DOCUMENT RESUME

ED 427 748 IR 019 299

AUTHOR Wills, Deborah
TITLE Making Sense of the World Wide Web: The Application of Library Practice.
PUB DATE 1998-11-00
NOTE 6p.; In: WebNet 98 World Conference of the WWW, Internet and Intranet Proceedings (3rd, Orlando, FL, November 7-12, 1998); see IR 019 231.
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Access to Information; Computer Literacy; Foreign Countries; Information Literacy; *Information Management; Information Retrieval; Information Scientists; *Information Seeking; Information Sources; Librarians; *Library Role; Library Services; Programming Languages; Researchers; Search Strategies; User Needs (Information); *World Wide Web
IDENTIFIERS Electronic Resources

ABSTRACT

The nature of the World Wide Web poses considerable challenges for the scholar or student who wishes to identify and locate Web resources of use to research. In contrast to the Web, the traditional library can be seen as a well-defined and organized physical space. An awareness of the successes, challenges and on-going projects involving library practice can help the non-librarian wishing to make sense of the Web for various purposes. The creator of Web resources may consider standards for metadata such as those developed by the Dublin Core. Web researchers may work cooperatively with librarians who have a long history of managing information resources. While the Web poses particular challenges to users and researchers at all levels, the basic principals of selection, organization, and access, as defined by library practice, continue to prove their relevance and adaptability. This paper focuses on some ways in which traditional library practice provides a methodology for approaching the Web, ranging from methods for organizing and accessing pre-existing resources to methods for enhancing resources at the point of creation. Contains 17 references. (Author/AEF)

Making Sense of the World Wide Web: The Application of Library Practice

Deborah Wills
Library, Wilfrid Laurier University, Waterloo, ON, Canada
E-mail: dwills@mach1.wlu.ca

Abstract: The chaotic nature of the Web confounds the researcher who wishes to identify useful Web resources. This paper focuses on some ways in which traditional library practice provides a methodology for approaching the Web: ranging from methods for organizing and accessing pre-existing resources to methods for enhancing resources at the point of creation. Library practice, which has informed the selecting, organizing and accessing of information, provides a methodology for introducing an authorial voice into the Web environment.

1. Introduction

The nature of the World Wide Web poses considerable challenges for the scholar or student who wishes to identify and locate Web resources of use to research. The rapid pace of growth; the wide range of quality, audience, and purpose; and the lack of standards for identifying the content of documents, are among the factors that confound the researcher. Equally important is the way that hypertext affects the organization and presentation of information. On the one hand, the endless flexibility of hypertext links, coupled with the Web's openness to user input, add to the democratization of communication. On the other hand, this same flexibility breaks down the traditional hierarchy among texts and subverts the authorial voice.

In contrast to the Web, the traditional library can be seen as a well-defined and organized physical space. In addition, its services do not stop at the walls of the building. Librarians have always been aware of the greater pool of published materials and of other collections--libraries, historical societies, etc.--which provide potential sources of information. Traditional library practice, which has informed the selecting, organizing, and accessing of information, also provide a methodology for approaching the Web and introducing an authorial voice into a chaotic environment. This paper will focus on a few select ways in which library practice has meaning for the Web: ranging from methods for organizing and accessing pre-existing resources to methods for enhancing resources at the point of creation.

2. Organizing Pre-Existing Resources

Librarians have always acted as intermediaries between information and users; they select resources based on predictions of user needs, and organize them for future identification and access. One important way intermediaries function on the Web is by creating hierarchical subject trees of selected, annotated resources. Users may browse these trees, moving from general to specific topics in a way that enhances the precision of their search. While there are no standard subject headings used by all trees, each has an internal consistency that guides the user to appropriate information.

Subject trees may focus on a narrow topic or may pull together resources from many topics in a form reminiscent of the traditional "bibliography of bibliographies." One of the best and most comprehensive of these trees is the *Argus Clearinghouse* [Argus 98] begun at the University of Michigan School of Library and Information Studies and maintained by a team of consultants who hold degrees in library and information science. Since 1993, this team has encouraged the creation of subject-specific guides to the Internet and has provided information architecture design services. While the Clearinghouse guides vary in their focus, organization, comprehensiveness, and timeliness, most provide good starting points for investigating a topic. The *Clearinghouse* also rates its guides to help users judge their value.

Another subject tree is the *WWW Virtual Library Project* [Virtual 98] begun at CERN in 1991 and

now maintained by volunteers, many of whom are librarians. This tree includes an experimental Library of Congress Classification System, a step toward the standardization of Web resources.

Many subject trees include a search tool for their selected resources, not to be confused with the comprehensive search engines that try to encompass the entire Internet. One good example is *Infomine* [Infomine 98] maintained by participants from all the University of California Libraries and Stanford University Library. *Infomine* provides annotated links to research and educational materials in a wide range of subjects. Its search mode allows Boolean searching within each of its broad subject categories. Search results also provide pointers to other categories of the *Infomine* tree which may include resources on the user's topic. This type of searching is somewhat analogous to performing a keyword search on a online catalogue; the results of the initial search indicate specific subject terms for use in subsequent searches.

The Web also contains bibliographies of resources for specific topics. Expanding on traditional print bibliographies, many provide citations to both print and Web resources with hypertext links to the Web documents. One excellent, frequently updated example is *The Scholarly Electronic Publishing Bibliography* [Bailey 98]. The *Bibliography* includes citations to articles, books, and electronic resources with appropriate links to full-text Web resources. The *Bibliography* may be browsed via its table of contents or searched using Boolean logic.

3. The Library Catalogue and the Web

The creation of a subject tree or bibliography is a systematic exercise imposing order and authority over a defined collection of resources. However, the need to integrate these resources into a single finding tool remains. The natural choice is the library catalogue, which is still the most persistent authorial voice for managing information resources. While subject trees may provide flexible keyword searching and subject classification, only the library catalogue provides rigorous authority control: the establishing of specific forms of names and subjects to be used throughout a set of bibliographic records.

The library catalogue, which has been online in most libraries for many years, uses standard rules to describe, classify and provide authority control for resources. Long before the emergence of the Web, library catalogues have provided rigorous access to information in a wide variety of formats: print, microform, audio tape, CD-ROM, etc., so the user need not predetermine the format of desirable resources. The catalogue ensures that a single item may be accessed in a number of ways: by author, title, and one or more subject headings, so that users with only partial information about a resource will be able to identify and locate it. Indeed for many library users, who may never approach an information desk or consult with a librarian, the catalogue performs a valuable public service. It is the job of the cataloguer to anticipate the needs of the user and the questions the user may have about the content, location, and usability of resources.

Keeping their individual user communities in mind, librarians have never attempted to catalogue all published information indiscriminately; instead, collections librarians select appropriate resources, and cataloguers work to make these resources as accessible as possible. Even within a given subset of published materials, cataloguers do not attempt to create all records from scratch. The sharing of catalogue records among libraries has been an established practice for many years, beginning in the days when major libraries published their catalogues in print form, and continuing through to the sophisticated, online, world-wide sharing that exists today. Cataloguers represent an established, objective, and cooperative body of professionals accustomed to addressing the growing world of information resources and the changing needs of their user communities.

Given the proven talents of cataloguers for responding to user needs, the library catalogue can be seen as a natural vehicle for organizing Web resources. This, it should be noted, is a far different thing from "cataloguing the Internet" in its entirety: an overwhelming and probably unnecessary task.

A growing number of libraries have begun enhancing their catalogues by adding URLs to existing records. For example, at Wilfrid Laurier University, we identify URLs for online versions of publications and add them to our records [TRELLIS 98]. We are particularly concerned with showing the connection between items previously available in print and now available electronically.

Other libraries are creating records specifically for Web resources. One catalogue that includes many such resources is Washburn University's [Washburn 98]. To isolate the Web records in this catalogue, search

for the subject "internet resource." The URLs embedded in the retrieved records link directly to the Web. More and more, libraries are providing web-based versions of their catalogues, thus simplifying the connection to Web resources [webCATS 98].

Of course, given the unique characteristics of Web resources, existing cataloguing standards must be expanded and modified. However, it is important to realize that cataloguing standards have been evolving for many years to allow for new subject areas and changing terminology as well as new formats of information. The first edition of the present standards, the *Anglo-American Cataloguing Rules*, was published in 1967, followed by a second edition (AACR2) in 1978 and revisions in 1982, 1983, 1985, 1988 and 1993. More revisions can be expected as cataloguers respond to changing needs.

As in most disciplines, theory follows practice. For the purpose of gaining this practice, a number of projects have been undertaken. Two of the most important have been the 1991/92 OCLC Internet Resources project and the 1994-96 OCLC Internet Cataloging project.

The first OCLC project focused on a sample of 300 resources to determine if Internet resources could be catalogued using USMARC format and AACR2 standards. With some exceptions, existing standards were found to be adequate. The second project solicited worldwide participation of librarians to select resources according to local collection development policies and then to catalogue them. As described in [Jul 98]: "By the end of the project, 231 participants representing nearly all types of libraries had selected and cataloged some 4,707 Internet resources." As of late 1997, more than 16,000 Internet resources had been catalogued by about 500 different OCLC-member libraries. Together, the OCLC projects have "demonstrated the applicability of the USMARC format and AACR2 cataloging rules to the creation of description and access records for Internet resources" [Dillon & Jul 96].

A Canadian project [see Campbell & Cox 97] exploring problems and solutions for cataloguing Internet resources is the Cataloguing Internet Resources Project (CIRP) initiated by the Faculty of Information Studies (FIS) at the University of Toronto. Participating academic, public, school, and special libraries select appropriate resources which are catalogued at FIS. The database of catalogue records can then be used by all participating libraries.

Projects such as those at OCLC and FIS have highlighted the challenges of cataloguing Internet resources and suggested solutions that involve changes in cataloguing standards. One result has been the publication of guides to aid cataloguers in their work. For example, [Olson 97] followed from the OCLC projects. The CIRP project used this manual and other resources to produce a draft of standards, which was distributed to the libraries participating in the project.

Some of the challenges faced by cataloguers suggest fundamental problems in the way that Internet resources are produced and identified. For example, one ongoing problem is the changeable nature of URLs: it is impossible to produce a definitive catalogue record for an item when its "call number" could change at any time. Work has been done by the Internet Engineering Task Force to develop a Uniform Resource Name (URN), though the process of implementing new standards is slow. In the meantime, OCLC, which is actively involved in developing URNs, has produced a Persistent Uniform Resource Locator (PURL): a naming and resolution service for Internet resources. As described in [Weibel et al. 97] this service associates a PURL with the actual URL of a resource and returns the URL to the client. PURLs have been developed to allow for a smooth transition to URNs once that architecture is in place.

The structure of a URL must be viewed with caution as a source of publisher information. As [Campbell & Cox 97] discovered, "The URL does not necessarily reflect the hierarchy of the organization that produced the site, and great care must be taken to distinguish the organization responsible for the content from the organization upon whose site the document is mounted."

Another problem is identifying where one document ends and the next begins in order to provide accurate and meaningful descriptions of specific documents. Cataloguers at FIS [see Campbell & Cox 97] determined that Web sites fall into two main types: independent documents with "their own self-enclosed integrity," and sites that "serve as a gateway for broader resources," such as computer programs which must be downloaded by the user. In the latter case, the document description must extend beyond the home page of the resource to include the computer program.

Another challenge stems from the constant updates and revisions that affect the content of many Internet resources. The FIS cataloguers [see Campbell & Cox 97] dealt with this problem by identifying two distinct types of site revision: those which apply incremental additions and those which maintain a basic skeletal structure but change individual elements. The former were determined to be analogous to serials,

while the latter were treated as loose-leaf services.

4. Metadata: The Dublin Core

A major ongoing problem for cataloguers is the lack of standards for identifying key information about a resource, such as author, publisher and subject. In addition, a method is needed to identify and access the variety of formats found on the Internet, such as images, sounds and video. One proposed solution is the use of metadata or "data about data": elements embedded within the META tag of Internet resources to enhance description and accessibility. It is important to realize that the concept of metadata has long been used in the creation of library catalogues and traditional indexes and abstracts. The challenge has been to apply this concept to a networked environment.

OCLC and the National Center for Supercomputing Applications are sponsoring a series of workshops to foster the use of metadata in networked resources. So far, five Metadata Workshops have brought together librarians and information technology professionals. At the first workshop, held in Dublin, Ohio in 1995, a core set of metadata elements was identified [Weibel et al. 95] "to describe the essential features of electronic documents that support resource discovery." This set of elements, known as the Dublin Core, allows authors and information providers to describe their resources for themselves, using a straightforward framework.

The Dublin Core standard has been stable since the third workshop and has been implemented in a number of projects: over 30 were reported [see Hakala 97] at the fifth workshop in October 1997. One of the most famous implementations is NORDINFO's Nordic Metadata Project [Nordic 98]. Slated for completion in 1998, the project is creating basic elements for the production and utilization of metadata. Software and documentation used for this project will be available in the public domain.

It is important to realize that while such systems as the Dublin Core have the advantage of being relatively easy to implement, the catalogue records created are still deficient when compared with full MARC-format records that have quality control at the point of creation. For example, while the Dublin Core contains the concept of author, it does not have a way to identify "main entry": a uniform access point for identifying and accessing an item. Also missing is the concept of "added entries": additional access points, such as joint author or editor. Dublin Core participants continue to debate the relative merits of enhancing the core elements versus keeping the system simple and accessible to all creators of resources.

The Dublin Core is intended to complement existing resource descriptions: both the relatively crude indexes generated by search engines and more sophisticated catalogue records. An important feature of the Dublin Core is that it is "syntax-independent," meaning that element descriptions are independent of encoding methods and should be mappable to other syntaxes, such as MARC [Crosswalk 97]. Given the limitations of the Dublin Core, such as the the lack of main and added entries, [Caplan & Guenther 96] describe how machine mapping has proven problematic. However, the Dublin Core provides a solid foundation for human cataloguers to apply MARC syntax and enhance records for use in library catalogues.

5. Conclusion

An important purpose of the Metadata Workshops has been to bring together relevant groups, including librarians, the Internet Engineering Task Force, and text encoding researchers, to help integrate their related activities. As librarians work cooperatively with those in related professions, library practice continues to evolve and to enhance the usability of Web resources.

An awareness of the successes, challenges and on-going projects involving library practice can help the non-librarian wishing to make sense of the Web for various purposes. The researcher attempting to navigate the Web may use the many subject trees and enhanced library catalogues and may suggest resources to be added to these tools. The creator of Web resources may consider standards for metadata such as those developed by the Dublin Core. Web researchers may work cooperatively with librarians who have a long history of managing information resources. While the Web poses particular challenges to users and researchers at all levels, the basic principles of selection, organization, and access, as defined by library

practice, continue to prove their relevance and adaptability.

6. References

[Argus 98] *Argus Clearinghouse*. http://www.clearinghouse.net/

[Bailey 98] Bailey, C.W. (1998). *Scholarly Electronic Publishing Bibliography*. http://info.lib.uh.edu/sepb/sepb.html

[Campbell & Cox 97] Campbell, D.G., & Cox, J. (1997). Cataloguing Internet Resources. *Feliciter*, 43 (5), 60-63.

[Caplan & Guenther 96] Caplan, P., & Guenther, R. (1996). Metadata for Internet Resources: The Dublin Core Metadata Elements Set and Its Mapping to USMARC. *Cataloguing & Classification Quarterly,* 22 (3-4), 43-58.

[Crosswalk 97] Network Development and MARC Standards Office. (1997). *Dublin Core/MARC/GILS Crosswalk*. http://lcweb.loc.gov/marc/dccross.html

[Dillon & Jul 96] Dillon, M., & Jul, E. (1996). Cataloging Internet Resources: The Convergence of Libraries and Internet Resources. *Cataloging & Classification Quarterly*, 22 (3), 197-238.

[Hakala 97] Hakala, J. (1997). *Dublin Core in 1997: A Report from Dublin Core Metadata Workshops 4 & 5*. http://linnea.helsinki.fi/meta/dcnord.html

[Infomine 98] *Infomine*. http://lib-www.ucr.edu/

[Jul 97] Jul, E. (1997). Cataloging Internet Resources: Survey and Prospectus. *Bulletin of the American Society for Information Science*, 24 (1), 6-9.

[Nordic 98] *Nordic Metadata Project*. http://linnea.helsinki.fi/meta/

[Olson 97] Olson, N.B., ed. (1997). *Cataloguing Internet Resources: A Manual and Practical Guide*. 2nd ed. Dublin, OH: OCLC. http://www.oclc.org/oclc/man/9256cat/toc.htm

[TRELLIS 98] *TRELLIS*: TriUniversity Group of Libraries Catalogue. http://www.tug-libraries.on.ca/trellis/

[Virtual 98] *WWW Virtual Library Project*. http://vlib.stanford.edu/Overview.html

[Washburn 98] Washburn University Library Catalogue. http://lib.wuacc.edu/

[Weibel et al. 95] Weibel, S. et al. (1995). *OCLC/NCSA Metadata Workshop Report*. http://www.oclc.org:5046/oclc/research/conferences/metadata/dublin_core_report.html

[webCATS 98] webCATS: Library Catalogues on the World Wide Web. http://www.lights.com/webcats/

[Weibel et al. 97] Weibel, S., Jul, E., & Shafer, K. (1997). *PURLs: Persistent Uniform Resource Locators*. http://purl.oclc.org/OCLC/PURL/SUMMARY.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

REPRODUCTION BASIS

☒ This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").