Organizing the World Wide Web:

Traditional and New Approaches

Carrie Preston

SI 662

December 2000

Organizing the World Wide Web:

Traditional and New Approaches

In the past, discussions of cataloging Internet resources have focused on whether it could or should be done at all (Jul, 1996). Arguments against cataloging Internet resources included the contention that "nothing on the Internet is worth cataloging," that Internet resources are too vague and ephemeral for accurate cataloging data to be maintained, and that "cataloging the Internet" would require too much time and effort.

Currently, though cataloging the *whole* Internet is not seriously considered, the need to catalog selected Internet resources is recognized. However, once the notion that "nothing on the Internet is worth cataloging" is dismissed, the other problems discussed above remain. Internet resources are indeed vaguely defined; their boundaries are difficult to mark and they may provide minimal or conflicting information about their creation and ownership. This is particularly true of non-commercial World Wide Web resources; created and maintained by individuals rather than more-or-less-stable corporations, they are even more likely to have loosely defined boundaries, changing locations, and vague or nonexistent descriptive data. And cataloging is expensive in terms of time, effort and monetary resources, raising the question of whether it is worth the effort to catalog ill-defined and unstable information resources. Nonetheless, some non-commercial World Wide Web resources provide excellent content and so might be considered to "deserve" the reliable access and historical record provided by cataloging.

These problems have led to the re-evaluation of traditional library cataloging approaches in relation to Internet resources. They have also led to the development of new, custom-made approaches to organizing these resources. The newer approaches are

intended to provide appropriate cataloging schemes for Internet resources (as opposed to schemes originally designed for books) and to remove the burden of effort from cataloging professionals by allowing document creators to provide their own metadata. At the same time, new approaches tend to borrow from or incorporate features of older library cataloging and classification schemes.

This paper compares and contrasts traditional and modern approaches to cataloging Internet resources, with a focus on the problems of cataloging non-commercial World Wide Web sites as opposed to commercially produced "online" resources. The discussion will be focused mainly on one traditional cataloging approach (AACR/MARC) and one modern approach (Dublin Core Metadata), with mention of other standards and schemes where appropriate. I will begin with an overview of each scheme (in the case of AACR/MARC, focusing on the features specific to cataloging of Internet resources). I will then discuss traditional and new approaches in relation to four major problems involved in organizing the World Wide Web: the problem of defining and describing non-physical items with indistinct boundaries; the problem of reducing the professional time and effort necessary to organize the Web; the problem of providing reliable subject access to the growing body of Web resources; and the problem of "locating" and tracking non-physical resources over time.

Two Approaches to Organizing the Web:
Traditional Cataloging versus Metadata Markup

Traditional Approaches

Cataloging. For most American libraries today, the "traditional" approach to cataloging involves use of the Anglo-American Cataloging Rules (AACR) (Joint Steering

Committee for Revision of AACR, 1998) encoded in Machine Readable Cataloging format (MARC) (Online Computer Library Center, 2000a). Internet resources are a subset of "Computer Files" and so fall within the domain of Chapter 9 of the current edition of the Anglo-American Cataloging Rules. AACR provides for the same seven general areas of description allotted to books and other resources, plus an additional File Characteristics Area (to be discussed below).

The examples of computer file cataloging data offered in the current edition of AACR tend to refer to physically contained, "local" computer resources (those stored on floppy disks, CD-ROM, etc.). However, the Online Computer Library Center publishes (on the web) a cataloging manual specific to Internet resources (Olson, 1997). This manual contains examples and specific recommendations for interpreting and expanding the AACR rules for computer files, as well as specifying which MARC tags should contain various pieces of information.

Subject assignment and classification. Traditional library approaches to subject assignment and classification include controlled vocabularies, such as Library of Congress Subject Headings (Library of Congress, 2000b), and classification schemes such as Library of Congress Classification (Library of Congress, 2000a) and Dewey Decimal Classification (OCLC Forest Press, 2000). Application of these schemes to World Wide Web resources will be discussed later in this paper.

Newer Approaches

Newer approaches to "cataloging" or providing access points to web pages involve markup languages and "meta tags." Information is applied directly to a web resource, enclosed within programming tags that define the information as bibliographic

and nature and may also specify the type of information (title, author, subject access point, etc.).

This paper will focus on the Dublin Core Metadata scheme, one of the more well-known and precisely defined standards for applying meta tags to web documents. The Dublin Core is designed to allow creators of web pages to apply access points to their documents, providing an alternative form of access to the statistical keyword searching provided by web search engines. It is less a cataloging standard in itself than a framework, like MARC, which allows cataloging data to be entered in a form conforming to AACR2 or other cataloging standards--or, conversely, according to loose or no standards, thus allowing its use by non-professionals. (The tension between standardization and ease of use in metadata schemes will be discussed later.)

I have provided examples of both AACR/MARC and Dublin Core cataloging records with this paper. The MARC examples were created based on the recommendations of the OCLC manual, and are limited to the areas of cataloging discussed in this paper (bibliographic description and subject assignment), omitting fixed fields, numbers and codes. The Dublin Core examples were created using the best practices recommended by the official guide to using Dublin Core. The examples are intended to demonstrate the "basics" of the cataloging schemes presented and thus were created using relatively simple World Wide Web resources of non- commercial origin: a library school class project consisting of a paper and brief online tutorial; a Javascript-based random number generator with accompanying documentation; and an archive of guitar music in ASCI-based "tablature" format.

Figure 1. MARC record, "A Web-Based Tutorial on OPAC Searching..."

```
10   1   Almy, Gretchen.
0
24   12  A web-based tutorial on OPAC searching with Library
5        of Congress Subject Headings and free text keywords
         : $b final project report / $c Gretchen Almy... [et
         al.]
24   1   Finding what you need : $b a guide to searching
6        MCAT and locating books on the shelves
25       Computer documents.
6
27       prestonc@umich.edu
0
50       Title from research paper.
0
50       Title on tutorial: Finding what you need : a guide
0        to searching MCAT and locating books on the
         shelves.
50       Authors/Creators: Gretchen Almy, Sally Horvath,
0        Betsy Paulson, Carrie Preston.
50       Files in HTML.
0
50       "SI 666 Winter 2000"
0
50       Graphics display in color.
0
50       Descriptive research paper with embedded links to
0        web-based tutorial.
50       Includes bibliographical references.
4
52       Paper describing and demonstrating a web-based
0        tutorial developed to aid undergraduate students in
         searching an online library catalog.
53       System requirements: Internet access, graphical
8        World Wide Web browser.
53       Mode of access: World Wide Web.
8
65    0  Online catalogs $x User education.
0
65    0  Subject headings, Library of Congress $x Study and
0        teaching.
65    0  MIRLYN (Computer file) $x Study and teaching.
0
85   40  $g http://www-
6        personal.umich.edu/~prestonc/tutpaper.htm
```

Figure 2. Dublin Core markup, "A Web-Based Tutorial on OPAC Searching..."

<META NAME="DC.Title" CONTENT="A web —based tutorial on OPAC searching with Library of Congress Subject Headings and free text keywords: final project report">
<META NAME="DC.Title.Alternative" CONTENT="Finding what you need: a guide to searching MCAT and locating books on the shelves">
<META NAME="DC.Creator" CONTENT="Almy, Gretchen">
<META NAME="DC.Creator" CONTENT="Horvath, Sally">
<META NAME="DC.Creator" CONTENT="Paulson, Betsy">
<META NAME="DC.Creator" CONTENT="Preston, Carrie">
<META NAME="DC.Subject.LCSH" CONTENT="Online catalogs--User education.">
<META NAME="DC.Subject.LCSH" CONTENT="Subject headings, Library of Congress--Study and teaching.">
<META NAME="DC.Subject.LCSH" CONTENT="MIRLYN (Computer file)--Study and teaching.">
<META NAME="DC.Subject" CONTENT="online catalogs, OPAC, library catalogs, user education, searching, keyword searching, LCSH, Library of Congress Subject Headings, MIRLYN, college students, undergraduates, undergraduate students, tutorial, web-based tutorial">
<META NAME="DC.Description" CONTENT="Paper describing and demonstrating a web-based tutorial developed to aid undergraduate students in searching an online library catalog.">
<META NAME="DC.Date.Created" CONTENT="2000">
<META NAME="DC.Type.DCMIType" CONTENT="text">
<META NAME="DC.Type.DCMIType" CONTENT="interactive resource">
<META NAME="DC.Format.IMT" CONTENT="HTML">
<META NAME="DC.Identifier.URL" CONTENT="http://www-personal.umich.edu/~prestonc/tutpaper.htm">
<META NAME="DC.Language.RFC1766" CONTENT="en-us">

Figure 3. MARC record, "Research Randomizer"

```
10   1   Urbaniak, Geoffrey C.
0
24   12  Research randomizer : $b instant random sampling
5        and random assignment / $c by Geoffrey C. Urbaniak
         and Scott Plous.
25       Computer data and program.
6
27       gurbaniak@wesleyan.edu
0
50       Files in HTML with JavaScript.
0
50       "Copyright 1997-2000"
0
50       Graphics display in color.
0
50       Random number generator with documentation.
0
52       Programmable random number generator with
0        documentation, including a tutorial on random
         sampling methods.
53       System requirements: Internet access, graphical
8        World Wide Web browser with JavaScript support.
53       Mode of access: World Wide Web.
8
65    0  Random number generators.
0
65    0  Sampling (Statistics).
0
70   1   Plous, Scott.
0
85   40  $g http://www.randomizer.org
6
```

Figure 4. Dublin Core markup, "Research Randomizer"

```
<META NAME="DC.Title" CONTENT="Research randomizer: instant random
sampling and random assignment">
<META NAME="DC.Creator" CONTENT="Urbaniak, Geoffrey C.">
<META NAME="DC.Creator" CONTENT="Plous, Scott">
<META NAME="DC.Subject.LCSH" CONTENT="Random number generators.">
<META NAME="DC.Subject.LCSH" CONTENT="Sampling (Statistics).">
<META NAME="DC.Subject" CONTENT="random number generator, random
sampling, random assignment, statistics">
<META NAME="DC.Description" CONTENT=" Programmable random number
generator with documentation, including a tutorial on random sampling methods.">
<META NAME="DC.Date.Created" CONTENT="1997">
<META NAME="DC.Date.Modified" CONTENT="2000">
<META NAME="DC.Type.DCMIType" CONTENT="interactive resource">
<META NAME="DC.Type.DCMIType" CONTENT="text">
<META NAME="DC.Format.IMT" CONTENT="HTML">
<META NAME="DC.Identifier.URL" CONTENT="http://www.randomizer.org">
<META NAME="DC.Language.RFC1766" CONTENT="en-us">
```

Figure 5. MARC record, "R.E.M. Guitar Archive"

```
10   1   Bray, Chris.
0
24   10  Welcome to the R.E.M. guitar archive / $c created
5        and maintained by Chris Bray.
24   3   R.E.M. guitar archive
6
24   1   R.E.M. chord archive
6
25       Computer documents.
6
27       cbray@comp.uark.edu
0
50       Title from home page graphic.
0
50       Title from title bar: R.E.M. chord archive.
0
50       Files in HTML.
0
50       "Since September 1994"
0
50       Graphics display in color.
0
50       Includes guitar music tablature.
0
52       Archive of guitar music for 15 record albums by the
0        band R.E.M., in tablature format.
53       System requirements: Internet access, graphical
8        World Wide Web browser.
53       Mode of access: World Wide Web.
8
61   20  R.E.M. (Musical group).
0
65    0  Guitar music.
0
85   40  $g http://bubblegum.uark.edu/rem/
6
```

Figure 6. Dublin Core markup, "R.E.M. Guitar Archive"

```
<META NAME="DC.Title" CONTENT="Welcome to the R.E.M. guitar archive">
<META NAME="DC.Title.Alternative" CONTENT="R.E.M. guitar archive">
<META NAME="DC.Title.Alternative" CONTENT="R.E.M. chord archive">
<META NAME="DC.Creator" CONTENT="Bray, Chris">
<META NAME="DC.Subject.LCSH" CONTENT="R.E.M. (Musical group).">
<META NAME="DC.Subject.LCSH" CONTENT="Guitar music.">
<META NAME="DC.Subject" CONTENT="R.E.M., guitar music, guitar chords, TAB,
tablature">
<META NAME="DC.Description" CONTENT="Archive of guitar music for 15 record
albums by the band R.E.M., in tablature format.">
<META NAME="DC.Date.Created" CONTENT="1994">
<META NAME="DC.Date.Modified" CONTENT="[1998]">
<META NAME="DC.Type.DCMIType" CONTENT="collection">
<META NAME="DC.Type.DCMIType" CONTENT="text">
<META NAME="DC.Format.IMT" CONTENT="HTML">
<META NAME="DC.Identifier.URL" CONTENT="http://bubblegum.uark.edu/rem/">
<META NAME="DC.Language.RFC1766" CONTENT="en-us">
```

Problems in Cataloging World Wide Web Resources:

Traditional and New Approaches

Defining and Describing the "Item at Hand"

As Taylor (1999) notes in relation to subject analysis, Web resources present a problem when it comes to defining the item to be described. "On the Internet there is so far no definition for an analyzable unit. Should it be the whole electronic journal; individual issues of that journal; the whole web site...or individual pieces of a web site?" (p. 137). Even when one decides to treat "the whole web site" as a unit, when following hypertext links it is often difficult to tell when one has in fact left the site; if site authors have used inconsistent design, it may not even be possible to determine whether one is still on a page created and maintained by the same responsible agent. Basic descriptive tasks like determining the title proper can be complicated by the presence of different heading, graphic, and title bar titles. Unlike books, Web resources do not have standardized "title pages," and non-commercial sites, even those with excellent content, may provide conflicting or no information to fill basic descriptive slots like title, author, and date of creation.

The nebulous, loosely defined, interconnected nature of Web resources creates problems for basic item description. AACR2 and OCLC recommend eschewing Field 260, the physical description, "because there is no physical item being cataloged" (Olson, 1997). However, just as a book has a page length and a 500-page book means something very different to a user from a 30-page book, Web resources do have "size" and other important characteristics in terms of information content. While traditional cataloging methods work well for recording such information as title, author, and subject content

(where these things are determinable from the item), current traditional methods are not as proficient in giving users the information they need in order to decide whether or not they can or want to use a resource, *after* they have discovered the resource via the above access points.

For AACR/MARC cataloging, this type of description starts with the File Description Area, MARC field 256. Developed for physically contained computer files (and not necessarily all that helpful for them), the file description area seems to add little useful information to records for Internet resources. The area is divided into two parts, the "designation" or type of file and the "number of records, statements, etc." AACR2 offers only two choices for the designation: "computer data" and "computer program(s)" (and combinations of the two); from the user perspective this distinction is almost meaninglessly broad and does not paint a picture about what the resource is like. The second part of the area allows for an expression of file size, such as "1 file : 83565 bytes" (example from Olson, 1997), but as noted above, it is often difficult to determine which files (e.g., web pages) are a part of a given resource (e.g. web site). Even if it is possible to determine the boundaries of a resource, counting the number of files can be complicated. In a web site, for example, do only the HTML pages count as files? What about any graphic images displayed with the text on those pages? Graphical images are indeed separate "files" on the web server, but they tend to have large file size in relation to their informational contribution and can inflate a byte total made by adding the sizes of various files. Other measures of resource size, such as the number of printout "pages," are also problematic because they vary depending on the size of the monitor and the font settings of the web browser.

OCLC improves on the designation area by suggesting that catalogers choose among 30 diverse and specific ISBD(ER) designations, such as "Computer bibliographic database," "Computer sound data," "Computer spreadsheet program(s)" and "Computer game(s)" (Olson, 1997). This makes the file characteristics area more informative to the user; even the distinction between "Computer document" and "Computer documents" might bring to mind a single web page versus a site with multiple pages. It is uncertain whether the web will remain in its current form long enough for "Web page" and "Web site" (or some equivalents) to become established as designations in their own right. Beyond this broad distinction, giving users an idea of the amount of information contained within a resource is still a problem, and is likely to remain so. Good summary notes can help to give users a more accurate idea of the amount of information available (for example, in my summary for the rock guitar music archive, I state that the archive contains music from 15 full albums).

Other important characteristics of Web resources are consigned to the note area. Field 538, "system requirements," allows a statement of what hardware and/or software are required to use the resource and of the mode of access, giving users a specific field to look at for this important information. However, the file format (HTML, PDF, RTF, etc.) and the presence of illustrations, maps, etc. (normally covered in the physically description area) are consigned to generic 500 notes, making the information more difficult to locate efficiently. In addition to adopting the ISBD file designations, AACR might consider allowing a version of the physical description area (that dealing with illustrations, etc.) to be utilized for non-physical resources, or to expand the File Characteristics area to include this information.

In some areas, newer approaches such as Dublin Core are only slightly better at providing users with practical information about Web resources. The file format is given its own element, "Format," with allowance for specifying file size, and another element, "Type," resembles AACR's General Material Designation. However, the problems with specifying file size noted above still apply, and there is no specific provision (not even in a note) for the presence of illustrations and other extras. The Dublin Core scheme also lacks a generic "notes" field for miscellaneous content. Good "Summary" content is therefore even more important here than in traditional cataloging.

On the other hand, the Dublin Core does make provisions for some Web-specific description problems, such as allowing, with the use of qualifiers, for separate "date created" and "date altered" fields. Future development of Dublin Core and the qualifiers may lead to further refinements of this sort.

Saving Professional Time and Effort

There is no denying that AACR/MARC cataloging is time consuming and expensive in terms of professional attention, and some may particularly resent devoting this effort to Web resources that are not directly controlled by the cataloging institution and may not have a stable existence (Jul, 1996). Newer approaches, particularly the Dublin Core, have therefore made an attempt to shift the burden of cataloging from professionals to the creators of Web resources. Basic Dublin Core is intended to be easy enough for nonprofessionals to apply to their own creations (Weibel, 1996), and its simplicity has often been attested (e.g. Desai, 1997; Medeiros, 1999). Creators of HTML web pages may already be familiar with generic HTML META tags, making for a relatively short step up to Dublin Core metadata.

However, there is a trade-off here between simplicity and flexibility in the cataloging scheme and its utility for locating and describing resources. Dublin Core has no set formats for the content of most elements, only "best practice" recommendations (Dublin Core Metadata Initiative, 2000a). While the "basics" are indeed easy to learn, the best practices and element qualifiers do take time and effort which web authors will be unwilling to devote. And it is the best practices and qualifiers, which often emphasize some form of bibliographic control, that greatly increase Dublin Core's value as a retrieval and description tool. Uneducated users may not realize the importance of simple rules such as entering "Creator" names in a consistent format, and they are certainly unlikely to fill "Subject" elements with Library of Congress Subject Headings. Where content is entered into Dublin Core elements in an uncontrolled manner, only "keyword within field" searching is possible, since consistent whole-field formatting cannot be assumed. Thus, unless "the masses" of document creators make a great effort to learn and conform to Dublin Core best practices, the possible contribution of widespread Dublin Core usage to general web search and retrieval will be reduced.

Within controlled cataloging environments, however, Dublin Core can (at least theoretically) be utilized to its full potential. Cooperative cataloging programs can also help lessen the burden of cataloging on individual institutions. For example, OCLC's Cooperative Online Resource Catalog (CORC) combines records produced by various member libraries in a single database (Online Computer Library Center, [2000b]). CORC records are created with access to OCLC's Authority File, offering a source of authority control, and searching the database for pre-existing records helps reduce duplication of effort. Software tools such as Nichols Advanced Technologies' MARCit

(Breeding, 1998) have also been developed to "read" data directly into MARC fields from a web document, creating at least a basic starting record with reduced professional effort.

Another potential source of effort duplication is the creation of MARC records and Dublin Core markups for the same resources (Medeiros, 1999). Therefore, efforts have been made to create methods of transforming MARC records into Dublin Core records, and vice versa. The Library of Congress has compiled a Dublin Core to MARC crosswalk (Library of Congress, 1999) to aid translation of Dublin Core metadata into AACR2/MARC format. Dublin Core's fields are generally less precisely defined than MARC's, so reverse translation of Dublin Core is problematic, but with careful use of Dublin Core qualifiers and best practices at least a rough translation in this direction can be achieved. OCLC's CORC database stores records in both formats (Online Computer Library Center, [2000b]).

Providing Subject Access

One area in which traditional library schemes may be particularly helpful is subject access. "Modern," commercially developed approaches to web subject access generally involve feeding keywords into more or less complex statistical retrieval equation (as in Web search engines) and/or organizing resources under a hierarchy of often loosely defined and overlapping subject categories (as in directories like Yahoo!). These systems have a reasonable success rate in delivering users at least a few resources of the type they are looking for (Drabenstott, 1999). However, search engines tend to have very low precision and force users to wade through large "junk" returns, while click-through subject hierarchies become unwieldy and confusing when large numbers of

documents are added. Traditional library methods of providing subject access--controlled vocabularies/subject headings and number classification--may thus provide a good alternative for subject access. This is particularly for professionally maintained collections of high-quality resources (such as OCLC's CORC or Iowa State University's CyberStacks) where the time and effort needed to provide good subject access would be considered commensurate with the quality of the resources.

Best practices for Dublin Core metadata recommend using a controlled vocabulary, such as Library of Congress Subject Headings, to fill at least some Subject fields (Dublin Core Metadata Initiative, 2000a). Dublin Core qualifiers allow the cataloger to specify that a Subject field has been filled with Library of Congress Subject Headings, Medical Subject Headings, or Library of Congress, Dewey Decimal, or Universal Decimal classification numbers (Dublin Core Metadata Initiative, 2000a). Additional Subject fields may be added to contain free text keywords, for the benefit of keyword searchers and web search engines; it is recommended that Dublin Core catalogers utilize both in order to provide a range of methods of subject access (ALCTS/CCS/SAC Subcommittee on Metadata and Subject Analysis, 1999). LCC, DDC and other classification schemes may even be utilized to provide a framework for Yahoo!-type click-through subject hierarchies (e.g. Mundie, 1995).

Existing projects that utilize traditional subject access schemes in a Web environment include CyberStacks (Iowa State University Library, 1998), which uses Library of Congress Classification; BUBL LINK (BUBL Information Service, 1999) and CyberDewey (Mundie, 1999), which use Dewey Decimal Classification; and OCLC's CORC which encourages the use of Library of Congress Subject Headings in MARC and

Dublin Core records. Other groups have created their own traditional-style schemes; for instance, the Internet Public Library has created a system of matched, pre-coordinated subject headings and classification codes which serve both to provide searchable subject access and to structure the library's click-through hierarchy of categories (Internet Public Library, 2000). All these systems provide an alternative form of subject access to Web resources when search engines or unplanned, unwieldy subject directories fail to deliver the required resources.

Locating the Non-Physical Item

One problem unique to Internet resources is the problem of "keeping track" of a resource that has no physical embodiment or real location (Lynch, 1998). The Uniform Resource Locators (URLs) familiar to Web users are a method of identifying and locating Web resources, but they are unstable; the document formerly associated with an URL may move to a different location, or simply disappear, at the whim of the creator or system administrator. In order to combat this problem, the concept of the Uniform Resource Identifier (URI) has been devised (Schwartz, 1997). The URI is not a standard in itself like the URL, but a set of guidelines for standard creation, stating that an effective resource identifier must have six qualities: global scope, global uniqueness, persistence, scalability, extensibility, and independence. Schwartz (1997) directly identifies the URI idea as an attempt to bring "bibliographic control" to the Web, emphasizing the contribution of traditional library ideas to the organization of the Web.

An example of the implementation of the URI guidelines is OCLC's Persistent Uniform Resource Locator, or PURL (Shafer et al., [1996]). Basically, a PURL is a stable document identifier which can be associated with different URLs as the document

changes locations over time. Rather than transporting the user directly to a (possibly out-of-date) location, clicking a PURL causes a lookup of the resource's current address in the PURL server's database, the takes the user to the current URL. A disadvantage of this (and all existing) uniform name standard is that it depends on user maintenance of the current address in the PURL server's database; as with Dublin Core metadata, the willingness of resource creators to devote time and effort become essential.

Currently, OCLC's implementation of MARC Field 856 allows for the entry of Uniform Resource Names, even providing a specific subfield (Olson, 1997). They may also be identified in the Dublin Core Identifier field. Hopefully, URN standards will continue to develop in order to fill the place provided for them in these cataloging schemes.

Conclusions

Overall, traditional cataloging schemes have much to offer for the organization of World Wide Web resources. Modern schemes such as Dublin Core improve on traditional cataloging by providing frameworks for applying cataloging data directly to documents. However, these schemes directly depend on traditional methods such as AACR in order to provide structure for the data within the metadata fields. Traditional subject headings and classification systems also offer a possibility for high-precision subject access not possible with the statistical keyword retrieval methods currently used on the web. Without the use of traditional methods to ensure bibliographic control, newer schemes such as Dublin Core will be of limited usefulness for organizing the Web. Even the idea of the Uniform Resource Identifier, a method of tracking the Internet

"location" of a resource over time, depends on the traditional concept of bibliographic control.

Since traditional schemes were developed with books in mind, they do have certain failings in relation to Web resources. Particularly, defining and describing the item to be cataloged--the equivalent of the "physical description" of a book--is sometimes problematic using traditional cataloging methods. Newer metadata schemes such as Dublin Core, as well as traditional cataloging-focused organizations such as OCLC, have taken steps to improve this type of description, but much territory remains to be covered. If World Wide Web resources continue to be a prominent source of information in the future, traditional cataloging schemes may be updated and revised to better accommodate the description of these types of resources. New methods of tracking resources, such as the Uniform Resource Name, also contribute to the problem of identifying, defining, and "keeping track" of the non-physical document.

While cataloging schemes can be altered and updated, the problem of the "information explosion" is more difficult to solve. Metadata schemes such as Dublin Core aspire to allow document creators to make their own contributions to the organization of the Web. However, good metadata creation requires time, effort and a commitment to learning best practices which the general Web user is unlikely to undertake. Currently, cooperative projects such as OCLC's CORC provide the main solution to the problem of organizing the Web without monopolizing too much professional time and effort. The traditional document selection skills of librarians also come into play here as information professionals attempt to determine which Web

resources should receive the limited amounts of professional time and effort that are available.

As high-quality Web resources grow in number and the Web becomes a more important source of information for both academic and general users, library and information science professionals must continue to develop schemes for organizing the Web. Traditional cataloging schemes provide a solid basis and structure for evolving methods of organizing World Wide Web resources.

References

ALCTS/CCS/SAC Subcommittee on Metadata and Subject Analysis (1999). Subject data in the metadata record: Recommendations and rationale [Online]. http://www.govst.edu/users/gddcasey/sac/MetadataReport.html [2000, November 18].

Breeding, M. (1998). Nicholas Advanced Technologies, Inc. Computers in Libraries, 18(8), 26.

BUBL Information Service (1999). BUBL LINK [Online]. http://bubl.ac.uk/link/index.html [2000, December 3].

Desai, B. P. (1997). Supporting discovery in virtual libraries. Journal of the American Society for Information Science, 48(3), 190-204.

Drabenstott, K. M. (1999). Web search strategies. Unpublished.

Dublin Core Metadata Initiative (2000a). Dublin Core Metadata Element Set, Version 1.1: Reference description [Online]. http://purl.org/dc/documents/rec-dces-19990702.htm [2000, November 18].

Dublin Core Metadata Initiative (2000b). Dublin Core Metadata Initiative [Online]. http://purl.org/dc/ [2000, November 18].

Dublin Core Metadata Initiative (2000c). Dublin Core qualifiers [Online]. http://purl.org/dc/documents/rec/dcmes-qualifiers-20000711.htm [2000, November 18].

Dublin Core Metadata Initiative (2000d). Using Dublin Core [Online]. http://purl.org/dc/documents/wd/usageguide-20000716.htm [2000, November 18].

Internet Public Library (2000). Collection development at the IPL [Online]. http://www.ipl.org:2000/backroom/col/coldev.html [2000, December 3].

Iowa State University Library (1998). CyberStacks [Online]. http://www.public.iastate.edu/~CYBERSTACKS/ [2000, December 3].

Joint Steering Committee for Revision of AACR (1998). Anglo-American Cataloguing Rules: Second edition, 1998 revision. Chicago: American Library Association.

Jul, E. (1996). Why catalog Internet resources? Computers in Libraries, 16(1), 8-10.

Library of Congress (1999). Dublin Core/MARC/GILS crosswalk [Online]. http://lcweb.loc.gov/marc/dccross.html [2000, November 21].

Library of Congress (2000a). Library of Congress Classification outline [Online]. http://lcweb.loc.gov/catdir/cpso/lcco/lcco.html [2000, November 21].

Library of Congress (2000b). Tools for authority control subject headings [Online]. http://lcweb.loc.gov/cds/lcsh.html [2000: November 22].

Lynch, C. (1998). Identifiers and their role in networked information applications. Bulletin of the America Society for Information Science, 24(2), 17-20.

Medeiros, N. (1999). Making room for MARC in a Dublin Core world. Online, 23(6), 57-60.

Mundie, D. A. (1995). CyberDewey [Online]. http://ivory.lm.com/~mundie/DDHC/CyberDewey.html [2000, December 3].

OCLC Forest Press (2000). Introduction to the Dewey Decimal Classification [Online]. http://www.oclc.org/dewey/about/about_the_ddc.htm [2000, November 22].

Olson, N. B., ed. (1997). Cataloging Internet resources: A manual and practical guide [Online]. 2nd ed. Dublin, OH: Online Computer Library Center. http://www.oclc.org/oclc/man/9256cat/toc.htm [2000: November 21].

Online Computer Library Center, 2000a. Bibliographic formats and standards [Online]. http://www.oclc.org/oclc/bib/about.htm [2000, November 21].

Online Computer Library Center, [2000b]. Cooperative Online Resource Catalog [Online]. http://www.oclc.org/oclc/corc/ [2000, November 19].

Schwartz, Ray (1997). Uniform resource identifiers and the effort to bring "bibliographic control" to the Web: an overview of current progress. Bulletin of the American Society for Information Science, 24(1), 12-14.

Shafer, K., Weibel, S., Jul, E., and Fausey, J. [1996]. Introduction to Persistent Uniform Resource Locators [Online]. http://purl.oclc.org/OCLC/PURL/INET96 [2000, December 3].

Taylor, A. G. (1999). The organization of information. Englewood, Col.: Libraries Unlimited.

Weibel, S. (1997). The Dublin Core: A simple content description model for electronic resources. Bulletin of the American Society for Information Science, 24(1), 9-11.