Cataloging Internet Resources

This essay describes an attempt to begin cataloging Internet resources. It is also available at http://www.lib.ncsu.edu/staff/morgan/cataloging-resources.html

The NCSU Libraries "Library Without Walls" and "Study Carrels"

When gopher was the big Internet protocol, the Libraries was giving Internet classes to the University's students, faculty and staff. Getting tired of carrying this large books to the class, I heard of a smaller library in the Midwest that thought of organizing Internet resources on a gopher server. Picking up on this idea, I structured the Gopher at the NCSU Libraries using a what I called the "Used Bookstore Model". In this model very broad subjects were created matching the same sorts of subject heading commonly found in used book stores.

Hunter Monroe and Alex

About this time we made contact with Hunter Monroe. His goal was/is to create a OPAC-type database of Internet resources. He called his database Alex. The NCSU Libraries was lucky enough to foster a relationship with Hunter, and consequently, the NCSU Libraries has hosted his data on our gopher server.

Because every information system has to be browsable as well as searchable, the NCSU Libraries indexed the Alex database using the jughead technology. Later, since the hypertext transfer protocol was becoming popular, the NCSU Libraries experimented with methods of providing access to the Alex database via WWW browsers.

Working with Monroe, we learned he was maintaining his collection through the use of a database application. Each record in his application contained fields describing each Internet resource in terms of title, author, date, location, and (eventually) subjects. Monroe used his database application to automatically create gopher link files, the substance of gopher servers. With just a bit of encouragement, Monroe was able to modify his database application and create hypertext markup language (HTML) documents as well. The Alex database has proven to be a model for the rest of our development work.

Tim Kambitsch and Searching OPACs with WWW Browsers

At the same time we had been working with Tim Kambitsch then of Butler University on scripts to search our OPACs with WWW browsers. These scripts allow the user to specify Boolean queries to be applied to user-selected databases.

Mr. Serials and the 856 Field

During the 1994 Annual Meeting of NASIG we became more aware of the proposed MARC 856 field. This field is intended to describe the locations and holdings of electronic documents. We thought it would be a good idea to catalog Mr. Serial's collection of electronic journals and newsletters, as well as include a Universal Resource Locator (URL) in an 856 field. We proceeded to download two MARC records from OCLC (*ALAWON* and *Public Access Computer Systems Review*, and edited them to include URLs in the 856 fields.

By adding these URL's and HREF's to our MARC records, and by searching our OPAC for these records, the user then has the opportunity to navigate directly to the electronic resource after located items of interest.

Alex Meets Alcuin

The NCSU Libraries has also wanted to create a database of Internet resources. To that end, using our DRA OPAC software we created a new database named Alcuin and employed the DRA-WWW gateway scripts Tim Kambitsch.

Monroe's output was then processed with a locally developed utility called Alcuin's Little Helper. Alcuin's Little Helper is a very simple MARC record editor. It can be used to create MARC records by hand or convert Monroe's tagged output to MARC records.

Records were then imported into a newly created DRA database and made available through Kambitsch's gateway scripts.

Simple Internet Database (SID)

In an effort to mirror Hunter Monroe's ideas, we created a database of our own. We called this database the "Simple Internet Database" or SID for short. Internet resources were added to the database after being evaluated in terms of their identifying elements. The database was designed to accommodate as many major and minor subject entries as needed.

Learning from our experiences taken from Hunter Monroe, reports can be created based on the contents of the database. At the present time, these reports are HTML documents. Each document contains all the records matching sets of major subject/minor subject pairs. After the documents have been created they are saved in a directory of the library's WWW server.

Revisiting the concept of a catalog

Just like any other information format, bibliographic records describing Internet-based electronic serials should be included in our OPACs because it is a tool designed to help a defined set of people locate information in a comprehensive collection of data.

Broken URLs and the hopes for URIs

There are a number of obstacles impeding a library's ability to effectively add bibliographic records of Internet-based serials to its OPAC. The first is the dynamic nature of the Internet.

Until the concept of a URI becomes a reality, we can imagine a number of short-term solutions to this problem. The least likely solution is the addition of a new feature to OPAC software by vendors. This feature would examine all the records in the database(s) containing 856 field(s) and check for the validity of the URL(s) found there. Invalid URLs would then be added to a list and regularly sent to a database maintenance team.

A more likely solution is to do this ourselves using the report generation services already included in our OPAC software. Another solution, and quite possibly the most implementable, is the creation of a separate, locally maintained database of Internet resources. This database is not necessarily MARC-based, but it would contain the fields essential to create a complete MARC record. More importantly, it would be able to extract the URL of a record and check for its validity. Then, on a regular basis, when all the URLs had been verified, this database would create a report in the form of MARC records, and these records would be imported into the OPAC overwriting any duplicates found there. Realistically, none of these solutions are ideal, but may be necessary for the short-term.

Enhancing the controlled vocabulary

The controlled vocabularies of our OPACs have always been hallmarks of their usefulness and integrity. Much of the North American academic libraries rely on the Library of Congress Subject Headings (LCSH) for their controlled vocabulary. LCSH was intended to be the vocabulary of the Library of Congress and not necessarily North America; the Library of Congress is not a national library and their vocabulary is designed for their particular needs. Consequently, LCSH does not always include the vocabulary to adequately describe items in our OPACs. This problem is magnified by the length of time necessary to introduce new terms into LCSH.

Labor intensitivity and new skills

Initially, the addition of Internet resources into our OPACs will be a labor intensive process since many of the records will require original cataloging. Since the Library of Congress is not currently producing 856-aware cataloging records, libraries who want to include these sorts of records will have to create their own. It will take time for our bibliographic utilities to obtain a critical mass of these sorts of bibliographic records.

End-user education

With the wide-spread addition of the bibliographic records describing Internet-resources into our OPACs will come a need to educate our populations on the existence of these records in our OPACs. Furthermore, libraries will have to try to distill from the population's mind set that an OPAC is "only a list of books."

A Catalog's Definition Refined

A library catalog is an organized list of information resources arranged in all or any number of schemes (author, title, subject, accession, size, type, etc.) and these resources are readily-available to the intended clientele of the organized list.

Formalizing a process

The ultimate goal (or end) of this project was to create a systematic method for cataloging and classifying electronic serials and Internet resources. At the North Carolina State University (NCSU) Libraries we believe we have successfully outlined the work flow of such a method. It does not differ very much from traditional acquisition strategies. This is how it may work:

1. The collection management department, in conjunction with "subject teams" will select resources for inclusion into a database of relevant items.
2. The URL describing the resource is passed on to our cataloging department who analyze the resource in terms of its author, title, notes, subjects, etc.
3. The results of this analysis is used to update a database program.
4. Reports, based on the content of the database, are then generated updating our dissemination tools namely, an online catalog and our World Wide Web (WWW) server.

Eric Lease Morgan, NCSU Libraries
eric_morgan@ncsu.edu
http://www.lib.ncsu.edu/staff/morgan/
April 8, 1996