Catalog Construction: Evolution to Excellence

The future of online catalogs is fraught with controversy. Research has shown that library patrons use the online catalog as a resource of last resort, behind the internet and asking friends. One factor in this underutilization of the catalog is the controlled vocabulary used. The sophistication of modern users, expecting catalogs to be like Google, puts added pressure on the evolution of library OPACS. Examining the catalogs of institutions that are getting a great deal right in serving modern patrons was the goal of this paper. After exploring and reading about online catalogs of academic and public libraries, a comparison was made among the catalogs of North Carolina State Libraries, Westerville (OH) Public Library, and Queens (NY) Public Library. All three institutions represent excellence in service to their patrons and lessons may be learned in how each uniquely provides information to their communities. Even though the Endeca-based catalog at NCSU is legion in the professional literature, much can be gained in examining it personally. For the test drive searches of all three OPACs, "handheld computer" was used as the keyword/subject search, "Grapes of Wrath" as a title search, and "Frederic Chopin" as the author search term.

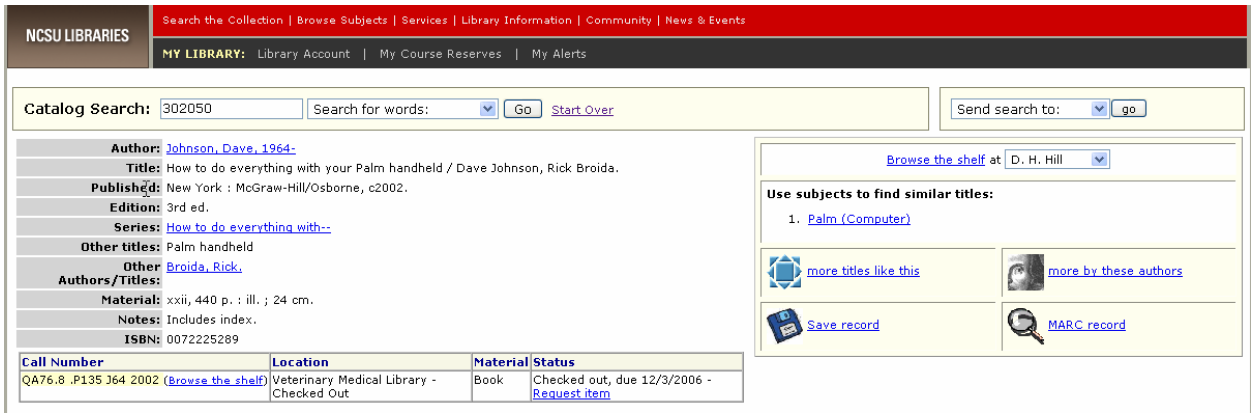
The Libraries and Their Online Catalogs

Serving 30,000 undergraduates and 7,200 graduate students, the North Carolina State University Libraries are composed of six libraries on the Raleigh campus. They circulate about one million items per year of the more than 3.6 million volumes. These libraries were the first to win the Academic College and Research Libraries' (ACRL) Excellence in Academic Libraries Award (North Carolina State University Libraries, 2006). The catalog is powered by Endeca, overlaying the Sirsi/Dynix Unicorn integrated library system, and is updated every night, taking seven hours to complete the process.

The initial interface is predominantly text and is utilitarian in design as befits a university library in contrast to a public library interface. More graphics with some changing content every time a patron enters the site would allow the patron to experience something new with every visit. The most prominent search box is a drop-down menu of popular "destinations" which is confusing. The catalog search box should be the most prominent feature. There is a website search box on the first page as well. The page is divided into six prominent areas for designated collections, information, services, subject browsing, community, and news and events. Choosing the "Search the Collection" section or pulling it down from the "destinations" menu, the catalog itself appears. Tabs along the top of the search boxes are Search (default), Advanced Search, and Browse. Terms are entered in the search box with Anywhere (keyword, as it turns out) as the default option, with title, author, subject heading, and ISBN/ISSN as additional parameters. The term "Anywhere" is confusing and could mean the physical location of an item across the six member libraries. The second search box is labeled "Search begins with. . . ." Drop-down menus for this search activity include "Subject begins with…" and provides a list of subject headings that could serve as a redirect to other subject headings. Some of the other choices seem less useful, for example: "Series begins with…" Andrew Pace explained the process by which these decisions were made in his Powerpoint presentation "Using Endeca for a Catalog Interface" (Pace, 2006). They worked to keep the interface as close to one search box as possible, debated about defaulting from brief record to full record, and ranked the ten dimensions they wanted to use for display on the results page (Subject: Topic, Subject: Genre, Format, Library, Subject: Region, and so forth).

When searching for "handheld computer" other dimensions of the topic appear down the left side of the page. Corresponding to the dimensions decided upon by the design team at

NCSU, it gives a range of choices for further exploration. The first subject offered was the Library of Congress subject heading “pocket computers”. The results list defaults to full view and was listed in relevancy order which a patron can change to sort by most popular, by publisher, by date (oldest first and newest first), by title or by call number. A patron can choose to limit the results by choosing only those titles available. Choosing a title lead to a screen where the patron can choose to return “more titles like this”, save the record, display the MARC format of the record, and find “more by the authors”. Other subjects were offered at this time as well and the patron can browse the shelf at the campus library of choice. This feature was very useful because if the title the patron found most interesting is in circulation and not available, the patron could browse for another title close by.



From this screen, there is the option of sending the search beyond this catalog to their consortium catalogs: WorldCat, Google Scholar, and Quick Article Search. These areas require authentication so could not be tried. If a patron tries a wildcard substitution in the term “handheld comput*”, the suggestion “Did you mean. . .”, with a drop-down menu of possibilities, is displayed to provide a stemming strategy. Boolean strategy was introduced in the Advanced Search mode, with choices to be made by format, library, language, or whether government documents and reference materials are to be included in the advanced search. A year range could

also be chosen as a search parameter. Using the browse tab from the search screen, a patron can drill down through the layers of the Library of Congress classification scheme to arrive at holdings by call number. The "gestalt" shaping of a subject drilled down into, increasing the precision of a topic, is useful.

Using "Grapes of Wrath" as a title, the results were defaulted to the full view of the record. A decision must have been made to use this default for a title search and use a brief record for a keyword search. The results list would be longer in a keyword search and so to avoid more scrolling than necessary, a brief record would be a better default choice. In the results list for the title search, the same sidebar for further options (faceting) and the same sort functions as the keyword search are presented. Again, they are sorted by relevance as a default. In a title search, a desirable feature would employ Functional Requirements for Bibliographic Records (FRBR) or "roll-ups" to reduce the number of records displayed for one item. Often the patron is not interested in five records on *The Grapes of Wrath* by five separate publishers. They just want the book! FRBR is on the wish list for this catalog as well as shopping cart features and email export options.

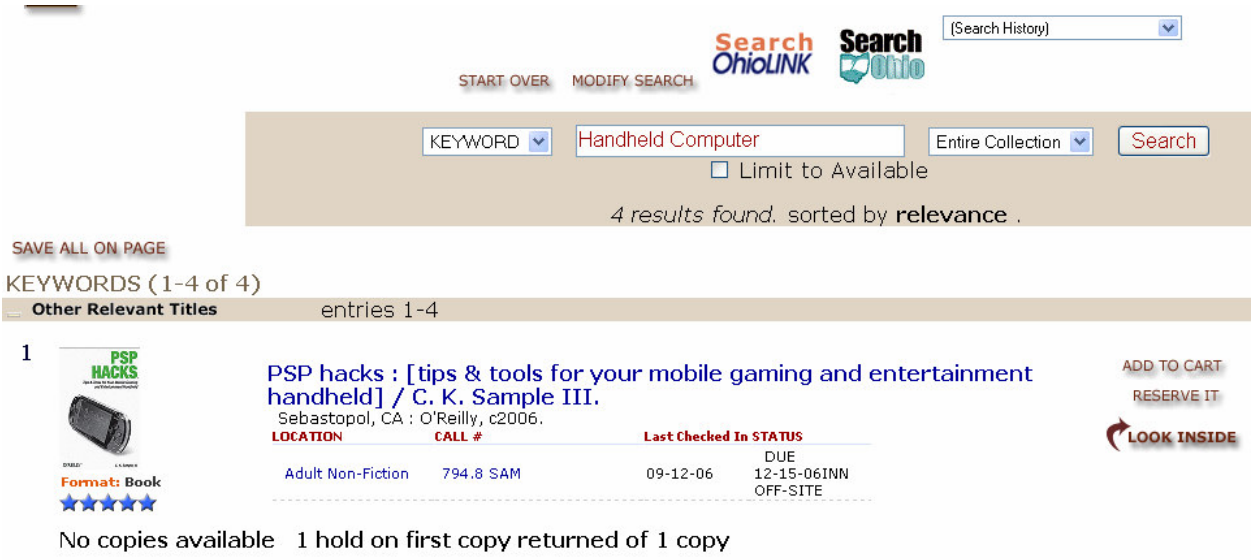
The third search was to test for alternate spellings of the word "Frederic". Searching by author, the search yielded only books. Re-forming the search using "Chopin, Frederic" as a keyword (Anywhere), the audio recordings appeared in the results list. Spelling "Frederick" with a final "k", yielded nineteen entries in contrast to the eighty-two records returned with the usual spelling of "Frederic". No redirect to the correct spelling (provided by an authority file) was offered. Quite a few records could be missed if the patron had not the correct spelling or if the patron missed the correct spelling in the subject list to the left of the results list. The same search in the Buley Library OPAC was conducted and a prompt redirect was made to the correct spelling

where the records were unified by authority control. To test this again, "Shostakovitch" was grossly misspelled and a search was made. The NCSU catalog searched for "Shostakovich", in addition to the misspelling, and returned unified results. So, one would wonder if the search term "Chopin, Frederick" corresponded to a composer that just did not have the first name listed in the 400 field of the authority file. The courtesy to go ahead and include the alternate spelling in the search is excellent and saves the patron a frustrating step of searching again by a corrected term.

Westerville Public Library (OH) is the smallest collection I examined at 307,950 volumes, circulating about 1,186,000 items per year. Serving 37,000 residents since 1930, this library has rated in Hennen's American Public Library ratings index within the top five libraries for eight years. Their catalog is delivered through the Millennium integrated library system by Innovative Interfaces, Inc. (Breeding, 2006). The library uses the *Castr* software for their podcast delivery system. They have archival video of selected programs from the library's events calendar. On a DSL connection, they load quickly and the playback is smooth. The library pages are supported with *Ajax* (asynchronous JavaScript and XML) to speed up the loading of the pages, behind the scenes, by loading only the "changed" data necessary for the page to return the data required. The catalog also uses Innovative's product *ResearchPro* to provide federated searching of local and national databases. These areas require authentication. To enable the catalog to be used on handheld devices, the library uses *AirPAC*, again offered by Innovative.

When first entering the site, the patron is immediately engaged in a contest that gets the "juices flowing". This particular contest is to win an iPod by writing a review of a book. This project reflects the social interactivity that is characteristic of this website. A podcast link is provided to stream video of missed events and programs. There is a link to library photos on *flickr* and links to featured services. From this front page, you can manage your library account and

pay your fines online. The catalog is the first item on the right side and it defaults to a title search of the entire collection. Pulling down the search box menu to SUBJECT and entered “handheld computer”. Expecting a results list, there appeared a message that the term “would be here”. This was not a very helpful message and no suggestions were offered in how to modify the original search terms. There was, at this point, the phrase” search as words” (keyword search). There is not even a sidebar with synonyms to guide the patron. From here, the search can be limited by availability, the search can be extended to the consortiums of OhioLink and SearchOhio, a new search can be started, or the search can be modified. The records are returned with no choice as to brief or full records. The results page is sorted by relevance and other sort parameters are not offered. Items can be saved by placing them in my “cart” and from the cart a patron can email the list or reserve items from that list. The cart feature and the “Look Inside” graphic are very much like features found at Amazon.com.



The “Look Inside” feature is not at all the table of contents, selected text, index and bibliography expected from experience with Amazon.com. This feature offers links to other library catalogs to extend the search and provides links to biographies, reviews, articles, and websites

about the search topic. The databases (articles) and biographies (reference) called for authentication. Following the links to Amazon.com, will yield the table of contents and other features about that item from that external webpage.

Along with a cover graphic which makes the catalog visually attractive, there are rating stars under each graphic. The patron's page for writing reviews is also authenticated. At this point, at least three authentication areas have been encountered. A "one pass" system that requires just one log-in for the patron's session, like the Athens structure, is highly recommended.

In *The Grapes of Wrath* search, a preliminary results list showed the first listing having six items. A video recording was the first entry and no designation was indicated as to the sort that was used. A criterion (date, alphabetical order, medium, etc.) for the sort could not be discerned. A patron can choose to sort by date by using the "Limit Search" option. A patron can choose that same process to limit the search for location of sections within the library, material type, language, and year of publication. If a patron wants to search for more items, the search terms are carried forth to the *OhioLink* and the *Search Ohio* consortium catalogs. The record itself is easy to read and very clean looking. The availability of items is clearly noted with due dates for those items in circulation. There is a curious "Last checked in" notation in each item record that may be system-generated and could be suppressed.

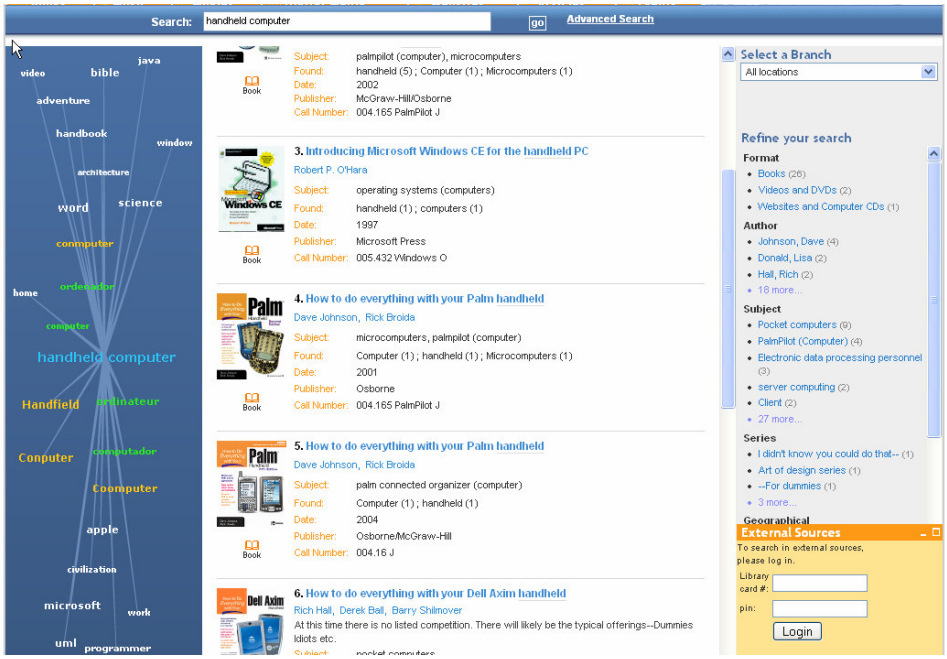
Spelling "Frederick" incorrectly in this catalog, produces an authority list for the patron to choose the correct spelling of the name "Chopin, Frederic." Again, there was no designation of the type of sort used as the default, but after examining the results list, the items were returned in alphabetical order. Large arrows appear to allow the patron to advance or retrace to previously visited pages of the results list.

The interface for the Queens Library is a visual delight. There are graphics to accompany news about the events in the library and the patron may choose a simple (featuring just the search box) or detailed homepage. Every time a patron visits the page it is different. There are images that change, featuring an activity in the library. Across the top are tabs of quick links to recommendations for music, videos, ebooks, digital media, events, and articles. The recommendation pages feature scrolling graphics (like those seen in audible.com and other commercial ventures), most popular picks, and library recommendations that are linked to the item's record. The search box is prominently displayed right over the largest graphic so it is eye-catching. The advanced search is on the same line. Two remarkable features are that the patron can choose the font size of the display and the character set for introductory text in English, French, Korean, Chinese, Russian, and Spanish. The "My Account" information is there for requesting and renewing items and "My Queens Library" allows the patron to customize the page for language, branch library, barcode and PIN, and list of events.

There was no default or pull-down menu to choose the type of search for the first search using "handheld computer". The return list is comprised of full records with no brief record choice. Graphics of the book covers are displayed and the list is sorted by relevance by default. Other choices are year, title, and author. Looking at the records, the search must default by keyword as the number of occurrences of the search term or synonyms are listed for each record under a "Found" line. Subjects are also listed for each record but are not linked to other items or a list of subject headings. The patron may also subscribe to an RSS feed for that search to be notified of acquisitions in the area of interest. The search can be narrowed by branch library as well.

The most startling feature, and one that introduces much interactivity to the catalog, is the *AquaBrowser* software. Using a search, discover, and refine approach to patron's search activ-

ity, once the basic search has yielded a results list, the "word cloud" interactive graphic allows the patron to reshape the search using the words offered. As a word is chosen in the "word cloud" the results list changes to reflect that reshaping. The faceting of the search is conducted in the right column by choosing the format, author, subject, series, geographical area, and year. There is a section for accessing external sources that requires authentication.



In the Advanced Search option, there are Boolean logic pull-downs on the tab "Author, Title, Subject". On the "Browse" tab, the patron has options to search with partial beginnings of title, author, and subject. While not a stemming function, it allows the patron to get help with spelling a search term. The "ISBN" and the "Call No." tab are simple searches by those parameters. There is also a "Database Control No." search that looks at the 001 field (control field) in the MARC record.

When searching for "Grapes of Wrath", full records were returned with no choice to view the results list as a brief record display. As in the subject/keyword search, the results list was sorted by relevance and the word cloud and Refine section was displayed. On this results page

and on some of the Advanced Search pages, there is a problem with graphic elements on the screen overlapping each other. You can read the text of some colored boxes but it is difficult. The html code looks to have incorrect height data for "containing" text in certain areas.

The search for the non-preferred spelling of "Chopin, Frederick" returned eleven items while the preferred spelling yielded 472 items. However, as compared to the return lists of both spellings in the North Carolina State Library catalog, the return list at Queens included the items from both spellings when the preferred spelling was used. The word cloud in *Aquabrowser* offered the suggestion of the preferred spelling and then re-formed the results list. This re-direct method is effective in having the patron salvage the search.

Desirable Features

In the ALA Techsource blog, Karen Schneider offers critieria for "features you wish your OPAC had" (Schneider, 2006). In summary of the features found in the comparison of the three catalogs, relevancy ranking is in place in all three catalogs. Patrons used to "page ranking" in Google and other web search engines make demands on the OPAC for similar result lists. Library OPACS are not a popularity contest that serves up data ranked by the most clicks gained by an item. Any progress in this area, however, is useful to the patron. These three catalogs are making strides with their search algorithms. Spell-checking, an essential feature that would make more searches successful, was utilized in the Queens Library and in the North Carolina State Library subject search. The Westerville Library indicated to the patron that their word "would be here" in the alphabetical list but offered no further help. Refining an original query was handled well by the NCSU Library and by the Queens Library. However, the word cloud in *Aquabrowser* at Queens often offered terms that were in titles of other items rather than conceptual subjects. Westerville gave the patron no help to reform the query except to search as word (keyword) or

look for subjects nearby. That is not a synonym approach to help the patron choose alternate search terms. Duplicate detection using FRBR was not evidenced in any of the catalogs, although it is in the plans at NCSU. Boolean logic is pushed to the back in these catalogs. NCSU presents Boolean operators in the advanced search mode only and the patron has to construct the expression by typing it into the search box. This demands prior knowledge of how the operators work and how a search expression is composed. Queens Library provides a tab in Advanced Search but provides pull-down menus for the operators. Westerville provided no Boolean search opportunities. Sorting result lists was evident in all three catalogs. Westerville's catalog allowed sorting only in the "limit search" area. This flexibility is an essential feature to allow patrons to order the data in the parameter that makes sense for a given purpose. Character sets were evident in the Queens Library but limited to the information sections of the webpage. The three catalogs were all in English. Faceting, or re-ordering a list to browse by category, was done best by the NCSU catalog and the Queens Library allowed faceting in the "refine the search" section. The Westerville Library offered some traditional parametric choices in the "Limit search" option. This traditional way to access data leads to "dead ends" in a search and is frustrating to the user (Rappoport, 2006). Interactivity was definitely a focus at the Westerville and Queens libraries. At the Queens Library, the patron can customize the page through the link "My Queens Library". The patron can contribute reviews to the Westerville catalog.

On weekdays, a suit might be chosen to wear to work. On Saturdays, jeans and a sweatshirt are perfectly fine. So it is with these three catalogs. Each is appropriate for the respective service community. All three OPACS reflect the information needs of the people they serve. The interface and the features of the catalog are chosen with that community in mind. The interfaces of the two public library sites almost hum with activity and their circulation statistics show a

pleased patronage. The university library interface supports the life of the student – even offering study help at LOBO (Library Online Basic Orientation). The effects of their patrons' increased sophistication in online searching and commerce has obviously influenced design of the interface and performance of the catalogs. These three libraries show passion in the quest for providing an efficient and enjoyable online catalog. From the energy exuded by the multitude of features and careful attention to attractively designed interfaces, these libraries will continue to serve as models for excellence.

References

Breeding, M. (2006). lib-web-cats: Westerville Public Library. Retrieved November 15, 2006, from http://www.librarytechnology.org/lwc-displaylibrary.pl?RC=2794

North Carolina State University Libraries. (2006). NCSU Libraries: fact sheet [Electronic Version]. Retrieved November 15, 2006 from http://www.lib.ncsu.edu/publications/factsheet/factsheet.html.

Pace, A. (2006). Using Endeca for a catalog interface. Retrieved November 1, 2006, from http://www.lib.ncsu.edu/endeca/presentations/200603-endeca-pace.ppt.

Rappoport, A. (2006). Faceted Metadata Search and Browse. Retrieved November 26, 2006, from http://www.searchtools.com/info/faceted-metadata.html

Schneider, K. (2006). How OPACS suck-part 2: the checklist of shame [Electronic Version]. *ALA Techsource*. Retrieved November 15, 2005 from http://www.techsource.ala.org/blog/2006/04/how-opacs-suck-part-2-the-checklist-of-shame.html.