The Catalog of the Future: Integrating Electronic Resources

Dana M. Caudle

Academic and research libraries spend millions to acquire electronic resources and to create their own digital resources. With that kind of investment, it is vital for libraries to provide useful access to these materials. The traditional method of providing access to print resources is the online catalog. Whether and how to include electronic resources such as web sites, aggregator databases, electronic journals, and digital materials in the catalog is perhaps the key issue in the future of cataloging. The purpose of the catalog was once to provide bibliographic control for a host of resources physically owned by an individual library. Libraries do not own and house many of the electronic resources for which they provide access. However, I believe the catalog should take on a new purpose: to provide systematic access to information in whatever form it takes, not just to inventory a particular library’s print resources.

Over time, as electronic resources grow, maintaining separate access to them is wasteful and minimizes the usefulness of the catalog. If libraries fail to integrate electronic resources with the catalog, it will become a historical curiosity instead of continuing to play a vital role in today’s library. My main reason for this is the simplest one of all: library users like “one-stop shopping.” Users do not care where the information is or who owns it, so long as they have access to it quickly and conveniently. They should not have to check both the catalog and separate lists of electronic resources, plus perform an Internet search using a search engine like Google, just because the information on their topic is available in more than one form. In fact, users might just find something already in the collection to go with whatever they found online. An online catalog is already an electronic resource. If the item in question is also an electronic resource, the user can just link to the resource from the catalog itself. In this way, the catalog serves as one portal to the Internet. Sarah Thomas enumerates some of the advantages of the catalog as portal:

The catalog is reinterpreted as an information service which registers in a systematic arrangement those publications and documents of interest to a particular community, regardless of the form in which they appear.
This discovery and access tool may exploit a variety of metadata schemes to locate materials, but it imparts unity, predictability, authority, and credibility to search results through the efforts of expert knowledge managers and the application of principles policies, and practices of their devising. In the short term, we can expand the catalog to be more inclusive and flexible.¹

The idea of incorporating electronic resources into the catalog serves the user far better than a web page list of the library's electronic resources. Why distinguish between resources because one is in print and another is not? Barbara Baruth claims our catalogs are not up to the task, that they should remain simple aids to what the library physically holds.² But that does not serve the user who must look in two places and it inevitably leads to short shrift for the physical collection. Not everything is online, even if users sometimes think so. Baruth also claims that library catalogs are ill-suited to the task and too limited for this purpose.³ They won't be if librarians tell ILS vendors this is what we want. SerialsSolutions originally offered to provide and maintain an up-to-date web page of all the individual journals in a library's aggregator packages.⁴ Now, because libraries are increasingly demanding it, they also offer to supply MARC records for all the journals in the aggregators, so that this material can be made available through the catalog. That is just one example of how customer demands drive the services vendors offer. If we need more powerful catalogs, the vendors will supply them. Baruth does have a valid point when she questions whether librarians manually selecting and cataloging web sites can keep up with the enormous and growing number of freely available sites.⁵ However, there is a big difference between trying to include the majority of free web sites and including those electronic resources for which the library has purchased access. The expense alone justifies the cataloging of the latter.

Secondly, it is possible with a little bit of simple programming to extract the titles of electronic journals in the catalog and generate that neat little list for the web site that SerialsSolutions and TDnet wish to supply as part of their journal management systems. This is precisely what Auburn has done with the titles in our aggregators and web sites we deem of particular value. We devised a code for each aggregator and inserted that code into a pseudo 710 in the catalog records for all the journals in that particular aggregator. Using a little programming with Visual Basic and SQL, we are able to run a search against our Voyager records, pull up the journals, and transform the results into a list formatted for the Web. Each journal, database, or portal in the list has a button which will take users to the closest appropriate screen within that electronic resource. It's a lot of work to keep the cataloged journals in the aggregators current, but we don't have to do it twice (once for the catalog and once for the web site) or abrogate our responsibilities as catalogers. Users then have their choice of access points. Integrating electronic resources requires greater cooperation between the bibliographers, the catalogers, and the systems staff, but the results are worth it for the user.

The biggest drawback to Auburn's system is having to manually check links and update records as the aggregators change. Auburn uses a single record approach which complicates matters. CONSER, the body that sets the rules for serials cataloging, has defined the concept of a definitive record for an electronic journal. While CONSER allows a single record approach, they are recommending this single electronic record and a separate record for the print version. It's this single electronic record that SerialsSolutions hopes to provide and keep updated for libraries. Auburn has combined print, electronic, and microform holdings on one record, so we would have a more difficult time taking advantage of both the new rules and outsourcing to someone like SerialsSolutions. A single record for all electronic versions and a separate record for print and microfiche may prove to be the better approach because of the time and labor required to maintain Auburn's single record for all versions approach.

This approach of integrating electronic resources works fine when the resources in question are journals in aggregators or librarian-selected web sites. Roy Tennant and others are correct when they state that not everything can be cataloged according to MARC and AACR2 and included in the online catalog. Some types of resources such as digital collections of images or archival finding aids may require other metadata standards such as Dublin Core or Encoded Archival Description (EAD). Dublin Core was developed by OCLC and others specifically so that non-catalogers...
could create metadata to be embedded in their web sites, and because of its generic nature, is used in a wide variety of ways inside and outside the library community. The archival community developed EAD specifically to create electronic finding aids. According to Roy Tennant, “As libraries digitize collections, metadata is required to organize and provide access to this content outside of, or in association with, the library catalog.”

Most of these other metadata standards are encoded using the mark-up language XML, which will almost certainly become the standard mark-up language for the Web in the next few years. Kyle Banerjee describes some of the many advantages of XML for libraries: it makes it much easier to share and search resources in different formats; it is quite useful for encoding metadata because it is a generalized tool that is not content specific; and it can be transmitted, manipulated, and arranged relatively easily by programmers making it easier to develop useful applications that involve specific resources. In fact, one project, the Open Archives Initiative, involves harvesting Dublin Core metadata from participating data providers and encoding it in XML for use by service providers. XML is also useful for encoding and exchanging commercial data. The publishing industry uses it for ONIX, their standard for encoding and managing metadata used in selling books online. ONIX raises the exciting possibility of buying at least some bibliographic data along with print items.

Because of its many uses, Tennant and others assume that XML is the wave of the future and MARC must make way for it. This is a rather misleading debate as it is somewhat like comparing apples and oranges. The point I am trying to make is that XML is a markup language used in all sorts of applications. You tell XML what metadata schema to use. For example, if your metadata is Dublin Core, you specify Dublin Core in the XML definitions. The computer then knows to apply the Dublin Core metadata. It can work the same way for MARC. So once you define a schema for the MARC format, it is a simple matter to wrap a MARC record in XML for use with that killer XML application Barbara Baruth suggests we prepare for. The debate should not be one of whether to replace MARC with XML, but rather how to define MARC as one more metadata schema that can be manipulated by XML.

In fact, the Library of Congress has already come up with the necessary elements for MARCXML, a version of MARC for use with XML and the necessary crosswalk software to convert MARC records into something usable in an XML-based catalog. OCLC’s CORC Pilot Project, which ran from 1999–2002, allowed librarians to use both Dublin Core and MARC to catalog thousands of web sites. To facilitate the translation of Dublin Core into MARC and vice versa, OCLC developed crosswalks which mapped one format to the other. That was one of the chief outcomes of the project. In fact, OCLC’s new Connexion service, which incorporated CORC and OCLC’s Worldcat bibliographic database, is described as an integrated cataloging and metadata service. The Connexion web interface allows the use of both Dublin Core and MARC according to what best meets the user’s needs. Records can then be exported in either MARC, Dublin Core HTML, or Dublin Core RDF/XML.

As a matter of fact, it is not necessary to rely on MARC, venerable as it is, as the only metadata standard for the library. Dublin Core, EAD, and TEI also have their uses, as does ONIX. David Dorman sums it up nicely in his report on a two day metadata preconference at the 2000 ALA Annual Meeting:

The presentations illustrated that traditional cataloging is now only one kind of metadata activity among many, and MARC is only one metadata standard among many. MARC will not soon die, but neither will it continue to reign supreme as the primary encoding standard for all cataloging. There will be an increasing number of metadata schemes: some flexible and extensible like Dublin Core, and some rich and structured like MARC/AACR2. And to tie them all together, there will be common structures like SGML and XML/RDF.

As Norm Medeiros also points out:

I do not advocate abandoning MARC. Nor do I think the online catalog has exhausted its usefulness. Yet one must admit the thought of XML-based records equaling the playing field between traditional and Internet resources is certainly intriguing.
I raise the whole subject of XML and other metadata schemes to present an alternative to putting absolutely everything into the catalog. That alternative is to make the catalog one database in a larger collection of databases all using metadata appropriate to the types of items they are representing. This is the approach advocated by Roy Tennant when he speaks of gateway catalogs. Endeavor’s EnCompass software is expressly designed to be the type of interface to search across and link the multiple metadata databases that Tennant envisions. Endeavor is the vendor behind the Voyager library system. EnCompass, which is based on XML, makes it possible to search through many different collections no matter what metadata scheme they use. So the library may have a collection of finding aids using EAD, of digital images using Dublin Core, and the resources in the online catalog using MARC. The user can search all three collections of records without ever having to know that they were constructed with different metadata. Using OpenURL-compliant reference linking, EnCompass can also search across multiple journal aggregator databases. EnCompass is configured by each library to provide only those sources the library has access to, although it is even theoretically possible to use Google through the EnCompass interface, provided the library sets it up under the extended services option. EnCompass truly achieves one-stop access for the user, only this time it is the catalog itself which is integrated with other resources. A library overseeing a large digitization project may find this to be one solution for providing access to all that newly digitized material. As Thomas puts it:

In the near future, however, we should expect a hybrid which will adopt some of the superior features of the catalog, but which will employ an increasingly sophisticated technological infrastructure to increase the yield for information seekers. This information management tool will have evolved from the catalog and will be influenced by what we today call the portal.

I believe that EnCompass is such a hybrid. In fact, depending on the library’s needs, integrating some electronic resources into the catalog and using EnCompass to integrate the catalog with other electronic resources could very well be the best method of all for providing seamless, integrated, one-stop access to electronic resources. To our users, it will be the catalog of the future.

Notes


11. OCLC Connexion web site <http://www.oclc.org/connexion/>


15. Endeavor EnCompass web site <http://encompass.endinfosys.com/>