THE EVOLVING DIGITAL LIBRARY AND WHAT MUST BE DONE TO STAY RELEVANT

George Machovec
Colorado Alliance of Research Libraries - Associate Director
george@coalliance.org

Abstract

The library and information communities are going through unprecedented changes. Libraries are rapidly evolving from a paper to digital presence and the physical library is redefining itself for a new role. This presentation will focus on some of the key transformational changes of our time including - the emergence of eBooks and what they mean for collection development and new ways of purchasing; the threat and opportunities of the digital distribution of eBooks and media which could leave the library as an outsider for popular materials; the growth of discovery interfaces and how they are changing the ways libraries present themselves; the role of digital repositories as the new "special collection" and university storehouse; the emergence of social networks and how we integrate with these tools; the popularity of mobile interfaces and what it means for our user interfaces; and the evolution of consortia as regional and national power brokers for purchasing.   In order for libraries to remain economically viable and relevant for user needs the institution must simultaneously change on many fronts.

Introduction

Libraries are going through substantive changes with the advent of the Internet, computer technology, social networking and changing in publishing patterns. The historic library roles of acquiring, organizing, preserving and distributing of printed materials is now changing in the digital world to licensing/managing electronic resources, providing discovery interfaces and advising users. However, libraries still have a huge body of legacy print material, special collections, physical facilities and unique expertise in consulting with their users.   The transition from print to digital is taking decades and each library must find its own balance on the migration from print to digital depending on its mission, nature of the collection, needs of users and budget. Some of the key issues libraries are now addressing

include how to purchase ebooks and balance this with the print monograph, the introduction of discovery interfaces/mobile/social networking, the move into digital repositories for library and institutional collections and leveraging the power of regional and national consortia to lower costs. The rethinking of the physical library and how it might be designed will be a major effort as current buildings age. A brief overview of selected trends and key issues affecting the modern academic library may help provide a perspective on the challenges.

Scholarly Journal Literature and Preservation

By the late 1990s it became clear that the Internet coupled with modern Web browsing was the preferred mechanism for distributing scholarly journal content to academic libraries. Virtually all major scholarly publishers began programs to digitize not only their current content but to also create digital backfiles. These provided readers with full-text searching, immediate delivery to the desktop, round-the-clock access and complete unbroken journal runs.

One leading example is the JSTOR project funded in 1995 by the Andrew W. Mellon Foundation. This project was based on a pilot at the University of Michigan to began building a trusted digital archive of journal backfile content and now numbers over 1,400 titles. This project had the initial benefit of allowing libraries to store offsite or discard long back runs of journals and is one of the great success stories in scholarly journal digitization (http://www.jstor.org).   In 2009 the service merged with ITHAKA (http://www.ithaka.org/) an organization that "helps the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways."   JSTOR was so successful that they have launched a "Current Scholarship" program that now offers a hosting service for several hundred current scholarly journals that are associated with some of the titles in the historic JSTOR backfile program. In addition, JSTOR is working with a group of scholarly academic publishers to host their ebook content for university presses from Harvard, Columbia, University of California and others.

As libraries move to clear their shelves of journal backfiles because of initiatives such as JSTOR, Project MUSE and publisher digitization programs; libraries continue to have a long concern about the long-term preservation of journal content. Although many publishers have digitized their own backfiles and are offering one-time or ongoing subscription access, librarians have deep concerns over ongoing long-term access especially if a publisher goes bankrupt, views backfile access as not profitable or perhaps some type of technical or natural disaster may interrupt or terminate access. To help assuage concerns about orphaned or lost content, initiatives such as LOCKSS and Portico have emerged as leading archiving and preservation programs.

In the mid-2000's Project LOCKSS (http://lockss.stanford.edu), Lots of Copies Keeps Stuff Safe, was begun at Stanford University as a peer-to-peer method to provide libraries with way to collect, store, preserve, and provide access to their own, local copy of authorized content they purchase. Running on standard desktop hardware and requiring minimal technical administration, LOCKSS converts a personal computer into a digital preservation appliance, creating low-cost, persistent, accessible copies of e-journal content as it is published. Since pages in these appliances are never flushed, the local community's access to that content is safeguarded. Accuracy and completeness of LOCKSS appliances is

assured through a secure, peer-to-peer polling and reputation system. The software was developed at Stanford and has won an ACM award for its innovative design. The project now claims that over 6,700 journals from 450 publishers are LOCKSS enabled. Because some publishers were uncomfortable with a grass-roots archiving program, a closed version of LOCKSS (CLOCKSS) was launched between selected publishers and research libraries to provide a dark archive preservation program which keeps these publishers happy and yet still provides archiving/preservation for the scholarly community.

Portico (http://www.portico.org) is an electronic archiving initiative that launched in October 2005 with the idea of creating a centralized dark repository for digital journal (and more recently ebooks) content. The Initial organizations involved in the project include JSTOR, The Andrew W. Mellon Foundation, Ithaka and the Library of Congress. Portico's goal is to create an economically viable technical infrastructure for archiving scholarly resources. This repository will only become accessible to libraries if a major trigger event causes a journal or publisher to become unavailable from the primary source. Portico current has archives for over 12,000 journals from 121 publishers; has over 65,000 ebooks and is now acting as cloud archival storage for 40 digital collections.

Even with due diligence in digital preservation, the long term archiving of paper journals is still of concern to many scholars. In North America, this need is being discussed through the CRL Print Archives program (http://www.crl.edu/print-archives).   A number of regional paper archiving initiatives are now being coordinated to make sure that key collections are available in paper for use by scholars or as a just-in-case copy if for some reason access to the digital version disappeared for some reason. The program is working on establishing a Print Archives and Preservation Registry where libraries can record their long-term commitment to "Development of and access to a system providing online data about print archiving programs and archived materials, and analysis of library collections to support archiving and retention decisions."

As academic/research libraries become more comfortable with secure digital preservation programs it opens the door for rethinking what to do with bound journal back runs. Once sacrosanct, with in-house use the norm, libraries are beginning to view these collections as lendable or even disposable.   Libraries are moving many of these bound journal archives into high density storage, remote warehouse locations or may even be disposing of volumes. Many of these volumes are now being lent to both local patrons as well as through interlibrary loan.   The logic is clear, as content because easily accessible in a digital environment and proper digital and print preservation initiatives are in place, local paper is declining in value.

eBooks and the Printed Monograph

Although journals took the lead in moving into the digital arena and becoming widely accepted by scholars, students and librarians; ebooks are now quickly becoming a popular content type to have wide acceptance with readers. The advent of popular ebook readers such as the Kindle (Amazon), Nook (Barnes & Noble), Sony ebook Reader, tablet computers (e.g. iPad), mobile phones and other devices are creating consumer demand for ebooks on which libraries can capitalize.

The first major scholarly ebook aggregator, NetLibary, launched in the late 1990's and was quickly followed by competitors such as ebrary, EBL and MyiLibrary. Consolidation in the marketplace has

brought NetLibrary under the EBSCO umbrella and ebrary has been purchased by ProQuest. This has allowed major journal aggregators to have both books and journals under their purview, a more complete package.

Some of the major primary publishers such as Elsevier and Springer have chosen to exclusively distribute their own content while others such as Oxford University Press and Cambridge offer their ebook content through all possible venues including virtually all ebook aggregators as well as through their own platform. Their view is that as long as their content is licensed the presentation platform is not the issue.

Until a couple of years ago, ebooks seemed to be viewed by academic libraries as a secondary format, still the stepchild of the printed monograph. With some exceptions (e.g. art) this perspective appears to be changing with both librarians and readers.

Futurist, Thomas Frey, Executive director and Senior Futurist at the DaVinci Institute (http://www.davinciinstitute.com/) in Colorado (USA) predicted at an April 2011 Innovative Users Group Conference in San Francisco that printed books will no longer be purchased by libraries within 10 years. Most of the printed monographs will be, if retained, moved to high density off-site storage and recalled when needed.

The purchasing of printed monographs in academic libraries is certainly in decline. For example, Colorado State University has reported in the last three years the number of printed monographs purchased by its library as fallen from 12,000 to 8,000 to 4,000 volumes respectively. Although there are other factors affecting this decline such as rising serial prices, a drop in funding due to the economic downturn and the need to acquire more electronic resources (including ebooks); the decline is being repeated in many academic libraries throughout the world.

A multi-year study of circulation statistics by eight academic libraries in the Colorado Alliance of Research Libraries found that for new monographs purchased that large percentages did not ever get used.

Purchase statistics for 8 academic libraries in the Colorado Alliance of Research Libraries, 1999-May 2008

- 566,401 titles
- 2.44 copies owned per title (high of 4.91, low of 1.60)
- 1,383,233 books
- 145,603 books purchased annually
- 59,621 titles purchased annually
- $107,892,173 spent (based on avg cost of $78.00)
- $11,357,070 spent annually
- 34.52% unused

For these same libraries circulation statistics were gathered to examine monographs which were never checked out

- Colorado State University   39.14% 278,650 titles
- University of Colorado at Denver   39.46% 129,914 titles

- University of Denver   47.77% 208,248 titles
- University of Colorado at Boulder   49.41% 348,181 titles
- Colorado College   50.89% 67,250 titles
- University of Northern Colorado   51.57% 116,799 titles
- Regis University   52.41% 55,848 titles
- University of Wyoming   59.62% 133,645 titles

These low use rates are not an exception in academic libraries but the norm. This is one of the factors causing libraries to look at Patron Driven Acquisitions (PDA) programs as a way to build collections that are actually used. No monograph is purchased unless it is used.

Ebook aggregators have large percentages of overlap in what titles they offer since primary publishers who distribute through these companies typically do not sign exclusive contracts. What this means is that librarians are selecting ebook aggregators not on content but other factors such as the flexibility in purchase options, the user interface and the ability to integrate with traditional approval plans to reduce duplication between print and digital editions.

Although much ebook purchasing has been done through collection set purchases, publisher packages or one-by-one selection by librarians, the popular technique being piloted by many libraries are patron driven acquisition (PDA) programs. Although this concept is not new as early programs such as this were launched in 1999 by NetLibrary, the idea has new energy and fervor among North American academic librarians.   In this type of program, a library will typically load MARC records from an aggregator (e.g. NeLibrary, ebrary, EBL or MyiLibrary) based on elements that might include publisher, subject, date, price cap or other criteria. The library will then let patrons access these titles and after some threshold of use is reached the library (or consortium if being done through a group plan) will either rent or purchase the ebook.

The advantage of such plans are many:
- A library only pays for what is used
- Pay at the point of need
- More titles can be exposed to the reader

In a recent 2010/2011 one-year pilot by the University of Denver with EBL the power of the demand driven approach becomes clear. The university library exposed just over 63,000 monographic titles through their online catalog and in their initial year 255 titles were actually purchased ($19,520 USD), 2,988 titles were viewed more than 5 minutes with EBL's Short Term Loan option ($39,831 USD) and 5,397 titles were viewed for less than 5 minutes with no purchase or short term loan triggered. In total, of the 8,640 titles actually used, the retail cost would have been $671,165 (USD) if they had been purchased by the library but only $59,341 (USD) was actually expended, a savings of $611,834 (USD). See the presentation by Michael Levine-Clark at http://www.coalliance.org/content/view/309/103/ at a May 6, 2011 eBook conference sponsored by the Colorado Alliance of Research Libraries.

One of the major dilemmas facing libraries regarding ebooks is that most publishers and aggregators do not allow interlibrary loan or resource sharing, a long standing tradition where libraries may lend books to other libraries. Springer is a notable exception and offers interlibrary loan options whereby a library can download a PDF of a book without DRM (digital rights management) and share it with a requesting library. More publishers should be encouraged to follow the Springer contract provisions.

One possible solution is the short term loan (STL) option now offered by several of the ebook aggregators. In this model, as currently conceived, if a library does not own a title, it is rented for a brief period (usually varying between 1 day to 30 days depending on the vendor and options selected by a library) and a small incremental rental fee is paid typically 10-20% the cost of the retail book. With a modified version of this short term loan program, one could envision the equivalent of interlibrary loan (ILL). ILL costs money to receive a request, pull it from the shelf, ship it to the requesting library and then to return it to the owning library. If funds would be redirected to an STL program the patron would get immediate access to a title and resource sharing could be retained. The challenge is to establish new financial models where libraries will pay for this ILL (i.e. STL) on behalf of a patron from another library requesting ILL.

The Role of Discovery Interfaces

As libraries content becomes increasingly digital in their content the role of discovery interfaces becomes crucial as they are the gateway for libraries to present to users what is licensed on their behalf. The library catalog and individual search interfaces from publishers, aggregators and other online services are often very powerful and tailored for specific needs, but a library's content becomes embedded in different silos making it difficult to find. The first solution by many vendors was to create metasearch (or federated) interfaces to broadcast searches to many targets and aggregate the results. The problem with this approach was that it was limited in how many targets could effectively be queried without a breakdown in performance.

In late 2004, Google Scholar was introduced and brought super fast searching across a whole body of scholarly information. This tool was then enhanced by allowing libraries to contribute their holdings data which would allow link resolvers to bridge patrons to library-subscribed content. Since then Google Scholar has continued to add much additional metadata, links to open access repositories, citation linking, patents, legal materials and other resources. As a free resource, Google Scholar is outstanding. However, there are problems – nobody actually knows what Google Scholar covers; it certainly does not have everything; what is the overlap between Google Scholar and other Google islands such as the main Google search, Google Books; the quality of metadata and searching is quite variable; the lack of advanced search capabilities; the inability to link back into specialized databases; lack of library branding; and the list goes on.

The success of Google Scholar with its recognized limitations has opened the door for the commercial information sector. The single search box paradigm in libraries was initially launched using metatasearch technology to bring together content from many targets. In the last few years, a new generation of discovery interfaces has moved libraries closer to a single search box experience by bringing together metadata under a single consolidated index, as is done in Google.   Initially, many of the discovery layer initiatives focused on the library catalog itself as the primary source of metadata. However, it was soon realized that MARC records are only one source of content and that journal literature, OAI harvesting from repositories and other data sources were also central to the mission of libraries.

A number of leading library vendors are competing to create a single search solution to capture this new market.   Some of the leading vendors that have released products include Summon (Serials Solutions),

EBSCO Discovery Service, Encore/Encore Synergy (Innovative Interfaces, Inc), Primo/Primo Central (Ex Libris) and OCLC WorldCat Local.

As with other information products, no solution has everything in terms of functionality or content and each uses different techniques to enhance their shortcomings. Since these services are from fierce competitors, metadata from some databases will probably never be shared between them. This is particularly acute between Summon (part of the ProQuest family) and EBSCO EDS. OCLC WorldCat Local, Encore Synergy, and PrimoCentral are from vendors who claim to be content neutral but they certainly have a bigger mountain to climb.

Do you need only one discovery layer or does it make sense to have a separate discovery layer for the traditional library catalog and one for other content? There are no right answers and different solutions will make sense depending on audience, local content, mission/needs, areas of subject emphasis, politics and funding. Of course, if no additional money is available, there's still Google Scholar.

At the same time as the deployment of the new generation of discovery interfaces, social networking and mobile devices are being widely used by consumer. Many of these library-oriented products are integrating components into their interfaces to facilitate social networking tools such as Facebook, Twitter and a host of other start-ups.

Mobile interfaces, particularly oriented to smart phones, have become the other leading user interface development issue for many publishers, vendors and libraries. There seem to be two major development approaches – create apps for different operating systems (e.g. iPhone, Android, Blackberry, Windows Mobile) or create mobile-friendly Websites for smart phone users. The opportunities and challenges of both approaches are obvious. Apps are widely used by the consumer and it puts access to different services directly on a user's device. However, companies will need to develop and maintain these apps in different operating systems. Mobile friendly Websites can be automatically deployed and if a user goes to your main site they can be automatically redirected to a mobile interface. Because this is neutral to the operating system it certainly is quicker to deploy. QR codes can also be used to easily direct users to these interfaces. What will library users prefer? It's not yet clear but it may take some experimentation, user studies and statistical use analysis to determine the right direction.

Web Scale, Digital Repositories and the Cloud

Many of the information services that libraries license and use on the Web are hosted elsewhere and as such are Web-scale or in the cloud. This is a marked shift from the late 1980s and 1990s when the trend among libraries was the local hosting of not just the integrated library system but also journal indexing/abstracting services. This was partially driven by the lowering cost of hardware and yet low bandwidth with the newly emerging Internet. By local hosting libraries could scoot around the problems of bandwidth and interruptions in the network. However, centralized hosting of journals, databases and other Web-services are now the norm and only a few organizations insist on the local hosting of content due to the massive IT overhead and need to get data feeds from providers (North American examples of sites still doing significant local hosting of journals and indexing/abstracting services include OhioLink, University of Toronto and Los Alamos National Labs).

The local hosting of integrated library systems (ILS) remains one of the key services still operated by many local libraries.   Many ILS vendors offer SaaS (software as a service) and provide local hosting for a fee but until recently it has been mostly used by smaller libraries who could not afford to operate their own system. In April 2009, OCLC announced their move into local services at Web scale. This announcement has far reaching implications and raises interesting questions about the expanding role of OCLC.

Historically OCLC was known for centralized services for cataloging and their interlibrary loan service and had developed a comfortable niche and working relationships with local ILS vendors. But the development of their centralized Web scale deployment of WorldCat Local, circulation, acquisitions, license management, digital repositories and identity management moves OCLC into the realm of commercial ILS companies. Could a library get rid of a local ILS and run the OCLC suite of services? OCLC certainly would like to see that happen.

There is a high-level of tension and ambiguity in the marketplace. At the same time as OCLC's big move, ILS vendors are looking to reinvent themselves and offer a broader suite of services. Weaker ILS companies have been acquired by others, some product lines are being terminated, and vendors are developing new products and tools. At the same time, the open source movement is picking up momentum with many libraries looking to open source software for services such as integrated libraries systems, institutional repositories, next generation interfaces, and other library management tools.

Despite OCLC's big successes, their history has also been littered with some failures, weak product offerings, periods of stagnation, and occasional lack of direction. In 2010, when Innovative Interfaces Inc. (III) deployed a Web-scale cataloging system, called SkyRiver, to compete with OCLC's move into their space, OCLC responded by requiring libraries to pay a huge mark-up to load their local SkyRiver holdings into OCLC's centralized database used by many libraries throughout the world. This action was seen as monopolistic and litigation in U.S courts is now underway to determine if OCLC is using its status in an inappropriate way.

In some ways, as libraries move into a more centralized approach to offering their users access to information over the Web, it means that libraries are becoming less diverse in their collections and offerings. What will make libraries unique are their special collections, research data and other local intellectual content that is not being distributed through the commercial marketplace. Institutional or Digital Repositories are the tools being used by libraries to store their unique digital content as it is either converted to digital form or born digital.

Although there are a number of popular commercial institutional repository packages such as ContentDM (OCLC), Digitool/Rosetta (ExLibris), Digital Commons (Berkeley Electronic Press) and others; there has been a surprising uptake of open source institutional repository software by many academic research libraries. The two most popular open source solutions are DSpace (http://www.dspace.org/) and Fedora (http://fedora-commons.org/) which can be freely downloaded, installed and maintained by anyone. Interestingly both projects began as separate initiatives but are now managed under a common umbrella organization called DuraSpace (http://DuraSpace.org/).   "DuraSpace is a 501(c)(3) not-for-profit organization. DuraSpace software and services are used worldwide as solutions for open access, institutional repositories, digital libraries, digital archives, data curation, virtual research environments, and more." A third project being led by the DuraSpace group is DuraCloud (http://DuraCloud.org) which will allow sites to store or mirror their digital content in cloud storage providers (e.g. Amazon, RackSpace).

Rethinking the Physical Library

What will the role of the physical library (bricks and mortar) be in the coming decades as less content needs to be housed and more is stored and accessed on the Web? Academic libraries are focusing on the library as a space for people – collaboration, research assistance, training, public meeting spaces, pools of PCs/Macs, high end workstations and specialized equipment for the creation of multi-media and even gaming.

Many academic libraries are beginning to move portions of their journal backfiles and monographs into high density remote storage facilities. Many of these facilities are regional being shared by multiple libraries in geographic proximity to reduce costs and perhaps even allow the de-duplication of unnecessary duplicate titles. Discovery interfaces are being used to search and recall the physical item when needed. The academic library seems to be shifting to the model where local access to the physical item is only offered for the high-use frequently requested materials. The irony is that materials stored in climate controlled storage facilities are better maintained and safer than when they used to reside in library stacks.

An example of this shift may be seen at the University of Denver Penrose Library (Denver, Colorado, USA). Beginning in April 2011, the library is undergoing a $32 million (USD) complete renovation and the entire collection will be in remote storage for about a year. Once completed, only about 20% of the original collection will be returned to the refurbished facility while the rest will be retained in high density storage about 10 miles off-campus for quick retrieval when needed by faculty, staff or students. How does the library sell this?

- Increased seating on the upper and lower levels
- Addition of deep quiet study areas
- More group study rooms in a variety of sizes
- ADA compliant book stacks and elevators
- Enhanced technology in all areas
- A new multimedia software support service
- A larger café with patio seating
- An academic events space
- Increased number of power outlets throughout the building
- Increased natural light
- Energy efficient lighting and mechanical systems

As can be imagined local media (http://www.9news.com/rss/story.aspx?storyid=197053) and some faculty and students on campus are very upset about these changes. A "Save the Stacks" group on Facebook is organizing protests. Library management will work with stakeholders to adjust what will be returned to the renovated library once it is finished, but in the end it will be a fraction of the 3+ million volumes housed in the original facility. Special collections will continue to be housed at the new library and digitization of the library's unique collection will continue to receive top priority.

Consortia and Group Purchasing

In the late 1990s a group of consortium leaders begin informally began to meet at the American Library Association to discuss how to license materials more inexpensively and get publishers and vendors to be more agreeable to group purchasing. The group, initially known as the Consortium of Consortia, eventually was named the International Coalition of Library Consortia (http://www.library.yale.edu/consortia/). ICOLC began biannual meetings which are now split between two continents – one held in Europe and one in North America.

The original purpose behind ICOLC was to get publishers and vendors to lower prices, have better contract provisions and work with regional and national consortia. The group also worked on best practices for licensing and guidelines for good vendor behavior.

Many libraries have found that working with one or more consortia is one of the best ways to save money in the digital environment. For example, the "big deal" for journals has brought libraries complete publisher packages of journals for what they used to pay for their historic printed subset (the downside, of course, is that a block of money is tied up in that one deal). Ebooks are a wonderful class of materials to purchase in a group. It is possible to obtain titles at perhaps 20% to 25% of the retail cost of purchasing them on a standalone basis.

Unknown to patrons, a library's participation in consortial activities is crucial to obtaining more content (and other services) thus making the library a richer and more relevant resource. If a library is not doing major licensing through consortia it is likely paying too much.

Conclusion

No one element defines a library but a suite of content, services and facilities characterizes a library and its relevancy to its community. Libraries must redefine themselves as the publishing industry evolves, new technology becomes available and user expectations change. One of the biggest mistakes to make is to stay mired in the past without recognizing these changes. The printed medium and legacy collections will be a part of libraries for many years to come, however, the shift to a digital future is upon us and libraries will become museums unless the digital future is embraced.