and large seen as important concerns in terms of digitizing books and journals.



Neither were concerns about creating channel conflict reflected in the answers to the degree that seemed predominant a few years ago.

Too great a risk of creating channel conflict with distributors and print retailers.	Response Percent
Important	18.2%
Somewhat Important	18.2%
Unimportant	**45.5%**

Nonetheless, an STM company explained how they are encountering friction with their print journal agents as the demand for print journals continues to fall. As the AAP/PSP's 2004–2005 Annual Survey reports:

- Global revenue grew considerably; $6,104 million in 2005, up 12.9 percent ($5,409 m) from 2004.
- Purchasing patterns continue to shift; "print only" revenue fell by 2.6 percent, only to be replaced by combined sales of "print and electronic" and "electronic only" titles.
- The industry's shift from print to electronic delivery continued with *just 54 journals remaining in a "Print Only" form.*

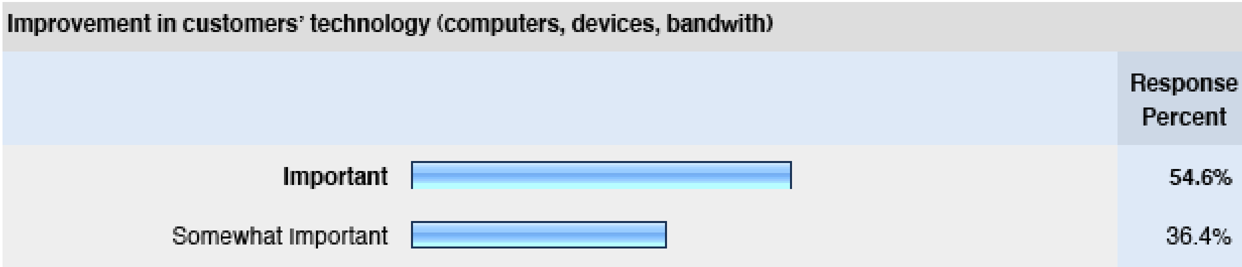
The potential for new revenue however, was high on the list of important considerations.

Improve top and bottom line performance	Response Percent
Important	**72.7%**
Somewhat Important	9.1%

Similarly 100% the publishers we spoke with felt that the generational shift and the growth in the cohort of born-digital customers will be a significant factor in growing the sales of digital product.

Generational shift as the 'born-digital' cohort matures	Response Percent
Important	**90.0%**
Somewhat Important	10.0%

Technology, moreover, must keep up with rising demand from these new customers.



Summing up the obstacles to success in digital commerce, with the explosion of search capabilities and the number of websites that have begun to request publishers' content, we find a daunting array of issues for publishers to deal with. Among these, the ongoing confusion that has resulted from copyright laws and business models based on print products has emerged as particularly vexing. A venerable if complex business, publishing is not perfectly adapted to the dynamics of modern corporate enterprise. Furthermore, with digital technology, we are following a moving target in that underlying software and physical devices are still evolving rapidly, in an environment where the technological capability of customers is very uneven. Finally, different publishers have different perspectives as to what is actually happening, and diverse motivations, which makes industry efforts on common issues more difficult.

Nothing remotely resembling this confusing situation was brought to light in the course of preparing the first two VISTA-sponsored reports on digital asset management and print on demand. The discussion of commerce in digital content has opened up a vast, new, uncertain vista.

LOOKING TO THE FUTURE

Yet, notwithstanding all the challenges that this report has highlighted, when asked what major changes they expected over the next three years, our respondents looked hopefully to a range of important improvements:

- interoperability among devices and files,
- color e-ink screens,
- Amazon's launch of e-book readers and content,
- digital content on many types of mobile devices, and
- some e-book bestsellers and big hits.

By the same token, some respondents anticipated developments that would definitely change the face of publishing as we know it. These included:

- the continuing growth of the power of Amazon and Google as distributors of publishers' digital products,
- content becoming secondary to the services provided along with it,
- journals becoming obsolete as smaller content types emerge, and
- content and services being delivered as customized products for individual consumers.

And all of this is occurring in an enduring atmosphere of uncertainty which, to sum up, includes the following:

- too many file formats;
- unstable, evolving business models;
- the mushrooming growth of wireless devices and the possibility of high-value content being delivered as easily as cell phone conversations or e-mail;
- lack of standards, and standard contracts;
- poor quality of metadata;
- immature and evolving technologies;
- inexperienced customers with naïve expectations;
- no "holy grail" reading device; and especially
- *a business that is riskier than that with which publishers are normally comfortable.*

However, it is inevitable that in the near term, the archetypal customer approach to acquiring both physical and digital products will be: "Search, Discover, Access, Acquire." Publishers are faced with the need to respond to two different demands: on the one hand, to create infrastructure, product formats, marketing and business models that support the sale of digital content itself, and on the other, to make all works available in digital form in order to maximize the revenue and marketing value of consumers' booming interest in search and discovery.

The implications of reconciling these two different demands lead to conclusions that are even more radical. It seems clear from these responses and conversations that an attempt to bolt a digital business onto a legacy ink-on-paper business will not be fruitful. To enjoy success with digital delivery of content, publishers must revise their specific publishing endeavor altogether. This includes putting in place:

- digital workflows,
- digital asset management,
- digital asset repositories,
- digital delivery of content, and particularly
- continuously evolving business models to optimize the customer experience and fit with the developing nature of digital products and services.

To profit from delivering text-based content in digital form, a publisher must let go of the "book/journal" business metaphor and understand that "fluid content" is king. It is necessary to embrace the reality that the ultimate customer is the *only* customer and intimately to understand, even anticipate, evolving consumer needs. As a consequence, radical customization becomes the driver of the underlying business model, however that specifically plays out going forward. As noted above, even the demand for "radical customization" to meet the needs of the 'customer of one', can now be satisfied by new technology tools that automatically extract and aggregate from a publisher's digital archive whatever content elements customers might request.

To facilitate this from a business perspective, however, each publisher must become a "holding company" whose job is to support lines of quickly evolving different businesses. To use a term from the world of software, publishers, at least for a time, must live in a state of “perpetual Beta”, having the agility to experiment, analyze and then move quickly in order to adjust as the demands of the marketplace change. To do this it is essential to re-conceive and re-connect all the elements in a company. A silo-based operation will not succeed in the digital world.

In sum, success with digital product will belong to those publishers who can become truly digital companies.

∞

Appendix
HARPERCOLLINS' DIGITAL STRATEGY: A CASE STUDY

The evolution of HarperCollins' digital content strategy is the outcome of several interacting conditions:

- being part of a dynamic global corporation,
- embracing the market-changing impact of digital technologies and search,
- pursuing ongoing process innovation, and
- maintaining, even in the digital world, a traditional publishing culture towards authors and readers.

This tradition stretches back to the company's origins in 1817, with authors like Mark Twain, Charles Dickens, and the Bronte sisters. Acquired (as Harper & Row) by News Corporation in 1987, HarperCollins was formed in 1990 when the News Corporation bought British publisher William Collins & Sons and merged the two entities. Publishers of literary and commercial fiction, business books, children's books, cookbooks, and mystery, romance, reference, religious, and spiritual books – with some two dozen imprints from Amistad to Zondervan – the company's revenues exceed $1 billion annually.

In terms of the sale of digital content in the trade book marketplace, and the use of digitized assets to promote sale of books in physical form, HarperCollins has staked out a leadership position. Over the past decade, the company has iteratively realized an evolving vision as articulated and supported by Jane Friedman, the company's often-honored worldwide president and CEO since 1997.

e-books – a reasonable first step

Like publishers large and small, HarperCollins was excited by the millennial expectations for e-books. In February 2001, the company announced a newly created imprint for all their e-book products, PerfectBound (now HarperCollins e-books). Their strategy was, and remains, to publish titles selectively, with full editorial design, marketing and publicity support. As Ms Friedman put it: "We will not just digitize books and release them into cyberspace." In 2003, HarperCollins made the decision to sell e-books to libraries, in the hopes that libraries would play a critical role in educating readers about the benefits of this developing format.

Notwithstanding HarperCollins's efforts and strategy, and those of many others, whether publishers, distributors or technology vendors, combined sales of trade e-books from all publishers are, in 2007, still only a small fraction of Arthur Andersen's millennial estimate of $2+ billion in sales by 2005. As discussed more in depth in the Executive Action part of this report, e-books account for about $25 million in annual revenues across the industry. This is due to a number of factors in a still under-developed marketplace that go far beyond the actions of any individual publisher.

"Search" appears on the digital horizon

In 2004, even as the hopes for e-book sales had dimmed, an important new player appeared on the digital horizon: "search." While search had long been a component of the Internet, led by companies like AltaVista, by 2004 Google, Yahoo, and Microsoft were in the forefront, collectively changing the game by focusing "millions of eyeball" on search results, and their attendant ads. Dozens of search engines, each with its own servers, were springing up all over the internet. Our conversations with a range of publishers point to the fact that, at this point in time, the only thing that unites commerce in digital content in general, across all of publishing, is "search".

Brian Murray, President, HarperCollins Worldwide, was one of a group of executives closely monitoring the implications of "search". In December of 2004, Google's announcement that it had reached an agreement to scan the collections of half a dozen major libraries, "surprised publishers, and raised red flags for us," Murray recalls. "We began asking ourselves: did we have the right digital strategy? Did sending copies of our books to dozen of different search engine partners, who were then making proprietary digital copies for their own commercial purposes, make any sense?"

During the Frankfurt Book Fair in 2005, they discussed internally the idea of writing an RFP: "why not invite companies to present ideas for what HarperCollins might do?" Working with Accenture, they created an RFP in about two months. Part of the strategy was to announce the project via the *Wall Street Journal*, and they invited WSJ reporter Jeffrey Trachtenberg to learn about the evolution of their digital strategy and the RFP process itself. "His jaw dropped when we showed him what we were doing," said Murray. The seven-figure project attracted some 30 vendors, among which were both Google and Amazon.

The response to search: a "Digital Warehouse"

HarperCollins' solution was to create a "digital warehouse" that would consist of digital copies of the active backlist, approximately 15,000-20,000 titles, to which would be added the digital versions of as many as 3,500 new books each year. HarperCollins was adamant that the word "warehouse" and not "repository" be used for the following reasons:

- to make the project instantly understandable to the rest of the publishing community; and
- eventually to fulfill their aim of recreating *all* the functionality of a physical warehouse – storage, distribution of content and marketing materials, protection of assets, and market transactions.

By creating such an asset, (in Trachtenberg's words) "HarperCollins would be able to produce digital copies of its books and then make them available to search services offered by such companies as Google, Yahoo, Microsoft and Amazon.com, while maintaining possession of the digital files." Another underlying rationale was clear as well: "HarperCollins's effort to make search companies use its digital copies is an aggressive response to anxieties felt by publishers worried that they will lose control over their intellectual property (on the web)." (WSJ 12/12/2005)

HarperCollins' strategy, however, was larger than simply protecting intellectual property. It contained a number of important dimensions:

- Getting books in front of the growing millions of digital consumers via search results and websites in order to grow sales.
- (As Friedman noted) "A digital infrastructure ... will allow us to protect our rights and the rights of our authors. We will make all of our books available digitally, but *we* will store the digital copies and license them out to those who want to use them."
- Building on how their "print sisters" (newspapers and magazines) handle the challenge of search engines, especially the "crawl" specifications and contracts that set the rules for how, and how much, content is discovered.
- Versioning control and tamper protection.
- Extending the reach of authors.
- Leveling the digital playing field for retailers.
- Creating a service for other companies to use, just as happens with the distribution of physical books.

In April 2006, HarperCollins announced that it had chosen NewsStand as its partner to develop the warehouse. In January 11, 2007, the publishers signed an agreement with the NewsStand division, LibreDigital, to provide publishing services for the industry.

As Murray and his LibreDigital partners outline it, however, the potential of the digital warehouse could turn out to be significantly broader and more nuanced than any physical warehouse. HarperCollins may one day be able, for example, simultaneously and interchangeably to license any content, to any website while controlling and customizing exactly the terms of the license, the amount of content that can be sampled, the countries in which it is visible, etc. on a site by site basis – a flexibility impossible in the physical world. Not surprisingly, this both creates, and is the result of, business dynamics that are very different from those of the familiar ink on paper world.

The Business Dynamics

One of the most important implications of our survey results is that a company cannot simply bolt a digital business onto its legacy ink on paper business. A re-structuring of the company, and its workflow, are essential for cost-effective publishing in the digital world, whether for marketing of traditional products online, or sale of digital ones.

The HarperCollins experience bears this out. The decision to respond to search by keeping control over digital content led to the strategic decision to create a digital warehouse. Even without an anticipated, calculable ROI, a range of other adjustments were then needed to support what turned out to be a fixed, seven figure investment in creating the warehouse itself. Thus, an unintended, but valuable, consequence of the digital warehouse were other changes, changes, as it turns out, that are necessary to support the full functioning of digital commerce. These included, among other:

- a fully digital workflow,
- digital text management, and
- digital archiving.

Murray notes, "We are lucky to be part of a global company that is relentlessly dedicated to spreading its infrastructure costs globally. The strategic review and decision about the warehouse accelerated a number of other projects that had been back-burnered. We got lots of 'green lights' which allowed us to scale up quickly."

To speed the transformation even more, Murray and his colleagues were not burdened by an overly-restrictive or bureaucratic development protocol. Rather than specific project management demands starting with project plans, tests, various gates for proceeding as the project unrolled, and ending with a fixed "launch", HarperCollins followed a process more like what the software industry calls scrum: doing what you need to do in the moment, changing and readjusting as conditions change, moving in short "sprints." "This was a new culture for us. We learned on the go, from our mistakes. It was a lot more 'internet' than traditional publishing. For example, we released a widget that people could put on their own sites to link back to our books, then realized we had to fix it. And so we did. We are in a state of continuous improvement, what has been termed 'perpetual beta.'"

Finding The ROI

As a consequence of the 'green-lighted' projects, once the final file of a book is created, HarperCollins can provide digital content to search engines and websites at zero incremental cost. This fact, as it turned out, began to point to an ROI: the cost of converting a single physical book, by having it re-keyed, and checking and editing it three times, is about $4,500. Therefore, if HarperCollins puts 3,500 new titles a year into its digital warehouse, this represents a savings of almost $16 million annually. Moreover, studies have shown that a title that is searchable with good metadata and marketing information, enjoys somewhere between 6% and 10% higher sales than a book which cannot be found via search or placed on websites. If you add to this the fact that printing costs represent 20% of a publishers overhead, the potential for digital product, as has been recognized for at least a decade, is considerable.

For all their success, Murray recognizes that HarperCollins is only at the beginning, and there is much to discover and resolve. "We worked closely with NewsStand

and LibreDigital, but getting the scans right was harder than anyone thought." However, now even seasoned book production managers are content with the outcome. By the same token, "new books moved through the process, through the digital workflow, better than we expected."

To the often-raised question as to whether HarperCollins or other publishers' digital warehouses will be able to meet the speed and accuracy of responding to search queries that Google or other search engines require, Murray counters with, "It's about the overall consumer experience. They question our speed, but we ask them about the quality of their scans, which are certainly not always up to our standards."

The Future

There is a significant difference between using digital technologies to market physical books online, and the sale of text-based content in digital form. Will the e-book market eventually take off? Will new digital trade products emerge? These are questions for which neither Murray nor anyone else in trade publishing have an answer. "Right now, we don't know what the right product is."

Comparing the level of commerce in digital products in trade publishing with that of STM publishing is like looking at a seesaw unbalanced completely in one direction. Today, if some $8 billion in annual sales from subscriptions to STM products (at least half of which are digital) is compared to $25 million in e-book sales, the contrast is stark. However, as statistics show, the part of the STM market that is like the trade e-book marketplace, i.e. single article pay per view downloads, has risen to about $8 million in 2005, which puts $25 million in annual sale of e-books in a more positive light. Moreover, as one survey respondent noted, the STM market could change radically once customers discover that they don't really need or want an entire subscription, just specific articles, and they are willing to pay for them.

Clearly such a profound transformation will not happen overnight. If, however, thanks to initiatives like HarperCollins' digital warehouse, digital trade books, and other new products were to find their 'groove,' while STM were to undergo a change like the one just described, the position of the seesaw might look much more balanced in the future.

In the meantime, and although they are still in early days, HarperCollins seems to have found a positive and productive path to its own digital future.

VISTA, a division of Publishing Technology, is the leading worldwide provider of solutions and services for the publishing industry. VISTA is a committed business partner to publishers, distributors and booksellers supported by our systems throughout the world. Our solutions fit all aspects of the publishing value chain, from author care through production, rights and royalties to warehouse and fulfillment – all integrated in an e-business environment, with full access to the analytical power of Publishing Intelligence.

To complement our comprehensive enterprise solutions, the full range of VISTA services include applications implementation, applications management and hosting services, and publishing-specific business consulting, together with lifetime customer support.

VISTA was founded in 1977 and has offices in Europe, North America and the Asia-Pacific region. In 2007, VISTA merged with Ingeta to form Publishing Technology. Visit us on the web at **www.vistacomp.com**.

Additional copies of this exclusive report may be obtained from VISTA:

Europe
Link House
19 Colonial Way
Watford, WD24 4JL
United Kingdom
T +44 1923 830200
F +44 1923 238789

North America
80 Cottontail Lane
2nd Floor
Somerset, NJ 08873
United States
T +1 732 563 9292
F +1 732 563 9044

Asia-Pacific
5 Alexander Street,
Crows Nest
Sydney, NSW 2065
Australia
T +61 2 9906 1222
F +61 2 9906 1441

Internet
www.vistacomp.com
solutions@vistacomp.com