Academic Libraries as Digital Gateways: Linking Students to the Burgeoning Wealth of Open Online Collections

Jane Hutton
West Chester University of Pennsylvania, jhutton@wcupa.edu

Follow this and additional works at: http://digitalcommons.wcupa.edu/lib_facpub
Part of the Library and Information Science Commons

Recommended Citation

This Article is brought to you for free and open access by the Library Services at Digital Commons @ West Chester University. It has been accepted for inclusion in Library Services Faculty Publications by an authorized administrator of Digital Commons @ West Chester University. For more information, please contact wcressler@wcupa.edu.
Academic Libraries as Digital Gateways: 
Linking Students to the Burgeoning Wealth of Open Online Collections

Jane Hutton
Electronic Resources/Reference Librarian
West Chester University of Pennsylvania

Abstract
Digital collections of full text e-books are proliferating on the Web and provide a wealth of open content for students. To examine whether academic libraries are providing a digital gateway to these resources, ten e-book titles from open digital collections were searched in the online catalogs and Web pages of ten academic libraries serving distance learners. Only three of the digital collection e-books were available from any of the library catalogs and none were found on library Web pages. Availability of the ten e-book titles through Google and other digital discovery tools also had mixed results. Continued projects for improved delivery of open online content are necessary. In order to fulfill their role as digital gateway for their academic communities, libraries must pursue metadata standards to support cross-searching, collaborative projects, and development of e-resource search software which integrates with the library catalog.

Introduction
Concurrent discussions in the literature focus on the development of digital libraries comprised of online open collections, the technical applications for cross-collection searching, and academic library services to distance learners. This paper examines current academic library practice in linking distance learners to full text e-books in open digital collections. Does the library serve as a gateway to guide students in an online environment to these freely available online academic sources?

Burgeoning Online Collections
Online collections of research materials are expanding rapidly on the Web. The American Memory collection from the Library of Congress now delivers digitized resources totaling more than 9,000,000 documents (Rosenzweig, 2007) and the Google book digitization project may ultimately deliver open full text content of over 30 million volumes (Quint, 2004). More than twenty major digitization projects in the Americas are listed on the International Federation of Library Associations and Institutions (IFLANET) Web site (http://www.ifla.org/II/diglib.htm) while national libraries in Australia, China, Scotland, Russia and library projects throughout Europe are offering their own unique digitized research resources to the Web searcher. OAIster, which catalogs digital resources, harvested 729 repositories of 9,950,256 records in January 2007, up nearly 50% from two years prior (http://www.OAIster.org/stats.html). An American history student can obtain primary materials from the Civil War era at The Valley of the Shadow archive from the University of Virginia (http://valley.vcdh.virginia.edu/) or from the World War II era in the University of California's Japanese American Relocation Digital Archives (JARDA) project (http://www.calisphere.universityofcalifornia.edu/jarda/) and the University of British Columbia Library's Japanese-Canadian Photograph Collection (http://angel.library.ubc.ca/cdm4/index_coll0610-6.php?CISOROOT=/coll0610-6). Architecture students can view drawings of Frank Lloyd Wright's buildings in the 1910-11 Wasmuth Portfolio from the University of Utah's Marriott Library Digital Collections (http://www.lib.utah.edu/digital/splash.php?CISOROOT=/FLWright-jp2). Freely available online full text books were pioneered by Project Gutenberg and Bartleby.com but have expanded at niche sites such as the university collections above and on mega-sites such as the Open-Access Text Archive at the Internet Archive (http://www.archive.org/details/texts). As of December 2007 the Text Archive provided access to over 312,000 text items from several major collections and a recent agreement between
the Open Content Alliance and the Boston Library Consortium will bring significant numbers of out-of-copyright volumes from the nineteen Consortium libraries into the Internet Archive content (Boston Library Consortium, Inc., 2007). As Rosenzweig points out, "The Web has given us a great gift—an unparalleled global digital library and archive that is growing bigger every day" (p. 48) but he adds a caution that such an abundance may be more overwhelming than enlightening. Academic libraries can support online learners by providing pathways to relevant resources in the burgeoning online collections.

In addition to the Internet Archive, several other projects have been created to funnel the Web searcher to full text books and other educational digital materials. INFOMINE has been in existence since 1999, constantly improving the findability of Internet scholarly resources in its "academic virtual library" (http://infomine.ucr.edu/about/about.shtml). A focused individual project by John Mark Ockerbloom, The Online Books Page (http://onlinebooks.library.upenn.edu/), provides title, author, and some subject access to over 30,000 books. With an automated approach, OAIster harvests metadata and provides access to digital resources based on the OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). At MyLibrary@Ockham another OAI harvesting program compiles its database of digital documents with an indexing tool to create searchability for 430,000 items "from the content of the 'hidden Web' for the purposes of facilitating teaching, learning, and research" (December 19, 2007; http://mylibrary.ockham.org/?cmd=about). Despite these enterprising initiatives a wealth of valuable research content from digitization projects still eludes the grasp of Web searchers and remains hidden as part of the "invisible Web." Lewandowski and Mayr (2006) estimate that 20 to 100 billion documents are part of the invisible Web which is not readily indexed and retrieved by search engines (p. 336). Dempsey (2006) describes digital resource discovery developments since 1996 but notes that "databases remain siloed...each relatively standalone with its user interface...which is increasingly problematic when library users have so many places where they can spend their 'attention'" (para. 19). Intner, Lazinger and Weihs (2006) also point out the difficulties of digital resource indexing due to lack of metadata standardization for resource descriptions and protocols for exchanging metadata. The technical details of metadata harvesting and indexing are beyond the scope of this paper; however the complexities of cross-collection automated finding tools impose impediments for access to digital documents within online collections. Unlike MARC records from a single library catalog which can be merged into a searchable database of consortial library holdings, the data fields for one set of digital document records may not match a set of digital document records from another source. As Liu (2007) notes, "Two things are urgently needed for the digital library: a standard for metadata content that is analogous to AACR2 and a standardized framework for holding and exchanging metadata that is analogous to MARC" (p. 151). Thus, despite the increasing magnitude of online full text research materials, effective access to these materials remains a challenge.

Academic Libraries as Digital Gateways

The term "digital library" is widely used in the literature with varying definitions. The Digital Library Federation 1998 working definition focuses on organizational structure:

Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities. (Digital Library Federation, 2004)

In seeking to delineate definitions of "digital library," Liu (2007) notes that "computer scientists and engineers might place more emphasis on access to and retrieval of the digital content of digital libraries, while library and information professionals might pay more attention to the digital collection and the services based on it" (p. 149). The directory of digital libraries at Academic Info (http://www.academicinfo.net/digital.html) includes a variety of organizational models. The California Digital Library (CDL) is a broad organization which serves the University of California libraries and their communities. Under a goal to "enable the UC libraries to effectively share their materials and provide greater access to digital content" (http://www.cdlib.org), CDL includes multiple digital collections. Other digital libraries such as the "Making of America Journals" (http://quod.lib.umich.edu/m/moaajnl/) and the Digital Collections at McGill University (http://www.mcgill.ca/dcp/) are subject-specific or single
institution models. Johnson and Magusin (2005), in writing about the digital library in the context of
teaching and learning, state that a commonly accepted usage in academic libraries is that a digital library is
"a hybrid of the print and the digital, including a gateway to online resources that extends the library
collection beyond its physical walls" (p. 8). Since most 21st century academic libraries now deliver content
and services to their users from the library Web site but provide print as well as digital materials, the term
"digital gateway" may be the best descriptor for the library's online presence and interface to the academic
community.

For students who study at a distance, online library services and research materials in digital form
have the advantage of being accessible 24/7 to any learner with Web access. United States college and
university data collected by Allen and Seaman (2007) indicate that the number of online students doubled
between 2002 and 2006 with 3.4 million students enrolled in at least one online course in fall 2006 (p. 5).
Whether studying at a distance or near their educational institutions, online students approach their learning
environments primarily via the Web and academic libraries must meet them there. "These learners are
largely dependent on the quality and academic usefulness of services that the digital library can offer
electronically" (Sharifabadi, 2006, p. 399).

In an OCLC survey of college students conducted in 2005, 85% of respondents stated that they
began a search for information with a Web search engine (De Rosa, Cantrell, Hawk, & Wilson, 2006, p. 1-7)
while only 2% would start such a search at the library Web site. Additionally, when asked how familiar
they were with electronic books, only 51% reported familiarity with the resources with 28% of those
responding as "somewhat familiar" rather than "very familiar" or "extremely familiar." Clearly the
academic library's primary clientele are not locating e-books at the library's Web site.

Dinkelman and Stacy-Bates (2007) studied access to e-books at 111 ARL (Association of
Research Libraries) libraries, finding that only 56% of the library Web sites offered Web pages devoted to
e-books, while searching online catalogs for electronic books was often problematic. Subscription e-books
from vendors such as NetLibrary and Safari were often listed on Web pages by database names, along with
open Web collections such as The Online Books Page. The authors concluded that more consistent Web
site descriptions of electronic resources, distinctions among categories such as databases and e-books, and
improved "functionality of search mechanisms" (p. 57) are necessary to improve access to e-books. By
improving "seamless access to all types of electronic resources, librarians increase the probability that
patrons will locate and use the most appropriate resources...to satisfy their information needs" (p. 57).

Case Study

To understand how academic libraries with distance learning programs are currently linking their
students with digital resources in open online collections, ten e-book titles from ten digital collections were
searched in the online catalogs and Web pages of ten academic libraries during the week of December 16,
2007. The educational institutions were selected to represent a range of institutional support for distance
learning, from West Chester University with few distance learning programs to Athabasca University in
Canada which is devoted to distance and online learning to Empire State College where there is no physical
library. The other university libraries examined were DePaul University Libraries, Northern Arizona
University's Cline Library, Nova Southeastern University's Alvin Library, Pennsylvania State University's
Libraries which serve the Penn State World Campus, Regis University Libraries, Sam Houston State
University's Newton Gresham Library, and University of Maryland University College Information &
Library Services.

Method

E-book titles and their respective digital collections are listed in Table 1. The online catalog at
each library was searched by title and by keywords from author and title. The same searches were then
performed on all library Web pages which listed e-resources or subject collections of Web sites or
databases. An e-book from other digital collections (e.g. NetLibrary) which occurred in multiple instances
for Burns' Poems and Songs and once for Alcott's Moods, was not counted as a positive result.
Results

Only three of ten e-books were listed in any of the ten library catalogs: Drew's *A North-Side View of Slavery* at one library; Twain's *Letters, 1853-1880* at four libraries; and Heydemann's *War, Institutions and Social Change in the Middle East* at six libraries. No title was listed on a library's subject or e-book Web pages. Seven titles were listed in the library catalog in print or microform formats, thus illustrating a disconnect between the needs of online distance learners and the monograph format available to them. Jenson's *Latter-Day Saint Biographical Encyclopedia* and Miller's *The Martinique Horror and St. Vincent Calamity* were not available in any format from any library. Several libraries featured links to one or more of the digital collections on the library's e-resources, e-books or subject guide Web pages but such listings seemed arbitrary (by including *The Online Books Page* but not *Bartleby.com*, for example).

Discussion

This descriptive survey of access to freely available e-books at ten academic library digital gateways confirms and expands the survey results of Dinkelman and Stacy-Bates by focusing on e-books from open digital collections rather than subscription vendors. Each study found that e-book title listings in library catalogs were limited. Links to digital collections on library Web pages provide only minimal assistance to the searcher since the resource descriptions are usually broad. Granularity of access is insufficient for guiding the user through the abundance of online resources. Online distance students in particular are disenfranchised by the lack of guidance to full text e-books and academic libraries serving them are not fulfilling their role as the students' online gateway to the wealth of scholarly resources.

If academic libraries offer insufficient access to free full text e-books, the Google search engine is well-positioned to direct users to the content of its digital book project and perhaps additional scholarly full text resources. Should the academic library hand over responsibility for developing workable search technologies and delivery of academic digital resources to Google and other commercial search engines? Each of the e-book titles in Table 1 was searched at http://www.google.com during the week of December 16, 2007. Three titles were returned as the 1st or 2nd search result and one title was the 16th search result. However, 50% of the titles were not returned in the first 20 search results. A notable search result for one of those titles was that the 2nd Google search result led the searcher to a dot-com Web site where the e-book could be purchased for $9.99. Students are ill-served by being directed first to fee-based content rather than led through a digital library gateway to freely-available online academic content. In warning about marginalization of libraries as digital portals, Lossau (2004) cautions against relying on commercial entities for access to digitized academic resources, noting that commercially driven search engines may rank search results by self-serving formulas, may concentrate on the most easily-indexed content and may not be interested in maintaining index content long-term.

Some academic libraries obtain e-book catalog records from vendors such as *NetLibrary*. Searches conducted for this study discovered several library catalogs which included *NetLibrary* records for e-books in Project Gutenberg, *Bartleby.com* or the University of California Press eScholarship collections. Access to the texts via *NetLibrary* presents the irony of libraries paying for freely available content. In addition vendors frequently limit the number of concurrent users for a title, paradoxically limiting access to open content. Another potential stumbling block could be unpredictable delays by *NetLibrary* in updating additional content from these digital collections.

Projects and Prospects for Linking to Online Open Content

As noted above, several projects such as INFOMINE, OAIster, *MyLibrary@Ockham*, and *The Online Books Page* offer access tools for discovering online academic resources in open digital collections. During December 2007 the ten e-book titles in Table 1 were searched in each of these four project sites. INFOMINE which is based at the Library of the University of California at Riverside and uses a combination of librarian and robot/crawler linking, provided title access to only the two e-books: Heydemann's *War, Institutions, and Social Change in the Middle East*, and Twain's *Letters, 1853-1880*. INFOMINE's strength currently appears to be subject linking to resources at the database or collection level. *MyLibrary@Ockham*, which contains only 1200 total e-book titles, included none of the ten titles.
The Online Books Page, primarily the work of one individual, included four of the e-books. OAIster, a self-described "union catalog of digital resources" (http://www.OAIster.org/), provided access to six of the titles. These discovery projects are being continually improved and academic library digital gateways should explore opportunities to incorporate their digital content linking.

Meanwhile academic libraries are developing their own unique applications for connecting users to open e-books and other Web resources. To provide a single digital gateway to subject-related online and print resources, East Carolina University's Joyner Library created a database application called Pirate Source (Nall & Lewis, 2005). Access to open access e-book titles could easily be included in such a finding tool. However, by existing independently from the library catalog, Pirate Source does not sufficiently address the problem of multiple library gateways to books in various formats. Another recent library project at the Columbia University Libraries (CUL) seeks to provide a single gateway to resources in all formats through the library catalog. By designing and implementing workflows around an automated cataloging form, CUL has streamlined creation of MARC records for resources in free remote online collections (Harcourt, Wacker, & Wolley, 2006). Although the CUL Internet Resource Cataloging Request form expedites entry into the library catalog and submission by selectors provides assured relevance to the university library's academic community, the continued human selection element may have difficulty keeping pace with the rapidly-expanding world of online digital collection resources.

Collaborative efforts have the potential to benefit both library users and library personnel seeking to provide academic e-resource titles. Lavoie, Connaway and O'Neill (2007) examined digital materials entered in the OCLC WorldCat database and noted the untapped potential for sharing resource records. Digital collections were frequently represented in WorldCat as single records for an entire collection. The authors propose that adding WorldCat records for digital manifestations of specific titles would benefit decision-making by libraries seeking to digitize unavailable titles while also "meeting the needs of users who increasingly operate in networked digital spaces" (p. 114).

Open source software applications such as iVia (http://ivia.ucr.edu) now used by INFOMINE and Data Fountains (http://datafountains.ucr.edu) continue to be developed as focused digital resource crawlers and metadata generators. To support the role of libraries as gateways to online digital content, the developers seek to provide standardized but flexible technologies. Mitchell (2006) notes that cooperation among the library community to provide such noncommercial software "will become as important as owning our buildings to house our physical collections has been heretofore" (para. 11).

Index Data (http://www.indexdata.com) has developed two software tools for searching open access digital collections: Pazpar2, an open source application; and MasterKey, a hosted service. As of December 19, 2007 MasterKey provided access to e-books and other digital resources at Project Gutenberg, The Internet Archive, Wikipedia, the Open Content Alliance, the Universal Library, the dmoz Open Directory Project, and OAIster. The MasterKey demo site (http://masterkey.indexdata.com/#) was used to search e-book titles in Table 1. Searches can be performed by author, titles, keyword, date and/or subject and results can be limited by source (collection), subject, author or date. Four of the titles were found and four other titles were located as e-books in different digital collections. Interestingly, because MasterKey includes the Library of Congress catalog in the demo search, the two remaining titles were located in print format. By including additional open online collections and providing searches in a single academic library catalog, the Index Data search software has the potential to connect searchers with book titles in multiple formats.

Conclusion

Given the growing abundance of digitized research resources and the complexities of resource discovery, how will academic libraries position themselves to provide these resources to their distance learning students? A study of undergraduates at the University of Rochester found that students' vision for the library Website was "a portal. They want everything they need to be pulled together into a single place...What they clearly did not want were information silos" (Smith & Clark, p. 38). Thus academic libraries should seek to provide a single access point digital gateway for books in all types of formats, including freely available e-books. From the service aspect, Slade (2004) describes a user study at
Athabasca University which found that ratings of library services by distance students were strongly influenced by how quickly and efficiently library materials were delivered. An additional frustration expressed by respondents was that "recommended titles were in use or not available from the library" (p. 29). By providing links to free online e-book titles, academic libraries can significantly improve service and reduce distance student frustration.

The Guidelines for Distance Learning Library Services (Association of College & Research Libraries, 2004) emphasizes that distance learners are entitled to services equivalent to traditional on-campus students, noting that "innovative approaches...to meet these needs is encouraged" (para. 12). This case study suggests that academic libraries currently are not serving as focused digital gateways to lead distance learners to the wealth of full text e-books in open digital collections. More than traditional students, distance learners benefit from increased availability of online academic content. By attention to metadata digital resource cataloging developments, collaborative projects supporting resource discovery tools, and early adoption of appropriate e-resource search software which integrate with the library catalog, academic libraries can fulfill the vision of a digital gateway for distance learners.
References


<table>
<thead>
<tr>
<th>E-book Title</th>
<th>Digital Collection</th>
<th>URL</th>
</tr>
</thead>
</table>