

## **A Role for Classification: The Organization of Resources on the Internet**

Susan J. Matveyeva

"Do we catalog only those items physically located in our libraries, or those items our patrons have access to? This question needs to be answered by each library as it discusses cataloging electronic resources" (Olson, 1997).

"The use of classification scheme gives some advantages to an Internet service... If an existing classification scheme is chosen, it will have a good chance of not becoming obsolete and will possibly be well-known to users" (Koch & Day, 1999).

### **Abstract**

The author explores a number of theoretical and practical issues concerning classification of remote electronic resources. The necessity to catalog electronic resources has raised questions regarding including and classifying them in library catalogs. Analysis of current cataloging decisions indicates that there is no single standard that has been formed. Some libraries assign classification numbers to remote electronic resources and others do not. Theoretical discussions of cataloging librarians show that there are a number of important issues related to using classification for the cataloging of electronic materials. The traditional purpose of classification for locating physical objects on the shelves loses its function in the case of remote electronic resources. The other function of classification, categorization, is more useful in the case of electronic resources, as it promotes the role of classification as a subject-organizing tool. Attempts at using library classification schemes in order to organize electronic resources has moved beyond the realm of libraries and their catalogs. Information scientists, database developers, and specialists in information retrieval have explored library classification abilities in organizing information on the Internet, in order to improve browsing and subject searching. Several projects have proven that such classification systems as Dewey Decimal Classification, Universal

Decimal Classification, and Library of Congress Classification can be useful in describing, organizing, and retrieving electronic resources.

## **Introduction**

The problem of classifying electronic resources may be approached in two different ways. One approach in classifying electronic resources is to focus on the library setting. Questions concerning inclusion and classification of Internet resources in library catalogs by catalogers and classification specialists raise many specialized issues. Another approach takes the same problem away from libraries' walls and directs it toward computer science specialists, web developers, database creators, etc. People who organize Internet resources must make decisions concerning the use of traditional library classification systems as one of the means for the organization of online information, or whether alternative approaches should be used.

In researching the problems of electronic resource classification, many sources were found to be useful. McRee's (2000) review of the AUTOCAT discussion concerning classification of Internet resources reflects the points of view and opinions of many experienced catalogers-practitioners. Koch & Day's (n.d.), "The Role of Classification Schemes in Internet Resource Description and Discovery," points out problems of practical usage of various classification schemes for organizing Internet materials. Diane Vizine-Goetz's (1999) "Using Library Classification Schemes for Internet Resources," illustrates the steps that are performed by librarian-researchers in implementing Dewey Decimal Classification (DDC) as a tool for organizing resources on the Internet, and their desire to adapt and spread this well-known traditional tool into a new environment. These and other sources, including standards, guidelines, reports, articles, and descriptions of projects, help in identifying some important aspects of the problem and their context. These materials document the ideas and points of view of international catalogers and the information science community.

### **Part 1. ELECTRONIC RESOURCES IN LIBRARY COLLECTIONS**

#### ***Internet resources in library catalogs: the current practice***

The question of how Internet resources should be classified is an interesting problem for theoretical discussion amongst catalogers. Partly as the result of these debates, librarians and other

information professionals have reached practical decisions. An observation of current practices regarding classification of Internet resources in library catalogs is helpful in understanding these practical issues. In order to understand these issues, some library catalogs were searched, beginning with the online catalog of the Library of Congress.

The Library of Congress assigns classification numbers (but not call numbers) to Internet resources. The "set search limit" button on their search menu is used in order to limit the type of materials to computer files only. Only "title" and "serial title" (not "keyword" or other fields) can be searched. This limits possibilities for researchers.

A search for the title "dialog" (search limited to "computer file") retrieved 5 titles. All results were in CD-ROM format. A search for the title "Yahoo" retrieved one title: "Yahoo!-education". The type of material was classified as "Computer file". Notes in the record stated "Title from home page as viewed on July 5, 2001" and "Mode of access: World Wide Web." LCC number "L11" was also in the bibliographic record, but the call number field stated "Electronic Resource". This reflects the fact that this material is not shelved and, accordingly, does not need a call number. Other information in the record included the URL, which is repeated in the field "Links."

The next search conducted was for the term "online" (search limited to "computer file"), which retrieved 72 titles. The results included materials that were and were not computer files. Other searches determined that Internet files and CD-ROMs could not be retrieved separately. Because of this, limiting by type of material seemed unreliable. If the material is available in print or CD-ROM format, as well as in the form of an Internet file, it has an assigned call number and a note stating "additional form available." If material is cataloged as an Internet file only, a class number is assigned and the call number field is noted "Electronic Resource."

The University at Buffalo Library catalog allows users to find computer files by using something called a "Command search". In order to find computer files, the search should include k (keyword) and d.fmt (data format) commands. After performing such a search, 11,655 records were retrieved. Fields included in these records were: LOCATION (Access through Internet) and no CALL NUMBER or CLASS NUMBER. There were many useful fields

including notes, URL, data of access, subject headings and even type of file, such as HTML-encoded text and graphics, or PDF.

The Columbia University Libraries catalog can be searched for computer files by using its Advanced Keyword Search function, where there is a "format" field. A keyword search for "computer" in any format retrieved 1,000 records. Most materials were in CD-ROM format, or in print format with a CD-ROM. Those that were in print format had assigned call numbers. The Internet files had assigned call numbers, as well.

The MIT Libraries catalog contains two limiting fields in its Advanced Search function for the retrieval of Internet resources: 1) format (electronic resources) and 2) location (Internet resources). Using both fields, 1,357 records were retrieved for "computer" (keyword anywhere). Using only the format field, the same 1,357 records were retrieved. A search limited to the location field retrieved 1,356 records. These results included those materials available online only (with no call or class numbers), and those available online and print (with call numbers for printed copies).

The University of Michigan Libraries catalog (MIRLYN) retrieved records for Internet files in its advanced search mode, when limiting in the format field to "computer file". The holding information for "Yahoo" included "ELECTRONIC ACCESS" in the location field and "See URL for access" in the call number field. Other computer files contained the same type of information. For these reasons, it appears that the University of Michigan does not classify the Internet resources.

Wayne State University uses an advanced technique in order to retrieve electronic resources. It consists of only one catalog, which allows cataloged Internet files to be browsed. Moreover, the field "Internet Address (URL) Browse" is one of the options in the catalog's basic search, as are numbers such as LCC, DCC, SuDOC, local numbers, and ISBNs, and ISSNs. The advanced search allows users to search materials by "Computer File Title" and "Internet Address (URL)". Electronic resources can also be retrieved with the help of two limitations: by format (the field has a very rich submenu, which make available e-books and electronic resources, as well as software packages and programs) and by location. In the retrieved records one can see "Location - WSU Libraries Electronic Resources" and "Call# - ELECTRONIC RESOURCE, Note - To access this title, please click on the above URL". Through these examples,

it appears that WSU provides a number for means of retrieving Internet files, but does not classify them.

The Library of Michigan catalogs Internet files. They can be retrieved by conducting keyword or subject searches, and can be searched by name or title fields if these parameters are known. A search is performed in two steps: a keyword, author, title, or subject search (for example, k= "electronic resource") brings statistics, retrieves some records, and opens a submenu, which includes the option of limiting the search. Clicking on this option opens a new submenu, which includes year of publication; material type; language; publisher; location; and words in author, title, or subject fields. Internet resources can be retrieved by limiting to location (electronic resources, e-book) and material type (computer file). When one accesses an Internet file by using its URL, the catalog remains in the background. Clicking on the button "return to catalog" allows for more searches of the catalog. The catalog is user-friendly and efficient. For example, it is easy to separate searches for CD-ROMs from searches for Internet files.

The Memorial University of Newfoundland Library catalog (SIRSI) has records for electronic resources that include multiple fields for electronic resource retrieval, but these resources are not classified. The default for call number and location fields is: "Call Numbers for: EXTERNAL, Location - Internet."

The San Francisco Public Library's catalog can be searched by material type. Though this is true, "Electronic Resources" or "Internet Files" can not be found in this list of material types. However, the catalog does include records of e-books from NetLibrary, which have classification numbers.

The Orange County Library System (Orlando, Florida) does not catalog online resources. It has an e-book collection available electronically via NetLibrary. The MARC records for these books include classification numbers. The Kansas City Library Consortium does not catalog online resources.

Through this investigation, it seems that a standard policy for classification of electronic resources has not yet been implemented in the libraries surveyed. Some libraries assign class numbers to Internet files, some assign call numbers, but more often call numbers are only assigned to materials available in more than one format. The majority of libraries do not classify Internet materials;

they just provide the location (online, electronic resource, "see URL" etc.) This is not surprising. Cataloging Standards, applied to the cataloging of electronic resources, allow catalogers to be flexible and feel free to make their own decisions. The Library of Congress's (n.d.) CONSER Cataloging Manual Module 31 Remote Access Computer File Serials Part 2 31:18 states, "While classification is not required in CONSER records, libraries that normally classify their serials are encouraged to also classify electronic serials. Though not needed as a location device, classification provides a useful tool for assessing the types of serials that are online and for many other purposes." However, examples presented in this document show that classification numbers are not included in the MARC records of electronic serials. Module 33.18 for electronic newspapers also did not include fields for classification. The Library of Congress's (n.d.) Draft Interim Guidelines for Cataloging Electronic Resources, reflects interim rules for treatment of electronic resources. Current practice shows that many catalogers (mostly in large university libraries) catalog electronic resources but do not classify them. Public libraries typically do not catalog Internet files. Some have e-book collections, which are classified according to rules of standard bibliographic descriptions.

### ***Theoretical discussions: pros and cons***

Library classification can be seen as a categorizing device used to organize knowledge into classes, or as a notation assigned to physical information packages (Taylor, 1999). Practical necessity to organize library materials, by way of shelving books and periodicals seems to have given priority to the second function of classification, notation. This leaves the first function (categorization) in the shadow. In a library setting, the main role of classification is to help users locate materials on shelves. If a user knows the call number, he/she will be capable of finding the necessary material quickly, even if a library collection includes millions of items. When users browse the shelves, they can find related subject material on the same shelf. In this case, library users do not need to know all the call numbers because all subject related books are easy to find next to each other. But the correlative importance of these two functions has changed with the development of the Internet. Library users no longer have the need to go to the library or check materials out, in order to find information. Instead, all that the user needs is a networked computer. In this situation, the function of classification as a

locator of a physical object at a certain place loses its importance. What about the categorizing function? Is it still important for Internet users? Can classification be useful as a tool for Internet subject retrieval? Questions concerning the usefulness of classification schemes for organizing Internet resources become questions about the significance of the function of categorization.

If the primary function of library classification is seen as a means of helping users find physical objects on the shelves, classification will not play a big role in the organization of Internet resources. But if one sees classification as a tool, which functions to bring together materials on related subjects, the role of classification of Internet resources becomes more important. Of course, the topic is not limited to determining what the role of classification is, exactly. For example, economical considerations are also important. Cataloging is a time consuming and expensive process. Do the profits of classifying electronic resources compensate for the time, energy, and money, which are likely to be spent on this type of classification? In order to answer these questions, practitioners' opinions are very important. What do those who routinely spend long hours classifying various kinds of materials think about the idea of electronic resources classifying?

The AUTOCAT 1999 discussion determined the main pros and cons of catalogers' opinions regarding the classification of electronic resources. Those who believed that electronic resources should be classified were mainly focused on library users, library collections, and librarians themselves. The first of their arguments in this specific discussion was that Internet resources should be classified because it is a means for libraries to extend their collections and provide access to online materials.

When libraries include these types of electronic resources in their catalogs, they extend their collections, often without spending any extra money. For users, remote access to these resources is even easier than access to traditional library materials. When a user finds a book in the catalog, he/she must go to the shelves, find the book, go to the circulation desk for check out, and so on. When a user finds the electronic file in the catalog, he/she can reach the source by simply clicking the URL. Traditional cataloging includes the assignment of call numbers. In the case of cataloging electronic resources nothing special has to be done. The bibliographic description should include all traditional fields such as author, title, publisher, and subject headings. Call numbers are assigned. Of

course, electronic resources do not shelve, and the main traditional meaning of call numbers--to lead to physical information packages on the shelves--loses any sense. Yet, call numbers are meaningful as information retrieval tools. Most OPACs can be searched by call numbers. When Internet files are included in OPACs, users then have the "Ability to browse all of these resources along with our physically available material together on 'virtual' shelves"(McRee, 2000).

The second argument made in the discussion on the side of classifying electronic resources was that doing so helps librarians in collection development, production of bibliographies, communication with vendors and supported services, and then that new material organization, indexes, discipline information, and list of available references could be compiled easily by using call numbers. According to McRee, "Many of the collection development software programs...use classification as a means of evaluating one's collection." Arguments against the classification of electronic resources in the discussion were based mostly on economic reasons. Some catalogers held the opinion that classification of electronic resources does not make much sense for two main reasons. First, only very few patrons search materials by call numbers. Statistics show that call number searches comprise only 1.27% of all searches (McRee, 2000). Because of this, it seems that the time, effort, and cost for classification would be poorly compensated by demand for this feature. The second main reason given during the discussion, against the classification of Internet resources, was that patrons would be confused to see call numbers for these types of resources, and would not be able to find the materials on the shelves.

Several other important ideas were pointed out in this discussion. First, it was clarified that Internet resources might be classified, but that it does not make much sense to assign to them a call number. Second, it was determined in the discussion that accessibility of Internet materials would be enhanced with the help of classification. Third, the participants of the discussion considered the problem of single or multiple records for materials, when a library owns them in both print and electronic formats.

Through the ideas presented in this AUTOCAT discussion, it can be concluded that the inclusion of classified Internet resources into OPACs has a number of positive consequences and useful functions. The classification could be helpful for users (access to



online information), librarians (collection development, bibliography, collection evaluation), and libraries (collection enrichment). The discussion also assisted catalogers in clarifying some practical and theoretical questions at a time when their libraries started to actively include Internet files into their catalogs.

## **Part II. ORGANIZING ELECTRONIC INFORMATION**

### ***Classification of Internet resources***

Another important part of this issue is the organization of online electronic information. Since the development of the Internet, the issues surrounding the organizing and retrieving of electronic information have been very important. This worldwide network of computers has created an electronic environment, which includes a huge, rapidly changing universe of texts, numbers, symbols, and images. In the organization of these resources and the separation of them from "information noise," demarcation of genres and forms has become the major obstacle.

This problem has concerned many experts in the field of information science: Web developers, database creators, computer engineers, linguists, librarians, and specialists in information retrieval. As often happens, there are two main approaches to this problem: the traditional and the non-traditional. The traditional approach supports adaptation of given resources for new purposes. An example of this would be use of traditional classification schemes for the organization of networked electronic information. The non-traditional approach offers new initiatives and inventions in hopes of solving the problems of classifying electronic resources. As often happens, supporters of the two approaches include specialists and individuals from different fields.

Information science and computer specialists find the non-traditional approach more natural. First, they are not trained librarians accustomed to bibliographic standards, including classifications. Second, they have their own methods, unfamiliar to librarians. Third, sometimes as representatives of an innovative field of knowledge, they are ideologically oriented toward advancement and invention. The non-traditional approach to organizing Internet resources offers automated non-cataloging methods of organization of these resources. Search engines, Web crawlers, keyword searching, automatic indexing, categorization, thesauri - these tools, among others, help to organize Internet

resources by using special software. The participants in these projects are typically computer specialists, librarians, and linguists. Clifford Lynch (1997), a computer science specialist who works in library setting, mentions the usefulness of cooperation between computer and library scientists in organizing the Internet: "The librarian's classification and selection skills must be complemented by the computer scientist's ability to automate the task of indexing and storing information. Only a synthesis of the differing perspectives brought by both professions will allow this new medium to remain viable".

Thus, librarians and their traditional tools - indexes, catalogs, and classifications - are necessary for success. One of the advantages of having a set of available traditional classification tools is the ability to adjust given standards to new electronic materials. It is labor intensive, and one can see many large projects, national and international initiatives with goals to add electronic materials to AACR2r, MARC, ISBD, CONSER and other cataloging standards. The guidelines for them have begun to be developed. Not only these, but entirely new standards have been created, as well. An example is the Dublin Core metadata schema, which developed from experimental initiative to an American national standard for cataloging of computer network resources and organization of information.

Librarians have good reasons to rely on classification, as it is a traditional, well-developed, useful tool for the organization of knowledge. Generations of librarians have put in much time and effort in order to create reliable, detailed, and effective classification systems. The combined efforts of many organizations, numerous projects, and countless hours of thousands of catalogers and classification specialists have been spent in order to improve classification systems and adjust them to the pace of growing knowledge. Common sense prompted the idea of examining the possibilities of classification for organizing the Internet.

In 1983 Elaine Svenonius expressed her opinions concerning the possible use of classification standards for retrieval of electronic resources. She believed that classification could improve precision or recall, provide context for search terms, enable browsing, and serve as a mechanism for switching between languages. These ideas were tested in the international project DESIRE. The authors of this project write:

The library community, over many years, had appeared to favor subject indexing systems (the use of a controlled vocabulary to assign indexing terms to documents) over the use of traditional classification schemes (grouping documents into a hierarchical structure of subject categories). During the first period of the development of networked information services, many specialists, especially those from the computing community, also questioned the value of library subject description systems in principle, pointing to the accomplishments of full-text indexing software. The increasing use of the Internet and the World Wide Web (WWW) for the storage and retrieval of vast amounts of information has, however, changed this perception. Two distinct ways of finding resources on the Internet emerged. One approach consisted of the development of robot based search engines which could be used for powerful keyword searches of the contents of the WWW. These are extremely useful tools, although they have a tendency to return large amounts of irrelevant information. The other approach started with producing "hotlists" which would encourage users to browse the WWW. The production of hierarchical browsing tools sometimes led to the adoption of library classification schemes to provide the subject hierarchy...proposed DESIRE test-bed services currently use a classification scheme which can be browsed. (Koch & Day, 1999)

The authors of this report investigate advantages and disadvantages of classification systems for description and retrieval of Internet resources. Their analysis of Internet services, which use different classification systems, allows them to reach the conclusion that classification can provide improved browsing and subject searching across databases. In particular, classification assists in browsing, particularly for inexperienced users or those who are not familiar enough with a certain subject; broadening and narrowing searches by following a hierarchical structure of classification; searching terms in context; and in providing multilingual access to resources.

The last point sounds especially promising. Because classification's notation is separate from a specified language, a searcher would be able to enter search terms in a given language and those terms might then relate to the relevant parts of the classification system

(as a switching language). They might then be used to retrieve resources in any given language on the subject (Koch & Day, 1999). Thus, many information professionals believe that library classification systems are useful as a means of organization electronic information. They can help to describe, organize, and retrieve Internet resources.

***Dewey Decimal, Universal Decimal Classification,  
and Library of Congress Classification on the web***

The results of the survey conducted show that classification systems can be useful for the organization of electronic resources. There are a number of Internet services that use classification systems. Classifications -- international, national, local, general and specialized, alphabetical, numeral or mixed -often coexist and compete with each other on the Internet. Some examples of Internet sites that have organized their information by using DDC, UDC and LCC classification schemes can be found on the Iowa State University "Beyond Bookmarks" website: (<http://www.public.iastate.edu/~CYBERSTACKS/CTW.htm>) [Opens a new window].

A survey of the sites included in this list reveals that the most popular classification scheme on the Internet is DDC. Many organizations and research institutions have tried to edit and develop DDC according to contemporary needs. Today, one of the most important needs is to make the system work in an electronic environment. For example, OCLC researchers have tried to promote DDC for many years. One of their goals has been to develop machine-readable form of DDC. As the result, Electronic Dewey and Dewey for Windows were created. Diane Vizine-Goetz and Joan Mitchell, editor in chief of the Dewey Decimal Classification, have outlined five research areas as especially important for increasing the role DDC as an online organizing tool: "Developing customizable views of the DDC, enhancing links to other thesauri, improving links to other editions, transforming the captions into end-user language, and decomposing numbers and using the parts for improved access" (Vizine-Goetz, 1999). These ideas were developed further in the OCLC CORC project, where DDC numbers were included for electronic resources. Vizine-Goetz has written about progress in research areas such as links between DDC and LCSH, the comparison between DDC vocabulary and other subject vocabularies, and end-user displays of DDC captions. All these and other enrichments and improvements serve the same goal: to

promote DDC's use into a new electronic environment. All these efforts increase DDC's abilities to serve as a categorizing aid.

### **Conclusion**

Classification systems are used in many library catalogs and in some Internet sites.

Libraries use classification in order to improve access to Internet files; to bring related subjects together; and to help librarians in collection development, bibliography compiling, communication with vendors, and supported services. Classification of Web sites assists users in browsing, broadening and narrowing searches, and in viewing information in context. In addition, classification creates the potential for multilingual access to a collection.

Theoretical analysis and practical usage demonstrate the importance of classification as a categorizing device and minimize its role as a notation that assists in locating materials on shelves. Classification can now be seen as a provider of subject access to information in a networked environment. Traditional library classification has survived in severe competition with non-traditional means of organization of information. Efforts of many librarians and information specialists in utilizing classification for new purposes and implement it into networked environments have been successful. This old tool of organization of information is very much alive.

### **References**

Jones, W., Ahronheim J.R., & Crawford, J. (Eds.). (2002). *Cataloging the Web: Metadata, AACR, and MARC 21*. Lanham, MD: Scarecrow Press.

Koch, T. & Day, M. (n.d.). *DESIRE - Development of a European Service for Information on Research and Education: The role of classification schemes in Internet resource description and discovery*. Retrieved from <http://www.ukoln.ac.uk/metadata/desire/classification>

Langridge, D.W. (1992) *Classification: Its kinds, elements, systems and applications*. London: Bowker.

Lynch, Clifford. (1997) The Internet: Bringing order from chaos [Electronic version]. *Scientific American Special Report: Searching*

*the Internet*. Retrieved from  
<http://www.sciam.com/0397issue/0397lynch.html>

McRee, E.J. (2000). Classification of Internet resources: an AUTOCAT discussion [Electronic version]. *Cataloging & Classification Quarterly*, 29(4), 19-38. Retrieved from  
<http://www.slc.bc.ca/mac/net.htm>

McKiernan, G. (2001). *Beyond bookmarks: Schemes for classifying the Web*. Retrieved from Iowa State University Web site:  
<http://www.public.iastate.edu/~CYBERSTACKS/CTW.htm>

Miksa, F.L. (1998) *The DDC, the universe of knowledge, and the post-modern library*. Albany, NY: Forest Press.

Olson, N.B. (Ed.). (1997). *Cataloging Internet resources: A manual and practical guide*. 2nd ed. [Electronic version]. Dublin, Ohio: OCLC. Retrieved from  
<http://www.purl.org/oclc/cataloging-internet/>

Svenonius, E. (1983). Use of classification in online retrieval. *Library Resources and Technical Services* 27(1), 76-80.

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

United States Library of Congress. (n.d.). *CONSER cataloging manual module 31: Remote access electronic serials, part 2*. Retrieved from  
<http://www.loc.gov/acq/conser/mod31pt2.html>

United States Library of Congress. (n.d.). *Draft interim guidelines for cataloging electronic resources B19.1.2*. Retrieved from  
[http://lcweb.loc.gov/catdir/cpso/dcmb19\\_4.html](http://lcweb.loc.gov/catdir/cpso/dcmb19_4.html)

Vizine-Goetz, D. (1999). OCLC research takes experience in classification to the Internet. *OCLC Newsletter*, 239. Retrieved from  
<http://www.oclc.org/oclc/new/n239/feature/08feature.htm>

Vizine-Goetz, D. (n.d.). Using library classification schemes for Internet resources. *OCLC Internet Cataloging Project Colloquium Position Paper*. Retrieved from  
<http://www.oclc.org/oclc/man/colloq/v-g.htm>