DOCUMENT RESUME

ED 212 286 IR 009 993

AUTHOR Mandel, Carol A.
TITLE Subject Access in the Online Catalog.
SPONS AGENCY Council on Library Resources, Inc., Washington, D.C.
PUB DATE Aug 81
NOTE 31p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cataloging; Databases; *Indexing; *Library Catalogs; Library Research; Man Machine Systems; *Online Systems; Subject Index Terms
IDENTIFIERS Failure Analysis; Free Text Searching; *Library of Congress Subject Headings; Preserved Context Indexing System; *User Needs

ABSTRACT

This review of the research on subject access to library collections focuses on the problems of and prospects for improved online subject access to library collections. Summaries of the general findings of studies on library catalog use and catalog users and some reasons for the frequent failure of subject searches in library catalogs are followed by a discussion of the use of "failure analysis" as a technique in studies of automated information retrieval systems. The advantages and disadvantages of free text searching are reviewed, the feasibility of using the Preserved Context Indexing System (PRECIS) to supplement Library of Congress Subject Headings is briefly considered, and some of the conclusions drawn from studies of library users' needs are presented. Research on the enrichment of cataloging records using free text descriptors, on enhancing currently used subject access systems such as Library of Congress Subject Headings, and on ensuring the effectiveness of the user interface with an online catalog is also discussed. Six recommendations are made for the improvement of subject access in online catalogs, and a 41-item reference list is included. (JL)

Subject Access In the Online Catalog

A report prepared for the
Council on Library Resources
by

Carol A. Mandel
with the assistance of
Judith Herschman

August 1981

ACKNOWLEDGEMENTS

The authors wish to thank the many individuals who took the time to talk with us and to share with us the results of their research and experience. We are particularly grateful to Pauline Atherton Cochrane, Laura Kassebaum, and Robert Zich for their help and advice.

TABLE OF CONTENTS

INTRODUCTION

New technologies are often adapted to traditional uses without fully exploiting added capabilities. To take full advantage of new developments, careful planning is needed. Librarians and information scientists are particularly conscious of the need to apply rapid changes in computer and communication technologies to expanding the ability to store, manipulate, and retrieve information.

Librarians recognized early in their use of online cataloging systems that the computer was not only a very sophisticated catalog card production machine but a device for retrieving bibliographic information in entirely new ways. The difficult question was, and still is, not whether to exploit the computer for the benefit of library users, but how. The question has been considered in relation to a variety of library catalog formats during the 1970's, but now it is clear that during the 1980's the most widely used format will be direct user interaction with online public access library catalogs. The nature of the use of the many bibliographic databases that have been made accessible online through commercial services indicates that computerized records and the software that manipulates them permit much more powerful searching strategies than do traditional card catalogs. This has prompted questions about the kinds of searching techniques necessary and desirable in online library catalogs and, more important, whether librarians are seriously limiting their new potential by loading only traditional bibliographic records into computerized catalogs.

These questions are particularly pressing in relation to searching for materials on a particular subject or for items for which the author or title are only dimly remembered. Such materials are sought by means of key words or subject terms; in

traditional American library catalogs this usually means the first word of a title or a Library of Congress Subject Heading (LCSH). The latter has been widely criticized by librarians and there is considerable concern that the present limitations of the LC subject system should not be carried over into the era of online library catalogs. An often-quoted statement of the problem was made by Bates in 1977:

> If we simply transfer the austerity-based LC subject heading approach to expensive computer systems, then we have used our computers merely to embalm the constraints that were imposed on library systems back before typewriters came into use![1]

This paper examines research on subject access in light of problems of and prospects for providing online subject access to library collections. "Successful" subject searching can only be defined in terms of the objectives of the access system and the expectations of the reader -- e.g., is an exhaustive bibliography desired, or only a few select books on a topic. Since the emphasis of this paper is on the kinds of access traditionally provided through library catalogs, it is assumed that the objective of a subject search in a library catalog would be: 1) to lead the reader from the topics he or she has in mind to the relevant vocabulary terms available in the catalog; 2) to provide the reader with records for most (80%) of the books in the system on the topic in question (but not necessarily to parts of books); and 3) to provide the reader with enough information to decide whether or not to call for the item identified by the search.

This is a very modest set of objectives and may not be acceptable to librarians who believe it is important to provide in-depth subject analysis to library users. However, this paper is concerned not only with what constitutes the best possible

means of subject access, but with political and economic considerations that are likely to affect decisions regarding future library catalogs. While a variety of methods for retrieving subject information have been used successfully in systems designed for specialized subject areas, they cannot be applied directly and immediately to library catalogs. The transition from present library methods of subject analysis to new forms of access accepted and applied by libraries will be a gradual process, accompanied by testing and experimentation. The purpose of the paper is to suggest areas where the Bibliographic Service Development Program might initiate or support efforts that will help research libraries improve subject access through online catalogs.

RESEARCH ON SUBJECT SEARCHING IN LIBRARY CATALOGS: FINDINGS AND METHODS

General Findings

Studies of catalog use and catalog users provide an overview of who uses the subject catalog, how often subject searches are successful (using varying definitions for "success"), and how persistent subject searchers are. The bulk of the research, of course, describes the manual catalogs that have been the major method of accessing library materials for over a century; however, some work has been done on machine files as well. Even the research done on manual card catalogs provides insights for planning online catalogs because it is important to understand the use made of the library bibliographic record. Library records contain a standardized, limited set of data elements rather than the descriptors, abstracts, and even full texts available in other kinds of files.

The basic findings may be summarized as follows.

1. Subject heading searches are sometimes used to identify items already known to the searcher;[2] conversely, some "known-item" searches are searches only for subject information.[3]

2. Although "known-item" searches account for more card catalog use than subject searches, the proportion of subject searches varies with the user population. Several studies demonstrate an inverse relationship between the amount of subject searching and the user's level of expertise. In a recent study at Dartmouth, only 28.6% of the faculty surveyed reported that the subject approach was the search method they used most often, as compared to 51.4% of the undergraduates questioned.[4] This may have changed significantly with the introduction of detailed subject searching in online catalogs at Dartmouth, but the new data have not yet been analyzed.[5]

3. Users often select terms that are either too broad or too narrow.[6] Separate subject heading lists, such as the LCSH, are rarely used to identify terms for searching, even when the lists are placed near an online catalog terminal.[7,8]

4. About half of the terms used by readers in their first try at the subject catalog correspond to either a heading or a reference found in the catalog. If subsequent tries are included, the success rate rises to about 70%.[9]

5. Not all users persist in subject searching until they are successful. Between two-thirds and three-quarters of the searches in manual subject catalogs, whether successful or not, do not continue beyond a single look-up.[10] There are some indications, however, that users might show greater perseverance when using an online catalog.[11]

6. Online searching in a small database is considerably more

successful, both in terms of number of relevant documents found and search time per useful document, when additional descriptive terms taken from indexes and tables of contents of books are added to the MARC record and made accessible.[12] This approach has not yet been tested in a large database.

7. Library catalog users think that more access points, both subject headings and key words, should be added to records for books.[13] Standard library cataloging practice currently results in an average of slightly under 1.5 LC subject headings per record.[14,15]

Reasons for Search Failures

Very little is known about the reasons why 50% of first attempts to seek a term in the subject catalog fail, although there is general consensus among librarians, and some evidence that the lack of specificity in LC subject terms and the lack of "see" references in library catalogs are the major contributing factors. The 50% "hit-rate" for terms used by the reader is, prima facie evidence that the entry vocabulary of library catalogs is inadequate. In other words, the natural language that expresses readers' requests is not mapped, either through cross references or sufficiently convenient displays in the thesaurus used, to the terms appearing in the library catalog. A rich entry vocabulary is not inexpensive to maintain, but has been demonstrated to be cost-effective because it greatly reduces the intellectual burden on both the cataloger and the searcher.[16]

Lack of specificity in LC subject terms as the cause of many subject search failures is more difficult to demonstrate conclusively. An early library catalog study and more recent studies of information retrieval systems have shown that, in general, material in subject areas with more abstract language

(e.g., education) is more difficult to access with precision than items in areas with relatively "hard" languages (e.g., chemistry).[17] A study done by Lipitz at Yale, where LCSH are used for subject access, found that "users engaged in subject searches frequently complained that subject sections in the catalog are much too large and general, rarely narrowed to cover only the particular subject aspect of interest to the user."[18] The complaints of Yale users are substantiated by a recent analysis of a sample of books classed with LC classification and given LC subject headings. The analysis demonstrated that in a number of classification ranges, subject headings did not add appreciably to discriminating among all of the items assigned the same class number. This led the investigator to conclude that "in these areas the reader could do just as well (or better) at the book shelf than they could in the library's catalog."[19]

However, the frequently-voiced complaints regarding lack of specificity in LCSH do not necessarily reveal that the actual vocabulary of the list is the cause of the search failure. For example, readers' requests often may be too specific to be met by monographs indexed as a whole, even though American library catalogs generally do not contain subject entries for parts of books. This difference between library policy and reader's requests could account, at least in part, for the demonstrated superiority of a system that adds information derived from the indexes and tables of contents of monographs. In addition to policy decisions, flaws in indexing practice may account for searching failures. After being shown a number of examples of overly-general terms assigned by LC catalogers, Edward Blume, (then Head of the LC Subject Cataloging Division) noted, "Subject headings can be created as needed, but often catalogers choose not to do so. Many of the bad examples of LC subject indexing cited by various speakers are not examples of the limitations of the system as such, but rather examples of extremely bad cataloging."[20]

Failure Analysis

The inability to specify the precise causes of failure in library subject searching illustrates a methodology problem in research on indexing systems recently described by Svenonius.[21] She points out that comparisons or evaluations of systems are studies of "aggregate variables," overlooking the separate elements that make differing contributions to the success or failure of the system. Even studies restricted to the indexing language miss specific features that may account for essential differences. While Svenonius's call for more theoretical research in this area is apt, existing methods for evaluating indexing languages by comparing terms in the system with actual request statements could also provide useful results in studies of subject access in online catalogs.[22] The online catalog provides the opportunity to monitor the frequency of use of different types of search terms, to observe which of the terms appearing in the database are used in searching, and to enumerate and analyze those terms used in searching that do not match terms accessible in the database.[23]

General performance measures used without analyzing the reasons for search failures do not provide the information needed to make decisions leading to improving a system. "Failure analysis" is commonly done in studies of automated information systems since machine searching can provide a step-by-step record of a search without inconveniencing the searcher. Although the results of such studies as reported in the literature are specific to the systems under scrutiny*, King has noted that search failure can be expected to fall into the following categories:

*It is easier to generalize the results of library catalog studies because subject access mechanisms in libraries are fairly standardized. However, differences among libraries, such as policy in providing cross references, may often be underestimated.

1. failures of policy or practice in indexing;

2. failures in the vocabulary used, usually due to lack of specificity or to ambiguous or spurious relationships between terms;

3. failures in searching strategy;

4. failure to reflect accurately the user's information need in the search request.[24]

Thorough failure analysis involves examination of the document missed, indexing records, requests, search strategies, and the users' relevance assessments. In-depth analysis also can be informative when both successful and unsuccessful searches are compared.

Free Text vs. Controlled Vocabulary Searching

Many of the other techniques and performance measures that are well developed for evaluating automated information systems, such as measures of precision, estimated recall, and search time per relevant document found, can be applied to evaluate subject searches in both online and manual library catalogs. Studies employing these methods might cast some light on the relative merits of free text vs. controlled vocabulary searching on bibliographic records for books, a question often debated by librarians. The MUMS and SCORPIO systems at the Library of Congress provide an ideal opportunity to compare the capabilities of two different software systems, one limited to searching exact subject headings (often phrases) and the other accessible by subject word. Both are used to search one database. However, comparisons of free text searching with controlled vocabulary searching have been applied to a variety of databases and systems and invariably lead to the same conclusions: a combination of

both is best, with the optimal mix dependent upon the specific features of the database, the system, and the user's requirements. Vendors of database search services have found that their customers demand the fullest range of search possibilities available; BRS, Inc. has found that this same request is made by the libraries for which it provides online public catalogs.[25] In this case, the test of the marketplace has indicated the desirability of flexible free text searching of records for books.

Developing software that permits free text searching is a simple task in comparison with providing the data elements to be searched. Library bibliographic records are not rich in searchable words. Other than the words in the title of a book, and very occasionally a contents note, few useful terms exist in the traditional library record beyond those added as specific access points by a cataloger. Atherton created a database of MARC records enriched by descriptive terms from the tables of contents and indexes of the books represented. She found that the enriched database, referred to as BOOKS, was clearly superior for subject searching. Atherton's work has been widely publicized and well received, yet libraries have not made efforts to enhance their bibliographic records as she suggested. One possible explanation for this inaction is that practitioners believe that further testing and demonstration of the value of such enrichment is necessary. Another is that the enhancement process adds a workload that cannot be absorbed economically. If this is the case, it would be useful to know whether it is possible to sacrifice the LC subject headings presently being applied <u>in exchange</u> for the uncontrolled vocabulary terms available in the BOOKS database. This could be done by analyzing the results of the BOOKS searches and eliminating matches made only on the LC subject heading portion of the record. Economic realities may make "trade-ins" a more realistic possibility than acquiring an additional vehicle for subject access.

PRECIS

A "trade-in" often debated by subject catalogers is the substitution of PRECIS strings for Library of Congress subject headings on catalog records produced at LC. At the request of the ALA Subject Analysis Committee, the Library conducted a study in 1977 to test the feasibility of adding PRECIS strings to LC records. The relative merits of the two systems for use by those seeking information were not addressed. Such a comparative study, although difficult to design, would be of interest to many librarians. However, the conclusions drawn by the Library of Congress in 1977 indicate there is no pressing need to conduct such a study. The Subject Cataloging Division determined that

> there has been no public demand that the Library of Congress either replace the traditional Library of Congress subject headings with PRECIS, nor to add PRECIS strings to traditional catalog cards or MARC tapes ... In view of the fact that the addition of PRECIS strings to all current cataloging would cost approximately $1,000,000 per year and that there has been no demand to do this, the Library of Congress will not seek money from Congress or from any other source to maintain two subject heading/indexing systems.[26]

The LC study is an important reminder that steps taken to improve subject access must realistically assess relevant economic and political considerations.

Studies of Users' Needs

Quantitative measures of existing systems for subject searching also need to be supplemented by behavioral science methodology to provide qualitative assessments of the needs,

perceptions, and level of satisfaction of the systems' users.[27] A major study being conducted by the OCLC Research Department has employed the focused-group interview technique to determine "library users' perceptions, expectations, and criteria for success in using the subject catalog."[28] The study is intended to provide the designers of online public catalogs with descriptions of the features that will support and enhance the present subject search tactics of library users. Although detailed analyses are presently being applied to the results of 200 individual interviews and 13 group interviews, the investigators have already been able to describe some of the desirable features of an online subject catalog. These include:

1. additional access points, including key words in titles and added subject headings describing both the whole book and its chapters;

2. online display of a thesaurus to help searchers choose broader, narrower, and related terms;

3. the ability to define searches with Boolean logic;

4. the ability to delimit searches by: a) date, b) inclusion or exclusion of conference proceedings, c) level of understanding required by the reader, d) fiction/nonfiction, e) language;

5. transparent (i.e., automatic) translation from the users' natural language to the terms used in the catalog;

6. additional descriptive information culled from the book (e.g., table of contents) that would permit browsing at the terminal rather than in the stacks to make relevance judgments.

It is worth noting that all of these features have been used

successfully to improve subject access in a variety of information databases. Librarians designing online catalogs are in the fortunate position of being able to capitalize on the developments and experiments made during the past decade in providing online access to specialized databases.

BUILDING ON WHAT IS KNOWN

Although there is much more research to be done on subject access, the preceding section also indicates that a great deal more is known than has been applied in library catalogs. A two-fold approach is needed to plan for the future: 1) continuing research to determine the most effective means of subject access that can be used in libraries, and 2) taking action to improve and enhance established methods of subject access.

Research on Enriched Records

Word-by-word (i.e., free text) searching throughout the record as well as subject heading (i.e., controlled vocabulary) searching of added entries has been enthusiastically used by most online catalog customers served by BRS. Delimiting searches by data elements in the record and applying Boolean and positional operators also make the most of data in bibliographic records. But many librarians argue that there simply are not enough descriptive words in standard records for monographs to permit adequate subject retrieval. The average bibliographic record for a monograph contains between one and two subject headings; journal articles indexed in common reference tools are given considerably more descriptors.[29] There is no rational intellectual justification for the discrepancy in the "index-term per page" ratio for books and articles; the explanation lies in

library economics and priorities. The most successful use of free-text searching occurs in databases containing abstracts or full texts, not just bibliographic citations. The possibilities for enriching databases of library records range in a continuum from methods requiring considerable effort or expense (e.g. preparing an abstract for each work cataloged) to minimal enhancements added automatically.

Atherton demonstrated that 300 additional descriptive words taken directly from the contents of a book and added to a MARC record considerably enhanced access to library materials. Although she devised an efficient method for this enrichment, it appears that no library or group of libraries is willing to pay the price of the added labor at this time. A much more modest method of generating additional descriptors might be to use a program to add to MARC records the appropriate terms from the LC classification schedules whenever certain class numbers appeared in the record.[30] As Atherton suggests, continued research is needed to test for the most effective and most practical enhancements to records. In addition, new ideas will need to be tested for acceptability in the library community and modified accordingly. Academic libraries, faced with budget cuts, are not likely to adopt cataloging practices that require additional labor. In fact, the last decade has shown greater reliance than in the past on standard Library of Congress records in most libraries. Convincing libraries that they need to enrich their records for subject access will require a large body of research and a well argued plan to show that the benefits of such enrichment outweigh the costs.

Research on Search Failure

As discussed in the first part of this paper, the reasons current library catalogs fail to respond to subject search requests are not adequately understood. For example, it may be

that the fundamental policy underlying library subject cataloging -- that of providing only subject headings coextensive with the entire book -- is at fault. Search failures must be analyzed (see section entitled "Failure Analysis") if users' actual demands on the library catalog are to be understood and systems to meet these demands designed.

Studies of online catalogs (such as those now under way) provide opportunities to gain insight into the causes for failure in subject searches through diagnostic analyses of a sample of search requests. This would require tracking the searches and results, asking two or three skilled searchers (e.g., reference librarians) to repeat the searches, showing any additional documents discovered to the reader for a relevance assessment, and analyzing the reasons why the additional documents were missed by the library user. The results will help individual libraries assess their own systems; if a pattern in the results pinpoints problems in indexing, indexing language, entry vocabulary, or general problems in searching strategies, the results will prove useful to the entire library community. If failure analysis is not built into the online catalog studies now in process, special studies should be undertaken.

Enhancing the Current-Method of Subject Access

Although its precise contribution to successful searching varies by system, a controlled vocabulary continues to be a valuable component in subject access systems. The controlled vocabularies of the many specialized bibliographic databases available online differ widely, a desirable condition for accessing material in special fields, but one that makes searching the databases difficult to master. Research library catalogs cover a wide range of subjects and most share a single, general vocabulary -- LC Subject Headings. An enormous investment has been made in these headings; there are literally

millions of these subject terms embedded in bibliographic record databases in the U.S. and Canada, and the Library of Congress is committed to supplying them for the bibliographic records it creates. (French-Canadian users of the University of Toronto Cataloging System, UTLAS, are even translating LC subject headings used in MARC records into French.) Online catalogs must help searchers take advantage of LC subject headings.

Librarians have learned that effective searching in subject catalogs requires consulting "the red book", the thesaurus of LC subject terms. Yet other catalog users rarely use this tool -- only 5% of SCORPIO searchers consult the list placed near LC's public access terminals.[31] Online interactive displays of thesaurus terms have proven successful in a number of retrieval systems and the OCLC study has found users to be enthusiastic about the idea of using such an aid.* Thus conversion of the LC subject list to an online thesaurus, mounted by networks, utilities, vendors, and other providers of online catalogs and capable of interrogation online by library users throughout the country, would add a powerful searching tool to online catalogs.

How might this conversion be accomplished? The answer depends partly on whether to aim for restructuring the terms in the LC list into a fully hierarchical arrangement or to settle for the more modest objective of a thorough editorial revision of the LCSH cross reference structure to bring it up to the standard of current thesaurus construction. Such an editorial revision was suggested by Angell in 1972 as a means of making the manual list more useful.[32] With the advent of the online catalog, the revision would lay the groundwork for an interactive display of

*Kaske and Sanders reported resistance by a few scholars to the idea. These few scholars feared the "rigid conceptual relationships" that they thought the online display would enforce. One suspects these scholars are reacting to the implications of the phrase "tree of knowledge" that was used by the researchers to describe the displays, and that the scholars would, in fact, find it useful to consult a "thesaurus."

narrower terms (NT, currently "see also" references in LCSH), related terms (RT, currently "see also" references in LCSH), and broader terms (BT, currently "see also from" notes in LCSH not made into references), in addition to "use" (currently "see") references and "used for" notes (currently "see from" notes). The addition of scope and history notes would greatly compensate for changes in terminology and for the common (and inevitable) practice of refining and narrowing subject headings without retrospectively subdividing files. By instructing searchers that earlier material on their specific topics may be found under certain broader terms, such notes would expand the recall power of the system. The revision might also include the addition of many specific references that are not now spelled out in the printed LCSH lists.

Restructuring LC terms into a hierarchical arrangement, a considerably more ambitious project, is also possible. A true hierarchical structure would facilitate experiments in vocabulary switching from LCSH to other thesauri as well as efforts to add terms to LCSH for use in special subject areas. It would also insure that the cross references displayed to users reflect accurate conceptual relationships between terms and would allow users to select terms from a hierarchical display.

The product of either a revision or a reconstruction (in either case, the actual subject terms used need not be changed) would be a database of subject authority information that would be displayed to both searchers and catalogers on online terminals. In addition to enhancing searching, the display would aid catalogers in identifying terms to apply and would ease future editing and maintenance of LCSH.*

*Catalogers find the UTLAS display of LC subject authorities so useful that a group of UTLAS participants cooperatively code and key all information from each issue of the LCSH Additions and Changes into the UTLAS subject authority file.[31]

Displaying of the existing LCSH headings and references will not, in itself, solve the problem of an entry vocabulary that matches only half of users' first tries. Even without making a single change in an LC subject term, access to the terms could be improved enormously by adding to the entry vocabulary, i.e., adding "see" references. This can only be accomplished through a system of continuous user-feedback; only the individuals at the user-end of a subject system can monitor and suggest those terms that are needed. This is not the fault of the cataloger, but is inherent in the process of working from a given book. In a study at the University of Chicago, library users were shown books and asked what subject heading they would supply for the titles. Participants did not check the catalog or a list of LC terms, yet very few of the terms they proposed were not on the LC list.[32] In other words, the list is adequately designed to match books in hand. But readers carry out research on problems rather than on books or topics,[33] coming to the catalog with an unpredictable vocabulary. Only a compilation of references drawn from actual requests will be rich enough to meet their needs. Online library catalogs provide an opportunity to capture samples of actual request language, although reference librarians are in a position to offer suggestions and feedback as well. This kind of feedback is collected regularly from the users of database search services: the ERIC Vocabulary Improvement Program provides a model for collecting and editing suggestions for new and revised terms as well as for references.[34]

Ensuring an Effective User Interface

Tracking and diagnosing search strategies may reveal some predictable patterns in user behavior, but the process of research itself is largely creative, intuitive, and unpredictable. Library users thinking "out loud" into tape recorders during searches of the card catalog at Ohio State University revealed "an amazing jumble of luck, inspiration, and

some knowledge ... triggered by flipping through cards ..."[37] Because the research process is one of building continually on the answers to questions posed and resolved by the researcher, Swanson has emphasized the importance of "creating a structure or framework within which searchers themselves can exercise maximum ingenuity and resourcefulness."[38] In designing online library catalogs, the best aid to subject (and other) searching may be the program through which the user interacts directly with the system, the "user-friendly interface." Descriptive information in the database and powerful software to search it cannot be used to full potential unless the interface is easy to learn. Currently, a researcher using more than one library (a common occurrence) does not need to learn how to open a catalog drawer or flip through the cards each time he or she enters a new system. But even "log-on" and "scrolling" procedures are likely to vary among online catalogs. While commercial online services suffer from unfortunate diversity,[39] libraries have a unique opportunity, right now, to standardize command language, search procedures, and other elements of the user interface for public catalogs before independent systems proliferate. The BSDP, presently coordinating reseach efforts by libraries employing or planning online catalogs, is in an excellent position to encourage standards development by the participating institutions.

There is no doubt that subject access will be better in the online catalog than in existing library card catalogs. The question is, how much better? In part, the answer depends on actions taken now.

SUMMARY OF RECOMMENDATIONS

1. Understanding the relationship between library bibliographic records and researchers' subject searches.

In American research libraries, the bibliographic record, including terms added as subject descriptors, is a standardized product. Yet the library profession has not documented how this product is used in subject searches or why searches succeed or fail. Gross surveys of rates of success or failure only begin to answer the questions and do not indicate how improvements can be made. Diagnostic studies of actual searches done by samples of reader populations (e.g. senior faculty in a social science discipline, graduate students in a humanities area, etc.) should be undertaken, using methodologies well developed for evaluations of information systems. There is no easy, speedy way to gather the large body of detailed data needed, but the enormous U.S. library investment in standard bibliographic records makes the need for such data pressing.

2. Enriching the search terms available in library bibliographic records.

Library bibliographic records do not include the abstracts and long lists of descriptors that are searchable in many information systems. Atherton's "Books are for Use Project" demonstrated that enriched MARC records hold considerable promise for improved subject access. Her recommendations for continued research to determine the optimum number and kinds of terms added to records should be followed. As methods for producing expanded MARC records are devised, plans for cooperative efforts to do so should be developed and tested for acceptability in the library community. The methodology used by Information Systems Consultants, Inc., in developing for ARL a cooperative plan to

expand access to microforms could be applied to planning for a cooperative plan to expand subject access.

The Subject Access Project, funded by the Council on Library Resources in 1978, demonstrated the potential of "enriched" bibliographic records for subject searching. Another CLR-funded project carried out by William Mischo in 1979 used enriched records to provide in-depth access to a reference collection through computer-generated KWOC indexes.[40] Council support for grant proposals in this area should continue and expand. BSDP could encourage such research by developing and publicizing guidelines encouraging researchers to submit proposals on pragmatic methods for enhanced subject retrieval.

3. Enriching the entry vocabulary for subject searches.

A rich entry vocabulary is an important component of a successful subject access system. The entry vocabulary currently used in most American libraries is based upon the cross references provided in the Library of Congress Subject Headings (LCSH). While any library is free to expand upon these references, few have allotted the extra staff time needed.

Any expanded entry vocabulary to LC subject terms should be made available for use in American library catalogs, especially online catalogs where the switching from entry term to actual term can be done automatically. The expanded vocabulary should be: 1) based on the actual language of reader's requests, and 2) developed cooperatively. One possible way to create a file of such references would be through user input to the Library of Congress.

The Library of Congress has indicated its interest in participating in a project to test a mechanism for and the utility of collecting user suggestions for new LCSH terms and additional cross references.[41] The test project would collect

suggested terms from several libraries over a set period of time. The terms would be collected in a uniform manner at each institution with project supervisors at each of the libraries to sort and edit suggestions before submitting them to LC. The suggested terms would then be reviewed and analyzed by the Library of Congress. An advisory committee might be formed to meet with LC staff to determine the conclusions and recommendations indicated by the results of the project.

4. Getting the most from the controlled vocabulary.

A controlled vocabulary is likely to remain an essential component of library subject access for some time to come. Trends in current library operations also indicate that LCSH will be the controlled vocabulary used by most American libraries as long as LCSH terms are provided on LC MARC records. While major changes in LCSH terms would be burdensome to American libraries already editing their catalogs to conform to AACR2, a reconfiguration of the LC list into an online thesaurus would create a reference tool of great benefit to both libraries and readers. In addition, the thesaurus could form the basis for expanding the list in special fields or for machine translations into other vocabularies.

The Library of Congress is interested in improving the LCSH cross reference structure in conjunction with the Library's plans to automate and distribute its subject authority file. As previously noted, a range of possibilities for improvement exists, running from a limited editorial project to a recasting of the list into a hierarchical thesaurus. The Library would like to participate in a small, working meeting of no more than ten experts in subject access and thesaurus construction for automated systems. The objective of the meeting would be to explore in detail alternatives for editing or restructuring LCSH (without revising the subject terms used) and to define the project(s) required to reach the most feasible and desirable end

product.

5. Developing and promoting standards for the user interface in the online library catalog.

It is recommended that the BSDP form a working group of librarians currently involved in the development of online public catalogs. The group members would define those elements of the interface that should be standardized (e.g. command language) and begin to develop standards to which their emerging catalogs could adhere. Further standards development might then be turned over to an ANSI.Z39 subcommittee, such as the Z39 group now working on terms and symbols in retrieval systems.

REFERENCES

1. Bates, Marcia. (1977). "Factors Affecting Subject Catalog Search Success." Journal of the American Society for Information Science 28, 161-169.

2. Krikelas, James S. (1980-81). "Searching The Library Catalog -- A Study of Users' Access." Library Research 2, 215-230.

3. Lipetz, Ben-Ami. (1970). "User Requirements in Identifying Desired Works in a Large Library." Yale University Library, New Haven, Conn. (ED 042 479).

4. Fayen, Emily Gallup. (1980). "A Survey of Dartmouth College Library Card Catalog Users." Unpublished report. Dartmouth College Library, Hanover, N.H.

5. Fayen, Emily, telephone interview, March 31, 1980.

6. Kaske, Neal K., and Sanders, Nancy P. (1980). "On-line Subject Access; The Human Side of the Problem." Reference Quarterly 20, 52-58.

7. Kaske, Op. cit.

8. Pritchard, Sarah M. (1981). "SCORPIO: A Study of the Public Users of the Library of Congress Information System." Unpublished report. Library of Congress, Washington, D.C.

9. Haftner, Ruth. (1979). "The Performance of Card Catalogs: A Review of Research." Library Research 1, 199-222.

10. Bates, Op. cit.

11. McFadden, Thomas. (1981). RIT Interface Project, Phase 2: Interim Progress Report. Unpublished report. Rochester Institute of Technology, Rochester, N.Y.
Fayen reports similar observations at Dartmouth.

12. Atherton, Pauline. (1978). "Books Are For Use: Final Report of the Subject Access Project to the Council on Library Resources." Syracuse University School of Information Studies, Syracuse, N.Y.

13. Kaske, Op. cit.

14. O'Neill, Edward T., and Aluri, Rao. (1981). "Library of Congress Subject Heading Patterns in OCLC Monographic Records." Library Resources and Technical Services, 25(1) 63-80. 63-80.

15. McClure, Charles R. (1976). "Subject and Added Entries as Access to Information." Journal of Academic Librarianship 2 (1), 9-14.

16. Lancaster, F. W. (1978). "The Cost-Effective Analysis of Information Retrieval and Dissemination Systems." In Key Papers in the Design and Evaluation of Information Systems (D. King, ed), pp. 23-38. Knowledge Industry Publications, White Plains, N.Y.

17. Lancaster, F. W. (1977). The Measurement and Evaluation of Library Services. Information Press, Washington, D.C., p.30.

18. Lipetz, Op. cit., p.65.

19. Atherton, Op. cit., p.35-36.

20. International PRECIS Workshop. (1977). The PRECIS Index System: Principles, Applications, and Prospects; Proceedings

of the International PRECIS Workshop, University of Maryland, October 15-17, 1976 (H. Wellish, ed.), p.197. H.W. Wilson Co., New York, N.Y.

21. Svenonius, Elaine. (1981). "Directions for Research in Indexing, Classification, and Cataloging." Library Resources and Technical Services 25(1), 88-103.

22. King, Donald W., and Bryant, Edward C. (1971). The Evaluation of Information Services and Products, pp. 150-152. Information Resources Press, Washington, D.C.

23. Brenner, Lisa P., et al. (1980-81). "User-Computer Interface Designs for Information Systems: A Review." Library Research 2, 63-73.

24. King, Op. cit., pp. 184-185.

25. Kassebaum, Laura, personal interview, February 17, 1981.

26. Library of Congress Subject Cataloging Division. (1978). PRECIS Project. Unpublished report. Library of Congress, Washington, D.C.

27. Swanson, Rowena Weiss. (1975). "Design and Evaluation of Information Systems." In Annual Review of Information Science and Technology 10 (C. Cuadra and A. Luke, eds.), pp. 43-101. American Society for Information Science, Washington, D.C.

28. Kaske, Neal K., and Sanders, Nancy P. (1980). "Evaluating the Effectiveness of Subject Access: the View of the Library Patron." In Communicating Information: Proceedings of the ASIS Annual Meeting 17, 1980, pp. 323-325. Knowledge Industry Publications, White Plains, N.Y.

29. McClure, Op. cit.

30. Wellish, Hans. (1972). "Subject Retrieval in the Seventies - Methods, Problems, Prospects," . In Subject Retrieval in the Seventies: New Directions; Proceedings of an International Symposium, University of Maryland, May 14-15, 1971 (H. Wellish and T.D. Wilson, eds.), pp. 2-27. Greenwood Publishing Co., Westport, Conn.

The utility of this suggestion can be easily tested by checking the language of a sample of reader requests against both LCSH and the classification schedules. Requests collected during the online catalog studies now being coordinated by the Council on Library Resources could be used.

31. Pritchard, Op. cit.

32. Angell, Richard S. (1972). "Library of Congress Subject Headings -- Review and Forecast." In Subject Retrieval in the Seventies: New Directions ; Proceedings of an International Symposium, University of Maryland, May 14-15, 1971 (H. Wellish and T.D. Wilson, eds.), pp. 143-167. Greenwood Publishing Co., Westport, Conn.

33. Rood, Joanna, Telephone interview, April 16, 1981.

34. Swanson, Don R. (1972). Requirements Study for Future Library Catalogs: Final Report to the National Science Foundation. University of Chicago Graduate Library School, Chicago, Ill.

35. Swanson Don R. (1979). "Libraries and the Growth of Knowledge." Library Quarterly 49 (1), 3-25.

36. Booth, Barbara. (1979). "A 'New' ERIC Thesauras, Fine-Tuned for Searching." Online, July 1979, 20-29.

37. Herndon, Gail A, and Van Pulis, Noelle. (1979). "The On-line Library: Problems and Prospects for User Education." In New Horizons for Academic Libraries; Papers Presented at the First National Conference of the Association of College and Research Libraries, Boston, Nov. 8-11, 1978 (R. Stueart and R. Johnson, eds.),, pp. 539-544. K. G. Saur, New York, N.Y.

38. Swanson, Don R. (1979) Op. cit.

39. Atherton, Pauline (1978). "Standards for a User-System Interface Language in On-line Retrieval Systems." Online Review 2 (1), 57-61.

40. Mischo, William H. (1979). "Expanded Access to Reference Collection Materials." Journal of Library Automation 12 (4), 338-354.

41. The recommendations in sections 3 and 4 are based on a meeting at the Library of Congress on April 15, 1981. Participants were Carol Mandel, Mary K. D. Pietris, Lucia Rather, and Robert Zich. At that meeting, Ms. Rather indicated her willingness to work with the Council on Library Resources in implementing the recommendations proposed.