

eContent

By Richard. W. Boss

eContent—which includes electronic versions of books, journals, maps, media, and archival materials—has become a significant part of public libraries’ resources. For most public libraries, the vast majority of eContent is made available from external sources; however, local eImage and eManuscripts collections are common. While most content has been digitized from other formats, there increasingly are original electronic publications, especially eJournals.

The major advantages of eContent are integrity of the collection, availability around the clock, remote access, and multiple simultaneous users. Unlike print, media, and archival materials; eContent is unlikely to be misplaced, stolen, or vandalized. It can be made available 24/7, rather than being available only during regular library hours. Except as there are licensing restrictions, eContent is available from anywhere and to multiple simultaneous users. However, electronic sources can be destroyed by system failures or hacking; therefore, almost all eContent providers maintain back-up files.

The TechNote focuses solely on eContent, not the tools required to create or maintain the content. The Library of Congress has inventoried tools and services for building collections of eContent. The index of tools and services is available at <http://www.digitalpreservation.gov/partners/resources/tools/index.html> or by searching for “LC index of digital preservation tools.”

The various types of eContent are discussed in the following paragraphs:

eJournals

eJournals have existed since the mid nineteen seventies, and have been made available by many public libraries since the early 1990’s, but they did not become a significant part of most public libraries’ collections until 2000. In the early years of

eJournals, they were held primarily by academic research libraries. They were not routinely deposited with national libraries. In 1994, the Koninklijke Bibliotheek of the Netherlands decided to include all eContent produced in the Netherlands in its deposit collection. Elsevier Science, the large Dutch science, technology, and medicine publisher; and one of the first to produce scores of e-Journals, deposited its titles. By 1995, the depository had 315 eJournal titles. In 2002, Elsevier Science signed an agreement with the Koninklijke Bibliotheek to have it become the archival agent for its eJournals. By 2007, the collection included 3,500 eJournals from scores of publishers and most other national libraries were accepting deposits of eContent.

There appears to be no accurate count of the number of eJournals available. Estimates range as high as 50,000, still only a small fraction of the 900,000 ISSNs that have been issued.

One of the reasons that eJournals became popular was the existence of electronic indexes and abstracts from which links could be made to the full-text. Another was that the average length of articles (8 pages) made it possible to quickly download or print them from a server. The widespread availability of PDF, a format developed by Adobe and made widely available without license fees, precluded the emergence of multiple proprietary formats. It also protected publishers because it captured the image of the articles, rather than making the articles subject to revision or reformatting. In recent years, more titles have become available in other formats, including HTML, a format that allows full-text searching. ASCII, an older standard for full-text also continues to be used.

There are scores of eJournal providers. The three offering the largest number of titles are Dialog, Ebsco Host, and OCLC FirstSearch.

Dialog (www.dialog.com) is a for-profit company that offers access to more than 11,000 eJournals, primarily in the fields of science, technology, and medicine. It provides links to the full-text of eJournals from its extensive indexing and abstracting

databases. While many of the links are to publishers' servers, there also are links to many titles on the Ebsco Host Electronic Journal Service.

Ebsco Host E Electronic Journal Service (<http://ejournals.ebsco.com>) is a for-profit subscription service that provides access to more than 8,000 eJournals stored on its servers. It is possible to search by journal title, subject, article author, or article title. There is also an option to store searches for a specified period of time and have an e-mail sent to a searcher's e-mail address whenever new articles matching the criteria are added.

OCLC Electronic Collections Online (www.oclc.org/electroniccollections) offers a not-for-profit subscription service for more than 5,000 journals from 70 scholarly publishers. Access is through citations found in *First Search*. Most of the titles are in PDF format, but some are in full-text. Many of the eJournals are available without additional charge to subscribers that also have print subscriptions. A library retains access to an archived journal even if it ceases to subscribe to it.

A useful finding tool when one is looking for the availability of a specific eJournal or many e-journals in a field is *e-journals.org* (<http://www.e-journals.org>). It is not an aggregator of e-journals, but provides links to electronic journals from around the world. Access is by keyword or subject area.

eBooks

There were approximately two million eBook titles available as of the end of 2008—still a small fraction of the estimated 35 million titles in the world's libraries. While the majority of the available eBooks are of interest primarily to academic libraries, at least 250,000 of them are offered by public libraries to their patrons.

The first major eBooks effort dates back earlier than the first eJournal effort, but has been slower to impact libraries and their users. *Project Gutenberg* (www.gutenberg.org) began in 1971 when Michael Hart, its founder, was given virtually

unlimited use of a Xerox Sigma V mainframe at the University of Illinois' Materials Research Lab. Rather than using his account for data processing, he decided to work on the electronic capture, storage, retrieval, and searching of the holdings of libraries. The focus was on having volunteers digitize books in the public domain. Despite its early start, there were just 29,000 free books in ASCII format in the Project Gutenberg Online Book Catalog as of early 2009. The emphasis is on classics. They are chosen by volunteers, keyboarded, and uploaded to one or more computer sites around the world known as FTP sites. The host site is limited to the index and links to the FTP sites

NetLibrary (www.netlibrary.com), now a division of OCLC, a not-for-profit organization, began in 1999. By the middle of 2001, it offered 11,000 titles; but by early 2009 the number had increased to more than 160,000. A library can purchase a collection of titles tailored to its needs and budget. Only one user at a time per title can be accommodated, but the library is free to set the check-out period. It can also choose to have multiple copies of a collection. MARC records are available to enter into a patron access catalog. The titles can be read on any computer. NetLibrary offers full-text searching, a dictionary with audio pronunciation, and personalization features, including bookmarks, annotations and "my favorites."

The **Million Books Project** (www.ulib.org) was launched by the Carnegie Mellon University Libraries in 2001 and set as its goal the digitizing of one million books by 2007. It fell 25 percent short of that goal, but more than 30 other research libraries also launched major digitizing initiatives between 2001 and 2005 and contributed to the MBP. By 2008 there were scores of scanning centers in operation around the globe. The only public library that launched a major digitizing program during this period was the New York Public Library. The MBP had digitized 1.5 million titles as of the first quarter of 2009. Approximately 90 percent of the titles are in the public domain. The titles are maintained on three sites: in China, India, and the United States.

Among the reasons why eBooks were slow to be adopted was the reliance on proprietary formats and the lack of good readers. Two early readers for which eBooks

were produced were the Rocket eBook reader and the SoftBook Reader. They were withdrawn from the market in 2003 when their sales dropped sharply after the introduction of lighter-weight readers with more attractive displays. Unfortunately, the titles purchased to read on them could not be read on the newer readers.

Even the newer readers, which used general purpose technology configured with special software for reading eBooks, were far from popular because they sacrificed screen size in order to reduce weight. One of the few exceptions was the Toshiba's Portege M2000, a tablet PC that weighed about the same as a book and had a screen that could accommodate an entire page with a font size that most people were able to read. It might well have become the most popular unit on the market were it not priced at nearly \$2,500.

The lack of success in the consumer market was a major factor in publishers' decisions to look to libraries. They began to work with Baker & Taylor, a major distributor to public and academic libraries, and Follett, a major distributor to school libraries, in mid-2003.

It was in 2003 that libraries began to purchase eBooks in significant numbers, not only because Baker & Taylor and Follett were promoting them, but because by then many of the titles could be downloaded to a desktop machine, laptop, or other computer device.

Initial circulation was low. One major public library tallied an average of four downloads per title in 2003. Another public library that introduced e-books in the third quarter of 2003 estimated an average of two downloads per title over a three month period. As there was no wear and tear as with print titles; and there was no labor cost for circulation charge and discharge, and reshelving; the libraries did not give up on eBooks.

By 2006, eBooks were circulating well. In part because of the increasing number of titles available, and in part because a library offers a single source of quality titles that

have been selected by professional librarians. There is no need to go to the Web sites of a score of publishers and distributors. Even more important, eBooks from a library's collection are usually free to library patrons.

Google Book Search (<http://books.google.com>) was launched in late 2004 as the Library Book Project with a goal of digitizing as many as 15 million books from a dozen major research libraries, including Harvard, New York Public, and Oxford. The program was an expansion of the Google Print program, which offered digital excerpts of books in copyright. Searchers using Google see links to relevant books. For those in copyright, there are brief excerpts and links to libraries and booksellers that have the titles available. For those in the public domain, there is full-text browsing and the option of downloading a PDF version. The program has been controversial because Google uses the "opt-out" approach with regard to copyrighted works. It digitizes books without seeking permission from the copyright holder. The copyright holder must specifically opt out of the program if it does not want to have its works digitized. Google argues that the brief excerpts that are available for copyrighted works actually helps sell books. The number of books available as of early 2009 was in excess of one million. Online advertising is the principal source of revenue for Google.

Microsoft's **MSN Book Search**, which was launched in 2006, lasted only two years because it could not successfully compete against Google despite the fact that it was committed to the "opt-in" approach, meaning that publishers have to specifically agree to have their titles digitized. That made it much more popular with publishers than Google's program. However, Microsoft was not willing to match Google's spending on digitization.

The other major eBook providers are eBooks.com, eBrary, Internet Archive, and OverDrive.

eBooks.com (www.ebooks.com) is a for-profit company that has offered popular fiction and non-fiction at prices averaging less than the print versions directly to

consumers since 2000. It entered the library market in 2004 with Ebook Library, a rental service for more than 140,000 scholarly, professional, and popular titles. A unique feature is its "Non-Linear Lending" program, one that limits the total number of lending days per year per title, but enables multiple-concurrent access. It offers its own PDF-based reader or enables downloading to a PC, laptop or PDA for offline use with Adobe Acrobat.

eBrary (www.ebrary.com) is a for-profit company that was founded in 1999 to sell eBooks. It offers subscriptions to some 100,000 titles and outright purchase to about five percent of them. It has developed its own reader software to access the eBooks on its hosted server. A library may also add its own documents in PDF format. The company claims more than 1,400 customers.

Internet Archive (www.archive.org) is a non-profit organization that has built a free and openly accessible online digital library of more than one million titles. It houses not only books, but several different media collections on servers at three California sites and a mirrored site at the Bibliotheca Alexandria in Egypt. Some 300,000 books scanned by Microsoft before it ceased its book digitization have been included. The Internet Archive administers the Open Content Alliance, a consortium of organizations that are building permanent, publicly accessible archives of digitized texts. It also administers Archive-It, a subscription service that allows institutions to build and preserve collections of born digital content. Over 65 institutions around the world are subscribers. Yahoo, a major competitor of Google, is a participant in the Open Content Alliance and has integrated the content of the Internet Archive into its index.

OverDrive (www.overdrive.com), a for-profit company, offers over 100,000 titles, including thousands of novels and general non-fiction titles. A library can purchase multiple copies of a title to accommodate simultaneous use of that title. While the titles are purchased, they remain on the vendor's server. Any PC or PDA may be used as a reader. The e-books automatically expire and check themselves back into the

collection. The vendor provides MARC records for inclusion in a library's patron access catalog.

The consumer market revived with the introduction of the Kindle Reader in 2007 and the aggressive marketing program of *Amazon* (www.amazon.com), a for-profit company. The reader uses an "e-ink" technology that makes the letters on the screen look print-like. Hundreds-of-thousands were sold despite the fact that the reader uses a proprietary format not supported by any other reader until iPhone/iPod Touch introduced support for the format. [However, the iPod's screen size is less than half the size of Kindle's six inches], The Kindle 2, introduced in 2009, increased sales even more despite its \$395 price because it offered built-in free wireless for downloading any one of 275,000 titles without a PC, a faster processor speed, and a memory capable of storing 1,500 e-Books. Most titles are priced at under \$10. Amazon also offered a conversion service from PDF to the Kindle 2.0 format for \$.10 a title. Amazon began taking orders for its Kindle DX in May of 2009. The reader has a 9.7-inch screen, one suitable for reading newspapers and magazines as well as books. The storage capacity is 3,500 eBooks. A new feature of the reader is the ability to support PDF as well as the Kindle format. The pre-release price of \$489 made it one of the most expensive readers on the market at the time of its introduction.

Of the PDF e-Book readers on the market as of the first quarter of 2009, the Sony PRS-505 was the most popular at a price of less than \$300. However, PDF files are harder to read because they are reduced to fit the screen and one cannot zoom (magnify) them. The majority of the libraries contacted by the author in the fourth quarter of 2008 were using PDF readers, including a number of models now discontinued, but still usable.

A list of a score of e-Book readers is available on Wikipedia at http://en.wikipedia.org/wiki/List_of_e-book_readers/

Since its release as an open standard in 2008 and published by the International Organization for Standardization as ISO 3200-2:2008, PDF has become the dominant

standard for two-dimensional documents, including eBooks. ASCII, a full-text standard for full-text on PCs, Macs, and handheld computers, is now little used. An emerging format is ADE (Adobe Digital Editions), a variant of PDF that protects eBooks from unlawful reproduction and distribution using Adobe DRM (Digital Rights Management).

eMaps

There are hundreds-of-thousands of maps in PDF format available online for viewing and/or downloading from scores of sources. The Perry-Castaneda Library Map Collection of the University of Texas at Austin has an extensive index of eMap sources at www.lib.utexas.edu/maps/map_sites/map_sites.html/. Most of the sources are academic research libraries, academic departments of major universities, and U.S. government agencies.

The leading online source of general maps is **Google Maps** (<http://maps.google.com>) and the leading source of geologic maps is the US Geological Survey (http://store.usgs.gov/b2c_usgs/catalog)

eMedia

It was only logical that digitizing efforts would go beyond the eJournal and the eBook. By 2005, there were digitizing projects underway to offer eMedia of audios and videos, art, and maps.

e-Audio's success is attributable to the 1995 introduction of MP3 (MPEG-1 Audio Layer 3), the digital audio encoding format that has become the standard for reducing the amount of data required (by approximately a factor of ten) to represent an audio recording and still sound like a faithful reproduction. However, many commercial music companies use proprietary formats that are encrypted in order to make it difficult to use purchased music files in ways not specifically authorized.

The most widely used player for audio is Microsoft's Windows Media Player (WMP), the latest release of which was WMP11 issued in 2006. It can only play audio files in certain formats; however, MP3 is one of them. It is possible to play other audio formats by installing additional software. While most public libraries have WMP11 or another MP3 Player on their computers, almost all also offer audio on CDs and DVDs. Patrons accessing audio using their own devices overwhelmingly use an Apple iPod.

Other formats for e-Audio are RealMedia and AAC (Advanced Audio Coding).

The *Internet Archive* has the largest collection of eAudiobooks (www.archive.org/details/audio) and music (www.archive.org/details/etree). As of the first quarter of 2009, there were more than 400,000 free digital recordings on the site.

NetLibrary (www.netlibrary.org) also offers thousands of eAudiobooks. These may be purchased individually or as collections. They may be used on any desktop or laptop running supported media software programs.

eBrary and *Project Gutenberg* have also added e-AudioBooks to their offerings..

eImages

JPEG (Joint Photographic Experts Group) and GIF (Graphic Interchange Format) are the two standards for digital images. JPEG is at its best on photographs and paintings. It is not well suited to line drawings or other textual or iconic graphics. These are best saved in the raw image GIF format.

Microsoft Windows Media Player11 is an all-in-one media player that supports not only eAudio, but also eImages and eVideo.

A significant number of academic and public libraries have created image collections, especially of local historic photographs. Thumbnails can usually be viewed without cost, but high-quality downloads images often require payment.

A unique subscription offering from *NetLibrary* is the Catalog of Art Museum Images Online (<http://camio.oclc.org>), a collection of more than 90,000 images of art objects and photographs from major museums around the world.

eVideo

MPEG1-4 (Moving Picture Experts Group) is the standard format for digital videos. The most widely used player for videos and moving pictures is Microsoft Windows Media Player (WMP11). However, VideoLAN's VLC Media Player 0.9.9 is an attractive alternative. It is available for free download from www.videolan.org/.

The largest collection of videos (www.archive.org/details/movies), some 179,000 as of the second quarter of 2009, appears to be that of the *Internet Archive*. Many of them are available for free download.

eArchives

eManuscripts

Manuscripts usually are captured digitally using PDF or GIF. Many libraries that hold archives of manuscripts have undertaken digitization projects in an effort to preserve them and to facilitate access.

The non-profit *Internet Archive* (<http://www.archive.org>), which has the largest number of records of any one site, relies on contributions of eManuscripts from the owners of the content. The majority of contributions have come from the special collections of academic and research libraries.

eWeb

An important component of the *Internet Archive* is the “Wayback Machine,” a 150-billion page Web archive dating back to 1996 at www.archive.org/web/web.php/. To surf the resources, one enters the Web address of a site or page where to start. Unfortunately, keyword searching was not yet supported as of the first quarter of 2009. However, there are broad collections of Web pages on such topics as September 11th, Hurricanes Katrina and Rita, the Asian Tsunami, and other major events.

eCollections

The Library of Congress and many academic and public libraries have undertaken digital collections of programs that are topical and multi-format. These often cannot be accessed by the titles of individual works, but only by topic. In the case of LC, that includes American History and Culture, a number of related digital collections of historic maps, photos, documents, audio, and video. Other topics include Advertising, African American History, Cities, Folklife, Immigration, Native American History, and Women’s History. Each of the topics is broken down further into collections, some of which consist entirely of books, journals, newspapers, maps, or media; and some of which are multi-format. Libraries interested in providing access to these resources for their patrons should consider cataloging the URLs.

Prepared by Richard W. Boss, May 30, 2009