Today's digital information landscape: What and how

The main point of this lecture is to bring home a single idea, namely, the what of library and information science has not changed so much as the how. Libraries are still about the processes of collection, preservation, organization, dissemination, and sometimes evaluation of data and information. While the mediums, environments, and tools have dramatically changed, the problems and services the profession addresses remain the same.

MARC and XML

MARC was a great technology that has all but outlived its usefulness. It was created in a time of computing scarcity and sequential access data files. Moreover, it has gone from a data structure used to print catalog cards to the fodder for entire library systems. MARC is a Gordian Knot that needs to be cut, and XML put into it's place.

There are many reasons why MARC is not the greatest technology for today's needs, but the most important has nothing to do with computing efficiency, language bias, or mathematical elegance. The most important reason is social. MARC is not a data structure familiar to any institutions beyond libraries. The rest of the world uses XML. If libraries are about disseminating data and information, then libraries need to speak the language of the intended audience. The audience understands XML but not MARC.

> "It is not so much the 'what' of library and information science that has to change as much as it is the 'how.'"

Databases and indexes

Databases are great for organizing and maintaining data but a poor technology for search. Indexes are great at search but really weak on maintenance. Use databases and indexes in conjunction with each other to create information systems. They are two sides of the same information retrieval coin.

A common technique for creating information systems is to combine databases and indexes. Use a database to organize and maintain the data. On a regular basis run a report from the database containing the content to be searched. Feed the report to an indexer to create an index. Provide an interface to search the index. Organizing and disseminating data and information are a lot of what libraries are about. In a digital environment, these are the tools do facilitate these tasks. We have organization and dissemination skills, we just do not exploit the best tools to accomplish these ends considering the current environment.

Exploiting the network and building relationships

Facebook, My Space, and LinkedIn are all the rage, and the reason they are is so popular is the networking. It is really easy to identify other people with your particular interests. Through tagging, reviewing, and sharing the people who spend time with social networking are building relationships between themselves and other people through common interfaces.

We, the library profession, can do this too as long as we try to put ourselves in the patrons' spaces. It is not as much about people physically visiting a library or even coming to a library website. It is more about making our content -- and expertise -- available at the time and place when the user's need it most.

The global network of computers supporting social networking and mash ups is also one of the fundamental ingredients to successful of open source software projects. The heart of open source is a process of building networks of like-minded individuals who desire to solve common computing problems. The process is not without leadership, norms of behavior, and conflict. At the same time, it is not a highly structured nor centralized process. It relies on a loosely joined network of computers as well a loosely joined network of people.

The process of computing has also taken advantage of the network. Web Services computing, in the form of SOAP and REST-ful applications, is the best example. AJAX is a nifty computing technique. OAI-PMH is a wonderful way of sharing metadata. Blogs, ATOM, and RSS lower the barrier for the sharing of ideas and publication. OpenSearch and SRU exploit the network to provide modern-day Z39.50-like search interfaces. Ask yourself how useful your computer is without it's Internet connection. The folks at Sun Microsystems surely got it right a number of years ago with their slogan, "The network is the computer."

Institutional repositories and open access

Institutional repositories seem to be yet another reaction to the dramatic and never-ending price increases in scholarly literature. Believe it or not OAI-PMH was one of the first reactions. SPARC was another. Sprinkle the idea of open source on institutional repositories and you get open access. Institutional repositories and open access publishing activities are here to stay, but so is commercial publishing. Just as open source software is not going to replace commercial software, institutional repositories and open access publishing will live side-by-side their commercial counterparts.

I attended the Charleston Conference a few weeks ago. A common question was, "What are we going to do for a job when and if everything becomes open access?" Again, this sort of question focuses too much on the how of the profession and less on the what. Acquisitions departments are not necessarily about buying content. If there were, then they would be working in the Purchasing Department. An acquisitions department is responsible for bringing collections into the library. Those things can be items from commercial publishers, open access sites, the hosting college/university, or the Web in general. How are you going to preserve the content if you don't bring it in locally?

"Next generation" library catalogs

A "next generation" library catalog needs to include content beyond the content a library owns. In a networked environment, ownership is not as important. The catalog needs to be defined as the content needed by the students, instructors, and scholars necessary to do their learning, teaching, and research. While it will never be possible to acquire all the necessary content, it will include bibliographic data from books & journals, theses & dissertations, government documents, images, movies, sounds, data sets, etc. By combining metadata with full text it will be better able to perform relevancy ranking and identify obscure facts.

The most significant difference between traditional library catalogs and the "next generation" library catalog lies in: 1) the enhancement of the discovery process and 2) providing services against the collection beyond simple identify. Putting the users' needs and characteristics at the center of the query process will greatly enhance the discovery process. By knowing more about the searcher -- placing the query in context with the searcher -- it will be possible to improve find significantly.

Ask yourself questions about the searcher and tailer the search results. Who are they? What is their level of skill or education? What classes are they taking? What is their major. How old are they? Are they new to the subject or an expert? Who are their peers and what are they using? Use those resources as a guide. Do they want help? To what degree to they desire privacy? By knowing the answers to these sorts of questions search results can be tailored to meet individual needs; search can be put into the user's context.

More importantly, a "next generation" library catalog will provide services against the things discovered. These services can be enumerated and described with action statements including but not limited to: get it, add it to my personal collection, tag & classify it, review it, buy it, delete it, edit it, share it, link it, compare & contrast it, search it, summarize it, extract all the images from it, cite it, trace it, delete it. Each of these tasks supplement the learning, teaching, and research process. They are tools and processes our students, instructors, and researchers use to accomplish their individual goals.

"…a 'next generation' library catalog will provide services against the things discovered."

I like to summarize the ideas about the catalog in this way. "Collections without services are useless. Services without collections are empty. Library catalogs lie at the intersection of collections and services."

Summary

The principles of collection, organization, preservation, and dissemination are extraordinarily relevant in today's digital landscape. The advent of the globally networked computers, Internet indexes, and mass digitization projects have not changed this fact. If anything, they highlight need for these processes even more. Libraries are just one of many players in the information universe. It is increasingly important to adapt to the changing landscape and at the same time bring new value to the collections and services we provide. It is not so much about the what we are doing. It is more about the how.

Eric Lease Morgan
University Libraries of Notre Dame

December 1, 2007