THE CHRISTIAN LIBRARIAN®
JOURNAL OF THE ASSOCIATION OF CHRISTIAN LIBRARIANS

Volume 44 | Issue 2 Article 3

2001

Search Engines Giving You Garbage? Put A "Corc" In It, Implementing The Cooperative Online Resource Catalog

Robert O. Ellett Jr.
Joint Forces Staff College

SEARCH ENGINES GIVING YOU GARBAGE? PUT A "CORC" IN IT

IMPLEMENTING THE COOPERATIVE ONLINE RESOURCE CATALOG

Robert O. Ellett, Jr. is the Catalog Librarian at the Joint Forces Staff College, Norfolk, Virginia.

He is a Ph.D. Student at Nova Southeastern University's School of Computer and Information Sciences.

This paper presents an implementation strategy for adding Internet resources to a library online catalog using OCLC's Cooperative Online Resource Catalog (CORC). Areas of consideration include deciding which electronic resources to include in the online catalog and how to select them. The value and importance of pathfinders in creating electronic bibliographies and the role of library staff in updating them is introduced. Using an electronic suggestion form as a means of Internet resource collection development is another innovative method of enriching library collections. Education and training for cataloging staff on Dublin Core elements is also needed. Attention should be paid to the needs of distance learners in providing access to Internet resources. The significance of evaluating the appropriateness of Internet resources for library collections is emphasized.

> *"The net is like a huge vandalized library. Someone has destroyed the catalog and removed the front matter, indexes, etc. from hundreds of thousands of books and torn and scattered what remains...."Surfing" is the process of sifting through this disorganized mess in the hope of coming across some useful fragments of text and images that can be related to other fragments. The net is even worse than a vandalized library because thousands of additional unorganized fragments are added daily by the myriad of cranks, sages, and persons with time on their hands who launch their unfiltered messages into cyberspace."*[1]

INTRODUCTION

The world of the Internet and its multitude of resources has created new challenges and opportunities for libraries. Threats of the Internet replacing libraries are commonplace. However, the library world is, in fact, changing and must change. The library must embrace Internet resources not in a spirit of competition with other information providers but as a means to continue its rich tradition as an information portal in society. Librarians are well equipped for organizing information. Cataloging is still viewed as a very valuable skill in managing the often chaotic world of the Internet.[2] The Internet offers libraries opportunities to rethink and ultimately redefine their role in the world of web-based learning environments.[3] In fact, in his review of the library literature, Beagle found that web-based learning environments force libraries to play a more active and facilitative role in the learning process.

Librarians are the key players in supporting the organizational skills required for information in Web-based environments.[4] Library catalogs in both the card and online formats solve the relevancy problem so prevalent in Internet searching.[5] Libraries, however, need to break from their tradition of ownership to the paradigm shift to focus on information access.[6]

People seeking information now prefer the often chaotic atmosphere of the Web to libraries that have carefully cataloged collections of relevant resources.[7] One of the most significant and oldest problems in the online learning environment of the Web is the lack of relevancy produced by Internet search engines. Online searchers want the most relevant resources. However, librarians have always dealt with this issue. Cataloging rules were developed and implemented to ensure that users receive accurate information that meets their needs. Relevancy and its synonyms of applicability, correspondence, and pertinence often require sifting through massive amounts of data when using the Internet to find information. Brandt succinctly defines determining relevancy in information searching basically as "answering the question 'Is what I've found closely related to what I need?'"[8] Library catalogs used to contain only entries of items that the library actually owned in its collection. However, with the paradigm shift from ownership to access, libraries now provide access to far more than their own collections. With the advent of the Internet, claims are now made that with search engines "everything" on the Internet can be searched. Of course, this is a gargantuan claim and search engines really did not deliver on this promise. Internet search engine companies actually do not even boast of their consistency or accuracy. Instead, they promote their strength as the ability to search the entire Web (which is virtually impossible). A major aspect of keyword searching that is problematic with Internet search engines is the location of the keyword searched. A keyword can appear in the header, abstract, credits, and HTML source code. Relevance depends greatly on where the keyword is located within the document. "Unfortunately, with the disparate nature of Web pages, wide variations in file sizes, and a complete spectrum of subjects, both scholarly and mundane, determining relevancy automatically is no easy task."[9] Generally, a search engine's relevancy factors are a closely guarded secret which is a definite disadvantage to the Internet searcher who attempts to comprehend the frequently unusual result sets of these searches.

The library must embrace Internet resources not in a spirit of competition with other information providers but as a means to continue its rich tradition as an information portal in society.

Many supporters of the Internet and Internet search engines like Yahoo and Altavista believe that these tools will eventually make information-providing institutions like libraries obsolete. They exclaim, "Why go to the library when information is at your fingertips (just a mouse click away), through Internet search engines that use keywords?" Sherman relates an analogy about Internet searching from Joel Truher, a vice president of technology for HotBot, who is also the architect of one of the most highly respected web search engines: "Currently, search is simply bad. It's like interacting with a snotty French waiter. The service is bad, you get served things you didn't ask for, you often have to order again and again, and you don't get things that are listed on the menu. People have learned to cope with it—they've internalized their frustrations."[10]

THE SIGNIFICANCE OF METADATA

Some Internet searchers believe that the Internet should be organized like a library. But should the Internet be cataloged and controlled like items in the library's online catalog? Experts do believe that some of the principles used in library cataloging should be applied and adapted to the indexing of the Internet. A current buzzword in libraries is *metadata*. Loosely, metadata is defined as structured data about data. Metadata, however, encompasses other principles as well. Librarians, catalogers specifically, have created metadata structures for decades. With the advent of the electronic world, the term, "metadata", came into vogue. Vellucci describes the origin of the term. Metadata was a term used frequently in the 1980s in reference to literature on database management systems (DBMS).[11] The term metadata was "used to describe the information that documented the characteristics of information contained within databases." Catalogers in the library world continued to use the terms "bibliographic data" or "cataloging data" when the object being cataloged was in a non-electronic form even when the bibliographic record migrated to the machine readable cataloging (MARC) formats. However, when catalogers began to describe networked electronic resources using the same type of bibliographic data, the MARC record became metadata. "The methods of organizing resources from the rather separate domains of library science, computer science, and information science all converged in this networked environment, and the term 'metadata' became a commonly accepted term in all disciplines."[12]

Metadata describes the attributes and contents of an original document or work. Some metadata is created specifically for computers to use. Often, metadata indicates the original format in which a work was created so that the computer can open both the

application needed and the document simultaneously. Milstead and Feldman describe the functions of metadata.[13]

Metadata acts as a surrogate for the item described. It characterizes the work so that the user can understand its contents as well as its purpose, source, and possibly even conditions or terms of use. Metadata can be incorporated into the structure of a Web document, or it can exist separately with a pointer to the document itself. This second model is similar to traditional cataloging methods. Cataloging MARC records that describe a book exist separately in a different location than the book, but the metadata points to the book's location with the use of a call number. Separate metadata files on the Web will "point" to the document they describe on the Web through the use of the URL.

Over 200 institutions have participated in the Founder's Phase of the CORC database project, including university and academic libraries, public libraries, museums, and government libraries.

However, a greater purpose of a metadata scheme is to establish and maintain standard structure and terminology. The fields used within the metadata scheme should be standardized. If the concepts of creator, author, sculptor, or composer all serve the same function, metadata attempts to map them to the single concept. This principle of standardization is very important within a metadata scheme. Metadata basically involves the rules on how to record the descriptive information and what to record. Vellucci discusses three vital characteristics of metadata: flexibility, interoperability, and extensibility.[14] Flexibility allows the metadata creator to include as much or as little detail desired to describe the item. This flexibility equates to various levels of description prescribed by the *Anglo-American Cataloging Rules.* In the traditional cataloging world, flexibility is also quite a buzzword. The establishment of the Program for Cooperative Cataloging (http:\\www.loc.gov\catdir\pcc) and the development of core-level cataloging records demonstrate this acceptance of flexibility in order to get catalog items to the user with timeliness as a value. Interoperabilty refers to the ability of information systems to interact in a useful manner on a real-time basis. The metadata record should contain a core set of data elements that would be common to all schemes. This commonality would facilitate the exchange and use of metadata in a variety of systems. Crosswalks often offer a way to map data elements from one metadata scheme to another. Several crosswalks currently exist to map the MARC format to and from other popular metadata schemes such as Dublin Core (DC), which will be discussed later in this paper. Extensibility includes the metadata scheme's ability to allow additional extensions of data elements and data qualifiers to accommodate specific users' needs. For example, the need to express that the author or creator is a corporate body would evoke a method of extending the data to include this additional information. Vellucci provides an excellent summation of the function of metadata. "Metadata are data that describe the attributes of a resource; characterize its relationships; support its discovery, management and effective use; and exist in an electronic environment."[15]

Vellucci explains the main difference between the MARC metadata and other metadata schemes such as the markup languages HTML (Hytertext Markup Language) or SGML (Standard General Markup Language). The MARC metadata record is separate from the document or resource it describes. However, with metadata schemes like HTML, markup languages can be used to encode both the descriptive data and the object itself.[16] Milstead and Feldman present various methods in which metadata can be created since it very logically can be embedded into the object or the item described. Metadata can be created at the time the object is created by the creator. Who knows better than the creator the nature and the scope of the object? Of course, metadata could be added later as part of the traditional cataloging process. Milstead and Feldman suggest that this first method,described as author-created metadata, will eventually proliferate because traditional cataloging methods cannot cope with the current explosion in electronic resources.[17] Federal agencies, in fact, have developed web-based forms for entry of metadata by the creator or owner of the site. The Federal Geographic Data Committee (FGDC) and the Environmental Protection Agency (EPA) both have respective programs in place (http//130.11.52.178/metaover.html; http://www.epa.gov/regional/epafield.html). For more information about author-created metadata, refer to the Web Development Virtual Library (http://wdvl.internet.com/Location/Meta/tag.html). This site, entitled "Meta Tagging for Search Engines," describes how each of the major search engines processes data in the metadata tag fields.

THE COOPERATIVE ONLINE RESOURCE CATALOG

The Cooperative Online Resource Catalog (CORC) (http://www.oclc.org/oclc/corc/index.html), developed by OCLC, Inc., began as a project on January 15, 1999. CORC was designed as a mechanism to create a cooperative database of both local and web-based resources by employing several library practices and standards such as the MARC format, the Dublin Core metadata standard, and authority control.[18] By using the principles of authority control, CORC promotes

relevancy in searching and eliminates irrelevant resources that are so prevalent with Internet search engine results today. The Internet resources input in the CORC database can then be imported into the library's local online catalog.

Since it is a web-based system, librarians need no special software to gain access to the CORC database (http://corc.oclc.org). All that is required is an Internet connection and an OCLC authorization number and password. The following types of materials are found in CORC: electronic documents, Web sites, electronic journals, electronic novels, music files, and other electronic files. Why should CORC be a cooperative venture? It would be impossible for a single library to catalog and maintain records of all the electronic resources available and needed by patrons due to the overwhelmingly vast amount of information available electronically. According to Hickey, Childress, and Watson, three of the major developers on the CORC system team at OCLC, "libraries of all types are finding it imperative to have a World Wide Web presence. The Web is the dominant method for library users to access online information ... libraries may be the best place to go for much information, but without a solid Web interface, they are no longer the easiest." The thoughts of these information scientists are well supported by the literature. These authors explain that with a Web presence, libraries are building their own portal pages. General Web resources are difficult to keep current because their links change so often. The purposes of CORC are to provide assistance with link selection, link maintenance, resource descriptions, and the creation of pages based on metadata supported by the World Wide Web. Hickey, Childress, and Watson, also attempt to answer the question as to why CORC is a product of OCLC.[19] CORC is very similar to OCLCs major activity—the provision of automated, shared cataloging databases. OCLC, with the cooperation of member libraries, has created WorldCat (a bibliographic database of over 40 million records), the OCLC FirstSearch system (an abstract and indexing service often providing full-text journal articles), and the OCLC cataloging service. The original records for the OCLC CORC database were extracted from two earlier OCLC projects, InterCat and NetFirst. InterCat was an OCLC project designed to encourage and investigate cataloging resources in WorldCat. This project of the early 1990s was solely a MARC (machine readable cataloging-based) initiative. Seventy-four thousand InterCat records were transferred into CORC using a metadata crosswalk. NetFirst was a database of web resources available via OCLC's FirstSearch system and the OCLC cataloging system. NetFirst records were basically abstracting and indexing records, although they do include Library of Congress Subject Headings and Dewey Decimal classification numbers. One hundred thousand NetFirst records were transferred into CORC. CORC, however, uses new technology such as Web-based metadata schemes for the input of data. Over 200 institutions have participated in the Founder's Phase of the CORC database project, including university and academic libraries, public libraries, museums, and government libraries. In fact, OCLC plans to market CORC as a FirstSearch database available to online users who subscribe to its massive collection of online database collections.

In implementing CORC, several of the following issues must be resolved:

WHY CATALOG INTERNET RESOURCES?

Internet resources have a variety of characteristics that are different from print resources. Sherman's research indicates that Internet searchers find the Internet poorly organized and varying in quality. Other searchers believe that the vast amount of data on the Internet is too difficult to sift through.[20] Material on the Internet is often prone to change in content and location, specifically the uniform resource locator or URL, also known as the Internet address. One of the most important incentives to catalog Internet resources is that many of the traditional print resources now even have Internet companion pages that are now solely available through the Internet. [21] A major rational in cataloging Internet resources and adding them to the library catalog is that frequently these resources are not available through general Internet search engines and directories. Often these scholarly research articles are not indexed by search engines that use only keywords as a means of identifying the resource.[22] The most compelling argument to catalog Internet resources is that many paper resources are now available only as digital documents on the Internet. Not only is it often easier to publish documents on the Internet, it is usually less expensive. The cost incentive seems to drive many publishers and agencies to publish solely on the Internet. Paper prices have skyrocketed. A recent illustration comes from the Government Printing Office (GPO). The GPO's Federal Depository Library Program (FDLP) found, in a 1996 study, that permanent public access to these federal documents was needed to ensure that the information was permanently and continually available to the public. The GPO, in recent years, has increased the number of titles it disseminates electronically. In fiscal year 1999, 46 percent of new titles were disseminated only electronically. In FY2000, over 50 percent of titles were only disseminated in electronic format. In May 2000, the House Appropriations Committee passed H.R. 4516, which cut the GPO FY20001 budget by 11 percent as more titles are being planned for electronic distribution only.[23] Government documents are just one case of why it is critical to begin cataloging Internet resources.

INFRASTRUCTURE ELEMENTS IN PLACE

Despite these measures to increase document distribution on the Internet, opponents to including Internet

Despite these measures to increase document distribution on the Internet, opponents to including Internet resources in library catalogs are vocal. Baruth raises several valid concerns when contemplating the addition of Internet resources to the library online catalog. How do libraries keep current with the overabundance of Internet resources that proliferate on the Web daily? Will additional staff, namely subject specialists and catalogers, be needed? Is the MARC format adequate for electronic resources? Will libraries duplicate the work of other libraries in creating bibliographic records for Internet resources? Can local automated online catalog systems handle the addition of Internet resources? Can library online catalogs compete with general search engines whose popularity causes the general public to question the future of libraries?[24] Ironically, all of these issues are addressed by the CORC system.

CORC offers several features that resolve these problems. Keeping current with the most up-to-date resources is crucial in the online world. CORC provides a variety of indexes to search the database, including traditional search keys such as author, title, and subject, and others such as uniform resource locator (URL) addresses, keywords, and numeric searches. A key factor to the success of CORC is that it is cooperative in nature. Like the WorldCat database, which now boasts more than 45 million bibliographic records, libraries around the world continually add resources to CORC. Over 7,000 library systems that participate in and use OCLC services cooperatively add and share resources. No additional staff should be needed, recognizing several factors. Libraries, using CORC, work cooperatively. The different libraries share the electronic resource cataloging records. In addition, virtual resources from the Internet do not require physical processing, so this is not another expense for libraries. Typically, Internet resources replace or supersede paper items, so print versions are no longer cataloged anyway. CORC solves the problem of URL or Internet address maintenance as URLs often change or are removed from the Internet. The CORC system provides participating libraries with lists of changed or removed URLs on a daily basis. The typical Internet search engines take months to update search indexes.[25] CORC not only includes Web sites but also electronic documents, electronic journals, electronic novels, music files, and other electronic files. As librarians create a record or attach their holding code to the individual Internet resource records, CORC keeps track that the particular library has cataloged that item. When the URL changes or is removed from the Internet, a message is sent to that particular library. The library can then modify the URL or delete the bibliographic record. This maintenance feature is beneficial not only to one library but to all the libraries that have used the bibliographic record for the Internet resource. Thus, one library could perform the URL maintenance for hundreds or thousands of libraries, reducing both time and effort spent in updating. Bibliographic maintenance is nothing new to libraries. They have updated their card and online catalogs with changed or discarded volumes for centuries.[26] Baruth also complains that no single library catalog can contain all the available Internet resources.[27] However, libraries have never collected or cataloged all published materials available. Librarians choose the most appropriate resources for their collection. It is understood that a given library will not collect all available resources.

Why should CORC be a cooperative venture? It would be impossible for a single library to catalog and maintain records of all the electronic resources available and needed by patrons due to the overwhelmingly vast amount of information available electronically.

The questions, however, remain as to how to implement CORC and the cataloging of Internet resources into the workflow of libraries.

COLLECTION DEVELOPMENT ISSUES

One of the basic questions to be answered is how to select Internet resources to catalog for the online catalog. Obviously, Internet resources are not published or made available to libraries in the same manner as traditional print materials. Several new factors must be examined in the selection process. Weber notes several areas of consideration: price, cancellation of comparable resources, appropriateness of electronic resources, stability of Internet resources, duplication or redundancy with other resources in the collection, licensing issues that govern access, copyright restrictions and access to information from previous releases.[28] Pearlmutter includes other factors in discussing the acquisition or collection development of electronic resources. A key concern to be remembered is that the Internet is primarily a world of self-publishing where, all too often, anything goes.[29] Librarians, more than ever, need to use information evaluation skills when selecting Internet resources. Ease of use is also recommended as part of the selection criteria. Pearlmutter also emphasizes the value of an effective collection development for the library's distance learners.[30] Distance learners often have different information needs than local patrons. Typically, ready reference sources when available on the Internet should be incorporated into the library's online catalog. The value of full-text sources as opposed to citations to print volumes should be recognized in selecting Internet resources. Cooperative collection development activities deserve special attention in the area of

electronic resources. CORC excels in this arena as librarians can view the electronic resource holdings of other libraries. The selections of other libraries should influence the library's electronic resource collection development policy. In light of distance education users, another matter to be resolved is whether the library's online connection will have the equipment and software to use the resource. Another question is whether the resource displays in the Web browser within a reasonable length of time.[31] It is useful to have Internet resources displayed in the online catalog. However, if users have difficulty accessing them because of equipment or software issues, it becomes a matter of frustration and disappointment. Librarians appreciate the value of accuracy, authority, objectivity, and currency of data and strive to promote these values in the selection of Internet resources to be described.

REFERENCE AND RESEARCH ISSUES

For the successful implementation of the CORC system, reference librarians and bibliographers play several significant roles. These key individuals often create print bibliographies of resources. CORC has a pathfinder feature through which electronic bibliographies of resources can be created. Frequently, these pathfinders supplement research needs and also update them with the most current available information as well. Pathfinders can be exported from the CORC system as HTML links with URL addresses. The author has created a pathfinder on Equal Opportunity Employment (EEO) information resources. Federal law requires that supervisors working in the federal government be knowledgeable of the principles of EEO. The pathfinder (http://purl.oclc.org/corc/system/Pathfinder/640:xid=AFQ) outlines the areas of prohibited discrimination and even includes a review quiz to test comprehension of these principles. It has also been used as a training tool by several Army libraries. Another advantage to using CORC pathfinders (which are basically bibliographies that hyperlink to resources on the Internet) available via the library's homepage is that academic computing staff or information systems staff need to add the pathfinder link to the library homepage only once. The maintenance of the individual pathfinder is then the responsibility of library staff. The feature is particularly important when library staffs do not have access to updating the library's homepage directly and must rely on systems staff outside of the library. Pathfinders, in this way, provide more control of updating and deleting information resource bibliographies.

Reference librarians also suggest resources to be added to the library's collection. The author has designed an electronic form (http://scis.nova.edu/~ellettro/form2.html) available via the Internet through which not only may library personnel suggest electronic resources to be added to the collect but also library users can contribute suggestions. This common gateway interface (CGI) form automatically sends an email to the author when the users submit the data. The user is asked to input a short description of the resource. The cataloger then can use this description in the cataloging process especially in the area of subject analysis. Distance education students would most definitely discover the importance of this form in suggesting electronic resources for the library's online catalog.

CATALOGING ISSUES

Catalogers will of course need to learn to use the CORC system and also how to input bibliographic records and pathfinders. The cataloger's role will be one of training the trainers of the CORC system. Catalogers will not only need to be well-versed in the intricacies of the machine-readable cataloging (MARC) format, but will also need to learn the elements of Dublin Core (DC), another metadata standard. The use of Dublin Core is a lower-cost simplified alternative to traditional MARC cataloging.[32] Used as part of descriptions within many hypertext markup language (HTML) forms, Dublin Core elements such as creator, title, language, and description further characterize the distinctive aspects of Internet resources.[33] Milstead and Feldman refer to the use of metadata (which is data about data) such as Dublin Core as crucial to the survival of library cataloging. Although in the traditional cataloging process, metadata is produced after the creation of the item (book, map, sound recording, etc.), Web page designers are currently using metadata to categorize the data.[34] Metadata is being created concurrently as Internet resources (online documents, Web pages, etc.) are being created. CORC will automatically harvest Internet sites with embedded metadata and supply library catalogers with a basic template.

THE IMPORTANCE OF AUTHORITY CONTROL

CORC has one of its greatest selling points, the feature of authority control, embedded within its structure. To describe the nature of authority control, a brief analysis of the Internet's problems with keyword searching is needed. Milstead and Feldman describe three problems that are inherent in natural keyword and Boolean searching.[35] Polysemy is the concept that most words have multiple meanings. These authors use the example of the word "springs". If an Internet searcher entered the keyword, "springs", information might be retrieved on fresh water springs, or on the season, or even on coil springs. Another relevant example to computer specialists or technologists would be an Internet search for the term, *ATM*. The search would retrieve results on banking and automatic teller machines and also on asynchronous transfer mode. Another problem, synonymy, is where many words represent the same

concept, although they always have slightly different meanings. An example of synonymy would be "ball" or "sphere". Finally, the problem of ambiguity in meanings is resolved with the use of metadata. A search engine must "understand" the meaning of the word, not just be able to match the spelling of the word. For example, a search on "cars" with an Internet search engine would result in information on automobiles, but it may also retrieve information on the musical group with that name. Milstead and Feldman assert that Web and Boolean search engines cannot determine these differences by the context of the words in the passage. These factors are remedied by what is called controlled vocabulary, or, to be more library-specific, authority control.[36] Authority control both in library catalogs and databases has two goals: to maintain consistency in the verbal form used to represent an access point and to provide the interconnections via relationships among works, words, and subjects. For the purpose of this discussion, an access point is an attribute of an item that a searcher is likely to use in locating the resource. Common access points are authors, titles, and subjects. Uniformity is the key concept in authority control. For example, authors who write under pseudonyms such as Samuel Clemens/ Mark Twain or Stephen King/Richard Bachman need to have cross references made from the names not used so that if the user searches under Twain, the database system will refer to Clemens.

Another advantage to the use of authority control is that the searcher is not responsible for figuring out which particular name or term the item is listed under. Other situations with personal names include married names, maiden names, or other changed names. Subject searching is another area that demands the benefits of authority control. In subject authority work, one term represent the concept (promoting uniformity and consistency) and the other terms that are related to it are linked (promoting linkage). For example, if the medical researcher searched for the term, "Sildenafil", if authority control existed, then all the entries containing the word, "Viagra" (the generic name for the drug Sildenafil) would be retrieved. The main problem with current Internet search engines is that they are void of these vital linkages. General Internet search engines cannot distinguish, for example Mercury the planet from Mercury the Greek god, or mercury, the chemical element. It is unlikely that a researcher desiring information on Mercury the planet would also want information on the chemical element. Internet searching engines through keyword searching are not capable of this distinction. The researcher wastes valuable time sorting through data that is irrelevant. Authority control is viewed as the ultimate customer service. Internet search engines lack this essential component of effective searching. Konovalov gives the example of searching for the "iron curtain" and receiving results that are littered with data on metallurgy or theater. "The best possible result of our work seems to this situation when our customers can easily find their needles in the haystacks of our collections." [37]

CORC uses as the basis for its authority control the controlled vocabulary tool of the *Library of Congress Subject Headings*. Embedded within the authority control of the CORC database are all of the see references and broader and narrower terms used in the *Library of Congress Subject Headings*. For example, if the CORC searcher entered Siamese cats as a subject term, the searcher would be prompted with the message "Siamese cats see also Cats". This feature is particularly helpful if the searcher needs assistance in defining or narrowing a search term. The cataloger using CORC can easily locate the correct authority records. For example, the term "online learning" and "distance education" both refer the user to "distance education". This authority control feature ensures consistency of terms and also hierarchical relationships between names and terms.[38] Thus, the cataloger benefits from these automated processes provided by the CORC system. Catalogers also need to work with their local system vendors to ensure that these bibliographic records for Internet resources are importable into their systems. System developers should be certain that their automated systems can handle both the MARC format and the Dublin Core format.

THE USEFULNESS OF DUBLIN CORE

Another feature also within the CORC database is the metadata scheme Dublin Core, which has gained popularity in library environments. Dublin Core or DC was named for the site of the first metadata workshop held in Dublin, Ohio. DC is maintained by OCLC and was developed in an international and interdisciplinary environment. The DC defines a set of fifteen basic data elements for resource description. These elements, many of which correspond to data in traditional catalog records, are divided among three categories: content (title, subject, description, source, language, relation, coverage), intellectual property (creator, publisher, contributor, rights), and Instantiation (date, type, format, identifier).[39] Ironically, all the elements within the DC metadata scheme are optional. All elements can also be qualified or fully described as needed. Included among the optional qualifiers are "personal" or "corporate" which qualifies or further describes the creator field. Weibel believes that DC is an effective alternative to MARC.[40] Chepesiuk explains the disadvantages of solely using MARC as the metadata scheme in cataloging Internet resources. Indepth cataloging costs a lot of money to produce.[41] Justifying the time and expense of performing MARC cataloging of Internet materials is difficult because Internet resources are so fluid. A Web site can be accessed one day and gone the next. The site can be totally removed or change its URL address.

The content can even be so changed or different that it is not recognized as the original site. DC will not replace MARC. It must "co-exist" with it. Dublin Core can be embedded into HTML documents to enhance retrieval in search engines as metatags. Many search engine producers admit to indexing keyword metatags.[42] DC metadata is viewed as a lower-cost, simplified alternative to traditional MARC cataloging. The CORC database represents a good blend of DC and MARC. In the CORC database, the underlying representation of the data is independent of either MARC or Dublin Core in that the data can be viewed in the manner which is most appropriate to the user's context.[43] Dublin Core is based upon the resource description framework (RDF). RDF is the product of the World Wide Web Consortium (W3C), which is a standards organization designed to provide the necessary components to create metadata schemes.

CONCLUSION

The implementation of CORC should be a team effort within the library. To appreciate fully and integrate all of the features of CORC, public and technical services personnel must work together to enhance the library catalog. The strengths of the various library functions will afford both the local patron and the remote user greater opportunities to satisfy their research needs. The future of the library depends upon its ability to include Internet resources into its collections. Internet resources are no longer just a luxury for library online catalogs. They are vital to the growth and development of a virtual library or information center. The role of the librarian as information evaluator has only expanded in asynchronous learning environments.

---

[1] Michael Gorman, "The Corruption of Cataloging," *Library Journal* 120 (Sept. 15, 1995): 34.
[2] Yuri Konovalov, "Cataloging as a Customer Service: Applying Knowledge to Technology Tools," *Information Outlook* 3, no. 9 (1999): 25.
[3] Donald Beagle, "Web-Based Learning Environments: Do Libraries Matter?," *College & Research Libraries*, 370.
[4] Ronald Chepesiuk," Organizing the Internet: the Core of the Challenge," *American Libraries* 30, (Jan. 1999): 61.
[5] Lori Leibovich, "Choosing Quick Hits Over The Card Catalog," Available http://www.nytimes.com/library/tech/00/08/circuits/articles/10thin.html Accessed December 14, 2000.
[6] D. Scott Brandt, "Relevancy and Searching the Internet," *Computers in Libraries* 16, no. 8 (1996): 35-36.
[7] Leibovich, "Choosing Quick Hits Over The Card Catalog."
[8] Brandt, "Relevancy and Searching the Internet," 36.
[9] Greg R. Notess, " On the Net: Rising Relevance in Search Engines. *Online* 23, no. 3 (1999):15-18. Available http://www.onlineinc.com/onlinemag/OL1999/net5.html Accessed December 14, 2000.
[10] Chris Sherman, "The Future of Web Search," *Online* 23, no. 3 (1999): 56.
[11] Sherry L. Vellucci, "Metadata and Authority Control," *Library Resources and Technical Services* 44, no. 1 (Jan. 2000): 36.
[12] *Ibid.*, 37.
[13] Jessica L. Milstead and Susan Feldman,"Metadata: Cataloging By Any Other Name," *Online* 23 (1999): 25-31.
[14] Vellucci, "Metadata and Authority Control," 38.
[15] *Ibid.*, 39.
[16] *Ibid.*, 41.
[17] Milstead and Feldman,."Metadata: Cataloging By Any Other Name," 28.
[18] Thomas B. Hickey, Eric Childress, and Bradley C. Watson, "The Genesis and Development of CORC as an OCLC Office of Research Project," *OCLC Newsletter* 239 (1999): 28.
[19] *Ibid.*, 30-31.
[20] Sherman, "The Future of Web Search," 57.
[21] Konovalov, "Cataloging as a Customer Service: Applying Knowledge to Technology Tools," 26.
[22] Greg R. Notess, "On the Net: Rising Relevance in Search Engines," 16.
[23] Timothy L. Coggins, "Print no more: U.S. Code, Code of Federal Regulations, and the Federal Register," *Virginia Lawyer* 49, no. 3 (Oct. 2000): 55.
[24] Barbara E. Baruth. "Is Your Catalog Big Enough to Handle the Web?," *American Libraries* 31, no. 7 (Aug. 2000): 57-58.
[25] Hickey, Childress, and Watson, "The Genesis and Development of CORC as an OCLC Office of Research project," 30-31.
[26] Karen Calhoun, "CORC at Cornell Project: Final Report". Available http://ivy.mannlib.cornell.edu/corc/corc-final.htm Accessed December 14, 2000.
[27] Baruth. "Is Your Catalog Big Enough to Handle the Web?," 58.
[28] Mary Beth Weber, "Factors To Be Considered in the Selection and Cataloging of Internet Resources," *Library Hi-Tech* 17, no.3 (1999): 298-300.
[29] Jane Pearlmutter, "Which Online Resources Are Right For Your Collection?," *School Library Journal* 45, no. 6 (June 1999): 27-28.
[30] *Ibid.*, 28.
[31] Diane K. Kovacs, "Building a Core Internet Reference Collection," *Reference & User Services Quarterly* 39, no. 3 (Spring 2000): 235.
[32] Eric Miller, "Resource Development Framework: Achieving Interoperable Metadata on the Internet," *OCLC Newsletter* 239 (1999): 32; Stuart Weibel, "CORC and the Dublin Core," *OCLC Newsletter* 239 (1999): 33.
[33] Norm Medeiros, "Making room for MARC in a Dublin Core world," *Online*, 23, no. 6(1999): 58.
[34] Milstead and Feldman,"Metadata: Cataloging By Any Other Name," 30.
[35] *Ibid.*, 31.
[36] *Ibid.*
[37] Konovalov, "Cataloging as a Customer Service: Applying Knowledge to Technology Tools," 26.
[38] Milstead and Feldman, "Metadata: Cataloging By Any Other Name," 31.
[39] Weibel, "CORC and the Dublin Core," 33.
[40] *Ibid.*
[41] Chepesiuk," Organizing the Internet: the Core of the Challenge," 61.
[42] Medeiros, "Making room for MARC in a Dublin Core World," 59.
[43] Weibel, "CORC & the Dublin Core," 33.