

INFORMATION SYSTEMS CONSULTANTS INCORPORATED

220 WEST READ STREET, BALTIMORE, MD 21201 TEL 410-528-1112 FAX 727-1262

eContent

eContent—which includes electronic versions of books, journals, media, and archival materials—have become a significant part of most libraries' resources. While most eContent 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.

History

eJournals

eJournals have existed since the mid nineteen seventies, and have been made available by many libraries since the early 1990's, but they did not become a significant part of most libraries' collections until 2000. In the early years of eJournals, 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 2004, the collection included 2,600 eJournal titles, 4.5 million journal

articles, and required 5.0 TBytes of storage. It was growing at the rate of 60,000 articles a day. As of early 2007, many national libraries were accepting deposits of eContent.

There appears to be no accurate count of the number of eJournals available. Estimates range from 25,000 to 50,000, still only a small fraction of the 900,000 ISSN's 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, 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 formats that allow full-text searching.

eBooks

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 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 focus on the storage, retrieval and searching of what was stored in libraries. The focus was on having volunteers digitize books in the public domain. Despite its early start, there were just 20,000 free books in the Project Gutenberg Online Book Catalog as of early 2007. However, they were being downloaded two million times a month.

NetLibrary, now a division of OCLC, began in 1999. By the middle of 2001, it offered 11,000 titles; by early 2007 it offered more than 130,000 titles, still a small fraction of the more than 32 million titles believed to exist, more than half of them in the public domain.

Carnegie Mellon launched its Million Books Project in 2001 and set as its goal the digitizing of one million books by 2007. More than 30 other research libraries also launched major digitizing initiatives between 2001 and 2005. The only public library that launched a major digitizing program during this period was the New York Public Library.

Among the reasons why eBooks were slow to be adopted was the reliance on proprietary formats and the lack of good readers. Two 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. The Zire 21 from palmOne, Inc. weighed 3.2 ounces, but has a screen the size of a PDA (personal digital assistant). Hewlett Packard's iPAQ Pocket PC H4150 had a screen that is only slightly larger. One of the few exceptions was the Toshiba's Portege M2000, a tablet PC that weighed about the same as a book and has a screen that could accommodate an entire page with a font size that most people were able to read. It might well have been the most popular unit on the market were it not priced at nearly \$2,500.

Publishers of eBooks focused on the consumer market and paid little attention to libraries before 2003. Consumer sales in 2003 were under \$10 million in 2003, representing just three-quarters of a million units. Barnes & Noble was so discouraged by poor sales, that it discontinued offering eBooks in the fourth quarter of 2003.

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, notebook, tablet PC, or PDA.

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 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.

eMedia and eManuscripts

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. Increasingly, manuscripts and other unpublished documents in archives were being captured digitally to preserve them and to facilitate access.

Formats

There are literally scores of formats for eContent, but there are a few widely adopted ones. For eJournals, these are ASCII for full-text and PDF (Adobe Personal Document Format) for image files.

For eBooks, these are ASCII for full-text; Adobe eBook Reader (PDF) for PCs, tablet PCs, laptops, Macs, and Palm handhelds; MobiPocket for PCs, tablet PCs, laptops, pocket PCs, Palm handhelds, and SmartPhones; and Microsoft Reader for PCs, tablet PCs, laptops, and pocket PCs. 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).

For eMedia, these are RealMedia for both audio and video; MP3, MP3 Pro, (MP3 is a subset of MPEG1) and AAC (Advanced Audio Coding) for digital audio; JPEG (Joint Photographic Experts Group) and GIF (Graphic Interchange Format) for digital images; and MPEG1-4 (Moving Picture Experts Group) for videos.

For eArchives, there are PDF and GIF.

Space limitations preclude discussing each of these formats in this TechNote. However, detailed information about each can be found on the Web.

In most cases, a provider of eContent will enable the download of the appropriate format. However, a library should be careful to determine that it can support a format before deciding on an acquisition.

Major providers of eContent are identified in the following paragraphs, with those which provide multiple types of eContent following those that provide only a single type.

Major eJournal Providers

There are scores of eJournal providers. The two offering the largest number of titles are Dialog and Ebsco Host.

Dialog (<http://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 journals from its extensive indexing and abstracting databases.

Ebsco Host (<http://ejournals.ebsco.com>) is a for-profit company that offers access to more than 15,000 eJournals stored on its servers. It is available by subscription. It is possible to search by journal title, subject, article author, or article title. There is also an option to store searches and have an e-mail sent whenever new articles matching the criteria are added.

e-journals.org (<http://www.e-journals.org>) is not an aggregator of e-journals, but provides links to electronic journals from around the world. Access is by broad subject area or keyword.

Major EBook Providers

The major eBook providers are eBooks.com, Google, Microsoft, the Million Book Project, OverDrive, and Project Gutenberg.

eBooks.com (<http://www.ebooks.com>) is a for-profit company that offers popular fiction and non-fiction at prices averaging less than the print versions. Titles are sold only individually. Several thousand titles are available. They are searchable by author, title, keyword, or subject. Three format options are available: Adobe eBook Reader, MobiPocket Reader, and Microsoft Reader.

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 is 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. As of early 2007, a number of publishers were in litigation with Google. The number of books available as of early 2007 was in excess of one million.

Microsoft ‘s *MSN Book Search* (www.msn.com until it creates a URL specifically for the service) was launched in 2006 with an initial plan to digitize 150,000 books by mid-2007. It was committed to the “opt-in” approach, meaning that publishers have to specifically agree to have their titles digitized. The service was also planned to include books in the public domain.

The *Million Book Project* was launched by Carnegie Mellon University (<http://www.library.cmu.edu/Libraries/MBP>) in 2001. Its goal was to digitize one million books by 2007. As of early 2007, more than 750,000 were in the database. More than half were scanned in China and more than a fourth in India---countries in which there are at least 40 scanning centers with sophisticated OCR (optical character recognition) equipment to enable full-text searching. A little more than 20 percent of the books are in English, primarily contributions from Indiana University, Pennsylvania State University, University of California/Berkeley, and University of Washington. . 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. The URL for the U.S. site is <http://www.ulib.org>).

OverDrive

Libraries that want popular titles should consider OverDrive (www.overdrive.com). It has over 30,000 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 to read an eBook. The e-books automatically expire and check themselves back into the collection. The vendor provides MARC records for inclusion a library's patron access catalog.

Project Gutenberg

Libraries with limited resources may wish to consider Project Gutenberg, a free site with the text of approximately 20,000 titles in ASCII format. The site grows by approximately 350 titles a year. Only titles no longer protected by copyright are included. As a rule of thumb, that is books copyrighted prior to 1923. 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 at www.gutenberg.net is limited to the index and links to the FTP sites. As of early 2007, Project Gutenberg planned to add eAudioBooks and other eMedia.

Major eContent Providers

The most important players in eContent are eBrary, Internet Archive, OCLC, and Questia.

eBrary (<http://www.ebrary.com>) is a for-profit company that was founded in 1999 to sell eBooks. It subsequently broadened its scope to include eJournals, and eMedia. It offers both subscriptions and outright purchase of the entire collection. It has developed its own reader software to access the eBooks on its hosted server. A library may also add its own eContent in PDF format. The company claims more than 1,000 customers, but has not provided statistics on the number of titles it has available.

The *Internet Archive* (<http://www.archive.org>) is a non-profit organization that was founded in 1996 to build an internet library. It relies on contributions of eJournals, eBooks, eMedia, and eArchives from the creators of the content. It includes eBooks from the Million Books Project. Other collaborators are the Library of Congress and the Smithsonian. Its Web site has links to many other eContent providers.

OCLC has two major programs, NetLibrary and Electronic Collections Online. Founded in 1998, NetLibrary (<http://library.netlibrary.com>) was purchased by OCLC in 2001. Since that time, the number of eBook titles has been increased from fewer than 40,000 to more than 130,000 scholarly and reference works in behavioral sciences, social sciences, physical sciences, management and public relations, law, and technology. 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.”

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

A unique subscription offering from NetLibrary is the Catalog of Art Museum Images Online (CAMIO), a collection of more than 90,000 images of art objects and photographs from major museums around the world.

OCLC’s Electronic Collections Online (<http://www.oclc.org/electroniccollections>) offers digital images of articles in more than 5,000 journals.

The Open Content Alliance

The Open Content Alliance (<http://www.opencontentalliance.org>) is a consortium of non-profit and for-profit groups committed to building a free archive of eBooks and eMedia. It was conceived in 2005 by Yahoo and the Internet Library as a response to the Google Library Project. The emphasis is on eBooks and eMedia. from throughout the world. Contributions are solicited from libraries, organizations and publishers. Contributors are required to obtain permission from copyright holders before submitting

titles not in the public domain. Among the major contributors are Columbia University, the Netherlands-based European Archive, the Internet Archive, the National Archives in the United Kingdom, University of California, University of Toronto, and University of Virginia,. Multiple formats are accepted, but PDF appears to be the most widely used by the contributors. There were approximately 100,000 eBooks in the archive as of early 2007 and several thousand eMedia. Yahoo provides the search engine.

Questia

Question is a for-profit company that offers both eJournals and eBooks on a monthly or annual subscription basis, primarily to individual students, faculty, and other researchers. An institution subscriber can provide simultaneous access to multiple users. As of early 2007, there were more than 1.5 million articles available in its eJournal service and 67,000 titles in its eBook service. Its subject matter strengths are the humanities and social sciences at the undergraduate level. Searching is by keyword, phrase, title, author, or subject.

Prepared by Richard W. Boss, March 19, 2007; revised April 4, 2007