Meta-What? : Metadata and Information Management For School Library Media Collections

Allison G. Kaplan
University of Delaware

Introduction

Perhaps this paper is better titled, "Metadata; or, bringing them back to the card catalog," because that is exactly what we want to do.

Think about the old days, the days before the technological advances of computer systems. Think about the hours of working with students to help them find the information they wanted by looking through the card catalog. Think about the extreme satisfaction on the part of both student and librarian as the desired book is located by using good library skills.

Things are different today. Although teaching students how to use the catalog is no less important now than it was in the pre-computer days, computer technology affects not only how our students get to the books in our libraries but also how they get to electronic information. Many of us have electronic databases that now comprise the first line of attack for research questions. However, too often the second (sometimes even the first) line of attack is "THE WEB." Internet access to information has lured students away from beginning first at the catalog despite our best efforts. As long as the catalog remains a secondary resource, no matter how many books are required for an assignment, students will not look at it as a gateway to information. We must make the catalog useful again; we must make it as alluring as the Internet. We must make it so that the OPAC (Online Public Access Catalog) itself takes the students beyond the library walls so that they no longer need to look first in the OPAC, then to the bookmarks, then to the databases, and then to the Internet. We must make the catalog the place for one stop information shopping. To do that, we must include in the OPAC records for electronic information packages.

The purpose of this paper is two-fold. First, is to familiarize the reader with some advances in cataloging that may not be so esoteric in the near future for school librarians. Second, is to provide hints for easy cataloging of traditionally uncataloged electronic information and thus enhance information access through the school library OPAC.

Information Packages and Metadata

In the beginning, we cataloged books, and only books. Then we started to include other items in our catalogs such as filmstrips and sound recordings. Catalogers today us the phrase "information package" to refer to anything being cataloged be it a book, a puppet, or even part of a web page. It is simply easier to say that we catalog information packages than it is to say we catalog all of these separate things. In fact, the phrase information package goes well with attempts to get catalogers away from thinking about cataloging types of items and into thinking about simply organizing information. If we think only about information organization, perhaps it will make the prospect of cataloging web sites a little less scary.

If organizing information is in fact our goal, then we should be interested in organizing all types of information. An informal survey (Kaplan 2004) of members of the American Association of School Librarians revealed that, although librarians thought that electronic information was an important part of the research process and that metadata, if not now, would soon be an important issue for school libraries. Despite these facts, on the whole, electronic information (web sites) was generally not included in their catalogs. There were two major factors for not cataloging web sites: 1.That web site URLs change all the time; and 2. There wasn't enough time to catalog web sites. With respect to changing URLs, I am reminded of the arguments given for not cataloging paperback books because they don't stay on the shelf very long. This may have been a viable argument in the days of hand cataloging but as automation took over and cataloging became a speedier process more and more librarians happily cataloged their paperback books. Time is always a factor in everything we do but it seems to be especially important in cataloging. With almost anything else to do, school librarians simply do not have the time to catalog information packages, especially those items that seem to be difficult to catalog. Not so long ago, those difficult materials were videotapes. No one wanted to catalog them because they were too hard to catalog and there was no time to do it. As we were once reluctant to catalog videotapes we are now reluctant to catalog web sites. As we once found that suddenly we could no longer remember all of the videotapes in our collections, we will

soon see that we will be unable to remember all of the sites for which we have created path finders or bookmarks. It is impossible to catalog all web sites but then we don't need to catalog all web sites. What we need to do is to have control of a few key sites that will help our students retrieve the information they need. As we select the books we have on our shelves, so too can we select the web sites we have in our catalogs. But first we must understand the nature of this electronic beast.

Metadata is, literally, data about data. In some libraries, we have been cataloging metadata for years. Creating a record for a part of a book, rather than the entire book itself, that is cataloging metadata. The *Reader's Guide to Periodical Literature* is an example of metadata cataloging. The way we retrieve clip art for our PowerPoint presentations is also a result of metadata cataloging. In today's cataloging world, metadata is usually used to refer to electronically based information most specifically that which is found on the Internet but it can be any digital information (e.g. clip art in a file). It is helpful to catalog, or in some way attempt to control, this information precisely because it is so seemingly uncontrollable. With declarations that AACR (Anglo-American Cataloguing Rules) and MARC (MAchine Readable Cataloging) are too oriented to book cataloging and not flexible enough to deal with the idiosyncrasies of electronic information packages (Tennant, 2004), information managers have worked hard to develop more flexible means of organizing electronic information. Standards, such as SGML (Standard Generalized Markkup Language), XML (e-Xtensible Markup Language), and TEI DTD (Text Encoding Initiative Document Type Definitions), have been developed to create a way for catalogers to transfer data about the information package into a syntax understood by computer programs for information retrieval purposes. Actually, MARC is an example of such a communication tool however archaic it may appear to be to cataloging theoreticians today.

It is exciting to see the changes in cataloging theory however; the practice of cataloging in school libraries is really dictated more by the behavior of the automated library systems than it is by theoretical advances. Therefore, in school libraries, at least for now, we are left to using the MARC format for information retrieval and AACR for information organization. For most of the general school purposes, I do not think that using MARC and AACR is such a bad idea. While it may be that in the not so distant future, our school library automation programs will accommodate more complicated cataloging formats, for the time being, it appears that the rumors of the demise of MARC are greatly exaggerated.

OPAC vs. Internet

The question you may be asking is, "Why mess at all with metadata? Why not depend on search engines, integrated databases, and such for access to electronic information packages?" The answer simply is that the catalog was developed on the basic premise of matching the user to the needed information. The Internet was not built on that premise. Also, we want students to know the whole array of information available to them, not just the web sites. But, students do not want to look first in the OPAC for the books and then on the Internet for electronic information. And who can blame them? We cannot possibly hope to catalog all Internet information, or even a fraction of it. But then, do students really need to get all of the information? Have they ever had access to all information? No! Our size and budgets have made it necessary to pick and choose what we have on our shelves. Similarly, our time, the reliability of the site, and the focus of our curriculum will dictate what Internet information we do and do not add to our catalogs. The point is that if the information is on the catalog then the students need to search only one place and will then spend a lot less time aimlessly surfing the Internet for information. Additionally, by cataloging web resources, we apply our techniques of collection development and avoid erroneous web sites that will lead students astray or, at best, give them poor information. Allow me to focus on one site as an example to illustrate my point.

Case in point

Mrs. Brown's eighth grade science class is studying the human body. In small groups, students are to find print and non-print information about various parts of the body. Group A has to find out about the skull. One question is how many bones there are in the skull. The students jump onto the computers and search using the phrase, "bones in skull," and the engines Google, Dogpile, and AskJeeves. One student looks on an online encyclopedia and one student looks for books on the school OPAC. The student who searches Google (http://www.google.com), comes up with over 400,000 sites. After reading through three pages of results, she finally comes up with two potential sites with information: one is a free encyclopedia (http://encyclopedia.thefreedictionary.com/skull) that has some good information and a black and white

drawing the students might be able to use. While she's there, she sees if she can link to the Classmates.com site that is advertised at the top of the page. The other site she finds a not very useful definition for bones from a Webster's online site (http://www.webster-dictionary.org/definition/bones). Nearly every other site listed is about some organization known as Skull and Bones and so, by the third page, she's had enough. The second student searched using Dogpile (http://www.dogpile.com). This student is buoyed by the fact that only 75 sites were retrieved, a much better hit than the 400,000 from Google. Plus, Dogpile offers a suggestion box and the student clicks on "What are the bones for the human skull?" This "refinement" yields two additional sites (77) but at least there are very few on the first page having to do with that Skull and Bone society. Dogpile is a more successful search with more relevant sites right on the first page. One site in particular, Skull Anatomy Tutorial from GateWay Community College in Phoenix, Arizona (http://www.gwc.maricopa.edu/class/bio201/skull/skulltt.htm) seemed to have the best and most easily understood information and it was second on the list so it was found fairly quickly. The student looking on AskJeeves, ended up with over 300,000 sites with the sponsored commercial sites listed first. He too made use of the refinement feature, "Bone Human Skull" and thus narrowed his search to 127,000 sites. Skipping the e-bay and Skull and Bones sites, he too ran into the skull tutorial as well as a very cool looking site from KidzWorld (http://www.kidzworld.com/site/p922.htm). Now the students have a problem because one site says there are 22 bones in the skull and another says there are 30. A search on World Book Online brought up 49 references that were so diverse, the students changed the search to just "skull." This time their search was much more productive. They found an article that confirmed there are 22, not 30, bones in the skull and they have a nice graphic to go with this finding. Finally, they searched the OPAC and found one book on the skeletal system. By now the students had really had enough. Their Internet searching was fun but in all the time they spent looking at the wrong sites, they found two useful sites but with conflicting information. Additionally, they has to sift through a lot of commercials to get to the helpful sites and one student thinks she is now signed up to get email for Classmates.com. They completely forgot that the librarian had a pathfinder for Mrs. Brown's science class and so they didn't get to the sites that way.

What could be done to help these students? I suggest that cataloging the useful sites instead of putting them on yet another place for students to go to is a good solution to this surfing problem. If these students only had to go to the OPAC to get to book and electronic information instead of searching in all of these different places, they might have more energy to complete their assignment. So now the librarian has to deal with metadata. However, in a way, he or she has already done that by creating pathfinders or bookmarks. Some bookmarks are more elaborate than others. Peter Milbury of Chico (California) High School has an extensive web site of helpful sites for his students (http://dewey.chs.chico.k12.ca.us/) (see figure 1). He has made use of HTML to organize his sites. Students in his school do not have to surf the Internet to find useful information, yet, they still have to search for books in a separate step using the school's OPAC. Since XML and HTML are ways of dealing with metadata, the website set up by the librarian has already set him up for the next step, that of taking the web sites off of the pathfinder or bookmark and into the OPAC.

Cataloging hints

Cataloging not just the parent site but also sites within the parent site is not a difficult matter. Most automated programs have a template option for electronic resources. What we need to do is focus on the positive aspects of making these resources available to our students from a single source (the catalog) rather than from multiple sources; bookmarks, subject pathfinders, or search engines. Yes we are dealing with some animal that the academics call metadata, but to our students we are just making a site accessible through the OPAC.

A few caveats: Be reasonable about the sites you select to catalog just as you are reasonable about those you include in your bookmarks. Government and education sites tend to have more stable addresses than do commercial sites. If changing URLs is a concern, stick to sites you are fairly certain will maintain current sites and stable locations. Most library automation programs allow the cataloger to select the type of information package being described. This is an important feature and should not be ignored in any kind of cataloging. Before beginning the cataloging process, be sure the record being created is for electronic resources. In cataloging electronic information packages, Taylor (2004) states that it is not reasonable to refer to a bibliographic record, the problem stemming from the root of the word, *biblio-*, meaning "book." She suggests instead the use of the phrase, "surrogate record" to refer to the record that stands in the place of the item itself. Here the item can be anything, a book, film, or web site.

I have often been asked how to catalog a videotape that contains two separate programs. Happily, in cataloging metadata that is no longer a problem. We can catalog the main NASA page as easily as we can catalog one of the links with in it. Because we don't have to worry about where the item will be located on the shelf or whether there will be duplicate call numbers, we are freer to catalog with more detail. Let us use the NASA site as a template for cataloging metadata using the MARC communications format.

Like a book, our surrogate record begins with the 245 tag. Consider the home page for NASA (see figure 2). Notice the prominence of the words, "National Aeronautics and Space Administration" with the NASA logo adjacent. We will take that as our main entry. Like most films, most web sites will not have personal or corporate names for main entries. We might want to consider the NASA logo as other title information. To include National Aeronautics and Space Administration as a statement of responsibility is redundant. Thus our 245 will look like this:

> 245 00 |aNational Aeronautics and Space Administration |h[electronic resource] :|bNASA.

We will also want an additional title entry for NASA:

> 246 30 |aNASA

Moving down the MARC record, we need to describe the publication data. A small variation from book cataloging is the reference to date of publication. If it is clear (or known) when the site came online, include that information in the |c of the 260 tag. If not, the |c is left blank.

> 260 __ |aWashington, D.C. :|bNASA.

Since the information package does not exist physically anywhere, there is no 300 tag. In some school library automation programs, this may be a problem. Some program demand information in the 300 tag or the surrogate record cannot be saved. If this is the case, one might just add a descriptor such as "1 web site" in the |a to take care of the problem, even though it is not technically correct. Sometimes, due to the inflexible nature of most school library automation systems, we have to be a little flexible ourselves. So if possible, we will skip the 300 tag and go on to the 5XX tags. The first tag to enter is the 538 for access information. Because we are focused only on Internet sites, our entry is simple:

> 538 __ |aMode of access: Internet.

The second 5XX tag is a 500 for source of title. Included here is the screen used for deciding on a title and when that screen was accessed. We are picking out the most stable site we can but even stable sites change their look now and then. Therefore it is important to inform the user when the site was accessed just in case we may need to account for changes in site information. We also include for the user's reading pleasure, the date the site was last updated. This can be very important information to the user. One wants to make sure the site is viable and wasn't just put up and forgotten. In our example site, we see that NASA is very good at keeping the site current.

> 500 __ |aTitle fromhome page (viewed on Oct. 13, 2004; last updated Oct. 13, 2004).

In cataloging web sites, a summary statement is particularly important, especially if the site contains a variety of information. Many educational and government sites have descriptions of the objectives and goals of the site. These statements, often found in the "about us" links, can be copied and pasted into the 520 tag. One should look for keywords that will aid in the retrieval of the surrogate record. One might also include a 505 contents note that maps out the parts of the site.

> 520 __ |aOfficial site of NASA, with extensive links to NASA projects and information about the United States space program.
>
> 505 0_ |aLife on Earth – Humans in space – Exploring the universe.

The subject area is where we really see the benefits of cataloging parts of the site separate from each other. Imagine the number of subject headings one would need to cover even half of the content of all of the sites put together. But, if the parts are cataloged separately, then a few broad headings are sufficient in describing the intellectual content of the parent page.

> 610 10 |aUnited States. |bNational Aeronautics and Space Administration.
>
> 650 _0 |aSpace flight |xHistory.
>
> 650 _0 |aSpace shuttles |zUnited States |xHistory.

The final part of the surrogate record is the 856 tag, Electronic Location and Access. As the title describes, this is where one enters the URL for the site. The first indicator will be 4; meaning that the information package being described is accessed through http. The second indicator is the relationship of the URL to the item as a whole. A value of 0 means that the item being described is the item at the URL. A value of 2 means the URL is for a related resource. If the second indicator is 2, there must be a

corresponding |3 to explain the relationship. To avoid mistakes in the URL, whenever possible, one should use copy and paste to enter the URL information.

856 40 |uhttp://www.nasa.gov/home/index.html

856 42 |3Spanish version |uhttp://www.nasa.gov/about/highlights/En_Espanol.html

Let us look at one part of the NASA site and what that surrogate record would look like. Suppose that every year, Mrs. Brown had her students investigate the planets. A site on the planet Mars would be particularly helpful, especially if the source was as authoritative as NASA. Figure 3 shows the front page of the Mars site and below is the full record. Notice the extended 245 tag: we must account for the parent site as well as the titles for the site itself. One of the problems with cataloging a part of a web site is deciding how to deal with the separate title. This is easily accommodated by using the |p for the part of the title. Notice too that a corporate entry has been added in the 710 tag for the Jet Propulsion Laboratory. Since JPL is part of the URL, the cataloger decided to make an entry for that entity.

245 00 |aNational Aeronautics and Space Administration |h[electronic resource] :|bsolar system exploration. |pPlanets. |pMars.

246 30 |aSolar system exploration

246 30 |aPlanets

246 30 |aMars

260 __ |aWashington, D.C. :|bNASA.

538 __ |aMode of access: Internet.

500 __ |aTitle taken from home page (viewed Oct. 13, 2004; last updated Jan. 9, 2004)

520 1_ |aPart of the Planets section of the NASA web site, this page focuses on the planet Mars with links to the Mars exploration and photographs.

505 0_ |aOverview – Moons – Gallery – Facts and figures – Kid's eye view – Resources.

610 10 |aUnited States. |bNational Aeronautics and Space Administration.

650 _0 |aMars (Planet)

650 _0 |aPlanets

650 _0 |aSolar system.

650 _0 |aAstronomy.

710 2_ |aJet Propulsion Laboratory (U.S.)

856 40 |ahttp://solarsystem.jpl.nasa.gov/planets/profile.cfm?Object=Mars

Conclusion

We have seen that it is not really so tough to catalog a web site. Indeed the benefits of having students rely mostly on the OPAC for information access truly outweighs the beginning difficulties of getting used to a new cataloging template. There are some vendors who offer monitoring programs for schools with sites on their catalogs. Programs such as Follett's "Monitor" allow the librarian to register sites with Follett. Follett then monitors the site and notifies the librarian of any URL changes.

It would be great if one day a student could type in "skull" and be lead to books, Internet sites, and database articles. Until such a time, the least we can do is connect our books with good Internet sites. The way to do that is by cataloging the sites that we already have bookmarked. Whether one calls this cataloging metadata or web sites, the point is that the job gets done and that we pull our students out of the Internet Ocean and back to the library catalog for one stop information access.

References

Kaplan, Allison G. (2004) Survey of members of the American Association of School Librarians discussion list for   the Affiliate Assembly.

Tennant, R. (2004). *Building a New Bibliographic Infrastructure*, from http://www.libraryjournal.com/article/CA371079

Taylor, A. G. (2004). *The Organization of Information* (2nd ed.). Westport, Conn.: Libraries Unlimited.