

“Next-Generation” Library Catalogues

The environment of globally networked and commodity priced computers has significantly altered the information landscape. Libraries, once a central player in this environment, have seen their “market share” dwindle. This presentation outlines one way this situation can be turned around, specifically, by re-inventing the definition of the venerable library catalogue.

In an analog environment information objects, such as books, require metadata surrogates in order to be presented to users. Things like MARC, AACR2, plus various controlled vocabularies and authority lists are the hallmark of such surrogates. In a digital environment where most of the available content is textual in nature, these metadata surrogates are not as necessary. Full-text indexing combined with relevancy ranked search results facilitate information discovery that is often cheaper and good enough (I didn’t say “perfect”) for most people. Hand-crafted metadata records are difficult to create. Full-text indexing is comparatively easy. Moreover, library catalogues traditionally inventory items in local collections, but with the advent of the Internet a person’s reach for content can easily extend beyond the walls of one’s own institution.

The traditional processes of librarianship include collection, preservation, organization, and re-distribution. These processes are manifested by the bibliographer, the archivist, the cataloguer, and the reference librarian. Considering our increasingly digital environment, these processes can be supplemented – not replaced – with computer techniques. The collection and organization of data and information can be facilitated through the creation and maintenance of relational database applications. The re-distribution of library content can be done by exploiting XML and indexing technologies. Archivists have the biggest challenge in our digital environment because their storage mediums are dependent on a rapidly changing technology. It is important to emphasize that these underlying principles of librarianship are not changing as much as the methods of librarianship. The “whats” of the profession are rather constant compared to the “hows”.

Consider the amount of storage space we typically have at our disposal. Desktop or portable computers typically come with 300 GB of hard disk space. iPods

with 8-20 GB. Flash drives with 2 GB. Think about ubiquitous CD-ROMs and DVDs. None of this counts the hundreds of giga-bytes of disk space one can use on centrally located institutional servers or through commercial entities such as Google, Yahoo, or Amazon.com. Again, the role of libraries – one of data and information storage – seems to be increasingly less relevant.

Considering such an environment, what is a library to do, and more specifically, what is the role of library catalogues? Actually, I believe the role of the library catalogue is quite bright as long as one opens their minds a little bit and learns how to take advantage of the environment. After all, a rising tide floats all boats.

I assert the library catalogue will be less about inventory and ownership. Instead it will be more about access and usefulness. Put another way, library catalogues will provide information that goes beyond “We have such and such an item” and move towards “Here is a set of items of interest, and these are the actions you can apply against them.” These actions are services and they can be divided into two types: 1) services against search results, and 2) services against content. All of these services are epitomized by action verbs, and they are akin to the types of things a student or a scholar applies to information objects. They include but are not limited to:

annotate • buy • cite • compare & contrast •
count occurrence of idea • create flip book •
create tag cloud • discuss • do morphology •
elaborate • find more like this one • find
similar & different • get • graph • highlight •
map • print • rank • reformat • remove from
my list • renew • save to my list • search
content of • search my list • share • summarize
• trace citation forward & backward • trace
idea forward & backward • translate

In other words, when a person searches the library catalogue they can be given opportunities to do things with the results. Some of these things have already manifested themselves, such as display a citation in any number of formats, or deliver items via document delivery. Other services might be more interesting. Ruminates on the following examples:

- Create tag cloud - Do a search. Find many full-text items. Select a few of them. Create a visual representation of word counts. Hyperlink the words to their locations in the texts.

- Create flip book - Do a search. Find many items containing images. Extract all the images and place them into an easily browsable list. Scroll through the list. Identify items of interest. Hyperlink to the place in document where the images exist.
- Annotate and share - Find a full-text item of interest. Add digital “marginalia” to the item. Save the item centrally complete with marginalia. Find an item that has marginalia. Turn on or off marginalia of specific individuals or types of people.

Libraries, by definition, are always a part of a larger community. Academic libraries are a part of a college or university. Public libraries are a part of municipalities. Special libraries work inside companies or governments. School libraries work in... schools. By extension, libraries are expected to know and understand the information needs of their hosting communities. Here lies an opportunity to be exploited.

When facilitating searches against your catalogue consider taking the searcher’s context into consideration. Are they a member of a particular population such as students, scientists, or senior citizens? Are they teaching or taking a particular class? Are they an expert or a novice on any particular subject area? Have they searched for similar items previously? What time of year is it, and what are the current hot topics in the news?

Based on the answers to these sorts of questions search results can be tailored. The search process does not have to be objective. Experts will probably want expert content and novices will probably want introductory materials. Students in particular classes will want items recommended by their instructors or relevant to the class. If the economy or elections are a particularly hot topic, then search results might be adapted accordingly.

Libraries are expected to know their clientele. Use that knowledge on behalf of the patron. Take it into consideration and build information services, such as access to the catalogue, around the idea. Such a service is not something Google is going to be able to provide nearly as easily as libraries. Here is a niche the profession can fill.

Most “next-generation” library catalogues have essentially the same functionality. In a sentence, they are more akin to indexes as opposed to databases,

and they all work very similarly. Identify content (metadata and/or full-text) for inclusion into the collection. Ingest content. Index it. Provide a user interface to search the index. Provide extra services (cite, rank, review, display cover art, etc.) against found items. A growing proportion of these systems are open source in nature.

Considering the current environment, here are a few things libraries can do to make the work of the catalogue increasingly meet user expectations:

- Index everything: books, journals, articles, images, institutional repository materials, special collections, etc. Full-text is better than metadata.
- Make sure your content (metadata and/or full-text) is easily accessible to the Big Three indexers/search engines. People are not necessarily going to visit your site first.
- Consider harvesting/ingesting things from open access journals and the Internet Archive. They are free, scholarly, and often meet the needs of users.
- When your system receives queries as input, analyze them before submitting them to the index. Evaluate them with your “librarian eye” in an effort to create educated guesses as to the user’s information need.
- Repurpose search results for many venues; exploit Service-Oriented Architectures and/or REST-full interfaces whenever possible.

There are enormous opportunities for libraries and librarianship as long as the profession understands how to exploit the changing environment. It requires a shift in attitudes and skills, not a total replacement of them. Spend time playing. Do informal research and development. Write down what you learn. Discuss it with your colleagues. Repeat. Many of your experiments will fail, but this is not a bad thing. It is to be expected. The experience will help you learn what is feasible and what is not. Through this process you as well as the profession will evolve and remain relevant for years to come.

Eric Lease Morgan
University of Notre Dame

October 15, 2008