

The Battle to Define the Future of the Book in the Digital World

by Clifford Lynch

Abstract

Commercial publishing interests are presenting the future of the book in the digital world through the promotion of e-book reading appliances and software. Implicit in this is a very complex and problematic agenda that re-establishes the book as a digital cultural artifact within a context of intellectual property rights management enforced by hardware and software systems. With the convergence of different types of content into a common digital bit-stream, developments in industries such as music are establishing precedents that may define our view of digital books. At the same time we find scholars exploring the ways in which the digital medium can enhance the traditional communication functions of the printed work, moving far beyond literal translations of the pages of printed books into the digital world. This paper examines competing visions for the future of the book in the digital environment, with particular attention to questions about the social implications of controls over intellectual property, such as continuity of cultural memory.

Contents

[Introduction: Hyped Machines, Hidden Agendas and Visions of the Future](#)

[Defining Digital Books and E-book Readers](#)

[Digital Books as Literal Translations of Printed Books](#)

[New Content Genres: Reconceptualizing Books in a Digital World](#)

[Converting Older Books to Digital Form: The Search for Critical Mass](#)

[The Control of Digital Books: A Hidden Agenda with Massive Consequences](#)

[Cautionary Tales from Other Content Industries](#)

[Consumer Expectations and Technological Controls on Content](#)

[The Global Marketplace: Rights Management, Control, and Censorship](#)

[Books Are Not Music: Reframing the Debate About Control Over Content](#)

[Restructuring the Publishing Value Chain and the Publishing Industry](#)

[Assessing E-book Readers](#)

[The Role of Standards](#)

[A Brave New World for Readers](#)

[The Uncertain Future of Digital Books in Libraries](#)

[Continuity of Access and the Preservation of Our Intellectual Heritage](#)

[Defining the Future of the Book](#)

Introduction: Hyped Machines, Hidden Agendas and Visions of the Future

Readers are legitimately confused as they try to understand the future of the book in the digital world. They somehow know that the inexorable advance of technology will likely eventually render the printed book obsolete, at least for many of the uses that it sees today. Indeed, the elimination or suppression of the book (often as a shorthand for ideas or history) has been a staple of science fiction for decades, and even fundamentally apolitical films such as the early *Star Trek* movies incidentally celebrate printed books as charmingly archaic collector's items. These portrayals have been absorbed into the public consciousness, leading to a sense that technology inevitably will supercede the printed work.

Many people have now seen, or at least heard about, the new consumer electronics appliances popularly called "E-books" or "electronic books" or (more accurately) "electronic book readers" - though few have actually been sold. Probably the best known example is the now-obsolete NuvoMedia, Inc.'s Rocket E-book™, but there are several others. Microsoft has produced a software product called Microsoft Reader, which turns a PC into an e-book reader, and Dick Brass, the Microsoft executive in charge of the product, is making predictions (supported by a rather flamboyant marketing videotape) that publishing will shift rapidly to electronic formats. The argument is a masterpiece of technological manifest destiny. Adobe has similar capabilities in their Acrobat and Ebook Reader products. The traditional print book publishing houses, online booksellers like Amazon.com and distributors such as Barnes and Noble are announcing an ever-shifting series of commercial ventures and alliances to produce material for electronic distribution. New players such as Fatbrain, Peanut Press, Netlibrary and Questia Media have entered the marketplace, promising sizeable commercial libraries of digital books [1]. Authors are also exploring digital books as a new means of reaching audiences, and one that may rearrange the economics of book publishing. With vast publicity, Stephen King gave away a novella called "Riding the Bullet" for downloading, and subsequently offered

installments of a novel called *The Plant* for paid downloading based on an honor system. This generated enough revenue for King to produce a number of installments before placing the project on indefinite hiatus. Accompanying all of this activity has been a chorus of predictions from the market research firms - Forrester Research, Jupiter, and Anderson Consulting (working in collaboration with the Association of American Publishers) excitedly predicting the emergence of a multi-billion dollar marketplace (though there is great variation in the predictions about how many billions, and how soon).

There's little point in trying to chronicle the product announcements and corporate alliances; these are changing from day to day, and any summary would be out of date by the time this paper is published. But it is worth observing that while two years ago most e-books came from small startups, the field is now dominated by very large companies like Microsoft, Adobe (which purchased a company called Glassbook that pioneered e-book reader software and rights management systems to support e-books), and Gemstar. Gemstar is a company that held a series of patents for interactive television programming directories and which then merged with *TV Guide*. In the e-books area they have purchased both NuvoMedia and Softbook, two of the major startup companies developing e-book readers, and have licensed production of e-book readers to the consumer electronics firm Thomson (which markets under the RCA brand, and should not be confused with the Canadian media and information giant Thomson). Thomson started shipping second-generation consumer e-book readers (the REB1100 and REB1200) in time for the Christmas 2000 shopping season.

Confronted with these developments, and the hype surrounding them, readers might reasonably wonder whether they are seeing the future of the book, at least in an early and as yet immature form, and if it isn't time to get aboard that future.

Every major newspaper and magazine seems to be running regular articles about e-books. I suspect more words are being published about the e-book phenomenon in print than have actually been placed into e-books so far. But the prospects for digital books and e-book readers are beginning to capture the public imagination. Much of the discussion seems to be about whether, and if so when, e-books will replace traditional print-on-paper books, and a great deal of the debate is infused with sentimental appeals to reading on the beach or in the bath, the joys of finely printed books, and of browsing in good bookstores. There's considerable speculation about how digital books may restructure the balance of power between authors and publishers, largely based on Stephen King's experiments, but little mass-media discussion of how digital books are really likely to change the world for the consumer or for society as a whole.

What's really happening is much more complex than the emergence of a new kind of consumer electronics device, or a new marketing channel for books enabled by these appliances. A whole group of disparate, long-simmering issues are converging around e-books, which serve as a sort of shorthand, or symbol, for the larger questions. The

sentimentally framed questions about digital books and electronic devices replacing printed books are largely irrelevant, an artificial and distracting controversy. Both can and undoubtedly will co-exist for a long time to come and will find their appropriate audiences and market niches. This will, I believe, sort itself out in the marketplace. The real issues are more fundamental: how do we think of books in the digital world, and how will books behave? How will we be able to use them, to share them, and to refer to them? In particular, what are our expectations about the persistence and permanence of human communication as embodied in books as we enter the brave new digital world? Will our thinking be dominated by the conventions and business models of print publishing (and the current power relationships among publishers, readers, and authors), and by our cultural practices, consumer expectations, legal frameworks and social norms related to books, or will we discard these traditions, perhaps in favor of evolving practices from industries such as music? These are questions about which I believe we need to think explicitly and deeply, and not just answer by default, as mere by-products of shifting trends in the consumer marketplace.

Concurrent with the rise of the e-book into the broad public consciousness has been the emergence of another series of controversies surrounding a technology called Napster and its use in the dissemination of digital music. The "content" industries have gone absolutely berserk over Napster, and have filed a number of lawsuits against it, as well as conducting a major public relations campaign to paint Napster and similar systems (e.g. Gnutella) as instruments of the devil. Jack Vallenti of the Motion Picture Association of America, Hillary Rosen of the Recording Industry Association of America, and Pat Schroeder of the Association of American Publishers have been busy working our nation's capital to marshal opposition to these developments, and promoting a set of technologies called "digital rights management systems" (otherwise known as "technical protection systems"). Legal tools such as the Digital Millennium Copyright Act (DMCA), which was passed (largely unnoticed by the public) in 1998 are being exploited to support lawsuits against new technologies and against consumers who employ them. The DMCA represents a massive change in the balance of control over content in digital form when compared to historical traditions surrounding the sale and use of intellectual works in the United States. A series of arcane and stunningly disingenuous legal assaults against a program called DeCSS, which permits legally purchased and paid-for DVD disks to be played on unsanctioned devices, are being used to attempt to establish new and unprecedented rights of control over consumer use of all types of digital content. If these activities are in one sense rear-guard actions against the implications of digital distribution for music and video, they are equally significant as activities to *establish* the legal framework that may well govern digital books, with e-book readers as the technology of choice for enforcing control.

In the public mind ideas like copyright are arcane and fuzzy; legislation like the DMCA mostly unknown and esoteric; and furthermore, there's a feeling that somehow books are *different* from ephemeral entertainment products like the movies and music that are the

subject of so many current lawsuits. Books are serious, they capture our knowledge, our intellectual heritage, our cultural discourse. Books have significance that transcends quarrels about who gets paid, and when, and how often, for playing popular tunes. But under the law they *aren't* that different, and what's happening in the music industry may well be establishing an important part of the future of the book - though that connection hasn't been sufficiently emphasized. Indeed, one might argue that the content industries don't want to stress it because it might well alarm the public about the broader agenda of technological control of intellectual property. Books *are* important, they should be different, but in a world of digital convergence where everything is reduced to sets of sequences of bits the precedents are being established in other spheres.

And completely left behind in the focus on reading technologies, control of intellectual property, and the economics of publishing (and all of their broader social implications) is the deep, important, and exciting question of how the digital medium may permit authors and readers to reconceptualize the acts of communication and documentation that have been embodied in the printed book for some or all of the purposes that the book has historically served. This may be the area with the greatest promise of truly transformative changes.

There are, then, at least three major (though sometimes subterranean) agendas implicated in all the hype over e-books:

- the nature of the book in the digital world as a form of communication;
- control of books in the digital world, including the relationships among authors, consumers/readers, and publishers, and by extension, the way we will manage our cultural heritage and intellectual record; and,
- the restructuring of the economics of authorship and publishing.

The purpose of this paper is to expose and explain the issues involved in these three agendas, and to connect them to related developments in other so-called "content" industries. After establishing these agendas and making these connections, I'll take a critical look at e-book readers and digital books, and also consider some of the crucial broad social and cultural issues at stake in the transition to the digital world, such as the preservation of our intellectual heritage, the role of libraries, and an assessment of what consumers can reasonably expect as they come to grips with digital books.



Defining Digital Books and E-book Readers

Imprecise and inconsistent terminology has been a major source of confusion in the hype over e-books, and an obstacle to disentangling the issues involved. It is essential to distinguish between the idea of a *digital book* and a *book-reading appliance*. A digital book is just a large structured collection of bits that can be transported on CD-ROM or other storage media or delivered over a network connection, and which is designed to be viewed on some combination of hardware and software ranging from dumb terminals to Web browsers on personal computers to the new book reading appliances [2]. Digital books cover a wide spectrum of material, ranging from literal translations of printed books, created by scanning pages or generating a PDF file, to complex digital works that are the intellectual successors of certain genres of book-length works, but which cannot be reasonably converted back into printed form. To a large extent, digital books exist (or at least should exist) independent of the devices that may be used to access, render and view them. A key role of standards (to be discussed later) is to make this independence formal, and to ensure that a digital book can be used with a wide range of viewing environments that may change over time.

Not every digital book can be viewed using every viewing technology. Some are highly targeted to specific viewing technologies, while others are versatile and can be easily delivered to many diverse viewing environments. Also, recognize that while it may be technically straightforward to deliver a book to a wide range of viewing environments, the publisher may deliberately *choose* to limit the environments a digital book can be delivered to. And of course, viewing technologies can be thought of as defining markets. Authors may choose to author for markets that they believe are large or easily reached or profitable, and as a consequence may choose to create works that deliver well to particular viewing technologies.

A book-reading appliance (like the Rocket eBook) is a new addition to the spectrum of devices that can be used to view digital books. It's typically a portable consumer electronics product priced at a few hundred dollars that includes a high-quality display, has a form and weight factor somewhere between a hardcover book and a laptop, runs for a long time on batteries, stores perhaps 5-20 books worth of content, and doesn't include a keyboard. There's considerable variation among different brands in the way digital books are loaded into the appliance. Some use connections to personal computers (mainly Windows machines, though there is some Macintosh support) that have previously downloaded books from the Internet; your "library" resides on your computer, but often isn't viewable there because the books are encrypted. Other book reader appliances use modems to download works directly from bookselling services over phone lines, or have Ethernet ports that allow them to be connected to the Internet for direct downloading of books. These connections, which eliminate dependence on a personal computer, seem to be the trend in more recent devices such as the REB1100 and REB1200. The appearance of wireless connections in e-book reader appliances seems to be only a matter of time, and only await improvements in wireless standards and infrastructure.

To make matters more confusing, at the same time that we've seen the emergence of book-reading appliances, we've also seen the introduction of general purpose *software book readers* that run on general purpose computers and that address the same functions of downloading and displaying books - products like Microsoft Reader and the newest Adobe Acrobat and Adobe Acrobat eBook Reader. One can think of these as software products that turn a general purpose desktop or laptop computer into a book-reading appliance. They emulate the functions of the specialized consumer electronics devices, though they can offer extra amenities because of the presence of the keyboard on a general purpose computer. While only a few tens of thousands of appliance book readers have been sold to date, the installed base for software book readers (if one includes Adobe Acrobat) probably numbers in the hundreds of millions.

Personal digital assistants (PDA) like the Palm Pilot are also being pressed into service as software-enabled book readers much like general purpose computers, though these represent a particularly constrained compromise because of their small displays and lack of a built-in keyboard or disk storage.

While appliances are relatively new as real commercial products, the idea of portable book reading appliances has been around for a long time. Alan Kay's vision of the Dynabook goes back to the 1970s [3]. Not long after the introduction of the first personal computers, various companies began to package content and software together to turn them into book readers. These are in some ways precursors to both today's appliances and software book readers. But the establishment publishing industry was not much engaged with these efforts except through "electronic information" or "new media" divisions or subsidiaries that were very much outside the mainline of traditional publishing. It's not until very recently that digital books have seemed poised to become a real industry competing with consumer books sales. There have been many barriers: the fragmentary nature of the market for reading environments and the lack of standards; the limited size of the installed base of reading environments; and, concerns about controlling intellectual property. All of these barriers are beginning to fall.

In this paper, I'll use the term *appliances* for specialized hardware devices, *software book readers* for products that run on general purpose computers or general purpose PDAs, and the more generic *e-book reader* to cover both, but not general purpose software like a Web browser that can also be used to view some types of digital books. While I'll try to be consistent here, be warned that this terminology is far from standardized in the industry or the media.

I want to highlight two important misconceptions that have been fostered by the rhetoric about e-book readers to date. First, an e-book reader isn't good just for reading *books*. It's for any kind of content that's moving from print to electronic form. Some of the most popular content being read on e-book readers today includes newspapers like the *New*

York Times or the *Wall Street Journal*, or popular general-circulation weekly magazines. "E-printed matter reader" would be a more accurate term, but it lacks the resonance of "e-book reader." E-book readers are going to be used for a lot more than reading books.

Second, there's a misperception based on a notion of substitutability. Many think of an appliance reader or a more general e-book reader, loaded with the appropriate content, as a *substitute* for a specific book that might also be available in print form. They talk about it and evaluate it, in that way, recognizing only that it has the marvelous chameleon-like quality that it can very quickly be made to substitute for a different printed work by simply loading different content. This is wrong. Even today's relatively primitive appliances can hold 10 or 20 books; a software book reader mounted on a high-end laptop can already store hundreds of books easily. Given the historic price-performance trajectories for storage, in a few years at least some high-end appliances will house hundreds, if not thousands, of books simultaneously, and certainly laptops with software book readers will house thousands or tens of thousands of books at once. Think of portable personal digital libraries, not portable electronic books, as the future role of these appliances.

And when we think about personal digital libraries, it's clear that the stakes are much higher; the capabilities and constraints of a reading appliance start to take on a powerful influence. And maybe this gets us thinking in some new directions. Imagine you have a portable device that can hold 5,000 books. You perhaps stop just purchasing individual books *to read*; instead, you think about selecting the best supplier of a *reference library* of books to have available to consult should you need to. You think about choosing a subscription service; in which case the choices that the service makes about which new books to add every week or every month begin to have a major influence on shaping the information and the views available to you. Issues of searching and selecting the right books and the right passages in the books becomes an important function that none of the current reader manufacturers seem to be thinking much about. We can think about purchasing books from multiple publishers when we think about an e-book reader as a surrogate for individual books. Is it equally reasonable and realistic to think about having it integrate subscriptions for reference libraries from multiple sources, or combining a reference library subscription with random purchases from specific publishers? These are all unexplored questions, and they have implications for standards, for digital rights management, and for issues such as individual privacy. Consider a few other implications: losing access to a single book may be a problem; losing access to an entire personal library built up over years is a problem of a different magnitude altogether. Can books be withdrawn from a digital library subscription, and if so under what terms and with what notification?

Another mental picture is one of continually adding books to a personal digital library housed on a portable device. But can external events cause books that we have purchased (or probably, more precisely, licensed) to be withdrawn from our collection without our

notification and consent? Such events are largely inconceivable with personal collections of print books.

Sometime in the 1980s I heard this statement about digital books:

"Here's a "view from the future," looking back at our "present," from Professor Marvin Minsky of MIT: "Can you imagine that they used to have libraries where the books didn't talk to each other?"" [4]

This is simultaneously provocative, asinine, and inspiring. Perhaps the idea was that digital books would somehow create hypertext linkages among each other; this is reasonable and useful. Perhaps there was a perception that books would become active knowledge structures, and would somehow enhance each other through some technique deeper than simply making links; this is a powerful and important idea, but the necessary knowledge representation structures have proved hard to develop and to populate. But as I think about personal digital libraries populated by publishers, I now find myself thinking about another memorable talk by Bob Lucky of Bell Labs [5]. Lucky described an imminent future where he discovers that his personal computer is swapping e-mail, expense reports, and other digital gossip with other corporate computers over the network; the computers know who is getting fired and promoted before the people involved. What might the books in a personal digital library be saying to each other? They might well be sending inventories of holdings and reading patterns upstream to each publisher that has provided books that are part of the collection, so that the owner of the personal digital library can be notified of new books to add to the collection; they might be trading statistics about how often each book is being consulted, and through what search terms. As e-book readers morph into personal digital libraries, we need to think about what information is being shared, and with whom. We also need to think about vulnerabilities. For example, imagine one's personal library being wiped out or corrupted because you've downloaded a virus-ridden digital book. Or, even more powerfully, an e-book full of crazy or maliciously false information that starts "talking" to your other books.

In a very real sense, presenting an e-book reader as a sort of substitute for a printed book underestimates and trivializes the future. One set of questions that e-book readers raise is about the future character and operation of personal digital libraries, and their relationship to commercial and non-commercial digital libraries and digital bookstores. Another is how these entities will be distributed across a mix of portable appliances, personal computers, personal storage on network servers, and institutionally or commercially controlled storage and services on the network. These are very large, complex, and serious questions that go far beyond asking whether a plastic-encased machine can satisfactorily substitute for paper pages bound in leather or cardboard.

Digital Books as Literal Translations of Printed Books

Electronic book-reading appliances are not the first vehicles for delivering books electronically. CD-ROMs, diskettes, and network-based delivery to terminals and workstations have been available for at least 20 years. With some notable exceptions, books - particularly versions of existing printed monographs, novels, textbooks, and other materials - delivered electronically through these channels have not been a great success, and the reasons are simple. Most importantly, current computer display technologies do not offer a pleasant environment for reading very long texts when compared to ink on paper. There are also problems relating to standards, rapid obsolescence of content due to technology changes, and the general complexity and instability of the personal computer environment, all of which seem ill-matched to the elegant simplicity of owning and reading bound printed books. For some people and some types of works, the ability to annotate, to highlight, and make margin notes as one reads is important though the importance of this for the general reader is perhaps overrated. These activities have been awkward for electronic books. And of course business and marketplace issues are also critical to success: availability of a critical mass of compelling content conveniently and under reasonable terms. Early efforts generally failed on these criteria as well, particularly because many publishers charged a large premium for access to the same or similar content in electronic rather than printed form, and also because of the general awkwardness and complexity of transacting commerce in electronic content.

Certain specific genres have been a great success in electronic forms, and these are rapidly displacing printed products. For example, bibliographies, abstracting and indexing guides, citation indexes, dictionaries, encyclopedias, directories, product catalogs, and maintenance manuals for complex systems such as aircraft work well in digital form. In a sense, by moving to the digital medium we have been able to understand these kinds of works more deeply, and to bring out their essence. It is not an accident that computerization was applied to the construction and editing of these works very early, and in a great irony we have now undone a perverse process where computers were used to construct these works and then reduce them to print because the infrastructure did not yet exist to make them available to the public as digital content directly.

All of these genres share several key properties: their readers want to find and then read relatively short chunks of specific text; they are frequently updated; and, in some cases they can be greatly enriched by the larger amounts of content and multimedia amenities that the electronic environment can inexpensively accommodate (the cost of increasing from 1,000 pages to 3,000 pages of content and adding large numbers of illustrations is much cheaper for an electronic work than for a printed work). They are more like reference databases than traditional books that are read sequentially from beginning to

end. In their digital versions, these items are often quite different than the printed works they superseded. These digital works are now well-established in business, consumer, and library marketplaces, both as networked information services and as CD-ROM-based content that is used with custom programs. Interestingly, they are not always a good fit with e-book readers because they aren't really read like books, and indeed in digital form aren't even presented like traditional books, but rather as databases to be searched or browsed. In this sense, they have succeeded precisely because they are not literal translations of the predecessor print products to the digital world. They have moved the content with little change, but radically restructured the presentation interface.

We can also learn from what has happened as scholarly journals, newspapers, and magazines have moved to electronic form. These shifts have been relatively successful in that the electronic versions have found substantial readership, but they aren't yet displacing the print products. The "unit" of reading in such works ranges from a page or so (a newspaper column) to a few dozen pages (for a typical journal article). Basically, the printed form has been translated rather literally into an electronic representation for these kinds of content. It's still formatted like print, and is intended to be read sequentially like print (in fact, one sees things translated from print that are truly abominable on screen, like multi-column formatting for journal articles). Numerous studies in university settings [6] have discovered what people do with these electronic offerings: They use the online (or other computer-based version) to browse, to do quick checking, to decide what they do and do not want to read carefully. But if the piece is over a few screens in length, *they print the article for reading*. In essence, they are using paper - a mature, robust, and exquisitely effective viewing technology - as their preferred user interface for reading [7]. Interestingly, studies by companies like Netlibrary and Questia Media who offer digital versions of printed books, and also by universities running instrumented digital books experiments suggest that digital books are being "read" online in rather similar ways - in short, randomly-accessed segments rather than sequentially. Users of these services do not seem to be reading texts linearly for hours at a time.

In cases where we literally translate a book into digital form by scanning it, or creating PFD as a byproduct of print production, or an HTML file, we face the same problems that are familiar from journals. With current display technology (at least until the appearance of reader appliances) people clearly want to print to do serious reading. It's interesting to note that one easily can find many PDF files of published books on the Internet, and these don't seem to cut into print sales significantly, but rather offer a preview function for potential buyers that probably enhances print sales. The National Academies Press has been offering their publications for free on the Internet for several years. These are typically a few hundred pages in length, and they are offered through a user interface that displays a page at a time and makes printing of large parts of the work awkward. The effect of this has been to increase print sales substantially by increasing the visibility of their publications. In essence, by making (free) printing on demand difficult, they rely on

the reader's aversion to online reading to drive print sales. A PDF file is also a nice amenity that serves as a complement to print, in that it allows searching and similar functions.

Even if printing on demand from the Net is available, crass, pragmatic, mundane issues come into play here that separate books from other, shorter materials. It's reasonable to print 20-30 pages on a desktop or office workgroup printer or on a public printer in a library (even at ten cents a page). You don't have to wait long, and you can bind the results with a paperclip or a staple. Only the truly desperate would actually demand-print a 300-page book, particularly on a personal printer that doesn't even issue a low-toner warning. Effective print-on-demand for book-length works requires specialized, high-end devices that can duplex print and bind the results, and can print very fast (and operators to feed them) - something like a Xerox DocuTech printer. Arranging access to such a device, routing printing to it, collecting the printout and paying for it defeats much of the immediacy and convenience of online access to materials. A number of universities and bookstores are experimenting with these kinds of print-on-demand systems, and companies like Bell and Howell Learning Systems (formerly University Microfilms (UMI)) have been providing access to specialized, low-use material like Ph.D. theses and out of print scholarly works for years via print-on-demand. Digital books as literal translations of printed books, delivered via print-on-demand (perhaps supplemented with online browsing) forms an established and viable market but a small one. Print-on-demand isn't cheap and it isn't particularly convenient - it's a lot like electronically ordering a printed book for physical delivery.

The central question for these kinds of digital books is whether new reader technologies can expand the marketplace beyond the niche of print-on-demand materials that can't really be published cost-effectively in traditional ways, or of e-books that serve as searchable and conveniently accessible supplements to printed texts. The answer will be determined largely by display readability, but convenience (you can carry the equivalent of multiple heavy printed tomes in one e-book reader), quality (electronic content can be more current and flexible than print content), and economics (how do the costs compare to purchasing printed books) will also be important factors. Also, recall that many appliance readers likely will be purchased on the basis of obtaining easy, portable access to more "hospitable" content such as newspapers and magazines, at least initially and PDAs are being purchased for functions like calendaring and note-taking. These machines provide a growing installed base that can allow users to experiment with the convenience of storing, carrying around, and consulting digitized books as an activity at the margins with little additional investment.

Leaving aside issues around the control of intellectual property that will be discussed in depth later, there's little economic risk or cost in generating a PDF file as a part of the publication of almost every new book today. This can be created as a simple by-product of the latter stages of the editorial and production process, and we will see these becoming

plentiful. There are still some problems with illustrations. Simple HTML markup versions are more complex, particularly if specialized character sets or illustrations are involved. So far, the evidence (mainly from major scientific journal publishers such as Elsevier) is that the farther one tried to push production of content for the digital environment upstream into the editorial process (for example, by marking up everything in SGML or XML) in order to produce highly differentiated print and digital versions, the more complex and expensive it gets. While journal publishers can justify these investments because of the rapid move of scholarly journals to the digital medium, it seems likely most book publishers will be content with digital books that are by-products and close analogs of their printed works for the near future.

New Content Genres: Reconceptualizing Books in a Digital World

We are also seeing the development of new *genres* of material that are highly adapted to the online reading environment, built on the early success of types of books that translated advantageously to the digital environment, such as encyclopedias. These new genres are designed to exploit the strengths of the digital medium. A scholarly Web site, for example, links and organizes many small chunks of text with multimedia content and provides the ability to search and navigate among them. It may also include interactive software components such as simulations, and use the communications capabilities of the Internet to build an interactive community around the work and its subject matter. It may also personalize its behavior for each individual user based on knowledge of that user's profile, interests, and history. The [Perseus Project](#) is one beautiful example of the possibilities of linking source material in ancient languages, translations, commentary, and multimedia into an extraordinary scholarly resource. There are many other examples, such as the [Valley of the Shadows](#) at the University of Virginia. These works are profoundly different from simply presenting digital versions of printed pages.

While the focus today is on network-based new genre works, companies like Voyager have been exploring the reconceptualization of authoring in the digital environment for at least a decade using CD-ROM technology. Voyager has published both classic works of literature re-presented for the digital environment and entirely new works of authorship, and serves as a particularly interesting case study. Today, network-based works are typically constrained by the limitations of both browser technology (in the user interface) and bandwidth available to readers (in terms of ability to provide rapid browsing through images, video material, and the like); custom programming coupled with CD-ROMs

offers a much more predictable and flexible environment for experimentation, and we can find Voyager offering products that blur the line between works like plays, movie scripts, and epic poetry as texts and as performances. Some of these works, perhaps more than current network-based projects, probe issues about how we will read and relate to texts in the digital world. They do not fit well into existing canons of either instructional materials or scholarly communications, and challenge readers raised within textual traditions [8].

The output of Voyager over the past decade offers a rich catalog of new genre developments, which will undoubtedly be recast into the networked information environment as that technology matures. It would be better to study the output of Voyager as a sort of display case of possible new genres than to re-invent them in the networked information environment [9].

Recently there has been a lot of thinking about how to devise intellectual successors to the scholarly monograph that specifically exploit the online environment. One key idea is that while the definitive and comprehensive version of the work will be digital, there will also be a sensible (though impoverished) "view" of the work that can be reduced to printed form as a traditional monograph. This is critical in providing scholarly legitimacy in an intensely conservative environment that still distrusts the validity of electronic works of scholarship, and will thus be important in encouraging authors to create these new types of works. It allows authors to exploit the greater expressiveness and flexibility of the digital medium without alienating colleagues who haven't yet embraced this medium. The Andrew W. Mellon Foundation is working with the American Council of Learned Societies (ACLS) to explore these possibilities in the area of history monographs [10]. This is a major, \$3 million initiative involving some five ACLS member societies and ten university presses which is targeting the production of about 85 new works and 500 backlist works into a Web site. In this sense it intends to produce a digital library rather than just individual new genre digital books while maintaining some level of uniformity among the digital books that might point the way towards future publishing initiatives. The American Historical Association is also running a program called [Gutenberg-e](#) for new scholars who want to publish digital books; here there is greater variation, with each book developing its own unique characteristics.

Similarly, there is serious work in education and consumer markets on training and instructional materials that are adapted to the network delivery environment. Authors and publishers are just beginning to explore the full range of possibilities here, and to understand how to develop combined print/online products. For example, travel guides might combine an easily portable paperback book with the comprehensiveness and timeliness of an online site offering 3D panoramas and walkthroughs with hypertext links, route computation from maps, and large amounts of more routine audio, video, and image multimedia content. We should also not overlook source code for computer programs, a form of text that has never been particularly useful in printed form (though there have been a number of books published over the years that consisted largely or entirely of

printouts of computer programs).

Many of the new genre works (and the genres of works from the earlier print tradition that now have achieved a more natural manifestation as digital databases) raise problematic intellectual and marketing questions about the scope of the work in a heavily hyperlinked world, about the preservability and integrity of the work, about fluidity of content, and the difficulty of identifying a reasonable number of fixed, individually capturable editions. For example, in the case of an encyclopedia or dictionary we stand to lose the ability to have snapshots of the state of knowledge and understanding, and of cultural biases that scholars can revisit from later epochs as we move away from editions to continually updated databases. Old editions of such works in print have become important research resources. Another very interesting problem is that in digital works that permit the reader to find his or her own pathway through the work it is often difficult to tell when one has "read" the work competely. This is problematic both for instructional works and for communication among scholars, though for different reasons. We will need time and experience to sort out these questions as we come to better understand the characteristics of the new genres.

Presumably, the compelling richness, flexibility, and timeliness of such multidimensional works will more than compensate the reader for the inconvenience of reading text on display screens. And the content is to some extent specifically designed with the strengths and limitations of display technology in mind. I believe we will ultimately find that some kinds of discourse - scholarly and otherwise - will be more effective using existing genres rooted in printed works (perhaps presented digitally as well as on paper) rather than in the new genres. Print still seems to be the medium of choice for longer texts intended for linear reading.

There is a developing "marketplace" in these kinds of works, which have great promise for enhancing scholarship and teaching and for providing more compelling content for some consumer needs, but these kinds of works are still experimental and often costly to produce, and they are usually not yet commercially viable (even in comparison to printed scholarly works). Many of the leading projects are subsidized academic works, which are justified by their contributions to scholarship, and the long-term economic sustenance of these efforts and their preservation are serious issues. The works in the new genres are being developed and distributed largely outside of the traditional consumer marketplace and commercial publishing framework (even the publishing framework for scholarly works). There is little concern, and certainly little obsession, with the control of intellectual property in their distribution. And very little work has yet been done to show how other popular or consumer-market-oriented print genres besides reference works - particularly and crucially, fiction - can be evolved successfully into new digitally based genres. If anything, we are seeing consumer-market developments coming from outside the book industry: for example, expanded special edition DVDs that include alternate cuts of an important film, plus extensive commentaries by actors, the director, etc. - as the first

entrants into a commercial marketplace in new genre works.

Fiction is a particularly interesting and important case. I have yet to see a work of print fiction as storytelling designed for the digital environment [11] which is compelling to me, though certainly there are examples of literature that has been moved to the digital world through critical editions (as works of scholarship) in ways that add a great deal of value. Perhaps fiction as storytelling will remain most effective as a genre targeted for printed books (including print- on- demand books), and the future of storytelling in the digital medium will become something different that is still to be invented - not film or video, which, like digitized print pages, can be delivered over the Net but which does not explicitly exploit the digital environment, but something else. Certain kinds of computer games may point towards one future here; see also the work of the renowned game designer Chris Crawford on the [Erasmatron](#). For another provocative view of a possible future for storytelling in the digital medium, see Scott McCloud's *Reinventing Comics* [12].

Michael Jenson [13] makes some other interesting points about how the entry of commercial interests may shape and constrain the development of the new genres. A commercial publisher probably will not want to encourage links to works that are not part of that publisher's catalog, for example. Each book is an island, as he says. I hope that this kind of thinking represents publisher immaturity in understanding the digital medium, however, and publishers will move beyond it. Indeed, this is already happening in scholarly journal publishing, where the publishers are creating consortia like Crossref to facilitate the construction of hypertext links between articles in journals from different publishers in response to reader demand.

Today, the vast majority of the new "book-like" digital genres are targeted for general purpose, network-connected computer workstations that have Web browsers and sometimes an array of other software that extend the browser's capabilities (such as movie viewers, or QuickTime VR). We shouldn't overlook the extent to which these new genres depend on computational and interactive capabilities, continual (and often high-speed) network connectivity, the ability to render high quality images, audio and video, and the availability of an extensible array of support software in order to work. They rely on links to materials throughout the Internet and on the characteristics of networked information resources that accommodate continual incremental updating. Will these digital materials function effectively (or even advantageously) within the much more constrained software and connectivity environment offered by (appliance or software) e-book readers, or are they inherently creatures of the general purpose, networked information environment? Will the new content genres and viewing technologies converge, or will they diverge, with the evolution of new genre content bypassing specialized viewing environments and requiring the most advanced state of the art that is found in general purpose computing environments? My guess is that at least in the near term we will see e-book readers mostly concerned with supporting digital books that are very similar to traditional print books

(perhaps with some very modest incremental enhancements, and with a more generous use of illustrations). The new genres will be targeted to the general purpose computing environment.

If this is true, the implications are important. In the digital world, the palette of capabilities available to the author has been vastly enlarged; there are new ways to communicate, to structure arguments, to provide insights. E-book readers significantly constrain this new palette; the priorities are not flexibility in authorship and reading, but control and familiarity through emulation of the printed work. We owe it to today's most innovative authors, and to our society, to make available the fullest potential of the environment that information technologists are developing, and not to limit these authors to the capabilities of book readers that are concerned with protecting and managing the works of those that have come before them. We must not allow these book readers to define in the public mind what is a book, and what is something else - something perhaps having less legitimacy as a cultural artifact based only on our ability to conveniently package it, market it, and control its use. From an economic perspective, we need to keep in mind that there are a lot of printed books, old and new, and a lot more authors who know how to write for the print medium, and that publishers understand what it costs to develop a book for the print medium. For commercial publishers generating a digital book revenue stream for these new technologies is likely to be more attractive in the near term than speculative ventures underwriting revolutionaries and experimenters who are trying to re-invent communication in the digital medium. There's nothing wrong with commercial interests behaving conservatively here, but we need to be sure to honor and empower the innovators, not marginalize or constrain them based on the dominance of commercial considerations.



Converting Older Books to Digital Form: The Search for Critical Mass

So far, I've discussed where *new* digital books (including new editions of old favorites) are likely to come from, and the conflicting forces shaping them. They'll come from innovative experiments in authoring within the digital medium, and they'll also come from direct translation and perhaps modest extrapolation of the print books that make up the current catalogs and active backlists of today's publishing industry. But history has also left us with some four centuries of books, the vast majority of which aren't currently in print. I suspect that much of the public believes that all of these works are, or will soon, be available in digital form - availability is waiting only on the deployment of high-bandwidth delivery systems, perhaps, and more scanning. This misconception has been

fed by a great deal of hype about the evolving network information environment [14].

If and when these older books are converted, almost all of them will become extremely literal translations of their printed forms. This kind of scanning-based conversion (which actually captures images of the print pages, rather than meaningful digital representations of the characters and words on those pages) is a fairly inexpensive manual process. More sophisticated conversions that capture the meaning of the content in computer-manipulable ways (at various levels of complexity) are much more expensive and involve extensive human intervention in the conversion process (with the extent of the intervention increasing with the richness of meaning being captured). We may see technological progress over time, which reduces these costs, and allows computational processes to be automatically applied to the page images to capture deeper meaning. Optical character recognition (OCR) technology is one important tool that has been slowly improving in quality over the past two decades, but accurate OCR still requires human review and editing. But all of these considerations speak solely to the mechanics of converting books to digital form.

The legalities of such conversions are a much more serious barrier, and one about which the public remains unaware. Roughly speaking, at least in the United States, any book published before the early 1920s is in the public domain (the details of precisely what is in the public domain are very complicated, and aren't crucial here). If you can find a copy, you can scan it, or, if you are willing to pay the labor costs, you can even re-keyboard it with added structural markup into a more sophisticated digital representation. Whether you obtain a new copyright for your converted digital version of the work seems to be legally murky [15], and seems to depend significantly on how much value you add in doing the conversion [16]. This is important because it has implications for the availability of investment capital to convert public domain materials, and for how these materials need to be protected as they are made available, if they need to generate a revenue stream.

Note also that for nearly all readers, really old works are more appealing in modern editions (with modern spelling, typography, and the like). While the original editions may be in the public domain, commercially acceptable ones may still be under copyright and cannot be converted without negotiating permissions from the copyright holder. [Octavo's](#) work with some of the great classics of publishing and intellectual history provides an excellent case study here; they had to go beyond simply imaging pages to also linking a parallel modern translation of the works to make them truly accessible and valueable to today's general readers.

After the early 1920s matters get much more complex. Some of the books published since then are now in the public domain [17]; the vast majority of them remain under copyright, however, and can't be converted to digital form without the copyright holder's permission. But who holds the copyright, and how do you find out? At least in mass-market

publishing, it's common for contracts between publishers and authors to stipulate that the rights to the book revert to the author a few years after the publisher takes the work out of print. When the author dies, these rights pass to his or her heirs as part of the author's estate. In other areas, such as scholarly journal publishing, it's more common to see all rights permanently transferred to the publisher as a condition of publication, which has greatly facilitated projects like [JSTOR](#), which are performing massive conversions of back runs of scholarly journals - all the rights for large amounts of material can be negotiated with a single owner.

The U.S. Congress has made this problem worse with the passage of the Sonny Bono Copyright Extension Act of 1998, which extended the term of copyright from life of the author plus 50 years to life of the author plus 70 years. This law was motivated mainly by the desire of a few media companies such as Disney to extend protection of a relative handful of commercially very profitable works. But it has had the effect of establishing a 20-year moratorium on the entrance of new works into the public domain, and greatly increasing the cost of converting a huge body of commercially fallow books into digital form. The cost of clearing rights for these works is likely to be hundreds of times greater than the costs of actually digitizing the works. For the vast majority, it's not clear that anyone will bother. While it might be possible to generate some return on the investment in digitizing, it's unlikely that the return will cover the additional costs of rights clearance.

The key point here is that for the vast majority of the enormous number of books published since 1920 or so conversion to and availability in digital form (even as direct translations of printed works) is far from guaranteed. In some relatively few cases, publishers hold the rights and can convert them - mainly in cases where the work is still in print on the publisher's backlist. But here too there is controversy and uncertainty. At least in the United States (the situation varies in other nations), publisher contracts with authors prior to the mid-1980s use language like "to print, publish and sell in book form", and it is unclear whether book form includes digital books. Currently there is litigation involving the publishing industry, authors, and third parties (notably Rosetta Ebooks) who want to license e-book rights from authors covered under such contracts with the print publishing industry. This litigation will determine whether print publishers also hold e-book rights for the books that are part of their active backlist published prior to the time when contracts became explicit about e-book rights [18]. Note that the in-print backlist is of particular importance for conversion to digital format because these are works where a marketplace demand has already been established - this is why the publishers are keeping them in print.

In the other and most commonplace situation where rights have reverted to the author when the book went out of print, e-publishing rights can at least in theory be cleared to permit works to be digitized, but this is going to be a case-by-case decision by one of the myriad of organizations (both commercial and nonprofit) and even individuals interested in making books available in digital form. Converting our literary heritage is going to be

slow, incremental, expensive, and somewhat haphazard. And in terms of making a really substantial corpus of material of high interest to the general public available quickly, the publishing industry, with its current in-print works and the out-of-print works it already holds rights to, is the only game in town. Whatever the shape of the future of the digital book, what happens with the availability of the current in-print corpus of material from the publishing industry is going to be important. Perhaps even important enough to keep the new genres marginalized, at least outside of the scholarly world, long enough for the traditional publishing industry to firmly establish new ground rules for digital books - even to define entirely the idea of what constitutes a digital book in the public mind. And this corpus will look very much like literal translations of printed works.



The Control of Digital Books: A Hidden Agenda with Massive Consequences

The questions about digital books and the role of e-book readers aren't simply about the responsiveness of new technologies to the needs of readers and authors; there is also a major agenda concerned with issues of control, economics, greed, and fear. These issues, rather than simple technological inevitability, may play a dominant role in shaping the digital book marketplace, and are central to understanding the promotion of digital books and e-book readers. Indeed, e-book readers may be the price that the publishing industry imposes, or tries to impose, on consumers, as part of the bargain that will make large numbers of interesting works available in electronic form. As a by-product, they may well constrain the widespread acceptance of the new genres of digital books and the extent to which they will be thought of as part of the canon of respectable digital "printed" works, as opposed to databases, video games, Web sites, and other things which are of interest to consumers or scholars but don't have the same legitimacy.

Some individual or noncommercial copying of copyrighted works is legal in the United States; some isn't. Reasonable people can disagree on the exact boundaries; fair use, for example, is a rather subjective matter. The law is complex, and considers factors such as the purpose of making the copy and the impact of making the copy. But activities such as making personal copies of a work, or of taking quotations for criticism or news reporting, are well established. The law, through the doctrine of first sale for copyrighted works, also ensures that people can keep copies of works that they've purchased as long as they wish, and can loan them to other people or resell them. Social norms and traditions of behavior, which are loosely correlated with the law, also set consumer expectations. Most book publishers haven't worried about this too much, at least until recently. They have viewed the ability of individuals and even libraries to lend works as a lost cause in the U.S.; this is

not true in some other nations. With the exception of a few areas, like the used textbook market, there isn't much revenue at stake. The situation is different for journal publishers, who are very concerned with copying of articles displacing subscription sales.

Historically, book publishers have relied primarily on law and economics rather than technological measures or prohibitions on technology to protect their revenue streams. They don't, for example, compromise readability of their books by using special inks and papers that don't photocopy well. Copying of a book is, in general, perfectly feasible. But copying more than a few pages of a book is inconvenient. The quality of the reproduction is sometimes poor. If one copies a whole book, one has to bind it somehow. If the book is in print, it's usually easier to purchase a copy than to photocopy one - and often cheaper though commercially published scholarly monographs priced at a dollar a page or more begin to shift this equation. Publishers didn't make a serious attempt, by and large, to outlaw Xerox machines, or to try to legislate the inclusion of special capabilities that recognize and refuse to copy pages from published works into copiers. Nor did they mount public relations campaigns suggesting that these devices represented the impending death of their industries, as occurred with the VCR. Publishers rely on the inconvenience, lack of cost effectiveness, and limitations of single-copy technology to constrain copying by individuals, along with users' respect for copyright law and intellectual property. Any attempt at large-scale piracy using the traditional mass-production technology of publishing (which addresses the cost-effectiveness and quality issues) is handled by legal mechanisms.

Some publishers have perhaps been uneasy about this compromise, particularly as rising prices and improving technology have made personal copies more attractive, and they have perhaps resented the lost revenue due to the ability to loan or resell books once purchased. They may have occasionally looked wistfully at the lending fees that some other nations require libraries to pay publishers. But this has been a consumer-friendly framework and one that has in worked remarkably well. It has been consistent with both the U.S. Constitution and political and social reality in the United States. And ways to change the ground rules just haven't been practical until now.

In the digital world, technology combined with new legal frameworks for doing business - contracts and license agreements, rather than the doctrine of first sale - allow publishers much more flexibility. They create new revenue opportunities, new capabilities for tracking and controlling the use of content, the potential to create new business models such as pay-per-view or limited-time subscription-based access, which generate ongoing revenue streams. In effect, by combining technology with the new business and legal framework, publishers can for practical purposes opt out of many of the requirements of copyright such as fair use. They can bypass first sale, subject only to being able to gain consumer acceptance and acquiescence in the marketplace. This is a promising prospect since each work is a monopoly in some sense, and since there is no meaningful balance of power between individual consumers and publishers in negotiating terms. Marketplace

rejection, a vast number of individual consumer choices that add up to a failed product or service, is the only obstacle, and this is one that is poorly understood and very hard to predict, creating an opportunity for persuasive, visionary marketers to launch new businesses.

And along with the promise of new revenue opportunities comes the threat of the digital environment - massive, cheap, perfect duplication and nearly free and instant worldwide distribution of copies, the placement of tools more effective than any pirate operation had in the print world into the hands of any individual connected to the Internet. New technologies for controlling content - such as e-book readers - can, with the support of recently-passed and pending legislative changes, offer publishers a way of addressing both the promise and the threat.

Best of all, book publishers have not been at the frontier of digital distribution, by and large. The music and video industries got there first, because in a very real sense their products are *intrinsically* electronic in form. While there are important differences between books and other types of content, at some level they are all bits in the digital world, and there's a great deal to be learned from the experiences of the pioneers.

Cautionary Tales from Other Content Industries

Music and video are different. Unlike printed text, mediation by technology-based "players" is an intrinsic part of using music and video. Playing, recording, and copying are often closely linked together in the technologies of the day that permit the performance of recorded music or video - tape recorders and VCRs, for example. It's only with digital books that books for the first time encounter the issues that have always been familiar to music and video publishers.

Technological mediation has given rise to rather different consumer expectations for nonprint materials. Consumers know that mediating technologies will become obsolete, perhaps sooner than the media that are played on the mediating technology. While they may own a piece of media - an 8-track tape or a vinyl record - for as long as they want to keep it, they don't expect any guarantees that the players will be available forever. Indeed, for films the entire idea of "owning" a copy is relatively new, dating to the availability of VCRs in the 1970s for the vast majority of consumers. Until then, films were purely experiential works for the general public.

Copying has always been an issue for the music industry - consumers recording songs off

the radio rather than purchasing recordings, making tapes of records, radio broadcasts and the like. But until recently quality and convenience limited the impact of these activities; though the music industry has always hated the idea that a consumer could tape a record to play it on the tape player in the car, or to give to a friend, there wasn't too much they could realistically do about it in an analog world. Convenience and quality factors kept it to acceptable limits in terms of economic impact.

Digital technologies like audio compact discs permit perfect-quality copies (as opposed to analog taping). Copying can also take place at processor speeds, rather than in real-time, where the audio is captured only at the rate it is played. As concern about copying increases, content owners increasingly view players and consumer recording devices as checkpoints that can be - and must be - designed to control individual copying. As an example, Digital Audio Tape (DAT) machines for consumer use include a mechanism that prevents the creation of second-generation copies of commercial works; this is legally mandated in the United States. There are various systems in use to prevent copying of commercial videotapes. And of course, these content industries also rely on legal mechanisms to control large-scale piracy, just as the book publishing industry does.

From the content provider perspective, things have gone very wrong with audio compact discs. These discs moved outside of a consumer electronics environment that could be managed by legislative initiatives into one characterized by open, general purpose, multi-application hardware and software that has proven impossible to control easily. Originally, audio CDs were playback-only consumer electronic devices. One could, of course, record an audio-CD onto a cassette tape or similar analog technology at diminished quality and at performance speeds. Consumers in the early days did not have the capabilities to digitally duplicate and distribute the contents of a CD. And they could not extract tracks from audio CDs in digital form. Over time, CD-ROM players, which could also play audio CDs, became commonplace peripherals for general purpose computers. While the audio CD player did not have a "digital out" port, the CD-ROM player did, in the sense that audio files could be moved into computer memory, and played, or saved to disk storage by means of a program called a "ripper". Combine the more efficient compression of MP3 audio and fast network connectivity with the availability of large amounts of cheap digital storage, and you have a great deal of music from commercial audio CDs moving to devices like MP-3 players and flowing across the Internet illegally. Services like Napster have been established to make this sharing systematic and convenient on a large scale. At least arguably, this is starting to cost the music industry real money, though it's hard to tell how much of the problem is because the music industry hasn't deployed competing for-fee products at reasonable prices and with reasonable levels of usage convenience [19].

The same phenomenon - now called "napsterizing" (a noun becoming a verb, much like Xeroxing), and showing how profoundly the effects of Napster have altered the thinking of the industry - is starting to happen with the digital video disk (DVD), except that films are much larger and call for more capable and capacious computers, disks, and network

connections, plus there is an encryption scheme that has to be bypassed. This scheme, while technically inept, is turning into a convenient justification for legal attacks not just on DVD copying itself, but also on the dissemination of research that makes it possible under the new provisions of the Digital Millennium Copyright Act. By contrast, the music industry has only been able to attack copying of compact discs, not the dissemination of knowledge about how to copy them or how they are encoded.

The music industry is responding to these developments with the Secure Digital Music Initiative (SDMI) standards activity (see www.sdmi.org) and the introduction of a number of other proprietary systems that are being tested by individual music companies. These systems will combine hardware and software technologies to control the ability to duplicate and distribute digital music. The details are complex and still under development, and there are a number of competing approaches. Further, there are real questions about how robust these systems will be against circumvention. But the outline is becoming clear (using SDMI here as a shorthand for whatever schemes the industry adopts for a protection system): future music content will include data that SDMI-compliant players will recognize. And SDMI devices will refuse to duplicate digital music carrying these markings under some circumstances, and refuse to play files with these markings under other circumstances (when they think that the file is an unauthorized copy). The effective use of a piece of music can be bound to a specific device that's authorized to play it. Consumers acquire content to be played only on a specific device and the system enforces this. This technology is reasonably easy to produce and reasonably difficult to circumvent, as long as all of the duplication and playing is on consumer electronic products that have specially designed features to enforce the policies, as long as the music moves within a closed system of devices that follow the rules laid down by the content industries. The hard parts, technically, come when the system tries to accommodate and protect "legacy" unprotected content taken off the existing base of audio CDs, or to allow some limited export into the general purpose computing environment.

The music industry is dreaming of new business models based on their ability to impose and enforce such rules. It's unclear whether the industry will keep the existing first-sale-based CD distribution or a networked analog as one of the business models, or how it will price any of the variations. Depending on how it structures the rules, SDMI or similar competing schemes may or may not prove acceptable to consumers. It may be so inconvenient, so at-odds with current consumer expectations, and so restrictive, that consumers may at least attempt to reject it, leading to an interesting power struggle between the music industry and its customers that echoes similar conflicts between the software industry and its customers over copy protection in the 1980s.

An additional advantage of the protection schemes is that they will bring music under the scope of the DMCA's provisions on technical protection systems. It will become a felony

to perform research or disseminate information on how to bypass the protection system.

It's very hard to exercise control over content when general purpose computers are involved - as players, copiers, or distribution devices though some companies, such as Intertrust, are developing products intended to address this issue in a general way, and we've recently seen a rash of announcements specific to various types of content - music (Liquid Audio and a2b), books (Adobe, Microsoft, Netlibrary) - promising control over content. The robustness of these protection mechanisms has yet to be tested in the marketplace, under the stress of ingenious software developers, cyptographic researchers, and hackers. The DMCA may have a powerful chilling effect, at least for those based in the United States. It's much easier to constrain the flow of content to closed consumer electronic systems with limited capabilities and carefully designed hardware and software restrictions, where users cannot introduce arbitrary programs to bypass controls and move collections of bits from one place to another. Periodically, there have been suggestions to legislate the incorporation of special-purpose hardware - an embedded piece of consumer electronics, if you will - in all *general purpose computing systems*. These proposals have gotten nowhere, so far. And without this kind of support, it remains to be seen whether content can really be controlled on the Internet and in a world that contains huge numbers of autonomously managed general purpose computing systems.

Obviously, the book publishing world, which has been slow to move to digital distribution, is watching these developments carefully (and no doubt thinking that "there but for the grace of God go I ..."). Book publishers do not want to have books follow the path of audio CD content onto the network as pirate digital files and to have to mount a catch-up technological initiative to try to regain control, facing uncertain acceptance with an uncontrolled digital alternative already in place. And they are probably relieved that they have not made more material available in digital form to date, and that for all the reasons already discussed, digital books often remain at a disadvantage to their printed counterparts. Yet the distribution of digital books and the new revenue streams they may offer under new business models are an increasingly attractive opportunity. The e-book appliance, as a closed consumer electronic system, may make publishers comfortable that they have addressed the threats and can exploit this business opportunity. Conceivably, even general purpose book reader software may provide a sufficient level of comfort about the threats to pave the way towards the new revenue streams. This has the extra advantage that it can build upon a very large installed base. But if publishers follow this path, it will change the way that consumers and society use books in the digital age.



Consumer Expectations and Technological Controls on

Content

There is a lack of consensus about what behaviors and activities we want the new technologies of content management to enable or guard against. Some content providers seem to have ambitions that are more appropriate for some Orwellian dystopia. I've emphasized the fear factor, which is being fueled by the experiences of the music industry; but there are also lucrative, unprecedented opportunities for new revenue streams - the greed element. Some content owners want to control in infinite detail all use and duplication of material, and to monitor that use, and possibly charge for it on a transactional basis if they don't block it out of hand. Indeed, these databases of consumer behavior may themselves become new business assets and offer new revenue opportunities.

Consider music again. Personal (private) copying - making a tape of a CD that one owns to play in the car - is something that most consumers find intuitively reasonable but that content providers might like to prevent. They'd rather require that you buy the same content as both a CD and a cassette tape. Perhaps you'd like to make a copy of one of your CDs to play in the car, or even a compendium of favorite tunes from your CD collection on a writable CD for your portable CD-player. Perhaps you'd like to download this compendium as a set of MP-3 audio files into a portable MP-3 player. Perhaps you'd like to transfer some of your aging LP collection to CDs before they stop manufacturing styluses for your record player. Or more to the point you'd like to convert your collection of SDMI compatible recordings to the new MPEG-2010 standard a few years hence. Or to take some music you've purchased to a friend's house and play it. And certainly, you'd like to be able to play music you've purchased on any of the many players you may have scattered around your home. Finally, I think that consumers have a strongly held notion that they should be able to purchase a sound recording and then play it as many times as they wish, without further charges - a flat rate - and that furthermore, nobody should know exactly how many times they played what parts of it. These are all reasonable consumer expectations that could run afoul of a technical protection and copyright management system.

Developing similar scenarios for digital books is more complex; text does not have the same kind of recombinant, omnipresent character that music has taken on. Certainly one might own several e-book readers, and want to be able to view one's digital books on any of those readers; to loan a digital book to a friend; to migrate it across generations of technology. One might want to print bits of a book on paper for any number of reasons - for annotation, for use in the kitchen, whatever. And of course most people are able to transcribe reasonably short pieces of text by hand from a viewing device to a piece of paper or a word processor. Cut and paste isn't essential, the words can move from screen to eye to brain to hand, whereas music and other media have to be duplicated using mechanical means by almost everyone. Technological constraints on copying do not

undermine cherished fair use principles for text to the extent that they do for nonprint materials.

But the key point here is that copy protection and content management systems track and control copying. They can't take into account why you are making the copy, or who else gets to see or use the copy; they can only control the making of copies and (at best) the number of copies of a work that are permitted to exist. Even these controls can likely only be accomplished within a trusted environment, which means that it will be very difficult to make the behaviors that content management systems can permit and prohibit conform to consumer expectations about copying. The number of permitted copies, or the amount of a work that can be copied is a poor surrogate for a full understanding of intent and behavior. When a copy protection system allows a user to make a copy of a work that is going outside of the protected environment, it's impossible to tell whether this is going to be played on an old car player or whether it's going to be distributed to thousands of people on the Internet as an act of piracy.

Copyright law permits copying under a fairly specific range of circumstances; it considers factors such as purpose and economic impact, which are virtually impossible to mechanize into hardware- or software-based testing criteria. Further, there's a large "gray" area involving personal or private copies, where many people believe the law isn't clear, and where most consumers evidently believe that it is permissible to make copies (such as many of the situations described above). New technologies can prevent the making of some legal copies, and certainly of many copies of ambiguous legality. Rights holders feel no obligation to deploy technology that is liberal in its willingness to permit the making of copies. There's no reason why technical protection systems have to facilitate even the making of clearly legal copies. This isn't a legal matter, at least the way the law is being interpreted today. It's purely a question of how restrictive the content suppliers can be while still gaining consumer acceptance.

The balance points among publisher fears, consumer desires, and technical capabilities have yet to be established for digital books. The debate begins with the desire to control copying, but it quickly expands to include control of use, usage monitoring, and new business models that emphasize pay-per-view and transient access rather than actual ownership of copies of works. The DIVX video disk system was one attempt to find such a balance for the video marketplace, and it failed. Whether one argues that it was too far in favor of the rights holder, or that it failed for other reasons, such as a lack of availability of enough compelling content, all we know for sure is that it failed.

It's worth trying to characterize the specific uses in the potential contest. One group of uses comes from consumer expectations established by the historic first sale doctrine: the ability to make unlimited use of something once purchased, to enjoy it for as long as the physical object lasts and technology is available to "play it," and the ability to resell or lend it. There's a good chance rights management systems can support most of these

functions. Whether content providers will continue to make content available to consumers under this bundle of terms is anyone's guess. If they don't there are some serious social consequences that I'll return to in the final sections of this paper. A second group of uses comes with expectations about the ability to make personal copies, and even to do some limited lending or redistribution of these copies. This is harder for rights management systems to deal with. It amounts to, or can be approximated by, a capability to limit the number of copies in existence, and this breaks down when copies move outside the boundaries of a trusted system. The third group of uses arises from another aspect of the public policy bargain that underlies copyright - the exchange of monopoly rights for a limited time subject to certain privileged uses - fair use being the most prominent example. Here questions of intent come into play and it's unlikely that rights management systems can identify legal uses. At best they can offer mechanistic approximations (such as we can't do fair use, but we will let you have up to three pages as "courtesy use"). Though, as discussed above, for text technological barriers to automated duplication of passages under fair use are merely an inconvenience, because almost anyone can transcribe text.

I have focused here on technological controls. The quest for technological controls has been paralleled by legislative initiatives intended to provide legal recognition and protection for these controls. For example, the recent U.S. Digital Millennium Copyright Act ([DMCA](#)) contains provisions protecting rights management information that is attached to content and making it a felony to attempt to circumvent technical protections on content under many circumstances. If a closed consumer electronics system that protects content can find marketplace acceptance, the DMCA will help ensure that its integrity can be maintained.

Complementing the DCMA is a set of proposed changes to the Uniform Commercial Code, which serves as a model for state law governing contracts. These changes - formerly called UCC2B, and now UCITA - would establish new state law giving strength to the idea that consumer transactions in information are controlled by licenses rather than the historic framework of purchase and first sale that have governed physical intellectual property goods. Opening a shrink-wrapped package or clicking an "I agree" button on a license agreement in an online purchase would be considered agreement to contract terms. License restrictions on the use of content that you might acquire, now under license rather than purchase - might prohibit you from making personal copies, loaning it to another person, or even criticizing it publicly, or only allow you to use the content for a limited period of time. Such license terms - many of which are not enforceable by technical protection systems (one cannot imagine a technical protection system that tries to block the writing of critical essays about a work for example) - may be equally or even more severely at odds with consumer expectations.

The e-book reader is fundamentally agnostic about the technological control of intellectual property. It can be used as a very powerful instrument for such control, but it need not

incorporate such features. It can be limited to serve only as a convenient portable reading device. Depending on what capabilities the book reader manufacturers choose to incorporate, publishers may be more or less willing to supply content. Depending on the policies that the publishers set for using the control capabilities that may be present, consumers may be more or less willing to buy. And, of course, by encouraging the transition to commerce in digital content under license agreements, e-book readers create the possibility of using license terms to restructure usage practices for content.

A final point about the first sale doctrine. While this has been valuable to consumers, it has been the lifeblood of libraries. First sale is the framework that has historically allowed libraries to operate in America. As we move to a world of digital books, licenses, and technical protection systems, there are very real questions about whether, how, and at what price libraries will continue to be able to provide access to this digital content. I will return to this point later.

The Global Marketplace: Rights Management, Control, and Censorship

Networked information creates a globalized information marketplace. Historically, the Internet has been a world without borders or customs checkpoints or geography. This is at odds with the very geographically based traditions of publishing, where companies obtain the rights to publish works in specific regional markets. There has always been leakage in this system; travelers purchasing books abroad and bringing them home, for example, or bookstores importing a few copies of works that haven't appeared in print locally and selling them at retail for a premium. There are specialty music dealers that import to the U.S. market audio CDs that have been released only in Japan. But none of this has any real economic significance. Further, the geographic distinctions have been steadily diminishing; it is now unusual to find books or CDs available in London that are not just as available in New York. With the recent capability to easily order works from network-based booksellers anywhere in the world this system has begun breaking down on a larger scale, to the extent that publishers are starting to feel economic effects when they do try to maintain geographically-defined markets. Perhaps the best example of this was the third *Harry Potter* book, which appeared in the United Kingdom several months before it was released in the United States. Large numbers of copies were ordered from channels like Amazon UK for shipment to the United States, to the considerable annoyance of the publisher that had purchased the U.S. rights to the work [20]. Net-based content - which can move across the globe without the inconveniences of customs or lengthy international

shipping delays - threatens to seriously upset some long standing business practices.

Books and audio content were effectively limited to regional markets by availability. For video works, incompatible regional standards were another barrier to the international flow of content outside of publisher sanctioned channels. A video cassette purchased in Europe and encoded according to European standards would not play on an American VCR player, for example. While these incompatible standards certainly weren't established to help keep regional markets in place, they have been convenient for that purpose. Someone in the United States interested in video content released in Europe not only needs to find a source for the content, but also a European VCR. But the consumer electronics firms don't want to produce different products for different markets. They'd much prefer global standards that let them develop a smaller number of products that can be sold everywhere.

Appliances can incorporate and enforce geographic market constraints and act as a bulwark against the tendency of digital content to easily jump national boundaries; for example, Digital Video Disk (DVD) players include a regional code. DVD disks are coded with the regions in which they are allowed to play. A tourist who purchases a DVD disk in Europe or Australia and brings it home to the United States will encounter difficulties playing it on his or her home hardware, even though the DVD content standard is consistent worldwide. Some DVD support for regional restrictions is in software, particularly for DVD players that are part of computer systems rather than stand-alone consumer devices - and makes it relatively easy to bypass the region code restrictions. DVD player software on general purpose computers will often allow you to set your region, and perhaps even change it a few times. Hacks are widely available on the Internet to help users either turn off region checking or allow unlimited region resetting. In response, manufacturers are moving this enforcement into firmware and hardware in the DVD players, more towards a closed system model, even for DVD players that are part of general purpose computer systems, since obviously the computers themselves can't be trusted.

Interestingly, NuvoMedia, which made the Rocket eBook, issued a 28 April 1999 press release (prior to being acquired by Gemstar) announcing support for what they call the "Territorial Rights Management System" to support geographic limitation. It's not clear what the granularity of this is, but if it is finer than the DVD regions - for example, if it's country by country - it is a powerful system of control not only for marketplace segmentation, but also for various forms of censorship and information control.

Part of the motivation for continuing to support regional markets is to preserve the economic arrangements that are structured around them - but part is also about honoring national policies. Texts, to the extent that they represent traffic in ideas, have always been seen as very dangerous imports, and many nations have chosen to control them. The controversies over the shipment of physical copies of Hitler's *Mein Kampf* (ordered

through Web-based booksellers) into Germany, in violation of German anti-Nazi statutes), may hint at problems to come, as may the recent judgment in the French courts against Yahoo for making Nazi memorabilia available at auction. If digital books become network-based information objects, the very nature of the Internet militates against any controls on where they can be delivered, though the most restrictive nations will try to control this in much the same way they control access to other information accessible on the Internet today [21]. E-book appliances can build in geographical sensitivities that reflect not only regional marketing constraints, but also national censorship policies. And large multinational corporations can be very accommodating on these issues: consider the responses to the issues about *Mein Kampf* and Nazi memorabilia. Or look at the history of News Corporation appeasing mainland Chinese interests about content on cable television, or even book publishing, such as the case of former Hong Kong governor Chris Patten's book on Hong Kong.

Such national "filters" can be used in several different ways. By legislating the use of readers that support such restrictions, with the cooperation of multinational content suppliers, a nation can go a long way to ensuring that undesirable content is kept out. It's no longer sufficient to smuggle in content, one has to smuggle in readers as well. In addition, with the cooperation of reader manufacturers, it's possible to keep locally developed content limited to the nation that created it. Finally, the monitoring capabilities of rights management systems can be used to inform governments, not just commercial content providers. As far as I know, there has been little examination of the extent to which international trade agreements and treaties may encourage or discourage such uses of the technology.

It is interesting to note that in the print world U.S. research libraries have spent a great deal of money and effort to create and maintain specialized research collections of the local literature and culture of foreign nations. Technological controls to enforce national boundaries and content policies or regional markets may well put an end to such activities, or at least make them much more difficult and costly.

Regional or national controls over the viewing of content are really just a specialized application of technical protection systems. For consumers, particularly in nations that do not control information flows, the controls represent a novel and annoying set of restrictions on the ability to acquire and use content, as well as the repudiation of the promise of the network as a global information marketplace. But they also represent a new and relatively unexamined locus of control on the use of digital information and have disturbing implications for restricting the international flow of information and for facilitating national censorship policies. Also, we should recognize that nation-states do not give up their borders easily and that the technological attempts to re-establish national control are in fact widespread. There is now great interest in so-called geo-location technologies (such as Digital Island's Traceware) which allow network servers to determine from what nation users are originating from as a means of incorporating

national policies into the services provided by networked information resources [22].

Books Are Not Music: Reframing the Debate About Control Over Content

Over the past few years, I had the privilege of serving on a U.S. National Research Council committee which published the report, *The Digital Dilemma: Intellectual Property in the Emerging Information Infrastructure*. I've drawn from some of the findings of this report in the last few sections, though I want to be clear that the opinions expressed above are mine, and do not necessarily represent the findings of this committee. This report is a rich and extensive examination of the intellectual property and technology issues that I've discussed, and is an excellent source for further reading on these topics [23].

In the past, miners used caged canaries to tell when the air in a mine was going bad; when the canaries stopped singing, it was time to get out before the air became unbreathable. *The Digital Dilemma* uses the metaphor of music as the canary in the digital coal mine; it argues that what happens to the music industry may be a bellwether for the broader array of intellectual property industries in the digital world.

I've tried to explain how developments in the music industry, which indeed serves as a canary in some senses, may be influencing the thinking of the publishing industry about its relationship with e-book readers. I believe that these influences are real. But I also believe that music is the wrong place to frame the public policy debate.

Music is ephemeral. It is widely viewed as entertainment. At some very real level, our society doesn't consider it to be important in the way that books are important. Books carry big ideas; they document history, politics, and intellectual currents. Books are dangerous; they cause wars, and governments over the years have banned, confiscated, and censored them. People die for writing books and for believing what is written in books. Books convey and illuminate religion and science. Our laws and the actions of our institutions are codified in books, or at least texts. Books are serious. Suggestions that government or commercial interests might control what we can read imply that they might also control what we can know and what we can think in a way that control over music could never achieve. Books represent our intellectual and cultural heritage. Censorship of books is a profound matter that implies censorship of ideas; censorship of music does not carry the same implications, for most people. The freedom of the written and spoken word

is enshrined in the U.S. Constitution and protected by courts and laws; this has been extended to other forms of communication, but it begins with words and texts. Restrictions on the sharing of books are tantamount to restrictions on the sharing of ideas. This is why libraries are so important to our society; it is one of the reasons we fund and honor them. The preservation of our books and other texts forms the core of the preservation of our intellectual record.

I believe that books, rather than music, are the right place to think about the implications of technological controls on content. This helps make it clear what's really at stake. Later in this paper, for example, I will discuss the implications of technological controls on our ability, as a society, to manage the record of our intellectual discourse, which is primarily textual. E-books, of course, form the nexus of the public policy debate about the future of textual content.

We must remember that the publishing industry is not the music industry though with mergers and acquisitions and the growth of media conglomerates over the past two decades, we may increasingly see the traditions and perspectives of the two industries struggling for ascendance within the same conglomerates. Relationships between creators and publishers vary substantially between print publishing and the music industry. While both industries have some tradition of defending free speech and opposing censorship, this tradition runs much deeper in publishing. Publishers also think in terms of permanence rather than ephemeral products. The music industry has been described as being at war with its customers, as viewing every customer as a criminal and proceeding from there. This is not true of the book publishing industry. For an excellent view of the peculiar and sometimes traumatizing copyright and economic history of the music industry, see Charles Mann's ["The Heavenly Jukebox,"](#).

The issues at stake here cut two ways. One is about whether consumers will accept a "print" publishing industry that pursues the same practices that the music industry seems eager to establish. But the other is the extent to which the publishing industry will follow the lead of the music industry in pursuing these practices. I think there's some reason to be hopeful that this won't happen, that the publishing industry will honor the importance of managing the cultural and intellectual record, and will ensure the free and transnational flow of ideas and the exchange and sharing of thinking among readers. Perhaps the publishing industry will even ultimately set a standard that other industries will follow.

And to the extent that all of the content industries, but particularly publishing, pursue and successfully market policies and practices that are at odds with consumer expectations and the broader public interests in such goals as preserving our intellectual heritage, I believe that texts are the right test case to use in formulating and evaluating public policy to remedy these problems.

Restructuring the Publishing Value Chain and the Publishing Industry

There are a number of structural changes that are taking place in the publishing industry. The 1980s and early 1990s were a troublesome time for those concerned with diversity in publishing. We saw the rise of national bookstore chains and the increased homogeneity of offerings from one bookstore to another in the retail marketplace. In publishing, there seemed to be a trend to blockbuster bestsellers that crowded out a much larger and more diverse range of works. It appears that fewer mid-list or niche books are being published, and those that are published are staying in print for shorter and shorter periods of time. There are lots of reasons for this. In brick-and-mortar-based bookselling, display space is at a premium. Publishers pay inventory taxes on warehoused books that they haven't sold. Large inventories of unsold books are a liability. The problem of inventories was aggravated by changes in tax and accounting practices, instigated by decision in *Thor Power Tool Co. v. Commissioner of Internal Revenue* [24].

Network-based bookselling (Amazon.com, Barnes and Noble, Borders Online, and a host of other players) is putting all publishers on a somewhat more equal footing as far as finding readers. There's an infinite amount of virtual "display space" available, and books can become visible to potential purchasers in new ways through searching or recommender systems [25]. A university press or small publisher can be as accessible as a major commercial publisher, or nearly so, through these Web sites. At least in theory, author self-publishing (the ultimate "small press") becomes more practical as these kinds of sales outlets are combined with electronic delivery (eliminating the up-front investment in producing physical books and arranging for order fulfillment), though there are still a number of barriers to this [26]. In the past few years we have seen the emergence of a large number of digital "vanity presses" to serve authors who don't want to take the final step to full self-publishing but who cannot find traditional publishers or don't want to work with traditional publishers. These self-publishing services often provide the authors with greater control and much more generous royalties than traditional publishers.

The major problem with self-publishing or vanity publishing is still finding readers, particularly when the quality of the offerings through these channels is so variable (and often poor) [27]. Self published or vanity published non fiction books probably have the advantage over other materials like fiction or music if they are to be discovered by searching content or reviews. Returning a work of non-fiction in response to a query about coffee growing in South America is likely to be less subjective and easier to assess than a

piece of music that is claimed to "sound like" John Coltrane or to appeal to Coltrane fans. Some self-published books can find at least some audience without the need for expensive advertising campaigns. There are still, of course, questions of authority and quality, but there are ways of at least partially addressing these concerns through reviews and recommender systems. Online bookselling using truly massive inventories of traditionally published works has been a great success with readers, though many of the companies involved are not yet profitable. The weaknesses of this model are the need (and cost and delay) of delivering the physical goods that are purchased, and the limited ability to browse (partially compensated for by the availability of tables of contents, reviews, reader comments, images of covers, sample chapters, and other surrogates). The success of author self-publishing, or digital vanity publishing, seems less clear, with perhaps a very few exceptions.

These trends should at least in theory lead to the publication of a greater diversity of books and a greater visibility of this diversity to the book-buying public - though hard data supporting these claims is still scarce. And as long as these electronic booksellers are delivering physical books, there's still the problem of keeping material in print. Network-based bookselling helped to address the problem of making a wider range of material available to readers; digital books will address the problems of inventory and delivery.

The cost of keeping material "in print" electronically for delivery or print on demand is small (although the tax and accounting implications have yet to be fully resolved, as far as I know), at least until the material must negotiate a format and standards transition, at which point an investment is necessary. Out-of-print material also seems to be coming back into print for electronic delivery through the efforts of tiny niche companies such as Boondock Books, as well as major players like Bell and Howell Learning Systems (formerly UMI) or Netlibrary. If it becomes possible to keep a bigger backlist alive for sale electronically without paying tax on it as inventory and having to treat it as an accounting liability, then the development of a market in these electronic materials will again reshape publishing in complex ways. While it will help publishers to make more works available for the long term, it may create new problems for authors. Authors who got return of copyright for their works when the publisher took them "out of print" will be out of luck in the new digital world of delivery on demand. They can remain in limbo forever, making pennies in royalties from the occasional electronic sale. Because of this, new contracts between authors and publishers are now often framed in terms of a specific length of time, rather than an indefinite period until a work goes out of print, and such terms are a hotly contested area of negotiation.

There are also some fascinating social questions about the nature of authorship and audience here as we think more broadly about digital books, as opposed to electronic distribution of print books. For example, what are the reader expectations about updating published work? Is an author ever really "finished" with a book (other than perhaps a novel) in a world of electronic distribution? Recent attempts to translate printed reference

works such as the *Electronic Encyclopedia Britannica* to the network environment are already encountering these issues, particularly reader demand for continuous updating of articles [28]. We are also seeing a series of experiments that create dialog between the author and his or her readers, either following the initial publication of a work in digital form [29] or as a part of a more extended publication "process" [30]. Similar experiments have also been conducted in electronic journal publishing.

A great deal of money is at stake in restructuring distribution channels. For a mass-market printed book, about half of the retail price goes to parties "downstream" from the publisher, that is retailers and wholesalers [31]. There's also the cost of accepting returns, an unusual and costly book industry practice under which a bookstore can return unsold books to the publisher for a refund. Internet-based bookselling and, later, digital book delivery will eliminate this cost. For Internet-based booksellers, there's still a lot of cost in obtaining and delivering the printed book, though the aggregate cost chain from publisher to consumer is reduced. Sales of digital books eliminate most or all of these costs, depending on the marketing model and how many intermediaries remain active in the sales chain [32].

As relationships among publishers, consumers, and retailers change, and sales, shipping, and delivery costs become much smaller with network sales and electronic delivery, the changing economics will mean greater profits to publishers or electronic retailers, larger royalty shares to authors, and even reduced prices to consumers [33]. We can expect to see major struggles around how the newly available dollars are divided, particularly in author-publisher relationships (where some of the major publishing houses are now offering very generous royalty percentages to their authors for digital publications) and with the growing alternatives of self-publishing and a multiplicity of upstart small publishers putting pressure on the large established industry players. There will clearly be a reconsideration of what value publishers add for various kinds of authors, and what authors should be willing to pay for that value. One of the most fascinating questions will be how the level of public recognition that an author enjoys relates to the potential value that a publisher offers. In addition, new claimants are emerging to demand a share of the revenues from the restructured e-book distribution chain. For example, Gemstar, which has made it clear that it wants to control not only reading devices but also the retailer services that offer content to these readers, speaks of collecting 10-20% of the revenue from e-books licensed to Gemstar reading devices.

We should be mindful that e-book readers are not just for books; they are for newspapers and magazines as well, where the daily or weekly printing and distribution of "disposable" paper is a very large cost. If these e-book readers permit newspapers to eliminate paper, printing, and delivery costs for large numbers of subscribers, this will have a big impact on profit margins.

The used book market has always annoyed publishers (and sometimes authors as well), because they don't receive any revenue from these sales due to the first-sale doctrine. For most types of books this isn't enough money to worry about, but there are a few niche markets where resale by book purchasers represents a significant economic impact for publishers, such as textbooks, where perhaps 20% of the sales are in the used market. Publishers do many things today to keep the used textbook market at bay, such as releasing new editions of popular textbooks every few years. Electronic delivery, in conjunction with technological control of content, could wipe out these resale markets overnight and yield significant revenue opportunities [34]. We can expect these types of books to be early targets for transition to digital forms, not only because of the enhancements that the digital medium offers the author for more effective communication, but for economic reasons as well.

Finally, e-books promise another kind of restructuring in the publishing markets. In general, publishers do not know their customers; a complex chain of wholesalers and retailers serve as intermediaries. Retailers accept cash, further contributing to the anonymity of readers. Publishers sell very few books direct to readers in the print world. In a world of e-books, particularly where there may be few cash transactions, publishers may get to know and track the behavior of their consumers for the first time. Certainly network-based retailers will be able to track their customers better because few will be anonymous. We may see more direct purchasing from publishers. Network digital book retailers may actually pass transactions through to publisher servers (along with purchaser identity information), or may simply report this information while supplying the books to readers directly from retailer servers. There may be compelling reasons why one wants to register ownership of an e-book with the publisher [35]. One can even imagine downloading a new e-book and having that e-book provide its publisher with an inventory of the other e-books stored in one's personal library. Another important point to recognize is that digital rights management systems can report actual viewing usage, which is a very different thing than purchase patterns.

The privacy implications here are substantial, particularly if one is skeptical about the confidentiality of the records of transactions with publishers and booksellers in a world where many more such records exist and may even be remarketed or sold as assets [36]. Recently there have also been a number of attempts by law enforcement agencies and prosecutors to obtain book purchasing habits, the most notorious perhaps being Ken Starr's pursuit of records of Monica Lewinsky's purchases at Kramerbooks in Washington D.C. There have been others dealing with purchase records for books detailing methods of manufacturing drugs, for example [37]. Libraries have been skeptical of legal protections for a long time, even though library circulation records have been protected under various state laws. Best practices in libraries keep circulation records for books only until borrowed items are returned; most libraries do not maintain a record of books that have been borrowed and returned, and thus cannot make such records available even under

subpoena.

Again, the culture of books may be a bit different, and may give rise to stronger commitments by publishers and retailers to protect consumer privacy, and even ultimately to support strong legislation protecting this privacy. All of the same issues apply to music as it comes to be marketed across the network - but people are likely to be far less concerned with the privacy of their listening habits than of their reading habits.

Assessing E-book Readers

We have discussed a number of challenges to the acceptance of digital books. Those that closely mimic printed books, or that represent the digitization of existing printed works have problems because they are not easy to read on-screen. Those representing a reconceptualization of the printed book face formidable challenges in their authoring, economics, and acceptance; these are emerging rather than mature forms. To the extent that digital books replace printed books in today's marketplace, control of digital works is clearly a central issue for publishers. They will be reluctant to make digital books available without confidence that they can't easily be duplicated and redistributed, and will perhaps seek much more control over digital works, say in the ability to implement new pay-per-view business models, to control and monitor use, and to control resale and geographic distribution of their materials. The techniques necessary to establish these controls may run strongly counter to consumer expectations and preferences. How does the emergence of the e-book reader (as appliance or software) address these challenges?

E-book readers are supposed to make on-screen reading of lengthy texts acceptable through improved display quality. While the resolution of book reader appliances is sometimes better than the 72 dots-per-inch (dpi) that is the industry standard for monitors, it isn't that much better in the devices on the market today. Display quality is also being enhanced by technologies such as Microsoft's ClearType, which exploit the properties of LCD displays to offer crisper text on the screen. Easy reading will probably require at least 200-300 dpi, plus some optical properties that are closer to paper than today's screens. Researchers at MIT, Xerox's Palo Alto Research Center, and other institutions - and commercial spin-offs such as E-Ink - are trying to invent digital paper, but this is a longer-term effort [38]. Prices for very high quality displays still need to come down. The solution to the on-screen reading problem isn't yet in place, but there's evidence it's coming.

If you can put a display of this quality on a consumer electronics-oriented book reader,

can you put it on a general purpose laptop computer? Yes, and the competition, both in price and functionality for the book readers will be the next couple generations of laptops. E-book readers won't be able to compete for long on display quality. Indeed, there are other display issues working against the appliance readers. The first generation of appliance book readers offered monochrome (black and white) or grayscale displays. This is fine when dealing with textual materials, but has limitations for illustrations, and is a major problem for the new genres of digital works, which often make heavy use of multimedia. This is much like the situation with early laptops, and I would expect that within a few years color displays will become the norm, but there will clearly be a period when inexpensive appliances will not be able to compete with software readers for general purpose computers in presenting certain types of content where color is important. Similarly, e-book readers today don't support video and audio well at a high level of quality (if at all), while laptops have the computational capability and graphics support to do so.

From a hardware point of view, in the long term I suspect it will be very hard to tell an appliance reader from a laptop, except for three differences. It will need fewer ports for connecting peripherals, it won't need a hard disk, and it won't need a keyboard. These translate into some significant size and weight advantages. Omitting the disk helps battery life. And appliances may be able to get by with a smaller display. Will it be worth purchasing both?

From the user's perspective, software is critical. E-book appliance software should be largely invisible. It has a single function, and it needs to be simple, reliable, and robust. Software for laptops (or general purpose computers) is still complex and fragile. To the extent that the appliances can avoid reliance on general purpose operating systems for personal computers - and particularly Windows with its Byzantine complexities - they may be able to retain an important competitive advantage. It will be a qualified one: the new digital genres rely on the richness of the complex personal computer environment, and will not be usable on appliance readers.

Many of today's appliances do not stand alone; they are used in conjunction with a library that is stored on a personal computer. In this sense, even if the appliance itself is simple and robust, the user must face the general purpose computing environment for obtaining works and maintaining his or her library. This is likely to be a disadvantage for some consumers, and we are already seeing an increasing emphasis on direct network connections (via modem or Ethernet port, and in future, wireless) in the latest generation of appliances to avoid this dependence. Bandwidth available for downloading books is an overlooked issue. To the extent that consumers have broadband connections so they can very quickly and easily download current newspapers and magazines, or books of interest, this is likely to change user behavior. Ubiquitous broadband access will also narrow the gap between what has been done historically with new genre works on media such as CD-ROM which offer predictable retrieval speeds and make it convenient to work with

continuous media (audio and video) and those that have been designed for the networked information environment, where such continuous media are temperamental and problematic, and not uniformly available to readers today, since many readers are still limited to dial-up connections. Broadly available high-speed wireless access, when it comes, will complicate matters further. It will bring much greater convergence between the niches currently occupied by appliance book readers, PDAs, and personal computers, and make it much less important that books actually be stored on the local device, with interesting implications for how the technologies to control intellectual property will operate.

One of the central issues is control over content; this was overlooked in much early thinking about electronic book readers, but has now emerged as a central issue. Here, appliance readers have an enormous potential advantage as discussed earlier. They represent a closed system where a reasonable level of control is relatively easy to establish. Content providers may offer content only for e-book appliance readers but not for general purpose computers, where providers can't be as confident they can control the content. And this may change the shape of the marketplace, and move the debate away from whether appliances or general purpose computers are better environments for reading new digital works to a take it or leave it proposition. Publishers may offer consumers the choice of existing print venues or appliances (or perhaps also a few specific software e-book readers) for digital works, and consumers will have to decide whether they are willing to accept these new marketplace offerings.

To the extent that e-book readers incorporate technical protection technologies, these technologies are in some sense neutral about what specific controls and limitations on use and copying will be put in place. They establish mechanisms that implement policies, and languages for defining these policies. The policies themselves will be set by content providers, not hardware or software manufacturers. The actual policies that publishers choose to attach to digital works will be critical. How far will they try to go? Are we, as consumers, willing to accept new constraints over the way we use the new digital books - to be unable to loan them to our friends, to consult them indefinitely? Are we willing to accept pay-per-view reading, or the "rental" of a work for a limited period of time? Can we accept ownership of a book even if we can't copy a few pages without hand transcription?

Control over use has many dimensions, not just the ability to control copying and the ability to meter and control reading, but also control over presentation. In theory e-book readers should be able to permit users to alter fonts and font sizes (providing large-print books on demand), facilitate use by people with various kinds of disabilities, and even provide an option to read your book out loud if you wish (and, with appropriate markup in the text, even allow you to potentially cast the people you wish to read the various parts of the book). These all represent real - and in some cases utterly compelling - added value to at least some readers. It remains to be seen how aggressively these capabilities are

incorporated into e-book readers, and how willing publishers will be to enable the use of such features with their content (and under what terms). There is an interesting balance here between artistic integrity (and publisher control) on one hand and the desires of readers for flexibility on the other. For example, in traditional publishing some authors retain control over *dramatis persona* that can participate in creating a "talking book" version of their work. We will need to re-negotiate these balances for digital books.

How well will e-book readers - either appliances or software for general purpose computers - be adaptable to new, emerging genres? Right now, it looks like they aren't compatible. Appliance readers are not really network devices, though they may download content from the network for reading. They are not portable Web browsers, and they don't come with radio modems that permit them to be continually connected to the Internet, at least today. E-book readers are designed to render a specific set of local content and let people navigate within it. They don't incorporate interactive database access and simulations, which are becoming routine for certain types of digital books. Software readers have more flexibility than appliances to potentially incorporate these capabilities, although they do this at the risk of further weakening their abilities to control content use.

Are we looking forward to the new digital genres, or backwards towards digitized printed pages as we think about digital books? I believe that current appliance readers look backward - but this is not as limiting as it sounds. There is a tremendous wealth of printed books that have already been written, and the vast majority of these can only be translated into the digital environment, not reconceptualized and rewritten for it. There are many genres and styles of discourse that are well suited to the printed book and this will not change. We will continue to produce printed books, and will want to transport these into the digital realm as well as having them in print. We may well see a market that embraces e-book readers for older materials transported into the digital environment and new works of authorship that wish to remain within the old genres, and also general purpose computers for access to works that are authored within the emerging inherently digital genres.

Is full access to works constructed for the networked information environment going to be compelling for the vast majority of readers, and if so, how soon? When will mainstream authors begin to exploit the possibilities of either e-book readers or more general digital books, and which option will they choose? And for what kinds of works? The new digital genres require rethinking and relearning the craft of authorship, and there are still many stories best told through the traditional linear book and many arguments best presented as lengthy textual passages.

The Role of Standards

Standards are going to be critical to building a marketplace in digital books. Severe problems are already emerging here. When purchasing an e-book, one has to specify what platform one is purchasing it for, and some e-books are available only for specific platforms (and not for other functionally equivalent platforms) simply because the vendor hasn't produced a file in the format specific to each available platform.

At a very high level, we see the new genres, as networked information objects, adopting the full range of evolving standards for multimedia information in the Web environment. They are tracking developments in browsers, multimedia, interactive software products, distributed database queries and the like. Intellectual property management for these types of works is mostly about access management and authentication, not about technical protection systems that control copying and use. They are often being positioned as services, and to the extent that they operate within a commercial framework there is a license or service agreement, access control, and little else.

E-book readers, both appliances and software for general purpose computers, are following a different track. In terms of content, there are two basic philosophical models. Both think about content as a digital object that is moved from place to place and that represents something intellectually closely akin to a printed book. One model is to define a subset of HTML/XML that includes text and some limited multimedia components; the Open Book (OEB) standard (see www.openebook.org) provides a blueprint for this. This is supported by many companies including Microsoft. Many of the products on the marketplace today support some sort of HTML subset (but all different) and this would standardize content for that class of readers. There are some questions here about the ability to incorporate additional character sets, particularly non-Roman ones, which aren't supported on current products. These are important for scholarly works, for libraries that need to serve non-English-speaking communities, and for global markets. Mathematical and scientific notation is also a problem. The other philosophical model is to use Adobe PDF, which can work at a page image level. This is supported by several vendors, including, not surprisingly, Adobe, and it has lots of advantages. It makes converting existing print materials cheap and easy. It handles mathematics, non-Roman fonts, and other kinds of content that are awkward and inefficient to support in HTML or XML. It is also ideal for scanned versions of older books. Neither of these approaches is particularly hospitable for the new genres, though the HTML approach, based on Web standards, is probably the better of the two. I would suspect that in the near term we will see both standards supported - and a growing disconnect with the new digital genres, which are operating in a totally different conceptual framework.

As standards in this area stabilize, it should not be hard to translate existing digital books that are representations of printed books into one of the standard formats for loading into

e-book readers, just as it isn't hard today to move from one reader format to another (except for PDF based formats to HTML). In this sense, standardization in this area won't be very disruptive and should be to the benefit of all players; I'd expect it to take place quickly. The war between PDF and HTML is more fundamental, and probably unresolvable in the near term due to the complementary strengths and weaknesses of the two approaches, which is why I think both will survive. But OEB and PDF are ultimately only important to the extent that they constrain the capabilities of authors in creating books for the digital environment.

The much more complex and controversial set of standards for e-book readers address the management, control, and protection of content. These standards are not on the radar screen for most interested parties, but probably constitute the core of many interoperability problems that are going to be genuinely hard to solve among the different e-book readers. Management standards include cryptographic algorithms and protocols, identity and attribute management for reading devices (including what geographical market the device is in), metadata formats to define duplication and use restrictions on content, and protocols for transferring content and metadata. The issues here are highly technical and complicated, and some also link to other, broader efforts such as the development of public key infrastructure for identity management. Many of the techniques needed here are still untested in commercial deployment, and some are encumbered by patents. There are still debates about how secure and how practical some of the proposed approaches are likely to be. Standards efforts are underway here - for example, the EBX work that had been led by Glassbook and has now been merged with the Open Ebook Forum work - but there are also vendors with proprietary systems that hope to dominate the market and become *de facto* standards. Consensus in this area is a long way away. The implications of choices made in these standards are profound. It's anyone's guess whether an e-book-specific rights management system standard will succeed. There are many competing digital rights management standardization efforts, all of which are coming from some vector leading into the digital convergence of all media, for example, SDMI or MPEG-21. There's also a lot of discussion about standardizing rights management definition languages, which are a component of any broader rights management system, and which open up an additional area for interoperability problems.

Content standards shape the characteristics of the content we can view on e-book readers. Management standards will help determine how content targeted for e-book readers can be used, shared, and even preserved, and thus such standards have major social, cultural, and economic implications. It is important that society as a whole, and not just publishers and technology developers, has a voice in the development of these standards, because it will be critical to the future of cultural institutions such as libraries. But it is equally important to recognize that management standards represent a vocabulary within which publishers will specify the actual controls on usage, and thus the issues around intellectual property cannot be resolved through standards alone.

A Brave New World for Readers

Consumers of all kinds - everyone who reads for pleasure or knowledge, or needs to use "books" in their work - face a confusing array of choices and questions. Should they buy an appliance book reader, either today or when the next generation comes out? Should they install book-reading software on their computers? When should they consider acquiring digital books rather than printed works?

In a way, the new genre is the easiest to evaluate. A general purpose computer, the software that comes with it, and a network connection is all the technology that's needed, and these new genres often don't directly compete with print, or they offer a sufficiently distinct set of services that print is a poor substitute. They will stand or fall on their own merits. Issues of preservation and continuity of access are largely unresolved; indeed, we don't even know what some of these questions really mean in this context.

For materials targeted to e-book readers, I believe that content will drive decision-making. If there's enough compelling material offered at prices that are competitive with, or better than, competing print formats, or that isn't available in print formats, consumers will purchase the reader technology that they need to get that content. Obviously, standards will play an important role in achieving critical mass of available content. Publisher decisions about usage constraints and control will also be important; if the electronic formats are much more restrictive than print, this will deter adoption.

But it is critical that we take a more thoughtful look at digital books and e-book readers. One of the key points we need to keep in mind is that today's e-book readers are only one step in what's likely to be a long technological evolution, and that the business environment will also evolve, with some existing players disappearing and new players emerging. Here are some questions that I think need to be clearly asked and clearly answered. And we need to think hard about what answers are and are not acceptable, and to be prepared to stay with print if we don't like the answers, because the long term risks may outweigh the short-term convenience.

- Can you loan or give an e-book (or access to a digital book) to someone else as you can a physical book? To what extent do digital books mimic (and perhaps even improve upon) physical books, and to what extent do they break with that tradition? What other constraints on usage (for example, printing) exist?
- Do you own *objects* or *access*? If your library of e-books is destroyed or stolen, can you replace it without purchasing the content again simply by providing proof

of license or purchase? One very interesting service is a registry that allows you to replace your e-books if you lose your appliance [39].

- From whom are you really obtaining content - the e-book reader vendor, a publisher, or some other party? Who has to stay in business in order to ensure your continued ability to use that content? What happens if the source of your content goes out of business?
- Can you copy an e-book for private, personal use? If you own two readers, can you move a digital book from one to the other without having to purchase it again?
- Do you have the right and the ability to reformat an e-book or a digital book in response to changes in standards or technologies or do you need to repurchase it? What happens when you upgrade or replace your e-book reader with another one? What happens when you replace the PC that might house your "library"? What happens if you replace one brand of e-book reader with another, perhaps because your reader vendor goes out of business?
- Do you have to obtain e-books on a pay-per-view or other limited time rental basis or do you buy a perpetual license to the content, or ownership of a copy?
- What are the policies of the content provider with regard to your privacy and to usage monitoring? What limitations does your book reader technology place on the ability of a content supplier to collect usage data?

The Uncertain Future of Digital Books in Libraries

Digital books and e-book reader appliances raise some serious issues for libraries. It's surprisingly hard to disentangle those questions that are specific to digital books and book readers from those that are generic to network-based information resources. To the extent that digital books are important works of scholarship, libraries - particularly research libraries - have little alternative but to purchase access for their patrons, though in cases where there is a print equivalent to the e-book they may choose to acquire the printed work instead (or in addition). Licensing digital books raises the same questions that arise for general electronic content. Is it being licensed for in-library workstations or for access by library patrons wherever they may be? Are costs based on the number of concurrent users, on the size of the user community, or on some other factors? Are traditional library interlibrary loan functions supported for these digital works, and if so how? Do the license terms recognize traditional library and education values such as fair use, and free speech and inquiry? Are there provisions to ensure the preservation of the material if the library wishes to preserve it?

The good news is that for content where libraries represent a significant part of the

market, meaningful negotiation of license terms can and usually does take place. Thus when one looks at materials like scholarly journals that are now licensed in electronic form, libraries have generally been able to negotiate reasonable license agreements. The bad news is that much of the content that is going to be available as e-books is targeted for consumer markets, and libraries do not represent a sufficiently large part of the market to get the attention of publishers in license negotiations; they will have to decide whether they can live with the terms offered to general consumers. There will be little meaningful negotiation about license terms for best-selling works, just as there is little meaningful negotiation about licensing terms for mass market software.

In the area of e-books, libraries are, I believe, confused about what they want, particularly in terms of business models. Assume that standards are resolved. E-book appliances (and probably even software readers) can potentially mimic the behavior of books - the library can acquire one, a single person at a time can view or borrow, and returns can even be automatic (access revocation). This is very different from site license to digital content, which has been the primary model for electronic information for the past few years. Do libraries want to regress to emulating the printed book? Or do they want to use digital books within a site license framework as an extension of current trends, treating e-book readers as just another display technology that their patrons may exploit? Or do libraries want some new hybrid solution that permits, for example, the acquisition of "peak load" copies of popular works for circulation for a limited time when they are popular and in high demand? Obviously, some of this depends on the terms and prices that libraries are offered. Libraries want to maximize access and service at minimal cost (in other words, get an unlimited use site license for roughly the cost of a single print copy), which in some sense is in direct opposition to publisher goals. There will be hard negotiations ahead. It is important for the library community to start talking about what they want from digital books, particularly in light of the ability of technical protection systems to enable new business models.

Specifically with regard to e-book readers, libraries have a host of more tactical questions to consider. There are a number of incompatible e-book reader formats. If critically important materials are available only for one specific e-book reader and not in any other format, libraries will have to deal with this today. Many libraries will quite reasonably wait until there is some convergence on standards in the marketplace before making any large-scale commitment to digital books that are targeted for e-book readers. They will then have to consider whether to assume that patrons will acquire e-book readers, or whether to purchase and circulate readers as well as content, much as some libraries loaned VCRs in the early days of video cassettes. Circulating e-book reader appliances will be a short-term decision, at any rate, while prices drop and the devices become more commonplace. In the longer run, they may offer a few e-book readers for in-library viewing and not much else.

If appropriate pricing schemes can be developed, digital books targeted for e-book readers

can offer some real benefits to libraries and their patrons. But in the quest for more responsive customer services, libraries must not overlook the longer-term societal goals and cultural missions.

Consider again the list of questions that were raised above for consumers, but consider them from the perspective of library values and library missions.

While patrons may be willing to make compromises about privacy on the basis of expedience and need (and have every right to make these compromises), libraries have historically been more thoughtful and principled. It will be important to clarify the privacy and use monitoring issues that surround digital books and e-books in a library context. And libraries will also be very sensitive to the terms of licensing or purchase agreements; these must preserve values such as free speech and discourse. Libraries want to ensure that they have copies of the works that can be incorporated in permanent collections for continued access, as opposed to pay-per-view or expiring versions of the works. Libraries need the ability to migrate these works in response to changing technologies. Libraries recently faced the threat of vanishing collections through license agreements for electronic information that had a finite term of use. They are now beginning to address this successfully in license negotiations with content providers. With e-books they may now face a new threat, planned content obsolescence. They also need the ability, for example, to incorporate into their collections works that were not necessarily intended for sale in their geographic marketplace, in order to provide patrons with a diversity of perspectives and voices, and to document world-wide cultural and intellectual developments.

Finally, with all of the hype about digital books, libraries must also continually remind their management, funders, and boards that e-book readers and digital books are unlikely to substantially reduce demand for shelf space to house physical books for the near future.

But beyond all of these issues, and more fundamentally, we must ask whether there will even be a role for digital books as part of library collections. Such a role seems clear for those works that are primarily targeted at the scholarly community, because libraries serve as major purchasers on behalf of that community. Given that many new genre works are appearing from that community today, it seems likely that these will be particularly well represented. But for consumer market e-books the answer is unclear, and is up to the publishers. If the model is that an e-book is downloaded and permanently locked to a single reader from a retailer, this means that libraries will be unable to loan such e-books in a meaningful way. They will either have to purchase a reader for each e-book (or small group of e-books) and loan the reader along with the content, or make the material available only for use within the library under a similar scheme. Unless publishers actually make a business and technical framework that allows libraries to circulate an e-book into a reader, reclaim it, and then circulate it into another reader, there will be little role for libraries here. And this also raises the possibility that publishers will want to charge a higher price for a "library version" e-book, or to charge the library a "per

circulation" fee, or both; this is entirely at the publisher's discretion. One might also see publishers forming alliances much like the film industry has done with Blockbuster Video to establish a system of commercial, pay-per-circulation "libraries" for e-books, while refusing to market e-books to traditional libraries in a form that can be circulated [40]. In the print world, and with the doctrine of first sale, libraries could acquire and subsequently circulate and preserve any works that were made available in the consumer marketplace. In a world of digital information, e-book readers, licenses, and rights management systems, libraries have no such automatic capability, and can function only at the pleasure of the publishers.

For large classes of content, libraries may not represent a large enough market to cause publishers to accommodate library requirements, or they may be asked to agree to prices and license terms that are intolerable. I believe that a future in which libraries are prevented from collecting and providing access to a substantial part of the output of the publishing industry would present a very serious social and public policy problem, both in terms of public access to information and the preservation of our intellectual and cultural heritage. Should such a situation come to pass, the potential solutions would be legal, legislative, and political, and based on the track record of legislative developments over the past few years, I think there is reason to be profoundly concerned about whether the problem would be adequately addressed.



Continuity of Access and the Preservation of Our Intellectual Heritage

Long-term access to content that one has acquired is very important, far more important, perhaps, than many of the content industries (which are focused on a world of hits, fads, and rapidly changing fashions in entertainment and content) realize. This is an issue of major concern for libraries and other cultural heritage organizations which bear so much of the responsibility for managing and preserving the social and cultural record, but also for consumers who love, value, and honor books, music, and other artifacts of creativity and intellect, and who want to share and pass on these works.

On one hand, digital convergence is happening. Everything is being reduced to uniform streams of bits that can be managed by digital rights management systems; all streams of bits are the same. On the other hand, when we think about what the various streams of bits represent, our expectations are not so straightforward or consistent. We view some works as experiential, with no strong expectation of ever being able to experience them a second

time (a play, a live performance). Others we regard as more permanent, but still perhaps ephemeral. They may not be obtainable forever, or we may know we would have to go to a vast amount of trouble to encounter them again after their season has past (television broadcasts, advertisements on television or radio, software). Some we regard as permanent. We acquire them with the assumption that we can keep them as long as we want. We further assume that they have been made part of our intellectual and cultural record, and that we could gain access to them through institutions such as libraries if we did not purchase them. Most books - indeed the vast majority of printed works - are in this final category.

We have been able to keep books for a lifetime, though they may be long out of print and out of fashion or commercial viability, to consult them again and again, though their pages may become brittle and their bindings fragile. We have been able to pass books on to our children, our nephews and nieces, to share them with friends, to keep them in libraries for future generations to learn from and enjoy. Can we accept a world where this is impossible?

Print has historically been extremely long-lived because it has enjoyed a unique lack of technology mediation and is one of the oldest media, certainly the oldest mass-produced and mass market medium. Paper - at least well-made paper - lasts a very long time. These properties are closely bound up with the unique cultural role and status of books. Recorded music has always been more fragile. The 78 RPM recordings of our parents and grandparents are only precariously accessible today, due to continually changing technologies, unless they have been reissued in current technology by a music company or transferred to a more modern medium (perhaps in violation of copyright) by someone who owned a copy of the original recording.

An audio CD player costs a few hundred dollars, as does a DVD player. But it can cost tens of thousands of dollars to replace an LP collection with audio CDs or a video cassette collection with DVDs, and this within a span of only a decade or two. Are we willing to burden our personal libraries (and our institutional libraries) of books, music, and films with such costs in order to make a technology transition every decade or two in order to satisfy the economic models of the content industries? And to lose some precious, but perhaps not widely popular, works with each technology transition because they are not made available using the new technology? Do we have, and will we continue to have, the rights and ability to preserve content that we have already acquired in the face of changing technology? For existing materials the answer isn't entirely clear. We certainly have the ability, but the legal rights of consumers and institutions such as libraries are less clear. In a future world of license agreements and digital rights management technology governed by the Digital Millennium Copyright Act, both rights and capabilities are questionable.

We know that we do not fully understand how to preserve digital content; today there is no "general theory", only techniques. The best understanding today is that preservation of

works like digital books or sound recordings will be accomplished by migrating content (bits) and translating from obsolete to current formats, rather than by relying on the longevity of specific storage media and the hardware that reads such media. It is essential that the legal and business frameworks for content honor preservation and that they permit such migration of content. This is a problem and a challenge of all types of digital content - for scholarly works cast in the new digital genres as well as for electronic representations of printed books intended for use on e-book readers. It will not be easy to ensure the continuity of our cultural and intellectual record in the digital age on purely technical and operational grounds.

But the need to be able to preserve can be very much at odds with the objectives and technologies of content control. E-book readers, because they can integrate such content control capabilities effectively and because they facilitate business and legal models that can render content obsolete and inaccessible, represent a particular danger. Legal and business problems may dwarf the technical problems of preserving digital content, both for individuals and for institutions. Libraries, which make systematic, institutional investments in content on behalf of society as a whole, must be particularly vocal and articulate advocates of the need for preservation. But it is important to consumers as well, though they often "speak" only through their decisions in the marketplace (with libraries, to some extent, acting as their advocates in public policy debates), and may not realize the longer-term implications of enormous numbers of individual marketplace choices until it's too late.

Note that it's not sufficient just to make "preservable" versions of materials available to cultural memory organizations like libraries, even if everything that was published was offered in such a preservable version to appropriate institutions (which is a huge and probably unrealistic assumption). Libraries are sometimes fallible or shortsighted in their collection policies, and their resources are always limited. If one looks at the history of the collecting and preservation of cultural materials over the last century, one finds again and again situations in which individual private collectors built up collections of contemporary materials that virtually all libraries overlooked or chose to ignore, and decades later libraries and archives built their collections of these materials (which proved to be vital raw materials for scholarship and cultural history) through acquisitions or donations from these private collectors. Research libraries are often driven by scholarly needs, and it can take decades to develop canons and to legitimize new fields of scholarly inquiry. Individual collectors preserve and protect potentially significant materials while these changes in scholarship occur. It is vital that individual consumers, and not just cultural memory institutions, have the capability to retain indefinitely the works that they acquire and to preserve them. This of course does not mean there is no role for pay-per-view in the economy of intellectual property. It means that for the vast majority of works, and certainly for digital works that will carry the same cultural role and importance that the printed book carries today, arrangements that go beyond pay-per-view to permanent access and "ownership" of copies will be essential as a matter of public policy.

Forced obsolescence of content - the need to repurchase it over and over again for changing technologies, to hope that the content will be made available in the new format and that money can be found to acquire it again - is only one threat to the cultural and intellectual record. There is another, which is fundamentally different in the digital world, and perhaps even more insidious. Even in a democratic society such as ours, which places a high value on free speech, writings are occasionally banned for various reasons. This may be the result of a legal action by the government, a private lawsuit (involving defamation, copyright infringement, or other issues), or the decision of a publisher to withdraw a book from marketing (in response to public pressure, the discovery of inaccuracies or plagiarism). Sometimes in hindsight the banning of a book proves to be a great mistake, a cowardly response to some sort of social or political pressure, and the ban is later reversed. In other cases the withdrawal of a book from the market may be justified, but years later people may still legitimately want to study the work that has been withdrawn to understand and analyze the events surrounding the withdrawal. The act of banning the work itself becomes important, and the banned work is part of the evidence that comes into play in documenting that act. Typically, at least in the United States, the banning of a book means that it is no longer for sale, and unsold copies are destroyed. It does not mean that attempts are made to locate and seize and destroy copies that may have already been sold to individuals or to libraries or that the possession (or sometimes even the private resale) of such a work is itself a crime. To the extent that the event of publication occurred, a limited number of copies have already become part of the permanent societal record. These copies have been irrevocably placed under the distributed, autonomous, and largely untraceable control of a wide variety of individuals and institutions as part of the publication process. This is the characteristic of our intellectual record which gives it such enormous strength, robustness, and integrity. In the digital world, with its much-enhanced capabilities to track the disposition of copies of works and where access so often substitutes for possession, it is all too easy to envision legal demands for post-publication withdrawal, editing, or censorship of works that would be able to reach every copy of that work in existence, utterly undermining assumptions about the integrity of our cultural and intellectual record and providing the courts and the government with unprecedented and dangerous capabilities to re-write that record at will. To be sure, the ability to bind an annotation to all copies of a work indicating plagiarism, scientific misconduct or fraud, or other attributes that the reader should be aware of may well be desirable, if used wisely. However, the ability to censor or withdraw works on a comprehensive basis is, to my mind, one of the most terrifying capabilities that may be enabled by the new digital environment.



Defining the Future of the Book

While contrary to all the hype, it seems clear that the future of the book isn't purely digital, and that in addition paper will be an important user interface via print-on-demand, many genres of books are rapidly migrating to digital form.

Two different and distinct things are happening to the book as it moves into the digital medium. It is being *translated* rather literally into a digital representation, and it is undergoing a transformative *evolution* into new genres of digitally-based discourse. Both of these developments, which can be viewed as two opposing endpoints of a spectrum of digital content, may legitimately lay claim to being digital books (along with everything in between).

These transformative evolutionary developments are not, at least today, heavily constrained by issues of control and protection of intellectual property, or of revenue and economic viability. They are still largely experimental, and are focused primarily on improving the ability of authors to communicate and document. Developments in this world are taking place largely outside the framework of the new technologies that are specific to these translated electronic books and the readers that support them. It is important that as we explore and exploit the capabilities of the new viewing technologies we also continue to nurture the development of the new genres that are evolving. These are an important part of the digital future.

New technologies - both in hardware appliances and in software for general purpose computers - are developing to facilitate the use of digital books. These technologies emphasize support of books that are translated, rather than reconceptualized, for the digital medium. These new technologies should make many digital books more convenient, more readable, and more useable.

But we must not let the hype about a technological update to the printed book - the move from printed book to e-book reader - trivialize the enormous social implications of the change that is starting to occur. These new technologies come with a potentially steep social price. They provide new levels of control, monitoring, and usage restrictions for digital books that may well go beyond what consumers are accustomed to with physical print books, and they create serious questions about our ability to manage, preserve, and provide access to our cultural and intellectual heritage. Without such capabilities for control, existing print books may not move to digital form, or at least not quickly and in large numbers. Appliance readers provide particularly powerful levels of control; the capabilities of software book readers to offer the same level of control remain to be validated in large-scale deployment. It is likely, though far from certain, that those publishers will favor appliance book readers. The precise balance points between marketplace acceptance by consumers and demands for control by publishers have yet to

be established.

It may be that consumers, and indeed society as a whole, are willing to agree to the various costs of adopting e-books. As we have seen, they raise serious questions about the future role of libraries. But thoughtful, informed consent is critical here. Hidden agendas and unforeseen consequences - that emerge only after e-books have become extensively established through a consumer marketing campaign that persuades the public that e-book readers and the content sold for them represents the future - do not serve us well.

Copyright, with all of its social as well as legal ramifications, has always been an explicit and carefully wrought bargain between creators and society at large. E-books can reshape this pact in complex and wide-reaching ways.

Issues of preservation, continuity of access, and the integrity of our cultural and intellectual record are particularly critical in the context of e-book readers and the works designed for them. These have enormous importance both for individual consumers and for society as a whole, and for libraries, which manage much of the intellectual archives of our society. Most fundamentally, we face the question of whether libraries can continue to collect books as they move to digital form, particularly in mass-market publishing. We must not overlook these issues in our rush to adopt e-book readers and content distributed for them, and libraries will have a special obligation to speak out on these issues and to educate society about them, while also trying to work out viable arrangements with the content industries.

Finally, we must continue to recognize that digital books, in the broadest sense, are at least potentially much more than simply digital content translated from the print framework that can be viewed by e-book readers promoted by today's publishing establishment and technology providers as part of an agenda of market share, new revenue opportunities, or control over content. Digital books, in all of their complexity and potential, are as yet only dimly defined, and will be a continued focus for the creativity and ingenuity of present and future generations of authors, teachers and scholars.

I have argued at length here that the printed word, and particularly its manifestation in the book, holds a very special and privileged place in our culture and our society. As we think about the migration of authoring to the digital medium, the book - rather than other cultural products such as musical works - should be the benchmark against which we measure and test our assumptions and beliefs about the roles and uses of intellectual property in the new environment. We must remain mindful of this distinction, and not constrain the virtually unlimited potential of the digital medium to the traditions and business interests that have coalesced around the printed book over the centuries and that may now seek both to define a new canon of "book" in the digital world, regaining the control of the digital printing press that they suddenly lost with the creation of the World Wide Web, and to surround these new ebooks with new technology-enabled controls on content. We need to be careful not to prematurely marginalize any of the new genres the

digital medium may enable. The most compelling case for ebooks as relatively literal of the printed book is based on greater convenience and ubiquity of access, and somewhat enhanced use. The case for digital books broadly, as new genres of works, is about more effective communication of ideas, enhanced teaching and learning, and renewed creativity. While the first case is a good one, if the price is not too high (in social as well as economic terms), the second case is truly compelling and inspiring. The future digital book will take us far beyond today's printed books and publishing industry, in many different and sometimes unexpected directions, though our points of departure will inevitably be an important influence. Let us welcome the journey and be open to many destinations; we will find treasures and wonderful surprises along the way. 

About the Author

Clifford Lynch is the Director of the Coalition for Networked Information ([CNI](#)).

E-mail: cliff@cni.org

Acknowledgments

An much earlier version of some parts of this paper appeared in "NetConnect," a supplement to *Library Journal*, on 15 October 1999 and January 2000, under the title "Electrifying The Book" (Parts I and II); I thank Francine Fialkoff of *Library Journal* for her help with those articles. Over the past two years I have had the opportunity to explore the ideas presented here in a number of talks, and I am particularly grateful to Cecilia Preston, Michael Buckland, Nancy Gusack, Shelley Sperry, Hal Varian, and the participants in the Buckland/Lynch Seminar at the School of Information Management and Systems at the University of California, Berkeley for their suggestions. On 10-11 May 2001, as I was finishing this paper, I had the opportunity to test and refine the ideas here and to gain many new insights at the excellent Conference on Electronic Books hosted in Ann Arbor by the University of Michigan and Blackwell Publishers; I thank the speakers and participants there. Finally, I am indebted to my colleagues on the National Research Council Committee on Intellectual Property in the Emerging Information Infrastructure as an important influence on the ideas presented here, though I want to stress that these ideas should be viewed as distinct from the findings of that committee as reflected in the report [The Digital Dilemma: Intellectual Property in the Emerging Information Infrastructure](#). Let me also state here that I am not an attorney and that nothing here is intended to be taken as legal advice.

Notes

1. Fatbrain has since been acquired by Barnes and Noble, Peanut Press by Palm computing, and Netlibrary seems to have shelved its planned IPO due to market conditions and scaled back its ambitions somewhat; Questia launched its product in early 2001.
2. Note that digital music, or digital video, is just another such collection of bits; in the wonderful new world of digital "convergence" they all look the same, and can be managed the same way.
3. See Alan Kay and Adele Goldberg, 1977. "Personal Dynamic Media," *IEEE Computer*, volume 10, number 3 (March), pp. 31-41; reprinted in Adele Goldberg (editor). *The History of Personal Workstations*. New York: ACM Press, pp. 254-263.
4. I believe that I heard this statement made by Alan Kay in a keynote address at an Educom conference. I'm indebted to Ed Valauskas for an actual reference to the same statement in print, though attributing it to Minsky rather than Kay. See Raymond Kurzweil, 1990. *The Age of Intelligent Machines*. Cambridge, Mass.: MIT Press, p. 328.
5. This vision is recounted, in a less elaborated form, in his book *Silicon Dreams* (New York: St. Martin's Press, 1989).
6. Examples include the joint American Chemical Society/Cornell University CORE project, the Elsevier Science Publishers TULIP project that was conducted with multiple university partners, or the Institute for Electrical and Electronic Engineering/University of California project.
7. In fairness, I should note that there is at least some anecdotal evidence that younger people, who have grown up with text on display screens from childhood, are less insistent about printing.
8. I am grateful to Richard Lanham, author of *The Electronic Word: Democracy, technology, and the Arts* (Chicago: University of Chicago Press, 1993), which explores some of these issues in depth, for making this point forcefully.
9. See, for example, Robert Darnton, 1999. "The New Age of the Book," *New York Review of Books* (18 March), available at www.nybooks.com/nyrev/. See also David Kirkpatrick, 2000. "The French Revolution Will Be Webcast," *Lingua Franca* (July/

August), pp. 15-16, for more on Darnton's vision.

10. See www.historyebook.org

11. See, for example, Matthew Mirapaul, 2001. "Beyond Hypertext: Novels with Interactive Animation," *New York Times* (5 March), p. B5 for a look at some experiments.

12. Scott McCloud, 2000. *Reinventing Comics*. New York: Paradox Press.

13. See Michael Jenson, 2000. "E-Books and Retro Glue Protect the Vested Interests of Publishing," *Chronicle of Higher Education* (23 June), available at <http://chronicle.com/free/v46/i42/42a06401.htm>

14. Consider, for example, the advertisements that the telecommunications company Qwest was running in 1999-2000 suggesting that through the new fiber-optic networks "every book, ever written, in any language" would be available in every bookstore and "every movie ever made" available for pay-per-view at every motel.

15. For example *The Bridgeman Art Library v Corel Corporation* (97 Civ.6232 (LAK) New York Southern District Court), case, which found that there was no new copyright in images of out-of-copyright artworks.

16. For example, by adding markup or annotation, or by doing very sophisticated imaging of a very old book in deteriorated condition to enhance its readability.

17. Depending on a number of factors, including, when the author died, which is not always easy or inexpensive to determine, or whether copyright was re-registered within certain time periods.

18. See, for example, Matthew Rose, 2001. "Definitions are Key in Publishers' Dispute Over Electronic Book Rights," *Wall Street Journal* (7 May); see also www.rosettabooks.com/pages/legal.html for documents relevant to the ongoing court case.

19. It's important to note here that it's not just the economic model for the service, but the content as well. It has become increasingly clear that there is a demand for more music than the industry has been willing or able to push through its retail channels. There is a brisk trading market in tapes of all of the concerts that the Grateful Dead ever performed. Miles Davis was unable to release all of the concert recordings that he captured in the early 1970s because his record company could not accommodate these recordings through the existing retail channels. The band Pearl Jam has released a series of live recordings of every concert that they played on their most recent tour. The age of scarcity of recorded

performances is coming to an end.

20. Note that *Harry Potter IV* was released concurrently in the United States and the United Kingdom.

21. Consider practices in mainland China, or some of the Middle Eastern nations, for example. See the report from Reporters Sans Frontières, 2001. *The Enemies of the Internet*, available at www.rsf.fr

22. See also Lisa Guernsey, 2001. "Welcome to the Web. Passport, Please," *New York Times* (15 March), and available at <http://www.nytimes.com/2001/03/15/technology/15BORD.html>

23. Committee on Intellectual Property Rights and the Emerging Information Infrastructure, Computer Science and Telecommunications Board Commission on Physical Sciences, Mathematics, and Applications, National Research Council. [The Digital Dilemma: Intellectual Property in the Emerging Information Infrastructure](#). Washington, D.C.: National Academy Press, 2000.

24. See *Thor Power Tool Co. v. Commissioner of Internal Revenue*, 439 U.S. 522, 532-33 (1979).

25. That is, software that suggests books based on the purchases of others who seem to share your tastes through a common purchasing pattern.

26. Not the least of which is the fact that many authors have neither the skill nor the desire to be publishers.

27. See, for example, D.T. Max, 2000. "No More Rejections," *New York Times Book Review* (16 July), p. 35.

28. See Alex Soojung-Kim Pang, 1998. "The Work of the Encyclopedia in the Age of Electronic Reproduction," *First Monday*, volume 3, number 9 (September), available at firstmonday.org/issues/issue3_9/pang/

29. See, for example, William Mitchell's *City of Bits* (Cambridge, Mass.: MIT Press, 1995) which was mounted on the [MIT Press](#) Web site for reader comments.

30. See Lisa Guernsey, 2001. "Evolving E-books Let Authors Answer Critics," *New York Times* (10 May), p. D4.

31. Note that for scholarly monographs digital publication and distribution helps with the inventory problem and hence with works going out of print rapidly, but has little effect on the crisis in the economics of monograph publishing. Restructuring the value chain doesn't help if the problem is first copy costs rather than distribution costs.

32. To the extent that publishers do not sell e-books directly to customers, there are also new issues about trust and auditability in the distribution chain. One is no longer moving physical copies that can be inventoried and counted, but rather a digital master that is distributed through copies.

33. There is a broad public perception that the music industry was extraordinarily greedy in its transition to audio CDs from vinyl records, where product costs seemed to go down and prices went up substantially. It will be particularly interesting to see how and if the publishing industry alters its pricing to the consumer as part of the restructuring of its internal economics.

34. For an example of developments in the textbook marketplace, see Thomas Weber, 2000. "New E-Book Technology Helps Protect Copyright," *Wall Street Journal* (11 September).

35. Not just to receive special offers and advertising, but for example, to be able to obtain a replacement, or to migrate from one e-book reader to another when a rights management system or a format change is involved.

36. Consider the Toysmart consumer database controversy (see cnnfn.cnn.com/2000/07/21/companies/toysmart/), and also the changes to the Amazon.com consumer privacy policies in this light.

37. Felicity Barringer, 2001. "Using Books as Evidence Against Their Readers," *New York Times* (8 April).

38. See, for example, Henry Jenkins, 2001. "Electronic Paper Turns the Page," *Technology Review* (March), available at www.technologyreview.com/magazine/mar01/mann.asp

39. NuvoMedia offered this.

40. For the history and economics of "for-profit" libraries old and new, see Richard Roehl and Hal R. Varian, 2001. "Circulating Libraries and Video Rental Stores," *First Monday*, volume 6, number 5 (May), available at firstmonday.org/issues/issue6_5/roehl/, and also Hal R. Varian, "Buying, Renting and Sharing Information Goods," *Journal of Industrial*

Economics, volume 48, number 4, pp. 473-488, and more generally Carl Shapiro and Hal R. Varian, 1999. *Information Rules: A Strategic Guide to the Network Economy*. Boston: Harvard Business School Press.

Editorial history

Paper received 29 April 2001; revised version received 13 May 2001; accepted 28 May 2001.

[Contents](#) [Index](#)

Copyright ©2001, First Monday

The Battle to Define the Future of the Book in the Digital World by Clifford Lynch
First Monday, volume 6, number 6 (June 2001),
URL: http://firstmonday.org/issues/issue6_6/lynch/index.html