

E-books and the Challenge of Preservation

Frank Romano
Rochester Institute of Technology

INTRODUCTION

The concept of electronic publishing was first articulated by Vannevar Bush of MIT in the seminal 1945 article "As We May Think." In 1991, Apple Computer introduced "Jurassic Park" as an electronic book for its Powerbook 100 laptop using the Adobe Acrobat portable document format (PDF). In 1998, the Rocket E-book was introduced, and in 1999, Simon & Schuster and Stephen King published an electronic novella that could be read on any Internet browser on virtually any computer, or downloaded to certain e-book devices. For the foreseeable future, most e-publishing will involve scientific, technical, professional, and academic information as well as some original fiction work. Librarians and others involved in digital asset management will have to preserve at least some of this material for future reference, since it is expected that original works will be created and many of these may exist only in electronic form. E-books are not a historical artifact or anomaly, but a new form of content conveyance—important for certain constituencies and also important from the perspective of maintaining a record, especially since many works will be in e-book format only. Growth, while steady, may be slow because of competing technical standards, digital rights management, definitional issues, and restructuring within traditional publishing as creators, existing publishing houses, and software companies position and reposition themselves in a changing market. A critical and perhaps underestimated set of issues concerns user acceptance.

The trend toward electronic publishing has been based on several factors:

1. Technological advancements that provide increased computing functionality at lower cost (Moore's Law).
2. The development of new channels of information distribution (Intranet and Internet).
3. The desire to reduce costs by eliminating paper, printing, and physical storage.
4. The ability to search electronic files efficiently and retrieve information quickly.
5. The ability to reuse information in other documents and other formats (with appropriate content rights management).
6. The acceptance of reading on-screen by segments of the population.
7. The convergence of text, imagery, audio, video, animation, and interactivity in new kinds of documents.
8. The ability of virtually anyone to become his or her own publisher.
9. The immediacy of content acquisition through electronic transactions and data downloading.
10. The demand for space in America's 122,000 libraries.

Since the advent of disk- and tape-based digital storage in the 1960s, we have seen the evolution and proliferation of more than 200 different data storage formats—from large- and small-diameter fixed disks, to flexible diskettes of every size, to compact and video disks. During this time, media have decreased in size and increased in storage capacity, from 1 kilobyte of data to 40 gigabytes of data, with the first terabyte disks imminent. No single format has existed for more than a decade, which has necessitated the recording and rerecording of information on new media to allow access by current

computing systems. This trend has also affected the entertainment industry as it evolved from records, to tapes in cassettes and cartridges, to CDs and now to DVDs.

At the Rochester Institute of Technology, files stored on 8-inch flexible diskettes from 1970s word processors are unreadable—not because of media condition, but because readers for that media are unavailable. Forty-four-megabyte Syquest disks from the 1990s are about to suffer the same fate. Libraries and information repositories face a continuing challenge in maintaining files on currently-supported storage hardware and media and in currently-supported file formats for currently-supported operating systems that require structured data organization.

DEFINITION

An "electronic book" or "e-book" is the presentation of electronic files on digital displays. Although the term "e-book" implies book-oriented information, other content is also displayable on such devices. Static text and images are typically displayed, but moving imagery and audio are also presented. E-book files can be provided as recorded units (disks) or downloaded from digital repositories (including Web sites) to desktop computer monitors, laptop screens, portable digital assistants (PDAs or "Palm" type devices), cell phones with expanded displays, pocket pagers, or dedicated digital reading devices (currently called e-books).

The e-book production cycle begins when an author creates an original work and submits it to a publisher. The publisher converts the work to one or more e-book formats and employs rights management encryption to electronically lock the file and generate a unique decoding key. (Initially a 40-bit encryption was used. The U.S. government now permits U.S.-only versions with 128-bit encryption which improves security.) An e-book distributor (who may be different than the publisher) manages the protected file. The e-book publisher or distributor transfers the work to an e-book retailer, who sells the protected e-book online and offers buyers a "key" to decrypt and read the work. A buyer connects with a retailer's Web site and purchases the work, after unlocking the file with the digital rights key and downloading it to read on an e-book reading device. Some of the digital rights solutions include Adobe PDF Merchant, WebBuy, Xerox ContentGuard, Reciprocal.com, SoftLock, netLibrary, InterTrust MetaTrust Utility, LockStream.com, and others. (Rights issues are discussed in more detail below.)

"E-book" is actually a misnomer. The device can display magazine content (e-magazine) and newspaper content (e-newspaper) as well as electronic directories, catalogs, and other material. The display device is independent of the content. However, a distinguishing characteristic of books, magazines, and newspapers is the size of the page—all must adjust to the device's screen size, which is presently about the size of a hardcover book page.

A Web site is a collection of HTML (HyperText Markup Language) coded files and other files (image, audio, video) in computer code that are displayed on a screen using a browser application program. The browser (e.g., Netscape Navigator or Internet Explorer) translates the coded data into displayable typographic and image elements and presents them to the viewer. An important aspect of such sites is the ability to click on defined elements which then automatically display other Web sites (hyperlinking). A computer linked to the Internet functions like an e-book and thus inherits many of the challenges associated with long term use and preservation of Web sites.

Consider the problem of how to identify—and find—a Web site. Web sites have addresses so that viewers can hyperlink—connect from one to another. Such addresses have been used as bibliographic references or identifiers. After only a few years those address may no longer be active. This presents another challenge to the preservation of information since it is not expected that most e-books will be delivered via media (disks, for example) but rather through connections to the Internet or proprietary sources—wired or wireless. Thus, the content may be unfindable or unavailable for downloading. While it may be expected that libraries and other information repositories might be "back-ups" for Web-based content acquisition, libraries and information repositories will have to store such information on some form of storage media, and may require a plethora of different reading devices unless standards evolve. An alternative scenario would have libraries serve as portals to any number of commercial sites. However, the likelihood of long-term preservation by commercial enterprises may not be as assured as that of certain libraries.

Thus, there are essentially three related challenges: the location of the stored information, the organization storing the information and its long-term viability and commitment to preservation, and technical issues. In addition, there are questions of digital rights management; possible definitions of new "artifacts," including the notion of an e-book itself; user acceptance; and a reconfiguration of interests and equities among authors, publishers, and software firms.

THE CHALLENGE OF PRESERVATION

Preservation of electronic content will be necessary as a practical matter (downloading current material) as well as for historical purposes. There are a number of scenarios for the delivery of this information:

1. The e-book device is connected to another computer which is linked to the Internet. The user goes to a specific Web site and selects the titles required. The Web site could be that of:
 - the e-book producer
 - a portal that represents several publishers
 - a specific publisher
 - an academic or corporate site
2. The e-book device has a built-in modem and is connected to the Internet via a phone line directly for downloading.
3. The e-book device is connected via a kiosk at bookstores, libraries, airports, or other locations for downloading.
4. The e-book connects via wireless modem to the selected Web site or other location for downloading.

In every case, the e-book "title" is stored on a remote storage system, and is then routed to the e-book directly or to a computer. No single data location of all e-book files will exist, and mergers and personnel changes at the hosting site may affect the long-term storage of the information. A company could decide, for example, to drop certain titles, or could go out of business. Thus, libraries and other data repositories hold the responsibility of long-term preservation.

Computer operating systems are usually aligned to structured storage systems that record coded data. Over time all of these items can change:

- the recording medium (e.g., magnetic tape, disk)

- the operating system (e.g., Windows)
- the storage format (e.g., binary, ascii, sound, video)
- the data coding system (e.g., HTML, XML)
- metadata (e.g., bibliographic or stylistic encoding)

DYNAMIC PRESERVATION

The storage of digital data will require a dynamic form of preservation, and a new definition of "archival" may have to be developed. The concept of long-term storage of a paper- or photographic-based item that remains unchanged over time may not be applicable with electronic publishing. Instead, the information will have to be rerecorded on new media to be used with existing file formats and computer operating systems as storage media degrade and systems, formats, and encoding systems are evolved.

There are, and will be, programs that convert from one encoding system to another. Over time, these will become more reliable and allow data to be reformatted to the current standard approach. But the conversion will have to take place in order to keep the information in a "current" format. Usually there is a two-year transition between one form of storage and its successor. This is both a management and a technical issue and tracks the organizational issues—the permanence and commitment of the archiving organization—cited in the previous section.

TECHNOLOGY ISSUES

As previously noted, the size of the page—or the screen—is the defining property of e-books. This was fundamentally enabled by the "portable-monitor" (higher definition liquid crystal display screens). Capabilities vary with price, which ranges from about \$150 to \$600. E-books like the Rocket Book (now the RCA Gemstar) attempt to emulate what a typical reader or student would do with a real book: highlight text, bookmark pages, browse indexes, or write notes in the margins. Most e-books (ranging in size from a pocket-sized Palm Pilot to roughly half the size of a laptop computer) are capable of downloading and storing text and displaying it in a prescribed format that is intended to mimic that of a typical printed book. The text is usually displayed one screenful at a time and in most models is advanced or regressed a screenful at a time with arrow buttons. Some do not have page numbers and a screenful of text may be considered a page. Page orientation can be adjusted with some brands and most electronic books also have advance features to move quickly ahead or back as if paging through a printed book. The books are battery-powered but also come with electrical adapters. Rechargeable batteries can last from 10 to 40 hours, depending on the brand and whether backlighting is used.

SCREEN ISSUES

The size and resolution capabilities of e-books vary. They can support text as well as black and white images such as graphs, line art, and newspaper-resolution photos. Grayscale images are not supported with most brands. All of the current e-books are black and white, with few color models. Within two years most models will display grayscale images, color, and also play sound and video. Most e-books come with proprietary software that is used to transfer data to and from the e-book as well as allow downloading from Internet-based or proprietary services.

The most significant advance toward a paperless world will be portable displays—lightweight, rugged, operating for hours using lightweight batteries, with high resolution and contrast. In the late 1980s, liquid crystal displays were incorporated in the first laptop computers and today the typical laptop computer includes a 12-inch to 14-inch full-color, liquid-crystal display with good resolution. LCD-based flat-panel displays are smaller, lighter, use less power, and discharge fewer electromagnetic emissions than CRT counterparts. There are currently experiments under way at Xerox Palo Alto Research Center (in cooperation with 3M) and E Ink (a spin-off from the MIT Media Lab in partnership with Lucent Technologies) and other variations on the notion of digital ink, digital paper, ultra thin screens, flexible displays, and such.

STANDARDS ISSUES

There are a number of issues and organizations involved in developing standards. These involve mark-up languages, identification, and metadata as well as hardware and software standards.

The widely-used HyperText Markup Language (HTML) and Portable Document Format (PDF) standards continue as dominant document formats on the Internet, but are not necessarily perfect standards for information delivered on handheld devices, like e-books. HTML displays can have difficulty with consistency and Acrobat displays the equivalent of printed pages, which may be oversized for most small devices. Both of these limitations are being addressed: HTML is metamorphosing into XML to allow more consistent reformatting on different screens and Adobe is integrating PDF and such reformatting into future versions of PDF. Microsoft has developed Clear Type font technology for clearer, more "paper-like reading" and has announced a standard text format and operating system for Microsoft Reader. Adobe has just released its version of a more readable screen font technology called CoolType.

A PDF file is truly a portable document. It can be generated from just about any application and keeps all typographic formatting, graphics, layout, and page integrity completely intact. Because the PDF embeds fonts, the recipient need not have the fonts that were used by the document creator. Graphics are compressed, which allows PDF files to be very small for transmission over networks. The reader software runs on most computers and is free—downloaded from Adobe's Web site.

In 1998, the National Institute for Standards and Technology (NIST) of the Department of Commerce formed the Open E-Book Standards Committee (OEBSC) to promote a standard e-book format. The Open E-book Publication Structure defines the format for content converted from print to electronic form. The Electronic Book Exchange (EBX) Working Group is establishing copyright protection and distribution standards. The Open eBook Forum (OEBF) is an international, non-profit trade organization whose mission is to promote the development of the e-publishing market. In January 2001, the Open eBook Publication Structure Specification Version 1.01 was placed before the OeBF membership for comment. OEB 1.0 uses HTML semantics, but XML-based syntaxes. The Open Ebook Authoring Group, made up of the major e-book reader manufacturers, a few large publishers, and Microsoft, among others, released the first Open Ebook Specification (OEB 1.0) in September, 1999—a specification based on XML. The Open eBook Forum, OeBF, and the Electronic Book Exchange Working Group, EBX, have agreed to merge efforts. There are also the Digital Audio-based Information System (DAISY) initiative, the Text Encoding Initiative (TEI) Consortium, NISO, W3C, DocBook, the International Publisher's Association, MPEG, the U.S. Copyright Office, the international DOI foundation, EDItEUR, and others.

The Open Ebook Standards Project, led by the Association of American Publishers (AAP), several leading publishers, and Anderson Consulting (now Accenture), released the results of an intensive effort to establish recommendations and voluntary standards. Experts have been working with AAP to develop standards for numbering and metadata, and to identify publisher requirements for digital rights management, three areas critical to the growth of the market. The new standards specify a numbering system based on the Digital Object Identifier, an internationally supported system suited for identifying digital content and discovering it through network services. The numbering recommendations allow for identification of ebooks in multiple formats and facilitate the sale of parts of e-books, and they also work with existing systems such as the ISBN to allow publishers to migrate to the new system.

The metadata standard has extended ONIX, the existing international publishing standard for content metadata, to include the information needed to support the new numbering system and to include e-book-specific data. With ONIX, publishers will be able to provide their metadata to (r)e-tailers, conversion houses, and digital rights partners. Indexing of the metadata will make e-books easier to find in online catalogs. AAP also released a comprehensive description of digital rights management (DRM) features needed to enable the variety of new products and business models publishers want to offer.

There are numerous proprietary software solutions being offered to translate digital e-book files for the many competing reader platforms. Most solutions incorporate security features to protect copyright owners (that is, the file cannot be printed or copied). It may be that reading devices may display some of all of these formats, but it is probable that one or two will become clear standards. Publishers have already restricted their market through the use of a reading device. Unless a very inexpensive reader is developed and becomes universally available, this market cannot evolve. The information for these readers must also be standardized and pervasive. It is not that we do not have standards—we may have too many of them.

USER ACCEPTANCE ISSUES

The AAP teamed with Accenture to evaluate the market for e-books and to define the basis of its publisher members entry into e-book publishing. In a study entitled "Reading in the New Millennium, A Bright Future for e-book Publishing," Accenture projects the e-book market at \$2.3 billion by 2005—10% of the estimated \$21.9 billion total consumer book market in 2005. This study also highlighted the importance of open standards to the success of electronic publishing because "it's easy for consumers: any book, any source, any device."

In January 2001, Forrester Research, an Internet research firm based in Cambridge, Massachusetts, released a report with the following projections:

1. Slow growth is expected for both e-books and e-book reader devices.
2. There will be strong sales for on-demand custom-printed trade books and digitized textbooks.
3. In five years, 17.5% of publishing industry revenues (\$7.8 billion) will come from the digital delivery of custom-printed books, textbooks, and e-books. Of this amount, only \$251 million will come from e-books for e-book devices.
4. As a result of the Web's distribution advantages, publishers will create a new publishing model that Forrester calls "multi-channel publishing," requiring publishers

to manage all of their content from a single, comprehensive repository containing modular book content and structure.

Virtually all recent studies predict a slow but continuous growth in the e-book market. Some are more optimistic than others, but few predict a decline.

PUBLISHER ISSUES

Publishers are implementing a range of strategies, partnerships, and experiments with delivery and packaging. AOL Time Warner Trade Publishing was one of the first traditional publishing houses to launch a digital division with the creation of ipublish.com. Random House and Simon & Schuster have also created electronic divisions. Barnes & Noble established an online e-book store on BN.com, and Amazon.com has also entered the market. Electronic publisher MightyWords signed distribution partners to sell its titles on Fatbrain.com and Barnesandnoble.com, and, in addition, consumers can browse, purchase, and download works at Adobe.com and other Web sites.

In 1995, book publishers produced thousands of multimedia computer CDs with interactive features, pictures and sounds, but consumers did not accept the new electronic works. Personal computers were not as pervasive; technical standards caused innumerable problems running the programs; and few PCs had CD-ROM drives. Multimedia has evolved into a market as standards evolved and the base of computer users expanded. Major book publishers, technology companies, online booksellers, and new electronic book middlemen are investing in the future market of digital books.

Authors may see electronic books as a way to sell books directly to consumers, freeing them from dependence on publishers. Publishers may see an opportunity to eliminate printers and bookstores. Online booksellers are moving into the publishers' business, printing digitized books themselves and selling their own electronic editions. Start-up companies sell the contents of books through digital archives of thousands of books and periodicals available online, liberated from the constraints of time and shelf space.

Publishers now see e-books as incremental sales to computer-savvy adults and the next generation of readers. A publisher's ultimate responsibility is to get the work to the largest possible audience and the Internet has that potential. But no one knows what an electronic book is worth. Some publishers are setting prices for e-books just below their printed equivalents, but others charge much less. Random House said that it would split equally with authors the wholesale revenue from selling or licensing electronic books, raising the author's share of the list price to 25% from 15%. Random House invested in Xlibris, a digital publisher that claims to issue more books in a year than Random House. After the success of King's e-novella, Bertelsmann, Simon & Schuster, and AOL Time Warner's book division approached agents for digital rights.

Digital publishing presents an opportunity for authors and publishers to develop a much closer connection to consumers. There will still be retailers, but certainly the middleman component may be smaller. Some publishers are already selling digital books directly to consumers as customized editions with modular contents, especially in the educational market. McGraw-Hill's Primis Custom Publishing division has a Web site that lets instructors select chapters and excerpts from a digital archive to build their own personalized electronic volumes. Instructors order directly and bypass campus bookstores.

Random House's Modern Library classics division sells electronic editions of its books directly to readers through links to literary Web sites such as those devoted to Shakespeare or Swift. Time Warner sells e-books through links to its own Web site. Barnesandnoble.com publishes and prints its own digital books. Barnes & Noble and Barnesandnoble.com have invested in several digital publishing and bookselling startups, including Fatbrain.com, iUniverse, and MightyWords.com. The company has installed print-on-demand systems in their warehouses so that they can begin printing and binding copies of books available from publishers as digital files. Book wholesaler Ingram Book Group's Lightning Source pioneered print-on-demand for runs as low as one book.

Amazon.com offers authors who want to self-publish either print or electronic editions a distribution channel. Start-ups are also building an alternative sales channel for the contents of digital books, as part of large online archives that let readers search through texts as well as browse their titles. Each of the main e-book contenders is pursuing a different strategy and competing for publishers' digital books.

NetLibrary sells electronic books to libraries via online access to the digital version of the book on their computer servers for a fee. Users can search the contents of books in the online collection but they cannot copy or print the books. Public and university libraries and some corporations are now customers. Questia and Ebrary, as well as other e-publishers, are negotiating with publishers and authors to enlarge their collections. Questia sells access to an archive of digital books for a subscription fee, with a variety of research tools, including links connecting footnotes in one book to text in another. Random House, McGraw-Hill, and Pearson's Viking-Penguin have invested in Ebrary, which lets readers search and browse freely through digital books and magazines, but charges a fee to print pages, copy text, or download content.

DIGITAL READER ISSUES

The future of digital publishing will also be shaped by the competition among three technology companies hoping to set the standards for publishing and reading books on screens (see earlier discussions of technology and screen issues). Microsoft, Adobe Systems, and Gemstar-TV Guide International are working to convince publishers and readers that their format is the most secure from copying, convenient to use, and easy on the eyes. Microsoft and Adobe Systems produce competing software programs intended to make reading on a screen easier on the eyes, and both have announced rival alliances intended to strengthen their respective positions.

Gemstar's format is used on portable appliances, like the Rocket e-book, instead of a laptop or desktop computer. Adobe Systems has by far the largest share of the digital publishing software market. Customers have downloaded over 180 million free copies of Acrobat Reader software for reading and printing digital documents. Gemstar holds patents on the technology to read digital books on specialized hand-held devices. Gemstar's latest generation, built under the RCA brand by Thomson Multimedia, is priced at \$300. Gemstar's system avoids both personal computers and the Internet. Online bookstores sell electronic books for Gemstar's format, but to download the digital texts consumers need to plug their devices into phone lines and dial directly into Gemstar's computer servers. Users of the devices can only store and retrieve their books on Gemstar's server. Devices that apply Gemstar's electronic book patents could become personal organizers, wireless pagers and phones, and generalized portable entertainment devices for text, video and sound, making the habit of reading an entry into the PDA and multimedia arena.

Microsoft and Amazon.com opened an electronic bookstore that distributes free copies of Microsoft's Reader software. Amazon.com sells electronic books for a variety of formats, including Adobe's. Microsoft makes no money from its Reader software but receives a small commission on the sale of electronic books in its software format. Microsoft started a similar cooperative marketing venture with Barnesandnoble.com with the release of a new version of its Reader software.

ON-DEMAND PRINTING

Publishers are applying print-on-demand methods, and it is starting to change their business. Xerox, IBM, and others now sell machines that in minutes can churn out single, bound copies of paperback or even hardcover books. The output is virtually indistinguishable from the products of traditional printing presses.

A traditional print run is cheaper only if hundreds of copies are printed at once. The print-run constraint does not hold for on-demand printing with the result that some low-selling books that would have passed out of print are staying in print longer, and a few books that might not have found publishers now have done so. The Perseus Books Group installed print-on-demand equipment in its warehouse near Boulder, Colorado, to print slow-selling titles in small batches instead of letting them fall out of print. The National Academy Press in Washington, D.C., did the same. New printing technology helps fulfill demand for special-interest titles created partly by online bookstores. Some publishers order print-on-demand editions of some of their books through Ingram's Lightning Source digital publishing division, and the bookseller Barnes & Noble has installed the machines in its warehouses to print books on demand.

The early indications are that electronic books are most likely to take off at the two extremes of the book market: with readers of popular genre fiction like romances and science fiction, and with readers of educational and professional business texts.

E-BOOK PUBLISHING

"E-book publishing" covers a business in which a provider enables authors to publish books through an online service. An author submits a manuscript, and it is published and printed as a book. A search of the Internet reveals over 100 e-publishers, most providing books in electronic form for on-screen reading using the computer's browser or PDF viewer.

| | |
|------------------|--|
| 1st Books | www.1stbooks.com |
| Artemis Books | www.artemispress.com |
| Books Just Books | www.booksjustbooks.com |
| Bookmice | www.zeus-publications.com |
| Books Onscreen | www.booksonscreen.com |
| BookSurge | www.booksurge.com |
| Booktech | www.booktech.com |
| Crossroads | www.crossroadspub.com |
| Digitz | www.digitz.net |
| Dissertation | www.dissertation.com |
| EBrary | www.ebrary.com |
| ElectricPress | www.electricpress.com |
| FirstPublish | www.firstpublish.com |

| | |
|-------------------------|---------------------------|
| GreatUnpublished | www.greatunpublished.com |
| Hard Shell Word Factory | www.hardshell.com |
| iPublish | www.ipublish.com |
| iUniverse | www.iuniverse.com |
| Lightning Source | www.lightningsource.com |
| Net Paper | www.netpaper.com |
| Ozone Books | www.ozonebooks.com |
| Universal Publishers | www.upublish.com |
| Zeus Publications | www.zeus-publications.com |

Stephen Riggio, vice-chairman of barnesandnoble.com, said, "You will see—very, very soon—authors become publishers. You will see publishers become booksellers. You will see booksellers become publishers, and you will see authors become booksellers." With the advent of e-publishing, book industry classifications are an anachronism.

RIGHTS, INFORMATION SECURITY, AND PRIVACY ISSUES

Replication and intellectual property risks exist because of the relative ease with which digital data can be copied, modified, and disseminated. An important industry concern is that digital content will emulate digital music and circulate free over the Internet. Technology companies are positioned to insert themselves into digital publishing as electronic wholesalers, taking the place occupied by distributors of traditional books. They provide protection from copying along with software and services to store and transmit digital books, in exchange for a percentage of revenue. These systems typically require four elements:

1. Authentication of transmissions and messages to determine if the originator is authentic, or that the recipient is eligible to receive the information.
2. Data integrity checks to determine that the data are unchanged from their original source.
3. Certification that the sender of data has delivered the data and that the receiver has received it, with evidence of the sender's identity.
4. Confidentiality to assure that information can be read only by authorized entities.

In the quest for security, publishers may be restricting growth of this new market. Let us use printed books as an example. The purchaser reads a book and passes it on to another reader, or sells it to a used book store, which then sells it again. (Many of us would not have been able to afford college without this system.) Although the publisher does not receive revenue from these subsequent uses or sales, the reader may develop an affinity for the author or subject, which stimulates future sales. Magazines are routinely passed around. Publication pages are often copied for distribution. In effect, we have had the "Napsterization" of the publishing market since printing was invented. But this practice may now be upset. Readers of e-publications who wish to save issues for future reference may not be able to do so (the archives of *The New York Times* and *The Washington Post*, for example, charge for access) and may find that the e-book readers do not have external storage.

From the publishers' and authors' point of view, there is cause for concern. Stephen King's *Riding the Bullet* was sold exclusively on the Internet. After 48 hours, *Riding the Bullet* sold more than 500,000 downloadable copies worldwide. The book was priced at \$2.50 per copy. Although many initial orders were delivered in free promotions, the financial implications of King's foray into e-books are still staggering. It took fewer than two days to sell 500,000 copies without printing, shipping, storage, wholesaler and

distribution middlemen, or other traditional publisher costs. However, within those same 48 hours, pirated copies were on the network.

In the report “eBooks: Publishing’s Next Wave or Just a Ripple?” from Cahners’ TrendWatch, an important point about balancing security and distribution was made: “Periodical publishers have an interesting problem with regard to digital rights management, and that is they want to protect their content, but advertising rates in periodicals is in large part based on “pass along” copies. For example, most ad rates for large consumer publications are premised on the assumption that a single copy is passed along to five other people. If you secure a digital version of that publication, you’ll ensure that someone pays for it, but you’ll also prevent them from passing it along. How do you determine your advertising rates based on that?”

CRACKING THE CODE

The Russian firm, Elcomsoft, released its Advanced eBook Processor software that enables users to convert copy-protected e-books into plain vanilla PDF documents that can be printed, copied and distributed easily. This software company received a cease and desist order from Adobe Systems, and had its Web site removed from the Internet. Adobe says that its eBook software copy protection isn’t applied by the end user, but rather by the copyright holder. The Russian programmer was imprisoned and eventually released—a release supported by Adobe. Publishers are fearful of e-book piracy and the thought that books could be swapped like MP3 files over the Internet. Adobe must demonstrate a secure option or they will lose the support of major publishers. But Elcomsoft also showed that they could break Microsoft protection systems. Many feel it is better to show the vulnerability of such systems in an open forum instead of driving it underground. For the Russian programmer, it was not a case of hacking, but a mathematical puzzle to be solved. This reflects a tension between the values of the research community and the values of the commercial community, and it isn’t clear how the conflict will be resolved.

WHAT IS A BOOK?

The question must be asked: why are e-book rights treated differently than printed book rights? In the case of Random House versus RosettaBooks, Judge Sydney H. Stein summarized the complex issues of the trial in one statement: "Show me why an e-book is a book." The result of ensuing argument and debate was a ruling that essentially defined e-books as a new medium of communication, like audio books. But, what happens when sophisticated software converts the e-book text to spoken words with the cadence and pronunciation of Anthony Hopkins. Is this analogous to the Kurzweil Optical Character Readers of the 1970s that scanned printed books into words and then "spoke" them to the blind with a voice synthesizer?

There is an interesting privacy issue in that book buyers are generally anonymous (at least those who use cash). Amazon attracted negative publicity when it used an individual’s book buying data for promotion purposes. In many cases, e-books will be sold only to a specific device assigned to a specific individual. Civil libertarians may see the irony in the complete democratization of publishing at the expense of privacy.

FROM BOOKS TO BYTES

Consider that more than 400 pounds and 2 million pages of printed text can be distributed on a one-ounce DVD, and it is clear why seven dental schools require course materials on DVD. The disk can be replaced with updated data and played on any computer with a reader. However, the search for security is tending toward a restricted Web site or database for access to the information and temporary storage on a portable device.

Text will remain a central element in electronic books. Text will be stored in the computer with the kinds of codes that can be used for searching and indexing. Structural elements of a book's contents will be tagged with codes that faithfully map the content's intellectual structure: chapters, sections, footnotes, and sidebars. But technologists dream of pages that sing and dance—a world beyond text. Multimedia illustrations would be helpful in subjects requiring complex illustration, such as the sciences. It is expected the future e-book devices will have TV-like functionality and that the text-based publication will be augmented with multimedia presentations.

But audio, video, and animation will increase the need for storage and require increasingly more sophisticated devices than mere text readers.

Libraries and other data repositories must take a more active role in shaping the future of e-publishing. Efforts are focused on standards, devices, delivery, security, and commerce—but almost no consideration is being given to preservation.