Volume 10,
Number 3

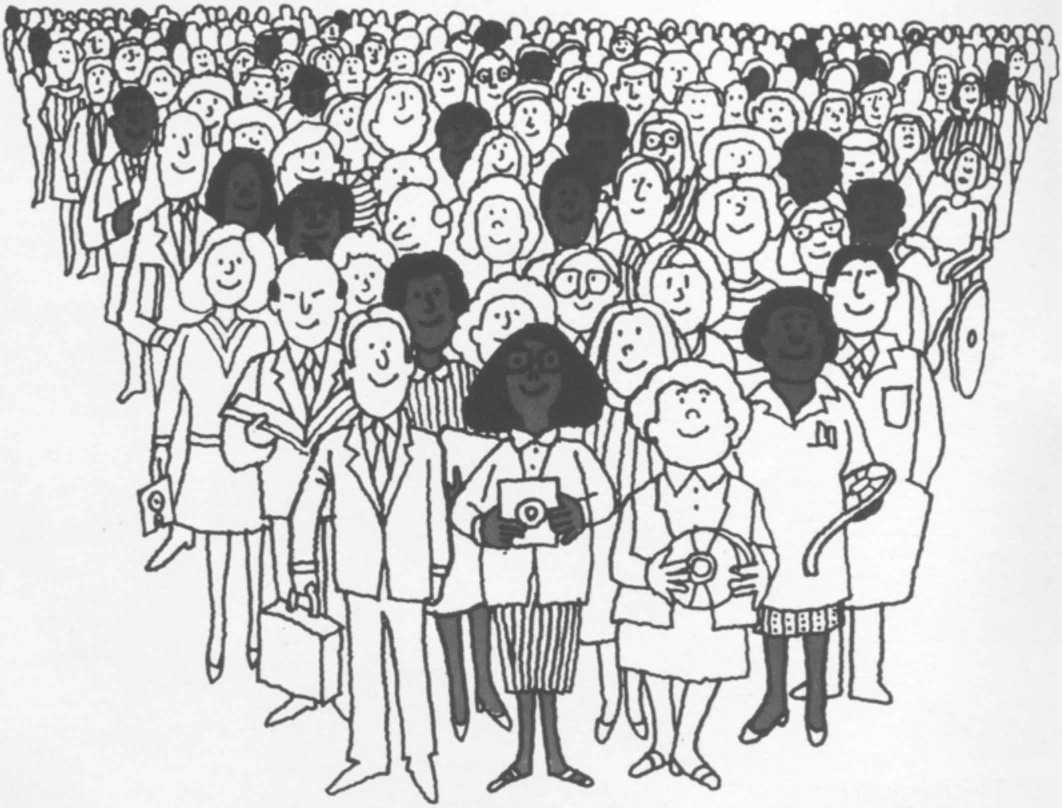
ISSN 0730-9295

Information Technology and Libraries

September 1991

CONTENTS

ITLBDC 10(3) 161–248 (1990); Inf Technol Libr



ISSN 0730-9295

Information Technology and Libraries

Volume 10, Number 3: September 1991

CONTENTS

ITLBDC 10(3) 161–248 (1990); Inf Technol Libr

Information Technology and Libraries

Information Technology and Libraries is the official publication of the Library and Information Technology Association, a division of the American Library Association, 50 E. Huron St., Chicago, IL 60611; *Executive Director:* Linda J. Knutson. The journal is issued quarterly in March, June, September, and December.

Information Technology and Libraries publishes material related to all aspects of library and information technology. Some specific topics of interest are: Automated Bibliographic Control, AV Techniques, Communications Technology, Cable Systems, Computerized Information Processing, Data Management, Facsimile Applications, File Organization, Legal and Regulatory Matters, Library Networks, Storage and Retrieval Systems, Systems Analysis, and Video Technologies. *ITAL* welcomes unsolicited manuscripts. Submissions should follow the guidelines stated under "Instructions to Authors" on page 84 of the March 1991 issue. Manuscripts of articles, communications, and news items should be addressed to the editor: Thomas W. Leonhardt, *Information Technology and Libraries*, University of the Pacific, Stockton, CA 95211. Copies of books for review should be addressed to: Susan B. Harrison, *ITAL Book Reviews*, The New York Public Library, 455 Fifth Ave., New York, NY 10016. Copies of software for review should be addressed to: George S. Machovec, *ITAL Software Reviews*, Arizona State University, Tempe, AZ 85287. Advertising arrangements should be made with Kurt Murphy, Hayden Library, Arizona State University, Tempe, AZ 85287.

Information Technology and Libraries is a perquisite of membership in the Library and Information Technology Association. Subscription price, $20, is included in membership dues. Nonmembers may subscribe for $45 per year in the U.S.; $50 in Canada, Mexico, Spain, and other PUAS countries; $55 in other foreign countries. Single copies, $15. Second-class postage paid at Chicago, Illinois, and at additional mailing offices. *Postmaster:* Send address changes to *Information Technology and Libraries*, 50 E. Huron St., Chicago, IL 60611.

Circulation and Production: ALA Publishing Services (Eileen Mahoney; Dianne Rooney; Amy Brown, Bruce Frausto, Josephine Gibson-Porter, Dan Lewis, and Donavan Vicha), American Library Association, 50 E. Huron St., Chicago, IL 60611.

Publication of material in *Information Technology and Libraries* does not constitute official endorsement by the Library and Information Technology Association or the American Library Association.

Abstracted in *Computer & Information Systems, Computing Reviews, Information Science Abstracts, Library & Information Science Abstracts, Referativnyi Zhurnal, Nauchnaya i Tekhnicheskaya Informatsiya, Otdyelnyi Vypusk*, and *Science Abstracts Publications*. Indexed in *CompuMath Citation Index, Computer Contents, Computer Literature Index, Current Contents/Health Services Administration, Current Contents/Social Behavioral Sciences, Current Index to Journals in Education, Education, Library Literature, Magazine Index, New Search*, and *Social Sciences Citation Index*. Microfilm copies available to subscribers from University Microfilms, Ann Arbor, Michigan.

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984. ♾

Intactness and Accuracy of Online and Card Catalogs: ALIS II vs. Card Catalog at Texas A&M University

Colleen Cook and Leila Payne

A research study was undertaken at the Evans Library (Texas A&M University) to analyze the comparative intactness and accuracy of the public card catalog and a first-generation online catalog. The results of the study indicated that the online file was marginally less intact and substantially more accurate than comparable card catalogs. Detailed results of the analysis are reported by MARC tag. Suggestions for future research are provided as is an exhortation to the library profession to be vigilant to the inaccuracies inherent in online files and to devise validation mechanisms appropriate to automated environments.

The Sterling C. Evans Library at Texas A&M University supports the research and curricular needs of 41,200 students and 2,200 faculty members in a library facility serving all disciplines with the exception of medicine and veterinary medicine. The Library houses 1.8 million volumes, has 14,000 standing orders, exclusive of GPO materials, and an annual materials budget of $3.6 million.

The Library began to construct a bibliographic database of MARC records in late 1975 when it contracted with OCLC for cataloging functions. In a retrospective conversion project (RCP) in 1979–80, OCLC converted 399,000 titles to MARC format. By the end of 1980, the Evans Library at Texas A&M had a complete MARC database of bibliographic records for all cataloged holdings.

OCLC archive and retrospective conversion project tapes were used to construct a local bibliographic database for the DataPhase ALIS II system in 1980–81. The Library used ALIS II for cataloging and circulation functions. Although ALIS terminals were located in the public catalog area, ALIS was not intended to replace the public card catalog. Public terminals were signed on in a technical mode in which search mechanics and screen displays were not user-friendly. Name, title, series, and call number searches were possible. However, because sufficient storage was not available, subject indices were not generated, and the ALIS public mode was not released to the public. With inherent flaws, ALIS rapidly outlived its usefulness, and the ALIS bibliographic and item databases were used to construct parallel databases in a local NOTIS system in 1988.

PROBLEM STATEMENT

The information-transfer process has always been susceptible to error—ancient scribes, monks in Medieval monasteries,

Colleen Cook is Head, Bibliographic Access Division and Leila Payne is Head, Database Management Division at the Sterling C. Evans Library, Texas A&M University.

The authors wish to acknowledge funding from the Office of Organized Research at Texas A&M University.

and twentieth-century librarians have all had to deal with inaccurately transcribed data. In creating online catalogs, modern day librarians have sought to avoid introducing a new set of errors inherent to computer technology into their already error-prone bibliographic files. To what extent have librarians been successful in these attempts? Despite the best efforts, bibliographic records can be missed in retrospective conversion projects; some programs and equipment cannot handle special characters, and some data may be lost through imperfect backup and recovery methods, to name only a few sources of error.

The need for accuracy and intactness of bibliographic databases increases in importance as libraries move towards integrated, automated library systems such as NOTIS. The quality of the core bibliographic database affects all aspects of library operations. Errors in the bibliographic database can result in inadvertent duplicate orders. Users are unable to find materials and lose confidence that either the automated system or the library can meet their needs. Ensuring quality control for a bibliographic database is difficult given the tools presently available. Authority control systems vary in complexity and ease of use. In the final analysis much data is reviewed for accuracy, to whatever extent feasible, only as it is created and input. While librarians know that bibliographic databases contain errors, attempts to ascertain error rates and to categorize types of errors have only recently been made.[1] Furthermore, librarians and users question how the intactness and accuracy of data in machine files compare to those in card files.

A research study was undertaken at the Evans Library at Texas A&M University to compare and contrast the intactness and accuracy of the bibliographic data in the public card catalog and the ALIS bibliographic database, and to analyze the bibliographic files for trends in errors. For the purpose of the study, intactness was defined as the expected presence or absence of a bibliographic record. The ALIS bibliographic database would be intact if all the titles in the card shelflist designated for conversion were present in the database. Withdrawn titles and titles held in the Medical Sciences Library (MSL), a separately administered campus library, for instance, were not converted in the RCP by design. The card catalog would be intact for a given bibliographic entity only if all main entry and tracings cards were found.

Accuracy was defined as present in a manner that permitted accessibility. For the ALIS machine file, a particular access point was considered inaccurate if it would not be accessible in an online searching environment. For the purposes of the study the following four conditions could cause inaccessibility:

1. misspelling
2. transcription error, e.g., words missing from a title
3. incorrect non-filing indicator
4. incorrect tagging if the error would cause the field to index improperly or not at all

In the card files, a given access point was accurate if a card for the access point was found in the search. Why a card was not found was not deemed significant to the study. Therefore, if a given title card was not found, that card was inaccessible and inaccurate, whether it had been stolen, never filed, misfiled, or contained typographical errors.

No attempt was made to identify and count cataloging errors, or errors resulting from changes over time in the cataloging code or MARC standards. Thus a missing delimiter h (GMD) on a microform record or a delimiter el on a uniform title were not considered errors on older records because they were not added to Books Format until February 1980.[2] The study investigated only the accuracy of information present. Missing records, fields, and tracings were not considered under the accuracy question, but rather were studied under the question of intactness. For instance, only traced series were reviewed in the analysis of card catalog intactness. Series that might have been traced or were not traced were not investigated. Other inconsistencies judged to be inconsequential to the accuracy question because they did not affect access included capitalizations, absence or presence of brackets, and insignificant punctuation. Choice and form of entry found on the shelflist card were considered correct for pre-*AACR2* headings. Conversely, enhancements to information on the shelflist, such as the addition of death dates and spelled out abbreviations in subject tracings, were not considered errors.

METHODOLOGY

Because a complete census would be prohibitive in terms of time and cost, the bibliographic files were studied through sampling.[3] A random sample from the shelflist was selected to compare to corresponding records in the public catalog and ALIS bibliographic files for intactness and accuracy. The shelflist was chosen to sample because it represented all cards in the card catalog, and it was used during the retrospective conversion project to create the online bibliographic file. The shelflist was, therefore, the singular, reliable point of comparison for the question of intactness, i.e., whether a record had been converted during the retrospective conversion project.

Sample Selection Description

The shelflist consisted of 977 file drawers containing approximately 1,000 titles each. Although most records were represented on a single card, many required two, three, or even more cards. In order to be valid the sample had to satisfy one of the following two criteria: (1) each title had an equal probability of being drawn or (2) the probabilities were known for the titles that were chosen. Since some titles were represented in the card file by multiple cards, it would have been impossible to satisfy condition 1. Condition 2, however, could be realized by choosing a random card and noting if it represented a title with one, two, or more cards. A title with two cards would have twice the probability of being chosen as a title with only one card, etc.

Precision required for the sample was ± 2%; thus a sample size of 1,500 was calculated to be sufficient. A somewhat larger sample of 1,954 records was ultimately selected for analysis, because it was expedient to choose two records from each shelflist drawer. Thus the sample was a modified, systematic sample, or a stratified sample with two observations from each stratum. The following procedure was used to select the records from the shelflist for the study.

Let N = number of file drawers

Let X_i = length of the ith drawer in 1/8 inches

Let $X^* = \max (X_i)$

Let D = random digit from 1 to X^*

The researchers measured the length of each drawer and recorded (X_i). Round tags numbered from one to X^* were placed in a box. A random digit from 1 to X^* was drawn. If the digit drawn was greater than X_i, the researchers replaced it and drew another tag, then picked a record in the interval from D-1 to D. The following information was recorded for each record: drawer number, length of drawer, and number of cards per record.

Each sample record was searched in the ALIS database. A staff member attached a copy of the shelflist card(s), and a screen print of the MARC bibliographic record from ALIS, when found, to an extensive workform. All fields accessible in the NOTIS keyword index were checked for errors.[4] Only by checking each keyword accessible field in NOTIS could a thorough analysis of the accuracy of the data be performed. In a keyword access environment a transcription error or misspelling in a field under keyword control would constitute an occasion for user failure. Errors were later tabulated only for indexed fields—for example, contents notes errors (field 505) were noted but not tabulated. An error was defined as the inability, for whatever reason, to call up a field. The researchers tabulated error according to the four categories mentioned earlier: misspellings, transcription errors, incorrect non-filing indicators, and incorrect taggings.

Clerks in the Catalog Maintenance Department and student workers searched the card catalog for all representative cards of each sample record. They checked each card not found a second time and if the second try was successful the access point was not considered lacking. However, an argument could be made that cards not found on the first try were, in fact, not valid access points. If a cross-reference pointed to a post-*AACR2* filing position for a heading, no error was counted.

From this data, the researchers established frequencies of total errors and of various categories of errors, and coded the data into a dBase III+ file for easy manipulation and cross-tabulation.

ASSUMPTIONS

The following assumptions were made for the study:

1. The shelflist was both intact and accurate.
2. The ALIS database was stable, i.e., no significant data loss or other alterations of data were made over the period of data collection.
3. Data missing from records by design did not constitute errors.

RESULTS

Intactness

ALIS MARC Bibliographic Database

Fourteen of the 1,954 records in the study were unexplainably missing from the ALIS database. At an error rate of .72%, the database was 99.28% intact. Despite the myriad conversion and reconversion projects, the ALIS database was less than 1% incomplete. It is interesting to note that 48 other records included in the study, (37 MSL titles and 11 withdrawn titles), were represented in the shelflist and accurately excluded from ALIS. Thus records were both converted and excluded from conversion with marginal error.

There were no discernible trends explaining the omission of 14 records from the ALIS database. For instance, the missing records, cataloged between November 1963 and May 1987, covered a broad time span in an apparently random fashion. Seven records were converted in the RCP; the other 7 had been cataloged on OCLC as acquired. Four were serials; 10 were monographs. Some educated guesses explaining the absence of the various records might be offered, however. One title was in a Cyrillic alphabet, and non-Roman alphabet titles had been excluded from the routine RCP, flagged in the shelflist, and corrected later. In another case, it was suspected that a serial record had been inadvertently overlaid through an OCLC number match in cataloging a title change.

Public Card Catalog

When comparing the card catalog with the ALIS database, the level of intactness was practically the same. The complete cardsets for 12 of the 1,954 sample records were missing from the public card catalog. Thus the card catalog was 99.38% intact with an error rate of .614%. With the exception of one record, all missing cardsets were for titles cataloged between late 1984 and 1987, when the Cataloging Division, facing huge filing backlogs and added workloads represented in maintaining the online catalog, set a low priority for filing cards, particularly for access points available in the ALIS system. The intactness of the card catalog deteriorated rapidly with the shift of emphasis from the manual card catalog to the online system.

Accuracy

Title Access Points

A title can be considered inaccessible in an automated file if the bibliographic information is improperly transcribed, the title is tagged in such a way as to be irretrievable (broken down in the study into non-filing indicator errors and other tagging errors), or if a word in the title index is misspelled. Errors that did not affect access were discounted. Titles omitted in missing bibliographic records (as determined in the intactness analysis) were not considered errors in the accuracy study. A title access point in the card catalog was lacking if the card could not be found by a typical library user.[5] No attempt was made in the study to identify or consider title access problems caused by cataloging errors or policy. The underlying assumption was that the information on the shelflist card was the standard against which errors were measured.

ALIS

Of the 2,303 title fields in the ALIS records,[6] 31 titles, or 1.35%, had errors which resulted in loss of access (see table 1). Other errors that were determined not to result in loss of access were not included in the error count. These included missing General Media Designators (GMD), non-filing indicators of blank or a fill character for titles that do not begin with an article, and the language delimiter el in uniform titles. Errors in the delimiter c were also omitted because few users read into the author statement,

Table 1. ALIS Title Accuracy

Field Tag	No. of Occurrences	Spelling Errors	Transcription Errors	Non-filing Errors	Tagging Errors
130	11	0	0	2	0
210	18	0	0	0	0
212	5	0	0	0	0
222	38	0	0	0	0
240	46	0	0	0	0
245	1,890	1	17	1	2
246	35	0	0	0	0
247	16	0	0	0	0
730	18	0	0	1	1
740	154	0	4	1	0
780	41	0	1	0	0
785	31	0	0	0	0
Total	2,303	1	22	5	3

and ALIS title access stopped after delimiters a and b.

Errors in main titles (245) illustrate the kinds of errors found. Of the 1,890 main titles (245 fields) in the sample, 21 (1.1%) were found to have errors of the four types mentioned earlier: 1 misspelling, 17 transcription errors, 1 incorrect non-filing indicator, and 2 occurrences of mistaggings. Transcription problems were of three types: errors in the creation of the MARC bibliographic records during the retrospective conversion project, errors resulting from the inability in ALIS to store special characters, and errors in current cataloging. Retrospective conversion errors included omission of subtitles, substitution of one word for another, and typographic errors. Often the errors were the result of the RCP operator accepting the OCLC database record without editing to reflect additions to the Texas A&M shelflist card. Other errors were intrinsic to the use of ALIS software. Letters or numbers were retained but not subscripted; for example, O_2 became O2. In one case all 245 information past the Cyrillic "myakhki znak" had been dropped. Non-filing indicators for titles beginning with articles had been set at zero. Two records had incorrect subtitle indicators (delimiter c rather than b), which resulted in the subtitles being omitted from the title index.

Public Card Catalog

The accuracy error rate for titles in the public card catalog was 5.89%; 125 title cards from 110 bibliographic records were missing from a total of 2,087 titles. Some missing cards could be explained by library filing priorities. As the library automated technical services functions, emphasis began to move toward maintaining the automated files rather than the card catalogs. In 1984 the decision was made to emphasize filing shelflist and subject heading cards and allow the author/title cards to backlog, thus using limited staff resources to provide card access to subjects and call numbers that were not accessible through ALIS. Few author/title cards were filed between 1985 and March 1989 when the card catalog was closed. Over half of the title cards which were missing were produced after 1985 when filing was a low priority (see table 2).

Subject Access Points

ALIS

All 6xx fields on the MARC records were considered in the study; locally assigned LC subject headings as well as MeSH were included in the sample.[7,8] There were 37 errors in the 3,347 subject fields in the ALIS database sample for a 1.05% error rate. Subject errors were categorized into three of the four groups mentioned earlier: incorrect taggings, transcription errors, and misspellings (see table 3).

Thirteen errors resulted from incorrect taggings. For example, a topical subject was mistagged as 630 (uniform title), which would index in some catalogs as a title. Some incorrect taggings were not counted as errors since they did not affect

Table 2. Card Catalog Title Accuracy

Date Cataloged	No. of Bib. Records with Missing Title Cards
pre–1975	5
1975–1983	10
1984	18
1985	57
1986	14
1987	6
Total	110

access. For example, a geographic subject heading (651), mistagged as a topical subject heading (650), was still fully accessible in a subject index, and was therefore not considered an error, although future, narrower qualifications by geographical unit would be precluded. There were 16 transcription errors in subject fields. Most of these errors reflected subject headings that were present on the catalog cards, but were not transcribed, for whatever reason, to the MARC record. It was noteworthy that in many cases subjects after the fourth were dropped from the MARC record, either because of machine manipulation error or because computer operators omitted the headings during the retrospective conversion project. Eight spelling errors were discovered in subject access points in the study: one personal name, five topical and two geographical subjects.[9]

Public Card Catalog

There were 207 subject heading errors (on 147 bibliographic records) found in the card catalog sample. As defined above, the absence of a card in the catalog, for whatever reason, constituted an error for the study. Given a total number of 3,465 subject tracings, the card catalog showed a 5.97% error rate in subject cards. As expected, most of the missing subject cards were ascribable to the period of transition between the card and online catalogs, when filing was not a high priority.

Name Access Points

ALIS

Of the 2,802 name headings in ALIS, 15 contained errors, for an error rate of 0.54% (see table 4).[10,11] One of the errors was a misspelling, 11 were transcription, and 3 were tagging errors. The misspelling was a typographical error, i.e., substituting the letter el for the number one in a birth date. Transcription errors included names from the RCP that did not match the form of name on the shelflist, and name headings that were dropped during the RCP. In addition, special characters in 4 names were dropped during the initial ALIS tapeload. For example, Hiort, Esbjørn became Hiort, Esbjrn. The tagging errors included two personal names incorrectly tagged as compound names, and a corporate name tagged as a title.

Public Card Catalog

Two hundred eighty-five catalog cards for name access points out of 2,931 were missing from the card catalog for a name access error rate of 9.72%.

Series Access Points

ALIS

Although the greatest error rate for ALIS access points was for series entries, the error rate remained lower for the online catalog than for the card catalog.[12] The ALIS error rate was 5.76%, reflecting 46 errors in 799 series access points (see table 5). There were no spelling errors in the ALIS sample. Tagging errors included untraced series tagged to provide series access, traced series tagged as untraced, a corporate name series tagged as a title, and a title series tagged as a corporate name. Three of the transcription errors were straightforward: one 440 series was omitted from the ALIS record, and one series included the letter O instead of a degree mark. The remaining series transcription errors occurred in records for serials and monographic sets in series. The errors were a result of the difference in controlling access to nonconsecutive series entries in the card and online environments. In the card catalog a series card was filed for each number in the series tracing, with consecutive numbers being grouped in a number span. For example, numbers 1–5, 7, 9 in the series statement would result in three series cards. In the ALIS catalog, a series tracing was needed for each number in the series regardless of whether the numbering was consecutive. An ALIS user interested in no. 4 of the series might not find it if the series tracing was for no. 1–5. Commonly the ALIS series tracing mentioned only the first series number followed by "etc." (Series ;

Table 3. ALIS Subject Accuracy

Field Tag	No. of Occurrences	Spelling Errors	Transcription Errors	Tagging Errors
600	233	1	0	4
610	70	0	1	3
611	3	0	0	0
630	8	0	0	0
650	2,641	5	10	5
651	387	2	5	1
653	2	0	0	0
690	3	0	0	0
Total	3,347	8	16	13

Table 4. ALIS Name Accuracy

Field Tag	No. of Occurrences	Spelling Errors	Transcription Errors	Tagging Errors
100	1,402	0	7	2
110	105	0	0	0
111	71	0	0	0
700	865	1	4	0
710	340	0	0	1
711	19	0	0	0
Total	2,802	1	11	3

Table 5. ALIS Series Accuracy

Field Tag	No. of Occurrences	Spelling Errors	Transcription Errors	Tagging Errors
400	1	0	0	1
410	34	0	0	1
411	1	0	0	0
440	228	0	1	7
490	375	0	1	2
800	12	0	5	0
810	103	0	15	2
811	0	0	0	0
830	28	0	0	1
840	17	0	10	0
Total	799	0	32	14

no. 1, etc.) In this case series access in ALIS was limited to the only number in the series tracing. The missing series access points for the other numbers in the span were considered transcription errors.

Public Card Catalog

Fifty-one series cards from the 466 series access points in the sample were not found in the card catalog, for a card error rate of 10.94%.

Summary of Accuracy Analysis (Table 6)

In considering the accuracy question, the following generalizations can be made: misspellings were primarily in topical LC subject headings (5 in 650 fields) and transcription errors were found most often in main titles (17 in 245 fields), topical LC subjects (10 in 650 fields), series entered under corporate name (15 in 810 fields), and title added entries (10 in 840 fields). Disregarding fill characters, incorrect non-filing indicators occurred most often, as expected, in uniform title fields (2 in 130 fields), where the non-filing indicator is first rather than second as in most title fields. Most tagging errors were in series and subject fields (7 in 440 fields, 5 in 650 fields, 4 in 600 fields). Other results of the accuracy study are noteworthy, although outside the narrow confines of the accuracy

Table 6. Summary of Accuracy Analysis: ALIS vs. Cards

Access Point	% Error ALIS	% Error Card
Title	1.10	5.98
Subject	1.05	5.97
Name	0.64	9.72
Series	5.51	10.94

question as stated. Special characters were problematic, e.g., every script el in the paging of 300 fields was lost and converted as a blank. There were also 142 records (7.27%) that were missing diacritics since ALIS could not handle them. Neither of these instances of special characters were considered errors resulting in loss of access, although the point may be debated, particularly with regard to loss of diacritics.

CONCLUSIONS AND FUTURE RESEARCH

Librarians charged with providing access to collections will always be concerned about the accuracy and intactness of the bibliographic files they produce. Wherever human manipulation and input of data are necessary, there will inevitably be error. One question posed in the Texas A&M study, i.e., whether the card or online catalog was the more intact and accurate, is a question significant only to the present period of transition in information technology, and, if the results of the study prove to be universal, is a question that can be safely put to rest. Librarians can probably be assured that their automated files are more accurate than the paper files which they replaced, but studies confirming the results of the Texas A&M and the Knutson study at the University of Illinois, Chicago Circle, should be pursued.

Despite library mythology to the contrary, the machine ALIS file was marginally less intact and substantially more accurate than comparable card catalogs at Texas A&M. Notwithstanding large scale projects and programs, such as the Retrospective Conversion Project and duplicate record resolution, the ALIS file at Texas A&M was more than acceptably intact and accurate. Moreover, some information was gained by the conversion from manual to automated systems, e.g., uniform titles were searchable on ALIS whether or not they were tagged to print on cards.

Machine files are not prone to atrophy over time as are card files. No user can randomly remove a title or subject access point from an automated file. Machine backups provide multiple copies of bibliographic files and equivalent replications of bibliographic data in a paper environment were simply not feasible. It can be predicted that automated files will not be as vulnerable to the ravages of time as manual card files.

Now that the question of the comparative accuracy and intactness of machine versus paper files can be put aside, new questions regarding the accuracy of machine bibliographic files arise, which are intrinsic to an automated environment. Some institutions, such as Texas A&M, that have already implemented a second-generation automated bibliographic system need to investigate the accuracy of files pre- and postconversion. Effective methodologies for such studies need to be devised. Consensus among the library community as to what constitutes a significant error in an online environment under keyword access needs to be reached. Finally, the significance of errors in integrated systems in which an inaccuracy in bibliographic data can cause problems from a multiplicity of viewpoints—from acquisitions to circulation and inventory control—needs to be determined.

It was very difficult for the investigators to compare the accuracy of a card and online catalog. The two are fundamentally different, and true comparisons can only be made to a limited degree. Information specialists generally agree that online systems provide greatly enhanced access capabilities over card files, even if keyword access is not considered. There is reason to believe that the essence of online systems may make them inherently less prone to the type of errors familiar to card environments. Librarians will need to be vigilant to the inaccuracies inherent in online systems and devise validation mechanisms appropriate to automated environments.

REFERENCES AND NOTES

1. Gunnar Knutson, "A Comparison of Online and Card Catalog Accuracy," *Library Resources and Technical Services* 34, no.1: 24–35 (January 1990)
2. *MARC Formats for Bibliographic Data (1980); History* (Washington, D.C.: Library of Congress, 1983), p.71, 75.
3. The authors gratefully acknowledge the assistance of Dr. Omer C. Jenkins, Associate Professor, Dept. of Statistics, Texas A&M University, in devising the sampling methodology used in the study.
4. MARC fields accessible in NOTIS keyword index: 100, 110, 111, 130, 210, 211, 212, 214, 222, 240, 242, 243, 245, 246, 247, 400, 410, 411, 440, 490, 580, 600, 610, 611, 630, 650, 651, 653, 655, 656, 657, 690, 700, 710, 711, 730, 740, 760, 762, 773, 777, 780, 785, 800, 810, 811, 830, and 840.
5. A&M had divided author and title catalogs and combined them into a single author/title catalog in December 1975. As a result, some cards existed for title added entries, even when main entry was by title. The absence of the duplicate title card was not considered an error.
6. MARC title fields in ALIS: 130, 210, 212, 222, 240, 245, 246, 247, 730, 740, 780, and 785, but not delimiter t subfields.
7. Name subjects were tallied as subjects.
8. MARC subject fields in ALIS: 600, 610, 611, 630, 650 (2d indicator zero), 651, 653, and 690.
9. Although three of the errors were in the first word of the subject string, the corresponding catalog cards were found for all five topical subjects. Card filers had simply filed the cards in the correct place regardless of the error.
10. Names include personal, conference, and corporate names, but not names used as subject or series names.
11. MARC name fields in ALIS: 100, 110, 111, 700, 710, and 711.
12. Name series were tallied as series. ■■

Information Communication Highways in the 1990s: An Analysis of Their Potential Impact on Library Automation

Harry M. Kibirige

The information environment of the 1990s is and will continue to be in a dynamic state of flux. Library automation will be markedly affected by major trends, some of which began in the 1970s and 1980s. One of the most significant change agents that will pervade all other trends is the establishment and regular use of high-speed, fiber optic communication highways. Five of the major features of the decade will be: the interchanging roles of computers and telecommunications systems; the increasing importance of information in national economies; the wired campus for local information sharing; the ubiquitous microcomputer with greatly increased computing power; and national and international information communication networks. These features will combine to make much more information available to the information user. Among the technological issues facing librarians will be the fact that library automation systems were not designed for operation in an internet environment and have thus to be redesigned. Library automation vendors must be prepared to retool to continue to be effective in the present decade. Librarians' knowledge of telecommunications must be enhanced so that they are able to communicate with technicians and thus contribute to the design of more viable automated systems. Some of the legal issues will center on copyright and the protection of sensitive personal files on the networks.

Library automation is presently being influenced by telecommunications more than ever before. By facilitating remote access to relevant data banks, telecommunications systems have modified the modus operandi of many libraries and information centers. As a few illustrations will show, the current trends have ushered in diversified information products and services which are now available to the patron. First, several indexes, abstracts, encyclopedias, handbooks, and dictionaries are now available online and in full text from remote sources. While some institutions have opted to have them locally on CD-ROM, the remote data banks continue to have the advantage of currency when the most up-to-date information is required. Second, some online catalogs' access terminals have been designed to reach out beyond the host library's confines, so that they now act as access devices to national and, in some

Harry M. Kibirige is on the faculty of the Graduate School of Library and Information Studies, Queens College, The City University of New York.

Information for this article was gathered through the generous funding of the Research Foundation of The City University of New York, which is hereby appreciated.

cases, international information resources. Third, the time it takes to execute an interlibrary loan request has been remarkably reduced by electronic identification of the source institution, the transmission of the request, and the final delivery of the document electronically. Finally, we are on the threshold of high-volume online delivery of full-text information among libraries and information centers and eventually to the user's desktop computer on demand. Such information will soon be replete with the requisite illustrations and, if need be, with sound explanations to boot. The main underlying technology for all these systems will be fiber optic-based telecommunications, the basis for modern information communication highways.

The 1990s will witness the maturity of more than a decade of research, development, and application of information communication projects based on fiber optics and related technologies. The use of fiber optic technology for telecommunications systems within the last decade or so has demonstrated several of its qualities as the most effective information transmission medium for the information age. It has been used in programs for transmission of large amounts of information at high speeds and with high levels of accuracy. The use of this technology in data communications for text processing has been relatively slow partly due to the size of the installed base of the alternative technologies, such as coaxial cable and twisted-pair copper wire. Recent experiments as well as market surveys indicate, however, that it is increasingly being used for information communication that may include data, video, voice and image. By the year 2000, much of the intercontinental and intracontinental information will be carried over fiber optic lines. Such trends are most welcome in an era when information is at a high premium for both national economies and corporate competitiveness. Librarians and other information specialists who intend to remain active participants in the information arena have an obligation to understand the potential for these systems for local and remote information access, sharing, and retrieval.

During this decade, a number of the perennial information issues for which technological solutions are needed will persist and continue to puzzle librarians and other information specialists. Five of the salient issues are worth mentioning at the outset. First, sophistication of user information needs and demands will be accentuated. Second, information technology will become very complex in response to the heightened levels of user demands. Third, the need for designing or redesigning of information systems using the new technologies will be the order of the day. Fourth, connectivity of information networks that have pervaded the information arena within the last few years will be one of the major tasks for information professionals. Finally, legal implications of the massive information interchange among the networks will prick the conscience of librarians and other information scientists, owners of creative works, and end users.

Among the technological breakthroughs that may alleviate some of the problems are the use of fiber optics, development and wide-scale deployment of the integrated services digital network (ISDN), transparent LAN-to-LAN connections, intercontinental and intracontinental broadband connectivity, and satellite-fiber optics digital interlinks.

This article focuses on the potential effects of fiber optic-based communication technology on information networks and systems design, with special emphasis on library automation. As is discussed in the rest of the article, most aspects of information processing will be influenced by this technology. Analysis of the major trends of the decade is presented to put fiber optics in its true perspective with respect to the operational information arena.

THE LIBRARY ENVIRONMENT IN THE 1990s

The information arena of the 1990s in which libraries are operating is destined to be in a perpetual state of flux. However, several of the trends started in the 1970s and the 1980s will be consolidated into viable enterprises. Five of the major features of the period are analyzed as follows: computers and telecommunications, information in national economies, the ubiquitous microcomputer, the wired campus for local information sharing, and national and international networks for remote information access.

Starting in the 1970s, there has been a gradual but steady blending of computers and telecommunications systems. This has resulted in powerful telecommunications and computer systems that have interchangeable roles. Currently there exist on the market telecommunication systems that can compute and computer systems that can communicate. Some of the examples include computerized branch exchanges (CBX) with considerable computing intelligence and sophisticated front-end processors with communication capabilities.[1] This has in effect established the logical as well as physical foundations for establishing intelligent networks of the 1990s. As Arnold Hieber explains, an intelligent network is essentially a network-scale computer capable of responding to a variety of user needs.[2] Such networks will be able to dynamically route the user from one device to another and alternate services among voice, data, image, and video, depending on user demand.

Another important trend that will persist in the 1990s is the awareness of the significance of information in the national economies. Since Daniel Bell promulgated the coming of the postindustrial society in the 1970s and the subsequent designation of the current era as the information age, national governments throughout the world are investing more resources in developing their information infrastructures. The U.S. government intends to spend $1.9 billion on a five-year federal High-Performance Computing Program aimed at linking research oriented supercomputers. In its October 1989 report, the Office of Technology Assessment discussed this program, though it called it the National Research Education Network (NREN). Among the policy issues covered were access rights and financing the access, establishment, and management of the network. Business corporations are also spending more funds on developing information products and services. Library systems are similarly spending more funds on computerized information projects, signifying awareness of the changing role of information in modern society. These trends will continue in the 1990s.

A third major trend that is a spillover from the 1980s is the proliferation of microcomputers (personal computers) in all sectors of society. Since becoming popular in the 1970s, micros have had a major role in spreading computer literacy to noncomputer scientists. They will continue to be the main access devices of the average individual to the universe of computerized data banks. Within the decade, desktop workstation processing speeds will surpass the fifty million instruction per second (MIPS) mark. At the same time, storage capacity of the micro-based disks is likely to reach gigabyte levels.[3] Use of the Reduced Instruction Set Computer (RISC) architecture with the associated software that increases clock speeds will put personal computers among the most powerful devices on the market for medium-sized operations. The increasing use of optical technology in computer storage systems will accelerate the rate of growth of the capacities of storage devices. Given such characteristics, the microcomputer of the decade will have higher performance than some of the main frames and minis of the 1970s and 1980s. For example the NeXT computer, manufactured by NeXT Computer, Inc., has already surpassed the processing speeds and versatility of many mini and some mainframe computers of five years ago. Libraries and information centers with online public access catalogs (OPACs) are presently encountering pressure from patrons to provide interfaces that will facilitate searching the OPACs from the office, the laboratory, or the home. Multiplexed information from powerful micro clusters will supersede the transmission capacity of current media and call for use of optic fiber as an alternative.

The networked campus or institutional compound for local information sharing based on local area networks (LANs) technology forms the fourth major trend of the decade. The practice of developing what is popularly known as "the wired campus" has spread to most major institutions of higher education in the United States.[4] Many campuses have updated or are in the process of upgrading their networks to fiber optics. Georgia Tech, the University of

California MELVYL system, and Carnegie-Mellon, to mention a few, have networks that have successfully implemented fiber optics. Library automation systems at these three institutions use fiber optics trunk lines. Other institutions, especially federal agencies and corporate compounds, have similar networks for optimizing local data sharing. Most of these networks are, in effect, networks of departmental LANs that are interlinked by a variety of protocols with TCP/IP being the most common among colleges. LANs will become increasingly important in government institutions, business corporations, and nonprofit organizations that are served by libraries and information centers. Between 1988 and 1989, the shipment of network interface adapter boards (interface cards) almost doubled from 3.4 million to 5.3 million.[5] As an information optimization structure, the LAN will continue its role of linking departmental/unit computing to external information resources using gateways and bridges.[6] Effective interlan connections demand more bandwidth, for which optic fiber for the backbone trunk lines is an ideal medium.

The last major trend to discuss that is continuing in the 1990s is national and international computer networks used for remote information access. Research-oriented networks like ARPANET and DARPA for defense contractors have been operational for some time. Some research networks are regional, like the Bay Area Regional Research Network (BARRNET) in the Southwest, or statewide, like the New York State Educational and Research Network (NYSERNET). The ultimate national research network, the National Science Foundation network (NSFNET), was initiated in 1986 to connect researchers to supercomputing sites but is in effect tending to interlink other networks like BITNET (a multi-purpose network for higher education) and ARPANET. For library-specific applications, the development of commercial data banks like Dow Jones, Dialog, Compuserve, The Source, LEXIS, and Westlaw has increased user appetite for remote access to information. While the library community has been traditionally served by value-added networks like TYMNET and TELENET, links will be needed via the research networks to reach the research community. The 1990s will provide more sophisticated information technology to simplify as well as accentuate access to relevant information. The implicit requirements of network speed, high volumes, and accuracy would entail the use of broadband and inevitably optic fiber as the transmission medium of choice.

This decade will certainly usher in highly developed information infrastructures and a relatively high demand for fast, accurate information from libraries and information centers. Due to the primacy of information products and services, most users will prefer full-text information packages with graphics and in some cases accompanying voice explanations. Linking the disparate devices needed for multimedia transmitters and receivers is likely to be ISDN, whose products will be deployed in larger numbers in the 1990s. Broadband ISDN or BISDN based on fiber optics will be a common internet link technology.

OPTICAL TECHNOLOGY IN INFORMATION PROCESSING

Large-scale commercial use of fiber optics in the United States first occurred in telecommunications in the early 1980s, when long-haul carriers used it in their cabling systems. By 1990, national carriers such as AT&T, MCI, and U.S. Sprint had completed most of their coast-to-coast route upgrades from copper to optic fiber-based cabling.[7] Information systems units in general and data processing shops in particular were later additions to the users of this technology. Some of the catalysts that have stimulated the upsurge in the use of optic fiber technology in the recent past include FCC electro-magnetic emission regulations enforced in October 1983,[8] increasing volume of systems sophistication and internet output that outstrip the copper distances and speed, transmission signal security, and electro-magnetic interference (noise) in industrial and medical environments.

The Technology

Variously known as lightwave guide, optic fiber, or fiber optics, the use of glass fiber in the propagation of data signals is a relatively new science. Glass fiber technology per se is, however, an old science. Creation of glass fibers can be traced as far back in history as the seventeenth century when Venetians heated glass to very high temperatures and pulled the molten glass to form thin strands of fibers.[9] The resultant fibers were used for decorating glassware. Practical applications of glass fiber as a signal-communication medium are attributed to the 1880 invention of the photophone by Alexander Graham Bell.[10] Using sunlight reflected by a thin mirror that was modulated by voice, the device carried sound signals. For a receiver, he used a photoconducting selenium cell that converted sound signals to electrical signals.

Although interesting, Bell's invention was not practical, because it used natural elements. Sunlight as the light source and the atmosphere as the propagation medium proved unpredictable and uncontrollable. Two basic elements needed—a stable light source and a transmission conduit—were not invented until the twentieth century. The first breakthrough was the invention of the light source, generically known as an optic oscillator, in the late 1950s. Successful parallel experiments were conducted at AT&T Bell Labs, IBM, General Electric, and MIT Lincoln Labs, which developed the laser (acronym for Light Amplification by Stimulated Emission of Radiation). AT&T was the first to create units that operated at room temperature in 1960. Bell Labs were also pioneers in developing an alternative to lasers in the form of light-emitting diodes (LEDs) in the 1970s. The search for a fiber for light propagation ended in 1970 when a Corning Glass research team (Robert Maurer, Don Keck, and Peter Schultz) developed a 20 dB/km attenuation rate optical fiber. Though this figure has been superseded (for example, a 4 dB/km attenuation rate is now common for high-quality fiber), it was the lowest signal attenuation threshold for the development of commercially viable cables.

At present there are essentially two types of fibers, multi-mode and single-mode (see figure 1). The latter was the first laboratory product, but multi-mode was first adopted for commercial use since it was easier to couple with LEDs or lasers. The main difference between the two is that the multi-mode has a larger central core (fifty micrometers) than the single-mode (eight micrometers). Most optic fiber installations are multi-mode utilizing LEDs. Use of single-mode fibers for future cabling will increase as the technology matures, because it exhibits less dispersion and a relatively lower attenuation rate.

Advantages over Other Media

Although optical fiber-based systems are not perfect, they have several advantages over most of the conventional communication media:

- The raw material for glass fiber is silicon dioxide or beach sand which is cheap and abundant in contrast to the cost of copper wire. It is lightweight, smaller in size, and thus easier to install in communications conduits which are already congested with copper cabling. Another factor of cost is the cost of cabling per meter per telephone channel(s). In this regard optic fiber is far superior to copper.
- Carrying capacity is better than that of most other media. A standard single telephone fiber cable has a 12.7mm diameter and contains 144 fibers. Each fiber operates at T3 rate or 672 channels and can transmit over 96,767 telephone calls. In contrast, a cable of copper wire with a diameter of 70mm contains 900 pairs, with each pair carrying 24 voice channels at T1 rate yielding a cable capacity of 21,600 telephone calls. Also, fiber optic cable can transmit either analog or digital information.
- The replacement lifespans of optic fibers is longer, especially in high-humidity environments. Unless it is very heavily shielded, copper will tend to corrode and may need replacement every two to three years. The normal period of operation for fiber optics is twenty-five to thirty years.[11]
- Due to the fact that optic fiber is an insulator, or dielectric, no electric current flows through it. Consequently, it is highly resistant to radio frequency interference (RFI) or electromagnetic interference (EMI). It is also not amenable to crosstalk among adjacent lines, thus allowing many fibers carrying several channels to be packed in a single encasement.

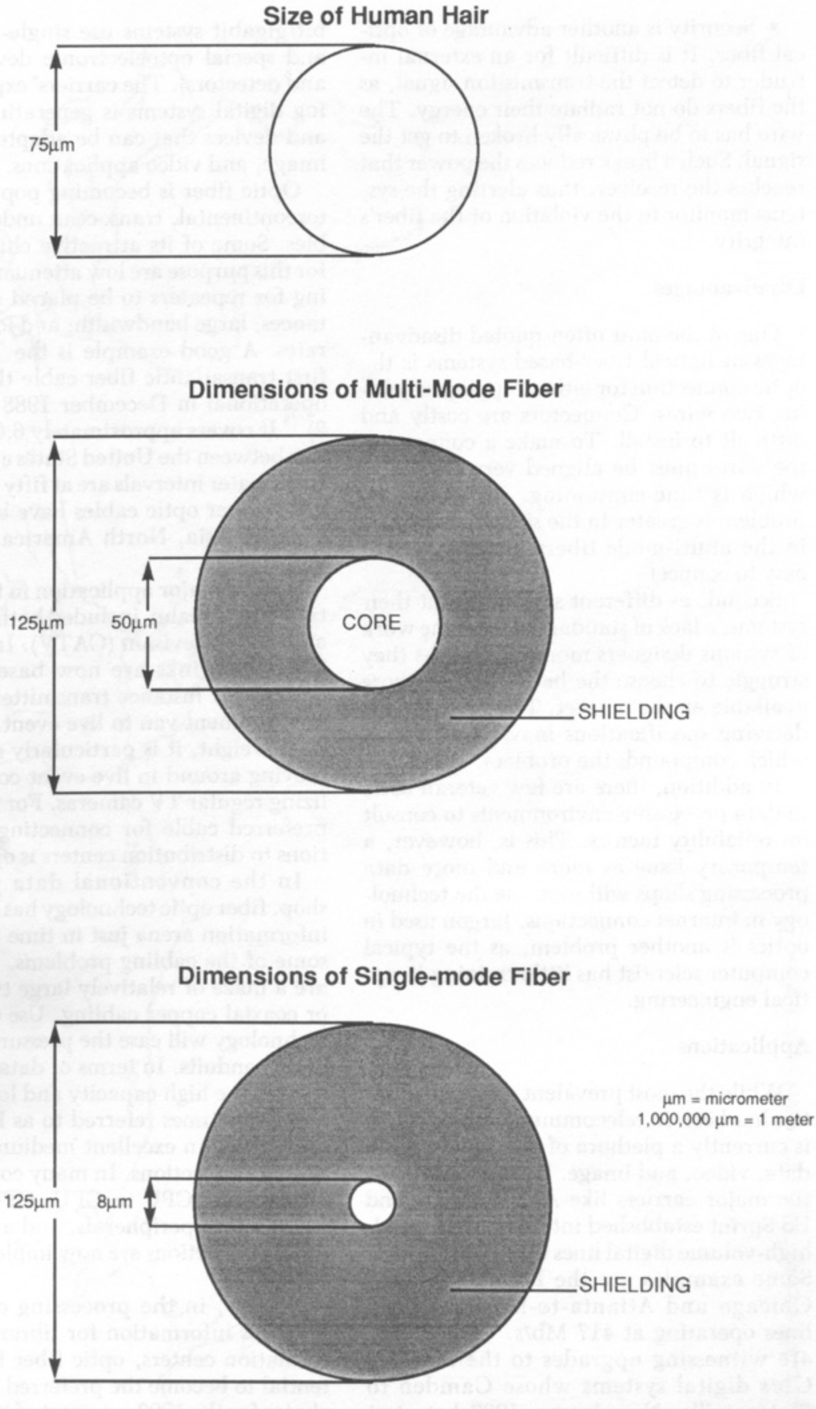


Figure 1. Dimensions of Optic Fibers.

• Security is another advantage of optical fiber. It is difficult for an external intruder to detect the transmission signal, as the fibers do not radiate their energy. The wire has to be physically broken to get the signal. Such a break reduces the power that reaches the receiver, thus alerting the systems monitor to the violation of the fiber's integrity.

Disadvantages

One of the most often-quoted disadvantages of optical fiber-based systems is the optic connection for either tapping or joining two wires. Connectors are costly and difficult to install. To make a connection the wires must be aligned very carefully, which is time-consuming. Although this problem is greater in the single-mode than in the multi-mode fibers, neither type is easy to connect.

Second, as different suppliers tout their systems, a lack of standards makes the work of systems designers more difficult as they struggle to choose the better among those available on the market. Test methods for deriving specifications may be different, which compounds the problem.

In addition, there are few veteran users in data processing environments to consult on reliability factors. This is, however, a temporary issue as more and more data processing shops will soon use the technology in internet connections. Jargon used in optics is another problem, as the typical computer scientist has little training in optical engineering.

Applications

While the most prevalent use of fiber optics has been in telecommunications, there is currently a plethora of users and uses in data, video, and image. During the 1980s, the major carriers like AT&T, MCI, and US Sprint established intercity high-speed, high-volume digital lines using fiber optics. Some examples are the Philadelphia-to-Chicago and Atlanta-to-Mosley AT&T lines operating at 417 Mb/s.[12] The 1990s, are witnessing upgrades to the new 1.7 Gb/s digital systems whose Camden to Pleasantville, New Jersey, 1987 beta test appears to have succeeded. These megabit/gigabit systems use single-mode fiber and special optoelectronic devices (lasers and detectors). The carriers' experience using digital systems is generating expertise and devices that can be adapted for data, image, and video applications.

Optic fiber is becoming popular for intercontinental, transocean underwater cables. Some of its attractive characteristics for this purpose are low attenuation, allowing for repeaters to be placed at long distances; large bandwidth; and low bit error rates. A good example is the TAT-8, the first transatlantic fiber cable that became operational in December 1988 (see figure 2).[13] It covers approximately 6,000 kilometers between the United States and Europe. Its repeater intervals are at fifty kilometers. Other fiber optic cables have been laid to connect Asia, North America, Oceania, and Europe.

A third major application in the interactive video realm includes both broadcast and cable television (CATV). In broadcast TV, short links are now based on fiber optics—for instance transmitter to studio, or equipment van to live event. Due to its light weight, it is particularly suitable for moving around in live event coverage utilizing regular TV cameras. For CATV, the preferred cable for connecting earth stations to distribution centers is optic fiber.

In the conventional data processing shop, fiber optic technology has entered the information arena just in time to alleviate some of the cabling problems. Most shops are a maze of relatively large twisted-pair or coaxial copper cabling. Use of this new technology will ease the pressure in the cabling conduits. In terms of data communication, the high capacity and low bit error rate (sometimes referred to as BER) make optic fiber an excellent medium for intersystem connections. In many computer environments, CPU to CPU, CPU to memory, CPU to peripherals, and a number of other connections are now implemented by optic fiber.

Finally, in the processing of text and graphics information for libraries and information centers, optic fiber has the potential to become the preferred medium of choice for the 1990s. As part of its preservation program, the Library of Congress is

Figure 2. Transocean Underwater Fiber Optic Cables.

currently converting some of its rare books to optically stored versions.[14] Technology is now in place to enable remote access to such collections, and pages can be turned without physically touching the items. Within the 1990s it should become economical to view the fifteenth-century illuminated manuscripts at the Library of Congress from a library in Seattle or Miami. In addition, the concept of interlibrary loans will be tremendously affected by the ability to transmit full text by telefacsimile and a variety of other electronic means.

THE ISDN FACTOR

The integrated services digital network (ISDN) concept will be utilized in many internet information systems designs. ISDN is the nearest to providing the end user with total information needs and requirements using the same devices. It has evolved from the telephone Integrated Digital Network (IDN) that provides end-to-end connectivity for supporting a broad range of services, such as video, voice, data, facsimile, and telemetry (see figure 3).

To understand fully the implications of the concept one needs to examine its elements separately. First, it is *integrated* in that the same structure can be used for a number of communications modes, including, data, voice, video, facsimile, and others. Second, it may be used to cater to the many *services* an institution may need to function effectively, for example, telephone message communication services, data processing services, and video or image processing services. In an ideal ISDN environment, all these services use the same transmission media and similar input, storage, and output devices. Third, it transmits *digital* as opposed to the conventional telephone's analog transmission. In addition, the switching mechanism is also digital whereby bits rather than volts are switched. Finally, it has been conceptualized and implemented as a *network* that permits devices to send, receive, and acknowledge messages to and from other devices.

ISDN coupled with fiber optics fits the specs of the network of the future as visualized by John Mayo and William Marx. Their view is that

> The information network of the future will give anyone, anywhere, anytime, access to voice, data, images, and video, in any combination. Much as we now plug an appliance into an electrical outlet, the telephone user of the future will plug a terminal into a universal port, taking for granted that it will operate, and link to a worldwide bank of facts, information, advice, entertainment, and thousands of other functions.[15]

The advent of broadband ISDN, which is based on fiber optics and known as BISDN, makes this feasible. When fully implemented, it will be a boon for library user access to information.

One of ISDN's problems is the lack of internationally accepted standards. However, the International Telecommunication Union's International Telegraph and Telephone Consultative Committee (CCITT) has proposed a number of ISDN standard interfaces that are gradually receiving international acceptance. In the United States, the American National Standards Institute (ANSI) has promoted the Fiber Distributed Data Interchange (FDDI) standard, which is nearing approval.[16] FDDI is a standard for 100Mb/s ring LAN using fiber optics. In addition, prototype standard and protocol-testing systems for checking conformance and performance are being perfected.[17]

TECHNICAL ISSUES

Several technical problems come to mind when one envisions the library as a network "information node" on national or regional networks.[18] Four salient issues will be discussed, as follows: automated library systems archipelago syndrome; redesigning user interfaces; reciprocal wide area/national access and patron identification; and the librarian's knowledge base of telecommunications.

First, the automated library systems archipelago syndrome afflicts most library automation programs in existence today. For decades, librarians and other information specialists have welcomed "turnkey" and custom-made, library-oriented hardware and software. The rationale for using such systems has traditionally been that they are easy to implement. With a few exceptions, the average library system has automation products that were designed to operate as individual units in stand-alone

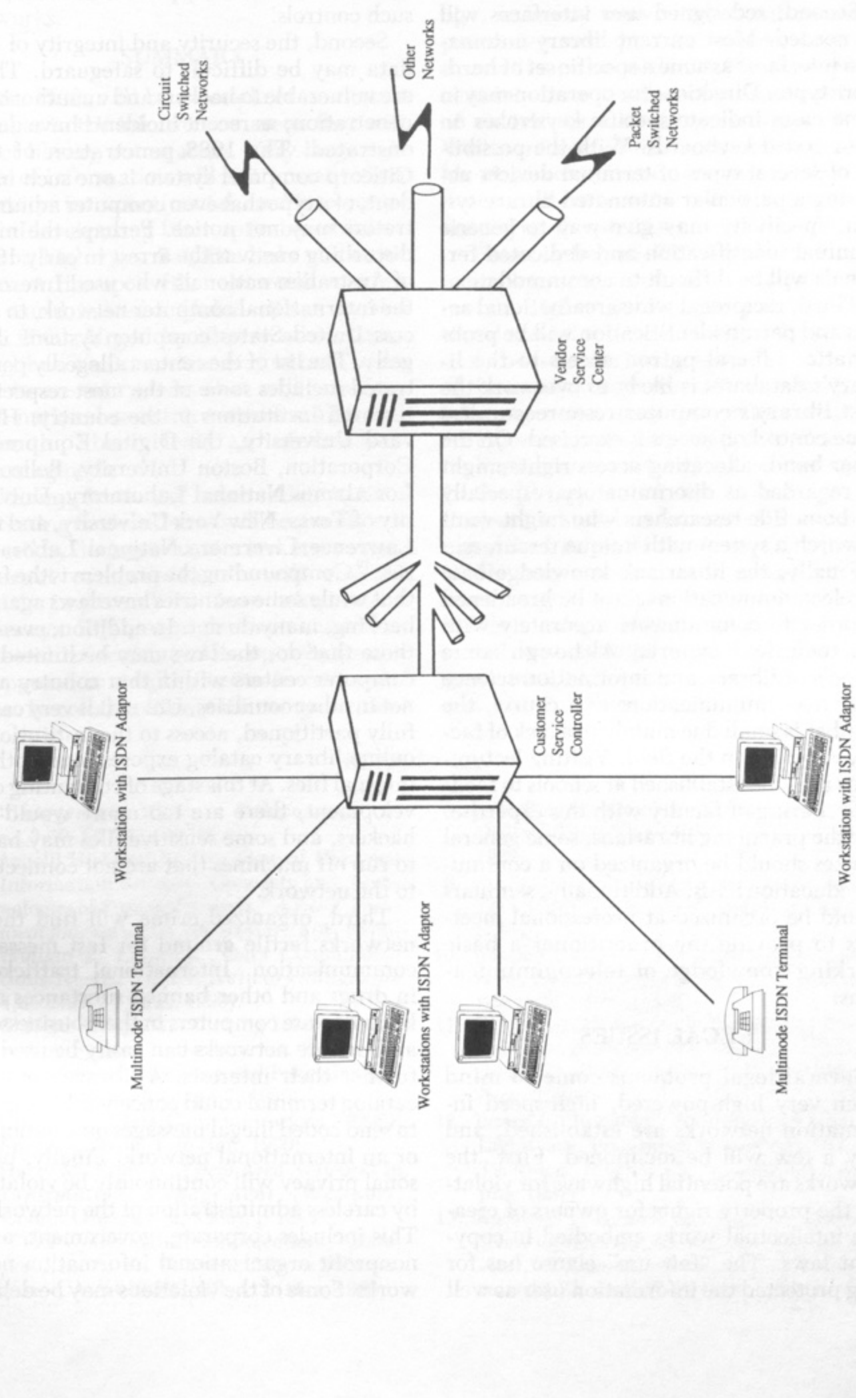


Figure 3. ISDN-based Network.

modes. Since they were not designed for internet operability they need a complex overhaul for libraries truly to be nodes on national networks using either the TCP/IP or OSI-based protocols.

Second, redesigned user interfaces will be needed. Most current library-automation interfaces assume a specific set of hardware types. Directions for operation may in some cases indicate specific keystrokes on color-coded keyboards. With the possibility of several types of terminal devices accessing a particular automated library system, specificity may give way to generic terminal identification and dedicated terminals will be difficult to accommodate.

Third, reciprocal wide area/national access and patron identification will be problematic. Liberal patron access to the library's databases is likely to overwork the host library's computer resources unless some control on access is exercised. On the other hand, allocating access rights might be regarded as discriminatory, especially for bona fide researchers who might want to search a system with unique resources.

Finally, the librarian's knowledge base of telecommunications must be broadened in order to communicate accurately with the technical experts. Although some schools of library and information science offer telecommunications as a course, the number is small due mainly to a lack of faculty expertise in the field. Visiting lectureships must be established at schools to facilitate sharing of faculty with this expertise. For the practicing librarians, some general courses should be organized on a continuing education basis. Additionally, seminars should be organized at professional meetings to provide the practitioner a basic working knowledge of telecommunications.

LEGAL ISSUES

Several legal problems come to mind when very high-powered, high-speed information networks are established, and only a few will be mentioned. First, the networks are potential highways for violating the property rights for owners of creative intellectual works embodied in copyright laws. The "fair use" clause has for long protected the information user as well as the librarian who might otherwise be deemed an accomplice. While national governments may be able to monitor violations on government-owned networks, there will be too many private networks for such controls.

Second, the security and integrity of the data may be difficult to safeguard. They are vulnerable to hackers and unauthorized penetration, as recent incidents have demonstrated. The 1988 penetration of the Citicorp computer system is one such incident, of a type that even computer administrators may not notice. Perhaps the most disturbing one was the arrest in early 1990 of Australian nationals who used Internet, the international computer network, to access United States computer systems illegally. The list of the centers allegedly penetrated includes some of the most respected research institutions in the country: Harvard University, the Digital Equipment Corporation, Boston University, Bellcore, Los Alamos National Laboratory, University of Texas, New York University, and the Lawrence Livermore National Laboratories.[19] Compounding the problem is the fact that while some countries have laws against hacking, many do not. In addition, even in those that do, the laws may be limited to computer centers within that country and not in other countries. Unless it is very carefully partitioned, access to the institution's online library catalog exposes many other internal files. At this stage of computing development, there are too many would-be hackers, and some sensitive files may have to run off machines that are not connected to the network.

Third, organized crime will find these networks fertile ground for fast message communication. International traffickers in drugs and other banned substances are known to use computers in their businesses, and private networks can easily be used to further their interests. A library's online catalog terminal could conceivably be used to send coded illegal messages on a national or an international network. Finally, personal privacy will continuously be violated by careless administration of the networks. This includes corporate, government, and nonprofit organizational information networks. Some of the violations may be deliberate, while others may be due to innocent human error. Again, an apparently innocent but sophisticated user can use the library's OPAC terminal as a window to access personal information on high-security networks.

SUMMARY

The 1990s will be a decade of numerous changes in the information arena that will affect library automation. Very large volumes of information will be transmitted on intercity, intracontinental, and intercontinental levels using fiber optic-based information highways. Satellite links will complement the optic fiber networks making the world indeed a global village by the year 2000. Urged by library administrators, library automation vendors need to retool to create systems that will interface with national research networks. This will make the library truly an information-accessing node.

Communication of information in data processing environments will be greatly enhanced by device-to-device low error rates. Internet communication will also be greatly improved by these information highways. Equipment designers may have to retool to accommodate optoelectronics as opposed to simply electronic device designs. In collaboration with computer center personnel, librarians should be able to identify general or specific systems that incorporate this new technology.

Information systems designers of this decade need to learn a lot of new technology, as many conventional information processing concepts, techniques, or practices will soon be outdated. The library will certainly have a different posture in terms of content as well as services provided. With so much full-text information accessible at remote locations, many of the volumes currently held will no longer be needed on the premises.

The coupling of ISDN and fiber optic technology will result in truly super information highways. Standards at both national and international levels will continuously be required. Judging by the activities of the national and international standardizing agencies, near-universal standards will soon be in place to effect the full implementation of these information highways. In the library world, the Linked Systems Project, which has successfully concatenated the Library of Congress, OCLC, and RLIN into a transparent, seamless network, augurs well for future networking. While legal problems will not be easily solved, especially in the private networks, governmental licensing should curb the excesses in violations.

REFERENCES

1. "Integrated PBXs," *Teleconnect* 9, no.3:50–66 (May 1991).
2. Arnold Hieber, "An Overview of Universal Information Services: Concepts and Technologies of Future Networks," *AT & T Technical Journal* 68,no.2:8–9 (Mar./Apr. 1989).
3. William P. Lidinsky, "Data Communications Needs," *IEEE Network Magazine* 4,no.2:29–30 (Mar. 1990).
4. Clifford A. Lynch, "Linking Library Automation Systems in the Internet: Functional Requirements, Planning and Policy Issues," *Library Hi Tech* 7,no.4:7 (1989).
5. Gary Stix, "Data Communications," *IEEE Spectrum* 27,no.1:35 (Jan. 1990).
6. Harry M. Kibirige, *Local Area Networks in Information Management* (Westport, Conn.: Greenwood, 1989), p.18–27.
7. "Fiber Optics and Communications," *Datapro Reports on Communications Alternatives*, CA40-010-101:101–112 (Mar. 1990).
8. Julie Pingry, "Problem Solving With Fiber Optics," *Digital Design* 13, no.3:77–78 (July 1983).
9. Michael Koenig, "Fiber Optics and Library Technology," *Library Hi Tech* 2, no.1:9–10 (1984).
10. Joseph C. Palais, *Fiber Optic Communications*, 2d ed. (Englewood Cliffs, N.J.: Prentice Hall, 1988), p.1–2.
11. Fred West, "Fiber Optic Technology Promises Dramatic Impact on Data Communications," *Data Management* 24, no.11:20–21 (Nov. 1986).
12. James E. Rogalski, "Evolution of Gigabit Lightwave Transmission Systems," *AT&T Technical Journal* 66, no.3:32–40 (May/June 1987).
13. Kathleen J. Hansell and Fran M. Putney, "New Options for Transatlantic Digital Services," *Telecommunications* (North American ed.) 23, no.6:27–28 (June 1989).

14. Jay Ward Brown, " The Once and Future Book: The Preservation Crisis," *Wilson Library Bulletin* 59, no.9:595–96 (May 1985).
15. John S. Mayo and William B. Marx, "Introduction," *AT&T Technical Journal* 68,no.2:2 (Mar./Apr. 1989).
16. Gary Stix, "Data Communications," *IEEE Spectrum* 27,no.1:35–36 (Jan. 1990).
17. W. B. Davis and others, "Architecture and Design of a Portable Protocol Tester," *IEEE Network Magazine* 4,no.2:24–31 (May 1991).
18. Lawrence E. Murr and James B. Williams, "The Role of the Future Library," *Library Hi Tech* 5, no.3:7–11 (Fall 1987).
19. John Morkoff, "Arrests in Computer Break-ins Show a Global Peril," *New York Times*, April 4, 1990, A1 & A16. ■■

More on Improved Browsable Displays for Online Subject Access

Dorothy McGarry and Elaine Svenonius

Key problems of subject searches in online public access catalogs are that too much may be retrieved and that the display may be arranged in what appears to be an unintuitive and unhelpful manner. A study was undertaken on one large database to determine how often the display of a subject and its modifications and subdivisions extended over more than two screens and what the results on display would be of compressing geographic subdivisions and other modifications. The study then addressed the question of how often the display of a term and its subdivisions was interrupted by the appearance in the display of inverted headings, phrase headings, and headings with parenthetical qualifiers.

STATEMENT OF PROBLEMS

A key problem of online subject searching is that too much may be retrieved. It is not unusual for a user, in attempting to browse an index of Library of Congress (LC) Subject Headings, to be presented with screen after screen of entries all representing modifications of one heading. This problem is exacerbated by a second problem—the entries presented are often arranged in what appears to be an unintuitive and unhelpful manner. In this paper we present empirical data that can be used to assess the severity of these two problems and suggest means for addressing them. The test data-bed used to supply the data is the index of LC subject headings used in ORION, the UCLA online catalog.

The Lengthy Display Problem

An excellent paper by Mia Massicotte sets forth the problem of browsable displays that are too long.[1] Massicotte contends that any subject heading display over one screen in length is too long. No doubt what is too long is in the eye of the searcher; however, some evidence from perseverance studies suggests that a majority of online catalog users do not scan more than thirty-five items.[2] Massicotte argues that conventional means to contain display size, e.g., the techniques of Boolean operation and word proximity, are not always helpful. This is because users may not know how to specify their searches; they may not know from what aspects or points of view a given subject is treated.

The solution Massicotte proposes to improve subject heading displays is to introduce a conceptual order into their presentation. She suggests that the size of the subject heading index could be "vastly reduced" if certain compression techniques were introduced. For instance, a routine could be created that would substitute for a class of subject headings, e.g., those subdivided geographically (indicated by subfield code

Dorothy McGarry is Head, Cataloging Division, Physical Sciences and Technology Libraries, and Elaine Svenonius is Professor, Graduate School of Library and Information Science, University of California at Los Angeles.

z), with a message to the effect that these headings could be found in such and such a location. For example, the last entry for the heading Islam and politics might read: "9 Islam and politics—subdivided by geographical area, e.g. Afghanistan." Keying in "9," the user, in a second retrieval step, would be shown all the geographic subdivisions of Islam and politics. Massicotte proposes that the same procedure might be followed for subfield y, which is used for chronological subdivisions. Subfield x, used for form and topic subdivisions, is not so amenable to compression by this technique. The solution Massicotte proposes here is to group form and topical subdivisions into broad conceptual categories representing different aspects or facets of a topic: graphic or illustrative, technical, legal, mathematical, scientific, economic, sociological, or political.

In our study we looked at the effectiveness of two kinds of compression in reducing the length of subject heading displays: (1) geographic compression, as defined by Massicotte, and (2) blanket compression, that is, compression that results from eliminating elements that (a) immediately follow a selected heading and (b) repeat.[3] For example, if "Art, French . . ." occurred ten times, in compression it would be reduced to one occurrence. Similarly treated would be "Art of children" and "Art—Research." To signify that the one heading remaining after compression is emblematic of a number of headings, a symbol such as an asterisk could be used; the asterisk would indicate that further instances of the modified heading can be found by keying in the number associated with the heading. While perhaps not as elegant as Massicotte's solution, we felt a semi-algorithmic approach to compression of headings would be simpler and, thus, less costly.

The Interrupted Display Problem

The second phase of our study looked at the problem of ordering entries within a subject heading display. This problem has been under scrutiny by a subcommittee of the ALA ALCTS CCS Subject Analysis Committee (SAC) that has been instructed to study the display of subject headings in subject indexes in online public access catalogs.[4] The subcommittee has looked at two alternatives for ordering or "filing" of subject heading entries: ordering as it is done following Library of Congress filing rules and as it is done following ALA filing rules.[5,6] The first filing method results in what the subcommittee calls a structured display (see figure 1). In a structured display the subdivisions of a given heading file before entries for the heading are modified in other ways, for example, by an inverted modifier, by words forming a phrase, or by a parenthetical qualifier. The second filing method results in a more strictly alphabetical display (see again figure 1).

The SAC Subcommittee deliberated the pros and cons of the two types of display—the structured and the alphabetical. The chief advantage of the alphabetical display is that users do not have to learn complex filing rules in order to understand the display. Moreover, an alphabetical display can help users who are unsure of the prevailing LCSH syntax (Is it "Water—Aeration" or "Water, Aeration of"?) Finally, an alphabetical display that requires relatively little in the way of special programs to be written is cheaper than a structured display. A structured display, on the other hand, has a singular advantage in its favor: it is logical. Given that one purpose of subdivision syntax is to break up large headings, it is illogical to separate the subdivisions of a heading. At times subdivisions can be widely separated. In the ORION database, under ART there is a case where more than 25 screens separate alphabetically proximate subdivisions. The unwary user scanning the entries under Art could hardly imagine that when he reached "Art—Missouri" that the next subdivision, "Art—Moldavian S.S.R." was 25 screens further on! In our second study phase, then, we examined how frequent and how large were interruptions in the sequencing of the subdivisions of a heading in a large catalog database like ORION. The ORION subject index generally follows the ALA rules in its arrangement and presents headings in alphabetical order without regard to punctuation. With a view to speculating on possible remedial action, we also looked at the incidence of different kinds of interruptions, for example, were sequences of sub-

Alphabetical Display	*Structured Display*
Reading—Ability testing	Reading—Ability testing
Reading—Abstracts—Periodicals	Reading—Abstracts—Periodicals
Reading (Adult education)	Reading—Africa—Congresses
Reading—Africa—Congresses	Reading—Case studies
Reading—Case studies	Reading—Color aids
Reading—Color aids	Reading—Computer programs
Reading comprehension	Reading—Congresses
Reading—Computer programs	Reading—France
Reading—Congresses	Reading—Great Britain
Reading disability	Reading—Remedial teaching
Reading (Elementary)	Reading—Research
Reading (England)	Reading, Psychology of
Reading—France	Reading (Adult education)
Reading games	Reading (Elementary)
Reading—Great Britain	Reading (England)
Reading (Higher education)	Reading (Higher education)
Reading, Psychology of	Reading comprehension
Reading readiness	Reading disability
Reading—Remedial teaching	Reading games
Reading—Research	Reading readiness

Figure 1. Alphabetical and Structured Displays.

divisions interrupted more by inversions, phrases, or parenthetical qualifiers?

METHODOLOGY

Samples

Our primary sample consisted of 64 subject headings selected from the subject heading index of the UCLA ORION database. The sample size was determined based on a standard error of 10 percent, a confidence level of 95 percent, and a desire to generalize about the percent reduction that could be achieved for 80 percent of our sample. Subject headings were chosen in the following manner: The first screen in the ORION subject heading index was examined to see if any of the headings there extended for more than two screens, that is, if there were two screens or more of headings that began with the same left-most main heading. If such a heading was found, it was included in our sample. If no such heading was found the enter key on the computer was pressed ten times (moving the display 10 screens forward) and the then current screen was examined for two-screen-plus headings. Whenever such a heading was found, the sampling of every tenth screen commenced following the last entry for that heading. While this method of sample selection proved satisfactory for generating the 64 subject headings that were to be analyzed for certain attributes, it was not valid for generating data to answer the question about how frequently in the database two-screen-plus headings in fact occurred. Therefore, a secondary sample was taken for the specific purpose of making this generalization. Ten randomly selected letters of the alphabet were chosen, and beginning with the first entry under that letter, every tenth screen was sampled and scanned for the presence or absence of two-screen-plus headings. A sample of this size was expected to support generalization about the frequency in the database of two-screen-plus headings at a confidence level of 95 percent and a precision of 8 percent.

Attributes Measured

From each subject heading in the 64-item and 100-item samples we eliminated LC Children's headings, MeSH headings, local headings, and headings that appeared to be obviously wrong, e.g., with typographical errors or direct geographic subdivision. Then, for each heading in the 64-item sample we collected the following data:

- The total number of entries
- The number of interruptions in the sequencing of the first subdivisions of a main heading caused by phrase headings, in-

verted headings, and parenthetical qualifiers

- The largest number of entries contained in any interruption
- The total number of headings minus those whose first subdivision is a geographic subdivision
- The total number of headings minus those whose first subdivision is a form subdivision
- The total number of headings minus those whose first subdivision is a geographic element or whose first element in phrase, inverted, parenthetical, or subdivided syntactic construction is repeating. (If a subject heading, e.g. "Abortion—Law and legislation," is further divided, say ten times, nine of the headings beginning "Abortion—Law and legislation" are said to be repeating. Similarly, if a subject heading, e.g. "Art, Modern," is further subdivided 773 times, 772 of them are said to be repeating.)

Once these counts were taken, simple descriptive statistics were used to summarize the data. Two measures of compression were defined:

1. Degree of geographic compression is the fraction whose denominator is the total number of headings in a subject heading set and whose numerator is the total number of headings minus those having first subdivisions that are geographic.

2. Degree of blanket compression is the fraction whose denominator is the total number of headings in a subject heading set and whose numerator is this number minus geographic first subdivisions and repeating first elements. Percentage reduction in display size resulting from the two different compression techniques was measured by subtracting the above fractions from 1.

RESULTS

Lengthy Displays

The first question our study sought to answer was how often lengthy subject headings occurred in the ORION subject heading index. A lengthy subject heading was understood to mean one whose extensions by subdivisions, phrases, inversions, or parenthetical qualifiers numbered more than 60, which is equivalent to approximately 2 or more screens. Sampling every tenth screen of 10 randomly selected letters of the alphabet, we found that, on the average, 53 percent of the ORION subject heading screens were lengthy. The largest heading in our primary sample was "Art," which contained 6,889 entries. This means that, at 30 lines per screen, headings beginning with the word "Art" extended more than 230 screens. Seven headings, more than a tenth of our sample, contained more than 1,000 entries. The mean number of entries per heading was 527 entries (17.6 screens). This number is suspect, however, given the occurrence in our sample of a few very large headings, such as "Art." For this sample, the median is a more reliable statistic than the mean. The median number of entries per heading was 214, or slightly over 7 screens. We cannot assume that the incidence in the ORION database of two-screen-plus headings perfectly mirrors the incidence with which users stumble on such headings in their searches. (Another study would be needed to establish such a correlation.) Nevertheless, simply on a probabilistic basis it would seem that the problem of subject-heading overload in a large ORION-like catalog is pervasive. Given that online catalog users tend to balk if presented with more than 35 entries to scan, it would seem also the problem of unsatisfied subject searchers must be pervasive.

We looked at two methods for reducing the incidence of lengthy headings. Following the suggestion of Massicotte, the first of these was deleting from the initial display of entries for a subject heading those whose first subdivision was geographic. We found that, by deleting such entries, a 33.65% reduction in display length could be achieved (see table 1). In other words, geographic compression reduced the average (median) heading to 66 percent of its original size. One heading, "Annotations and citations (Law)," which was divided exclusively by geographic subdivision, was reduced from 70 entries to one, a reduction of more than 98 percent. Overall 22 (34 percent) of the 64 headings in our sample were reduced to less than two screens after geographic compression. While substantial, this reduction was not as much as was hoped. One reason for this was that certain headings could not be reduced in size by geographic compres-

Table 1. Effects of Geographic and Blanket Compression

	Initially	After Compression Geog.	Blanket
Entries per heading			
Mean	526.92	431.02	65.81
Median	214	108.5	48.5
Reduction per heading			
Mean	0	40.36%	79.26%
Median	0	33.65%	81.28%
Headings less than two screens	0	22 (34%)	35 (55%)

sion either because they could not be divided geographically (15 fell into this class) or because they were divided thus not at the first, but at the second subdivision, e.g., "Bibliography—Bibliography—China."

The second method we looked at for reducing the number of entries displaying under subject headings, called "blanket compression," was to delete from the initial display of entries for a subject heading not only those with geographic (first) subdivisions but also those with repeating elements in subdivisions, phrases, inversions, or parenthetical qualifiers immediately following the heading. The results of blanket compression were dramatic, even better than expected: display length was reduced over 81 percent. In other words, as a result of blanket compression the average (median) heading was reduced to less than 19 percent of its original size. Of the 64 lengthy (two-screen-plus) headings in the sample, 35 (55 percent) were reduced to less than 2 screens. More than 80 percent were reduced by two-thirds or more. Certainly it would seem that blanket compression is an effective means to improve the browsability of subject heading displays in large databases like ORION.

The idea of compression combined with a two-step retrieval process is not new and some systems currently in operation do, in fact, operate in such a manner. In SCORPIO and (the former) LS 2000, the searcher controls whether or not to display the subdivisions of a heading. In SCORPIO the subdivisions of a heading, e.g., "Art," are shown only if a hyphen is typed after the heading (Art-). In LS 2000, headings display initially without subdivisions, but, on command, any heading can be expanded so all of its subdivisions display. Thus, in both systems subdivisions are suppressed in the first step of retrieval. While effective in terms of economical display, it can be argued that suppression of this sort hides information that might be useful to users without their being conscious of it. A second criticism that can be leveled at suppressing subdivisions is that it incurs a user burden insofar as a second retrieval step is always called for to achieve specification. The method we propose—blanket compression—aims for a middle ground between no suppression and total suppression and is less subject to such criticisms.

The effectiveness of different compression techniques in reducing the incidence of lengthy displays in subject heading indexes needs to be explored further. In our study we looked at only two possible types of compressions— geographic and blanket. The latter is semi-automatic, requiring only that a computer algorithm be defined to recognize a repeating first element. Other compression techniques, involving more intellectual effort, include the chronological and faceted-topic compression suggested by Massicotte. Form compression, in the manner of the geographic compression described above, is also a possibility. This would require differentiating form headings from topical ones (to the extent this is possible) and giving them different subfield codes.[7] In collecting our data we noted how many headings were immediately followed by form subdivisions. We calculated what the effect would be if these subdivisions were consolidated in the manner done for geographic subdivisions. The result was not impressive: headings size was reduced only 10 percent.

Compression is not the only means to reduce lengthy displays. Another approach would be simply to segregate certain subject heading elements, relegating them to separate MARC fields.[8] Chronological, form, and geographic subdivisions, for example, might be eliminated from the 6xx fields, letting fields 045 and 655 and an expanded 043 or 052 field shoulder responsibility for these facets of subject access. The reduction achieved in length of display would be the same, whether compression or segregation techniques were used, at least as regards the geographic, form, and chronological facets of subject headings. Compression algorithmically based on eliminating repeating elements does not have a segregation analogue.

Interrupted Displays

The question we addressed in our second study phase was how often the displays of subdivisions of headings were interrupted by the intrusion of different kinds of syntactical construction, in particular, phrases, inverted headings, and parenthetical qualifiers. We found that the largest heading in our sample, "Art," suffered interruption of its subdivisions 222 times. On the average, the subdivisions of a heading were interrupted 11.27 times; most of the interruptions were due to inversions (77.39 percent), relatively few to phrases (21.22 percent) and hardly any at all due to parenthetical modifiers (1.39 percent) (see table 2). Some of the interruptions were very large indeed: The subdivisions of "Art," as noted earlier, were broken in one place by a block of 773 entries (25 screens); those of "Architecture" by 515 entries (17 screens) and "Agriculture" by 413 entries (14 screens). A full 10 percent of the headings in our sample suffered breaks in subdivision display of more than three screens (100 entries).

While the above figures suggest that the interruption problem is serious, once again summary statistics expressed in terms of means are misleading when there are large outlying figures, such as those for "Art" and "Architecture." For a heading of median size, consisting of 214 entries, the median number of interruptions was only 2. It is of interest to note that half of these were caused not by inverted headings, which appeared to be the main culprit when considering means, but by phrases.

There are those who would agree that the interruption of subdivisions by alternative syntactic forms constitutes a problem, but would not be willing to pay the price of altering filing rules to eliminate it. The question thus arises whether it might be possible to reduce the number of subdivision interruptions and at the same time adhere to a strictly alphabetical order. An obvious step toward realizing such a possibility would be to reduce the number of syntactic constructions that interrupt subdivision displays. As was noted, it was revealed by our data that for headings of average size (where by "average" is meant median), the most frequent interruption to the logical sequencing of subdivisions was caused by phrase headings. A characteristic of some phrase headings is that they can be expressed in the form of a main heading together with a subdivision, e.g., "Physics research" can be expressed as "Physics—Research." Introducing a guideline that would encourage the use, when possible, of subdivision rather than phrase syntax would be one method of reducing the number of interruptions in subdivision sequencing. (Eliminating the 151 phrases occurring in our sample had the result of more than doubling the number of headings in our sample that underwent no interruptions, from 14 to 29.) Lois Mai Chan writes that

Table 2. Interruptions to Subdivision Sequencing

	No.	(%)
Headings with no interruptions	13	(22%)
Total no. of interruptions	721	
Mean no. of interruptions (per heading)	11.27	
Median no. of interruptions (per heading)	2	
Interruptions caused by:		
Inversions	558	(77.39%)
Phrases	153	(21.22%)
Parentheticals	10	(1.39%)

"According to a current Library of Congress policy, unless the proposed phrase heading is very well known by the informed public in exactly that form, it is considered more useful to establish the proposed new concept as a topical subdivision under the generic heading."[9] A second method to reduce interruptions in subdivision sequencing would be to limit the conditions under which an inverted syntax can be used. It would appear that guidelines to achieve this have been introduced at the Library of Congress in recent years. As is often the case in index language design, however, there are tradeoffs. The danger of overly favoring a subdivision syntax is that the resultant index language, while possibly generating better displays, may at the same time be rendered more artificial and less user friendly.

REFERENCES AND NOTES

1. Mia Massicotte, "Improved Browsable Displays for Online Subject Access," *Information Technology and Libraries* 7:373–80 (Dec. 1988).
2. Stephen E. Wiberley, Jr., and Robert Allen Daugherty, "Users' Persistence in Scanning Lists of References," *College & Research Libraries* 49:149–56 (Mar. 1988).
3. Any extension of a selected heading by a subdivision, by a parenthetical qualifier, by an inverted qualifier, or by words completing a phrase we regarded as a modification of the heading. We then defined "blanket compression" as the elimination of those subject headings whose first modifying element repeats that of a preceding heading. Thus, the blanket compression of "Art, Modern," which in the ORION database occurs as the initial string of 773 subject headings, would result in the elimination of 772 headings.
4. American Library Association, Association for Library Collections and Technical Services, Subject Analysis Committee, Subcommittee on the Display of Subject Indexes in Online Public Access Catalogs, "Guide to the Display of Subject Headings in Subject Indexes in Online Public Access Catalogs" (draft report dated April 16, 1990, and prepared for discussion at the Annual Conference of the American Library Association, Chicago, June 23 and 24, 1990.)
5. John C. Rather and Susan C. Biebel, *Library of Congress Filing Rules* (Washington, D.C.: Library of Congress, 1980).
6. American Library Association, Resources and Technical Services Division, Filing Committee, *ALA Filing Rules* (Chicago: American Library Association, 1980).
7. Separately subfielding form subdivision is a recommendation discussed by the above-mentioned SAC Subcommittee at its June 1990 meetings in Chicago.
8. One of several proposals drafted by Lucia Rather (Library of Congress), Chair of the Planning Committee for the May 1991 Subject Subdivisions Conference. This proposal is mentioned in a memo dated 3 August 1990 from Martha Conway, Library of Congress, Office for Technical Processes Research, to Planning Committee, Subject Subdivisions Conference.
9. Lois Mai Chan, *Library of Congress Subject Headings: Principles and Application*, 2d ed. (Littleton, Colo.: Libraries Unlimited, 1986), p.49. ■■

Reader Use of a Nationwide Research Library Network: Local OPAC vs. Remote Files

Susan S. Lazinger and Bluma C. Peritz

The primary objective of the present study was to examine whether readers conducting bibliographic searches in ALEPH—Israel's research library network—tend to search only within the OPAC of the library within which they are working or whether they access the remote OPACs of other libraries. The ALEPH network has a decentralized database. Therefore, it was possible to examine this question because each library has its own access code and each database can be searched separately. The data were collected by means of a one-page questionnaire left beside each terminal in the library of the Graduate School of Library and Archive Studies of The Hebrew University of Jerusalem during an entire academic year. Results of analysis of the data collected in this survey are presented in six tables.

GOAL OF THE RESEARCH

Israel's research library network, ALEPH (Automated Library Expandable Program Hebrew University), is unique among the world's major bibliographic utilities in that it has a decentralized database. Each of its twenty-six files, representing all of the country's university libraries plus several special files (e.g., a Union List of Serials) is held and maintained separately, with both format and authority control for its own online public access catalog (OPAC) delegated to each individual library. Because each library has its own access code and each database can be searched separately, it was felt that ALEPH presented a special opportunity for examining how extensively readers in a specific library utilize the network. That is, it was decided to examine whether readers conducting bibliographic searches in a specific library—the library of the Graduate School of Library and Archive Studies of the Hebrew University of Jerusalem—tend to search only within the OPAC of the library within which they are working or whether they access the remote OPACs of other libraries.

PREVIOUS STUDIES

A survey of the literature indicates that use studies of online systems in libraries fall into three categories. First, there have been numerous use studies of individual library OPACs, particularly in the first half of the 1980s. Into the second category fall studies of search activity in commercial databases within libraries, while the third category is comprised of surveys of staff reactions to bibliographic network use in their libraries, i.e., to reactions of the intermediate users (library staff), rather than the end users (library patrons).

The first category of use studies of online systems—OPAC use studies—can also be divided into two sub-categories: studies of

---

Susan S. Lazinger is Head of the Library and part-time lecturer, and Bluma C. Peritz is Director, Graduate School of Library and Archive Studies, The Hebrew University of Jerusalem.

OPAC use (i.e., primarily whether users conducted known-item searches or subject searches) and studies of OPAC user satisfaction, with some of the studies covering both aspects. Among studies of the first type, Norden and Lawrence in 1981 utilized transaction logs to study use patterns in the OPAC at Ohio State University, finding a 35 percent higher use of title searching than studies of card catalogs had led them to expect.[1]

In an early study of multiple institutions, also published in 1981, Moore used a combination of observation and a questionnaire to study the use of OPACs in four university libraries. Results showed that most searches were for known items but that, nonetheless, a significant proportion of users attempted subject searches even when the systems they were using had poorly developed subject capabilities. Not surprisingly, Moore also found that known-item searches were more successful than subject searches.[2] In a study that combined statistics on OPAC use with observations on user reactions, Pawley at the University of Guelph, using a questionnaire, discovered that not only were patrons using the title access as a substitute for subject searching (which was not available), but also they seemed satisfied with their search results.[3] A series of studies of OPAC use was conducted under the auspices of the Council on Library Resources (CLR) in the early 1980s, the results of which were published in three reports in 1983.[4] Surveys were conducted at thirty libraries, with results showing that users did more subject searching than they had done using other forms of the catalog, but with less success than they had searching known items. The user characteristic most frequently associated with success in searching the OPAC was shown to be experience.

Some of the later surveys on OPAC use place more emphasis on user reactions to the system than on use patterns. Scharf and Ward in 1989 report on a survey conducted to assess user reactions to a second online public access catalog, which operated concurrently with an earlier system. Their object was to determine how users learned both systems, how satisfied they were with both systems, and to what extent they had difficulty in identifying and understanding the difference between the two systems.[5] In the same issue of *Library Trends*, Magrath reports on two surveys examining user reaction to another second generation online catalog, a Home Users Survey, mailed to 400 dial-up users (of which 128 responded), and a Public Access Catalog Survey, conducted by surveyors who approached every person using the terminals at random times of the day.

Typical of the second category of use studies—those which examine the benefits and disadvantages of online searching in commercial databases in the context of cooperative reference network services—are the studies of Byrd et al. (1979) and Roose (1982). Byrd summarized use statistics for the first year of search activity for the Kansas City Libraries Metropolitan Information Network (MINET), including databases searched, types of participating libraries, patron status, search systems, charges, nonpaid searches, and referrals.[6] Roose discussed the cost effectiveness of centralization of online information retrieval services, problems associated with doing searches at the network level, and subsidizing search services.[7]

User reactions to bibliographic networks, as opposed to single OPACs, make up the third category of use studies of online systems in libraries. Bibliographic network use has been examined primarily at the level of the intermediate user—the library staff. Logic dictates that an overview of network function can be obtained only at the level of the librarian, rather than the patron, since the patron concerns him/herself only with specific searches in a specific OPAC. Markuson in 1976 and Baker and Kluegel in 1982 presented reports of surveys conducted among OCLC users.[8] The first attempted to answer the questions, Why do libraries join networks? How have on-line operations affected costs, staffing, and workflow? and Do staff like or dislike the use of terminals for file access? The second reported the results of a telephone survey of academic libraries concerning the availability and accessibility of OCLC for use by reference personnel. Other studies of library staff satisfaction with bibliographic networks include a survey by the Midcontinental Region of the National Library of

Medicine's Regional Medical Library programme (MCRMLP) of 307 Octanet users to determine both user satisfaction and intensivity of usage,[9] and a recent survey carried out by the Special Libraries Association (SLA) Networking Committee to identify important networking issues and concerns, and to list problems they had had with networking.[10]

Having presented a sample of the literature on use of online systems, the authors would like to state that the current study falls into none of these categories nor, as far as we know, is there any other survey of patron use which explores the precise question examined here: does it really make a difference to a library user if the library's online system provides access to the files of other libraries, or is the user primarily interested in searching the catalog of the library in which he or she is working? That no other such study exists in the literature can probably be explained by the fact that, again to the best of the authors' knowledge, ALEPH is the only decentralized library network—i.e., in which each library's file is held and searched separately — in use today. Whatever the disadvantages of such a decentralized database, one of its advantages is that it enabled examination of the question of local versus remote file use.

POPULATION STUDIED

The library of the Graduate School of Library and Archive Studies is a small departmental library housing some 10,000 books and 100 journal titles. While it is intended primarily as a study and research library for faculty and approximately 120 students i n the two-year program for the Certificate in Library Science, and the three-year program (including thesis) for the M.L.S., it also serves as the chief library and information center in Library and Information Science for anyone doing research in these fields in Israel. The entire collection, including journals and masters' theses, has been converted to machine-readable records and is available in ALEPH.

STUDY OBJECTIVES

The objectives of the present study were to examine the following questions:

1. What categories of user (first year Library School student, second year Library School student, etc.) are more likely to search only in the local OPAC and what categories of users in remote OPACs.

2. Whether the amount of search experience a user has influences his/her tendency to search in remote OPACs rather than only in the local OPAC of the library in which he/she is working.

3. Whether the type of search—i.e., known-item vs. subject—influences a user's tendency to search in remote OPACs rather than only in the local OPAC.

4. Whether user satisfaction with the system's response time, with the ability to access remote files, and with the search results is influenced by whether the user is searching in the local OPAC only or in one or more remote files.

User satisfaction was studied by means of a ranking scale, from 1 = poor to 4 = Excellent, for three factors: (1) response time, (2) success in accessing other libraries in network (if applicable), and (3) search results (see appended questionnaire).

In consequence, the following variables were studied:

- whether the search was in the local OPAC only or in remote files as well;
- user category;
- user experience (0–5 previous searches, or more than 5);
- type of search (known item or subject);
- user satisfaction with response time (four categories);
- user satisfaction with access to other libraries (this variable was ascertained only among those searching in remote files);
- user satisfaction with search results.

It needs to be noted here that the ALEPH network is composed of files held in a number of VAX computers located on each of the campuses of Israel's seven universities and connected by dedicated telephone lines. Each library's OPAC sits in a computer resident on its own campus, and access to its own file and any others residing in the same computer is usually quick and dependable. On the other hand, access to files located in computers in other cities, or even on other campuses in the same city (Hebrew University, for example, has three campuses in Jerusalem) is not always as dependable, due to Israel's rather problematic telephone system. This is why it was decided to examine, as one of the variables, user satisfaction with the system's ability to allow *actual* access to files to which they had "theoretical" access.

METHODOLOGY

The data were collected by means of a one-page questionnaire, the English version of which is reproduced in the appendix A. There was also a Hebrew version of the same questionnaire. Questionnaires in both languages were left beside each of the library's two terminals, along with pens and a list of the library access codes (i.e., the codes needed to access each individual file). A large sign near each terminal explained that the questionnaires were part of a research project and requested that users of the ALEPH system fill out *one questionnaire per search session*, in either language. Since both user category and user experience were indicated on each questionnaire, it was felt that collecting results *per search* rather than *per user* was valid. The data were collected for an entire academic year—from November 1, 1989, to June 30, 1990. A total of 156 questionnaires were filled out and deposited in the box designated for this purpose. In sorting the questionnaires, it was decided that the one category that was essential for validating the questionnaire for use in the research was the indication of user category. Thus, if this category was not indicated on the questionnaire, it was eliminated from the sample. If other variables were missing from the answers on the questionnaires, they were simply not counted when that variable was being tallied, and the number of questionnaires for which an answer for that variable was missing was indicated below the appropriate table for the variable. On this basis, 6 of the 156 questionnaires were invalidated, leaving 150 usable questionnaires on which to base our analysis.

RESULTS AND DISCUSSION

The single consistent trend observed in every category of results was the overwhelming tendency of all types of users doing all types of searches to access remote files. In all the categories of user type and search type surveyed, only two categories revealed a higher percent of usage of the local OPAC only than of the files of other libraries as well. Table 1, which analyzes searches by user category, shows that 55% of the users in the third category—Library School Graduates researching their M.L.S. theses—searched only in the file of the Graduate Library School. This result conforms with the logic of the situation: these users are working on theses specifically in areas of material likely to be found only in the library of the library school. The number of users in this category is small, however, and the statistical significance of the search patterns of a group of eleven users may not be conclusive. The other categories of users show that between 59% and 82% conducted searches in the online catalogs of other libraries from the terminals of the library school. Of these, the lowest percentage of searches in remote files was found among the second-year library school students. This was a somewhat unexpected result, since again logic would dictate that the lowest percentage of outside searches, after the M.L.S. researchers, would be found among the first-year library school students, who use primarily reserved material concentrated on the shelves by course, while the second year students do more independent research. The highest percentage of remote file searching was conducted by users in the miscellaneous category termed "Other" (faculty, staff, out-of-town researchers, students from other disciplines) and conformed with the expectation that this varied group would be likely to search for material in multiple disciplines from multiple libraries.

Table 2 analyzes the searches by user experience. Here there is a clear trend toward remote file searching by experienced users, with inexperienced users tending to search only in the local OPAC. Thus, only 45% of users who had conducted 0–5 previous searches ventured into the remote files, while 76% of the more experienced users searched in remote library files. An interesting side trend demonstrated in the user experience category is the overwhelming majority of experienced over inexperienced users (84% experienced, 16% inexperienced), an indication of the sophistication

Table 1. Searches in Local OPAC only vs. Remote Files, by User Category

	1st-Year Library School Student		2nd-Year Library School Student		Library School Graduate Researching M.L.S. Thesis		Librarian Non-Student		Other		Total Searches by Local OPAC only vs. Remote Files	
	No. of searches	% of searches	No. of searches	% of searches	No. of searches	% of searches	No. of searches	% of searches	No. of searches	% of searches	No. of searches	% of searches
Local OPAC only	14	24	15	41	6	55	3	27	6	18	44	29
Remote files	44	76	22	59	5	45	8	73	27	82	106	71
Total searches by user category	58	100	37	100	11	100	11	100	33	100	150	100

of the library school's users (and perhaps of today's users in general) with regard to automated bibliographic systems.

Analyzing the data for table 3 proved the most complex of all the categories examined. Because it was feared that a choice of search types termed "Known item" or "Subject" would confuse the respondees, search types were broken down into seven types (see appendix A): (1) Author, (2) Title, (3) LC Subject Heading, (4) Series, (5) LC Classification Number, (6) Word,[11] (7) Boolean. It was realized afterward that this definition of categories does not allow for a 100% accurate division into Known-item search versus Subject search categories. Therefore, the following decisions were reached with regard to this division: (a) a search by author and/or title was considered a known-item search; (b) a search that included a search by subject as one of its parameters was considered a subject search; (c) a Boolean search combined with an author/title search was considered a known-item search, but a Boolean search alone or combined with a search by LC number, LC subject heading, or word (i.e., keyword) was considered a subject search. The results, summarized in table 3, show that users searching by subject are slightly more likely to search in remote files than users searching for a known item, but the difference is only 7%. What is more interesting—again a trend indirectly brought to light by the analysis required for table 3—is that our study, like virtually all the previous studies cited above, showed a distinct tendency among ALEPH's users to search by known item—73%. This is all the more surprising since ALEPH possesses considerable subject searching capabilities, permitting searches by LC subject heading, by LC call number, by keyword, and by a Boolean combination of any of the above.

Tables 4–6 analyze user satisfaction with three search factors: (a) system response time, (b) ease of access to other library files (where applicable), and (c) search results. User response, in each category, is graded as follows: 1 = Poor, 2 = Fair, 3 = Good, and 4 = Excellent. Table 4, which analyzes response time satisfaction, shows that users searching only in the local OPAC are considerably more positive in their scaling of sys-

*Table 2. Searches in Local OPAC only vs. Remote Files, by User Experience**

	Inexperienced (0–5 searches) No. of searches	% of searches	Experienced (more than 5 searches) No. of searches	% of searches	Total Searches
Local OPAC only	12	55	28	24	40
Remote files	10	45	90	76	100
Total searches by user experience	22	100	118	100	140

*User experience not indicated on questionnaire = 10
Inexperienced = 16% of users
Experienced = 84% of users

*Table 3. Searches in Local OPAC only vs. Remote Files, by Type of Search**

	Known-item Search No. of searches	% of searches	Subject Search No. of searches	% of searches	Total Searches
Local OPAC only	34	33	10	26	44
Remote files	68	67	28	74	96
Total searches by type of search	102	100	38	100	140

*Type of search not indicated on questionnaire = 10
Known-item searches = 73%
Subject searches = 27%

tem response time than users accessing the remote files. This clearly indicates that accessing remote files is a relatively time-consuming process which considerably lowers user response time satisfaction. Table 5, which explores user satisfaction with the ease of access to remote files (a related but not identical variable), shows an even less positive response from the segment of the user population that attempted to access the files of remote libraries. While the two extremes of the scale—Poor and Excellent—receive 26% and 31% respectively for response time, their respective rating is 33% and 28% for ease of access, the lowest of all the user satisfaction ratings.

Likewise, table 6 shows that user satisfaction with search results from searches in the local OPAC is significantly higher than with search results in remote files. Here, searchers in the local OPAC rated 17% of their results as poor and 45% of their results as excellent, while searchers in the remote files rated 24% of their results as poor and only 28% as excellent.

CONCLUSIONS

Results of the survey answer the questions posed by this study as follows:

1. All user categories except the small group of M.L.S. candidates researching their theses conduct more searches in the network's remote files than in the local OPAC of the library from which they are searching.

2. Experienced users conduct the overwhelming majority of their searches in remote files, while inexperienced users tend to search to a much greater extent in the local OPAC only.

3. Users searching by subject are slightly more likely to search in remote files than us-

*Table 4. Searches by Users' Satisfaction with Response Time; Local OPAC Only vs. Remote Files**

	Local OPAC Only		Remote Files		Total Searches by User Satisfaction	
	No. of searches	% of searches	No. of searches	% of searches	No. of searches	% of searches
1 = Poor	3	7	27	26	30	21
2 = Fair	3	7	15	15	18	12
3 = Good	14	32	28	28	42	29
4 = Excellent	24	54	31	31	55	38
Total Searches by local vs. remote	44	100	101	100	145	100

*Users' satisfaction with response time not indicated on questionnaire = 5

*Table 5. Searches by Users' Satisfaction with the Ease of Access to Other Libraries (Remote Files Only)**

	Local OPAC Only		Remote Files	
	No. of searches	% of searches	No. of searches	% of searches
1 = Poor	N/A	N/A	30	33
2 = Fair	N/A	N/A	14	16
3 = Good	N/A	N/A	21	23
4 = Excellent	N/A	N/A	25	28
Total searches by local vs. remote	N/A	N/A	90	100

*Users' satisfaction with ease of access not indicated on questionnaire = 16

*Table 6. Searches by Users' Satisfaction with Search Results; Local OPAC Only vs. Remote Files**

	Local OPAC Only		Remote Files		Total Searches by User Satisfaction	
	No. of searches	% of searches	No. of searches	% of searches	No. of searches	% of searches
1 = Poor	7	17	27	27	34	24
2 = Fair	4	9	16	16	20	14
3 = Good	12	29	36	36	48	34
4 = Excellent	19	45	21	21	40	28
Total searches by local vs. remote	42	100	100	100	142	100

*Users' satisfaction with search results not indicated on questionnaire = 8

ers searching for a known item, but the difference is only 7%, so this does not seem to be a highly significant factor.

4. User satisfaction with the system's response time, with the ability to access remote files, and with the search results is strongly influenced by whether the user is searching in the local OPAC only or in one or more remote files. Users accessing only the local OPAC registered higher levels of satisfaction in all three areas than users accessing the remote files.

The implications of the answers to the questions posed are wide-ranging and, in some cases, the serendipity factor—the results of statistics obtained indirectly in order to obtain statistics required for the answers to the proposed questions—supplied results with highly interesting and unexpected implications. The first implication is that users of an automated library system that permits access to remote library online catalogs, and is not simply an automated version of a specific library's card catalog, will utilize that access and search in the remote files. Thus, readers do seem to want an online system to be more than an automated card catalog, at least from the point of view of remote search capabilities. On the other hand, patrons of the Library School made relatively little use of the subject access that ALEPH permits and that is another trait that ideally distinguishes an online catalog from a manual one. In spite of the fact that the overwhelming majority of the respondees were experienced in the use of ALEPH, this experience did not foster creative search techniques. Essentially, while experience gave them the sophistication to take advantage of the network's remote search capabilities, it did not provide them with sophistication in using the system's powerful subject searching capabilities within each individual file. An explanation may be that accessing remote library files is a simple process to learn, requiring only one command (LB/ + the library's three-letter access code), whereas Boolean searches, for example, are a little more complex, and searching by LC subject heading or call number requires looking up the heading or number in printed guides. On the other hand, searching by keyword is extremely simple and may be underutilized simply because users have not yet become accustomed to using this search capability.

With regard to user satisfaction, the study clearly indicated that user desire to access remote files, at least in the ALEPH system, significantly exceeds user satisfaction with the results. The discrepancy between the two variables can probably be explained by technical problems specific to Israel and/or the ALEPH system—primarily to the relatively primitive state of Israeli telecommunications—and, therefore, the implications of the user satisfaction portions of the survey are essentially of local interest, as an indicator to the system's developers, ALEPH Yissum, that impressive software that is bogged down by faulty telecommunications causes user frustration.

REFERENCES AND NOTES

1. David J. Norden and Gail Herndon Lawrence, "Public Terminal Use in an Online Catalog: Some Preliminary Results," *College & Research Libraries* 42:308–316 (July 1981).
2. Carole Weiss Moore, "User Reactions to Online Catalogs: An Exploratory Study," *College & Research Libraries* 42:295–302 (July 1981).
3. Carolyn Pawley, "Online Access: User Reaction," *College & Research Libraries* 43:473–77 (Nov. 1982).
4. Joseph R. Matthews, Gary S. Lawrence, and Douglas K. Ferguson, *Using Online Catalogs: A National Survey* (New York: Neal Schuman, 1983); Neal K. Kaske and Nancy P. Sanders, *A Comprehensive Study of Online Public Access Catalogs: An Overview and Applications of Findings, Final Report to the Council on Library Resources,* v.3, Report no. OCLC/OPR/RR-83/4 (Dublin, Ohio: OCLC Office of Research, 1983); Karen Markey, *Online Catalog Use: Results of Surveys and Focus Group Interviews in Several Libraries, Final Report to the Council on Library Resources,* v.2, Report no. OCLC/OPR/RR-83/3 (Dublin, Ohio: OCLC Office of Research, 1983).
5. M. K. Scharf and J. A. Ward, "Side by Side: Users React to a Second Online Public Access Catalog," *Library Trends* 37:402–13 (Spring 1989).
6. Gary D. Byrd, et al., "Minet in K.C.: A Report on the Kansas City Libraries Metropolitan Online Bibliographic Network," *Library Journal* 104:2044–47 (Oct. 1, 1979).
7. Tina Roose, "Online Bibliographic Searching in Cooperative Reference Networks," *RQ* 21:325–29 (Summer 1982).
8. Barbara Evans Markuson, "The Ohio College Library Center System: A Study of Factors Affecting the Adaptation of Libraries to On-line Networks," *Library Technology Reports* 12:11–32 (Jan. 1976); Betsy Baker and

Kathleen Kluegel, "Availability and Use of OCLC for Reference in Large Academic Libraries," *RQ* 21:379–83 (Summer 1982).

9. Claire Gadzikowski, "Octanet—An Electronic Library Network; The First Four Years: Summary and Evaluation," *Information Technology and Libraries* 6:138–43 (June 1987).
10. Sharyn Ladner, SLA Networking Committee Survey. Final Report, Paper presented at the Annual Conference of the Special Libraries Association (80th, New York, June 10–15, 1989).
11. ALEPH automatically indexes every word, excluding a stop list, in any access file an individual library chooses when its file is set up. In the case of the Library School, the Word search option searches the Title and LC Subject Heading fields. ■■

APPENDIX A. ALEPH USER SURVEY

Date:
Time:

I. Please circle the number of appropriate response:

A. How much experience searching ALEPH do you have?
1. 0–5 searches
2. More than 5 searches

B. In which libraries did you search? (may circle 1 *and* 2)
1. Library School (JLB)
2. Other libraries (Please list Aleph codes):

a. ________ f. ________
b. ________ g. ________
c. ________ h. ________
d. ________ i. ________
e. ________ j. ________

C. What types of searches did you do? (may circle more than one choice)

1. Author
2. Title
3. LC Subject Heading
4. Series
5. LC Classification Number
6. Word
7. Boolean

D. What type of user are you?
1. 1st-year library school student
2. 2nd-year library school student
3. Library school graduate researching M.A. thesis
4. Librarian (non-student at library school)
5. Other (Please define): ________

II. Please respond on a scale of 1–4, as follows: 1 = poor, 2 = fair, 3 = good, 4 = excellent

A. How satisfied were you with the following factors?
1. Response time: ________
2. Success in accessing other libraries in network (if applicable): ________
3. Search results: ________

Controlled and Uncontrolled Vocabulary Subject Searching in an Academic Library Online Catalog

Thomas A. Peters and Martin Kurth

An analysis of transaction logs from an academic library online catalog describes instances in which users have tried both controlled and uncontrolled (title keyword) vocabulary subject access during the same search session. Eight hypotheses were tested. Over 6.6% of all dial access search sessions contained both methods of subject access. Over 58% of the isolated sessions began with an uncontrolled vocabulary attempt. Over 76% contained only one vocabulary shift. On average, user persistence was greater during controlled vocabulary search legs, but search output was greater during uncontrolled vocabulary search legs. Several recommendations regarding catalog design and instruction are made.

A debate continues in the library and information science professions over the relative merit, appropriateness, and efficiency of uncontrolled vocabulary subject access in online catalog systems. Much of the discussion tends to be prescriptive rather than descriptive. Often the situation is presented as an either/or decision, both for system designers and users of systems. The present study does not focus on uncontrolled and controlled vocabulary subject access as a design issue, but rather as a use issue. This study offers descriptive data related to patron use of both controlled and uncontrolled vocabulary subject access. An analysis of seven months of dial access transaction logs of the online catalog at the University of Missouri–Kansas City describes the instances in which users have tried both types of subject access during the same search session.

Methods of uncontrolled vocabulary subject access vary among online catalog systems. Some systems offer keyword access to controlled vocabulary subject headings. Other systems offer a keyword search of a combined index culled from subject and title fields. Keyword access to these fields seems to be popular among users. For example, the study of the SULIRS system at Syracuse by Martin, Wyman, and Madhok found that the combined title/subject keyword search accounted for over 35% of all searches, the highest for any search option.[1] In the system studied here, only titles, series, and corporate authors can be searched by keywords. Not surprisingly, however, logs of search transactions have consistently shown that patrons have used title keyword searches to provide uncontrolled vocabulary subject access to bibliographic records. Before we can begin to speculate with confidence about what causes users to shift from one form of subject access to another, the situational characteristics surrounding the combined use of controlled and uncontrolled vocabulary subject access need to be examined. In the present study, search sessions containing controlled and uncontrolled vocabulary subject searching have been identified and isolated in order to describe the user persistence and search out-

Thomas A. Peters is Collection Development Librarian, Mankato State University, Mankato, Minnesota, and Martin Kurth is Head Serials Librarian, Eastern Washington University, Cheney.

put, in terms of the number of bibliographic records retrieved, associated with each type of subject search.

DEFINITIONS AND SYSTEM CHARACTERISTICS

The dial access online catalog activity at the Kansas City campus of the University of Missouri provided the raw data for this study. In the catalog, the following menu-mode search options are presented to public users: one title keyword search, two author searches, three controlled vocabulary subject searches, four series searches, ten Boolean search options, and a call number search.[2] The main menu screen is reproduced in figure 1.

Dial access search sessions were selected for this study because, in dial access transaction logs, it was possible to determine when users logged on and logged off. Determining when in-house public access search sessions began and ended would have been more difficult, if not impossible. In this study a dial access session from logon to logoff was called a *search session.* It can be argued that dial access sessions and search sessions do not always coincide. For example, sometimes several subject search sessions occur during the same dial access session. Conversely, a single subject search session can extend over several dial access sessions, perhaps separated by days or weeks. Both types of discrepancies between dial access session boundaries and subject session boundaries were found in the transaction logs examined. In general, we sought to achieve a balance between the desire to develop definitions requiring little interpretation and few judgment calls, and the need to reveal the search as experienced by the searcher.

In order to describe the elements of individual search transactions easily, several session characteristics needed to be isolated and defined. For the purposes of this study, a *subject search session* was defined as a search session containing at least one *controlled vocabulary subject search statement,* that is, a subject browse, medical subject browse, official subject term search, or Boolean search involving two controlled vocabulary subject components. Call number browse searches were not considered to be controlled vocabulary subject search statements because, although they can provide subject access to the database, they do not utilize a controlled vocabulary.

Two basic tests were used to isolate title keyword searches that evidently were used as *uncontrolled vocabulary subject search statements.* If the terms entered as a title keyword search statement were similar or identical to the terms entered in a controlled vocabulary subject search statement during the same search session, or if the

LUMIN MASTER SEARCH MENU

TITLE SEARCH:
t = title keyword search
AUTHOR SEARCHES:
a = author search (personal and corporate/governmental bodies)
k = corporate/governmental bodies keyword term search
SUBJECT SEARCHES:
s = subject browse search (use if unsure of heading)
m = medical subject browse search (use if unsure of NLM heading)
o = official subject term search (use if sure of heading)
OTHER SEARCHES:
w = series searches
b = boolean searches
c = call number browse search
q = quick-facts about the library and LUMIN

TYPE THE LETTER FOR YOUR SEARCH CHOICE & PRESS ENTER:

Figure 1. LUMIN Master Search Menu Screen.

terms entered in a sequence of title keyword search statements contained synonyms, the title keyword search statements were assumed to be used for uncontrolled vocabulary subject access. Because only search sessions containing at least one controlled vocabulary subject search statement were analyzed, not all attempts at uncontrolled vocabulary subject access were found.

Data-entry search statements were defined as those search statements where the user actually entered information brought to a search, such as whole or partial titles, author names, subjects, call numbers, series titles, and the like. The other types of search statements in the system studied were *option-selection search statements*, in which the user merely selected a number or letter option listed on the terminal screen. Most search sessions contained a combination of data-entry and option-selection search statements.

A *unique, data-entry search statement* was defined as a data-entry search statement preceded and succeeded by different data-entry search statements. Sometimes users repeated the same data-entry search statement without changing the keyed characters. Such a sequence of identical data-entry search statements was counted as only one unique, data-entry search statement. This study counted and examined the unique, data-entry search statements within isolated dial access search sessions.

A *vocabulary shift* was that point in a search session when the user switched from controlled vocabulary subject access to uncontrolled, or vice versa. A vocabulary shift was not necessarily an instance in which the user changed the search term he or she used.

A *search leg* was defined as a portion of a subject search session that began and ended with a vocabulary shift, a logon, or a logoff. With these boundaries established, a *controlled vocabulary search leg* contained at least one controlled vocabulary subject search statement. An *uncontrolled vocabulary search leg* contained at least one uncontrolled vocabulary subject search statement. Additionally, the search legs were labelled by their position within the total subject search session as first, second, or third and subsequent legs. The search sessions selected for further study contained at least one controlled vocabulary search leg and one uncontrolled vocabulary search leg. To make referring to them easier, these sessions were called *mixed access subject search sessions*.

Assuming that most users wanted to retrieve a satisfactory number of pertinent bibliographic records from the online catalog, we identified two forms of *user persistence* for quantification and study. On the one hand, persistence was defined as the number of searches for bibliographic records initiated by the user. On the other hand, the number of search legs per search session was seen as another measure of user persistence.

This study defined *search output* as the number of bibliographic records retrieved, but not necessarily displayed, during each search leg and the search session as a whole. As defined, search output does not measure the relevance of the retrieved set of bibliographic records to the subject pursued. Relevance of a retrieved set can be subjectively assessed by replicating a given search and examining the records retrieved, assuming that the database has not changed significantly since the original search. Although such search replications provide a valuable microcosmic view of patron searching, they were not attempted in this study.

In spite of the fact that public users choose from among the menu-mode search options listed above, from the vantage point of the system, only five system functions are used by the public: TERM, BROWSE, FIND, SELECT, and DISPLAY LOCAL HOLDINGS.[3] These functions, as translated by the system from the menu-mode search options, search the databases and indexes within the online catalog system. In the transaction logs studied, all menu-mode activity was represented in terms of these system functions.

The TERM and BROWSE functions provide access to the authority files in the database. The TERM function retrieves the set of authority records containing the search phrase entered. Unless right-hand truncation is used, the entire phrase must match an entire utilized main heading or subheading. The TERM search is not a keyword search. In the example below, the transaction log indicates that at 1:50 p.m. on February 28, 1990, the user retrieved from the authority file 16 authority headings which contained the search statement entered. He or she then performed a FIND search on 15 of the utilized subject headings, retrieving 62 bibliographic records. As defined above, these statements constitute a controlled vocabulary search leg.

16	0228	135039XV04	TERM SUBJECT employee fringe benefits—united states
62	0228	135242XV04	FIND 1–15

Public users of the online catalog studied can BROWSE alphabetized lists of utilized LCSH headings, MeSH headings, series titles, and call numbers. As they appear in figure 1, the s, m, w, and c searches are all browse searches. In the following example, the user browsed through the list of utilized subject headings beginning with "employee benefits." The transaction log shows that 17 headings were retrieved because 17 headings are the maximum number of headings that can be displayed on a screen. The user then followed with a FIND search on the first heading, retrieving one bibliographic record. The following is another type of controlled vocabulary search leg.

17	0228	134437XV04	BROWSE SUBJECT employee benefits
1	0228	134517XV04	FIND 1

The SELECT function is used to display portions of retrieved sets. If a FIND search or a TERM search results in a large set (over 30 or 50 records, depending on whether bibliographic or authority records are being retrieved), the SELECT command can be used to display portions of the retrieved set. In this sense and in this situation the SELECT command can be used to browse through a set of retrieved records.

Only the FIND function retrieves bibliographic records from the database. While pursuing any of the menu-mode search options, users can initiate FIND searches of the bibliographic database. Controlled vocabulary searching and title keyword searching, however, utilize the FIND function differently. The FIND function during a title keyword search always is a data-entry search, in which the user enters data brought to the search. The following is an example of an uncontrolled vocabulary search leg in which the user performed a title keyword search that retrieved a set of 34 bibliographic records. The user then displayed 30 of the records and selected the tenth record for full display.

34	0228	134659XV04	FIND TITLE employee benefits
30	0228	134724XV04	SELECT 1–30
1	0228	134922XV04	SELECT 10 $,F

When the user enters title keywords, the system provides direct access to the bibliographic records in the database. On the other hand, during controlled vocabulary subject searching the FIND search is used after the user has gained access to the subject authority file. A FIND search during a controlled vocabulary subject search involves selecting an alphanumeric option rather than entering search terms. As in the example of the TERM search above, the search terms have been entered at an earlier stage of the search.

A typical search session contains several system functions. The examples above all came from the same search session. The session consisted of three search legs and two vocabulary shifts, thus providing an example of the mixed access subject search sessions selected for further study.

17	0228	134437XV04	BROWSE SUBJECT employee benefits
1	0228	134517XV04	FIND 1
34	0228	134659XV04	FIND TITLE employee benefits
30	0228	134724XV04	SELECT 1–30
1	0228	134922XV04	SELECT 10 $,F
16	0228	135039XV04	TERM SUBJECT employee fringe benefits—united states
62	0228	135242XV04	FIND 1–15
0	0228	135343XV04	LOGOFF

HYPOTHESES

Eight hypotheses were tested.

1. In search sessions that include controlled vocabulary subject search statements and title keyword search statements, some of the title keyword searches will be used to provide uncontrolled vocabulary subject access.

Confirming the first hypothesis was necessary before any other hypotheses could be tested. To confirm or negate this hypothesis, we isolated for further analysis the subject search sessions containing both controlled vocabulary subject search statements and title keyword search statements.

Based on the criteria previously described, title keyword search statements were examined to determine whether they were uncontrolled vocabulary subject search statements. After the existence of uncontrolled vocabulary subject searching activity had been confirmed, the type of subject access occurring first and the average number of legs in the mixed access subject search sessions were then measured. Hypotheses 2 and 3 address these conditions.

2. Most subject search sessions containing both controlled and uncontrolled vocabulary subject search statements will begin with a controlled vocabulary subject search attempt.

3. Most subject search sessions containing both controlled and uncontrolled vocabulary subject search statements will contain only two legs (i.e., only one vocabulary shift).

The legs of the mixed access subject search sessions were measured to determine their search output. Search legs were compared according to their sequential position within the search session as a whole (i.e., first leg, second leg, etc.) and according to their vocabulary type (i.e., controlled vocabulary legs vs. uncontrolled vocabulary legs). The persistence of users, in terms of the number of FIND searches attempted during the legs, was quantified. The next four hypotheses state the anticipated conditions.

4. Persistence, in terms of the number of FIND searches initiated, will increase on average as the search session progresses from the first leg to the second leg and on to any subsequent legs.

5. Persistence will not vary on average between controlled and uncontrolled vocabulary search legs.

6. Search output, as measured by the number of bibliographic records retrieved, will increase on average as the search session progresses from the first leg to the second leg and on to any subsequent legs.

7. Search output will not vary on average between controlled and uncontrolled vocabulary search legs.

For the final hypothesis, the set of mixed access subject search sessions was compared to the set of subject search sessions that did not exhibit evidence of uncontrolled vocabulary subject searching. During controlled vocabulary subject searches, the ratio of unique, data-entry searches to subsequent FIND searches may be seen to reflect the relevance of the subject headings retrieved to the topic searched.[4] If a user requests that the bibliographic records attached to a displayed subject heading be retrieved, then it is likely that that heading has some relevance to the topic of the user's search. If a user's dissatisfaction with the controlled vocabulary subject headings displayed is a contributing cause for switching to uncontrolled vocabulary subject access, then the ratio of unique, data-entry searches to subsequent FIND searches during controlled vocabulary first legs of mixed access subject search sessions should be closer to 1:1 than the same ratio during purely controlled vocabulary subject search sessions, because the user's dissatisfaction with the headings displayed should result in fewer FIND search attempts. The last hypothesis to be tested posits the existence of this situation.

8. The ratio of unique, data-entry searches to FIND searches during subject search sessions not exhibiting evidence of uncontrolled vocabulary subject searching will differ from the same ratio during controlled vocabulary first search legs of search sessions also containing uncontrolled vocabulary subject searching.

MACROANALYSIS: THE PHENOMENA IN CONTEXT

The macroanalysis of the mixed access subject search sessions tried to place them in a larger context of all dial access search sessions during the months of the study. The study did not attempt to identify all use of the title keyword search for uncontrolled vocabulary subject access. Only the search sessions containing at least one controlled vocabulary subject search were selected for further analysis. This methodology, therefore, cannot provide data on the total use of title keyword search statements for subject access, even among only dial access users.

The results of the macroanalysis appear in table 1, in which the sessions labelled T & CV Subject Sessions are those subject search sessions that contained both title keyword and controlled vocabulary subject search statements. The percentages shown compare each column to the column at its left.

In the macroanalysis, the calculation of the total number of search sessions, the number of subject search sessions, and the number of subject search sessions containing at least one title keyword search statement required no interpretation by the analyzers. Determining the use of the title keyword search as an uncontrolled vocabulary subject search, however, often required speculation and judgment calls. For example, there was some evidence, particularly in the May 1989 dial access transaction logs, that users sometimes combined title keyword and controlled vocabulary subject searching to try to locate known or partially known items. This situation was the polar opposite of the situation under study. Occasionally it was very difficult to determine if the user was searching for a known item or items on a topic.

Approximately 22 percent of all dial access search sessions during the seven months studied contained at least one controlled vocabulary subject search statement. In comparison, Borgman found that approximately one-third of all search sessions contain at least one subject search statement.[5] The percentage of subject search sessions found in the present study may be lower than that found in other online catalog studies because several unaffiliated public, college, and special libraries in the study area use dial access only to verify holdings for known items prior to initiating interlibrary loan requests, so subject access is rarely, if ever, used by them.

Slightly less than half of the subject search sessions examined also contained at least one title keyword search statement. Sixty-three percent of those sessions manifested the use of title keyword search statements for uncontrolled vocabulary subject access. This set of 131 mixed access subject search sessions accounted for over 6.6 percent of all dial access sessions during the seven months of the study.

MICROANALYSIS: INTERNAL CHARACTERISTICS OF THE PHENOMENA

The microanalysis stage of the study examined the internal characteristics and dynamics of the mixed access subject search sessions. In table 2 the mixed access sessions are categorized into four types based on the method of subject access used first during a search session and the number of vocabulary shifts during the session. TYPE A search sessions began with an uncontrolled vocabulary search leg and contained only one vocabulary shift to a controlled vocabulary subject search leg. TYPE B search sessions began with a controlled vocabulary subject search leg and contained only one vocabulary shift to a uncontrolled vocabulary search leg. TYPE C search sessions began with an uncontrolled vocabulary search leg and contained more than one vocabulary shift between uncontrolled and controlled vocabulary search legs. TYPE D search sessions began with a controlled vocabulary search leg and contained more than one vocabulary shift between controlled vocabulary and uncontrolled vocabulary search legs.

The next phase of the microanalysis involved measuring the persistence, in terms of the number of FIND searches initiated, of every search leg in the mixed access subject search sessions. The search output, in terms of the number of bibliographic records retrieved, also was measured. The results appear in tables 3–5. The transaction logs for February and March 1990, which contained data on the number of bibliographic and authority records retrieved, were used for this phase. In order to study the persistence and search output of the search sessions, only the portions of the log directly related to subject access were analyzed. Portions of search sessions involving other types of search statements (e.g., author, call number) were not included in this phase of the analysis.

RESULTS

The data compiled in tables 1–5 enabled us to confirm or refute the eight hypotheses stated above. Hypothesis 1 was confirmed. Table 1 indicates that at least 6.6 percent of

Table 1. Isolating the Mixed Access Subject Search Sessions

Month	Total Search Sessions	Subject Search Sessions	%	T & CV Subject Sessions	%	Mixed Access Sessions	%
11/88	276	43	15.6	19	44.2	13	68.4
4/89	253	60	23.7	21	35.0	18	85.7
5/89	240	58	24.2	31	53.4	23	74.2
6/89	245	48	19.6	28	58.3	15	53.6
7/89	206	30	14.6	16	53.3	14	87.5
2/90	410	124	30.2	52	41.9	23	44.2
3/90	342	72	21.1	41	56.9	25	61.0
Total	1,972	435	22.1	208	47.8	131	63.0

Table 2. Characteristics of the Mixed Access Subject Search Sessions

Month	Mixed Access Sessions	Type A	Type B	Type C	Type D
11/88	13	7	4	0	2
4/89	18	12	2	2	2
5/89	23	15	3	0	5
6/89	15	6	5	1	3
7/89	14	8	4	0	2
2/90	23	9	5	5	4
3/90	25	10	10	2	3
Total	131	67	33	10	21
Percent	100	51.1	25.2	7.6	16.0

Table 3. Average Persistence and Search Output of First Search Legs in Mixed Access Subject Search Sessions

Month	First Legs: Total FIND Attempts	First Legs: No. of Bib. Records Retrieved	CV as First Leg: Total FIND Attempts	CV as First Leg: No. of Bib. Records Retrieved	UCV as First Leg: Total FIND Attempts	UCV as First Leg: No. of Bib. Records Retrieved
2/90	45 (1.96) n = 23	1,615 (70.22) n = 23	19 (2.11) n = 9	12 (1.33) n = 9	26 (1.86) n = 14	1,603 (114.50) n = 14
3/90	57 (2.28) n = 25	1,492 (59.68) n = 25	35 (2.69) n = 13	134 (10.31) n = 13	22 (1.83) n = 12	1,358 (113.17) n = 12
Total	102 (2.13) n = 48	3,107 (64.73) n = 48	54 (2.45) n = 22	146 (6.64) n = 22	48 (1.85) n = 26	2,961 (113.88) n = 26

n = the number of search legs in the set. Averages per leg are given in parentheses.

all dial access search sessions during the seven months of the study contained at least one title keyword search functioning to provide uncontrolled vocabulary subject access. This percentage probably would increase if search sessions with no controlled vocabulary subject search statements were examined.

Hypothesis 2 was refuted. As shown in table 2, almost 59 percent of the mixed access subject search sessions began with an uncontrolled vocabulary search leg. Hypothesis 3, however, was confirmed. Table 2 indicates that over 76 percent of the mixed access sessions contained only two search legs, thus only one vocabulary shift. Of those sessions that contained more than one vocabulary shift, however, more than two-thirds (21 of 31) began with a controlled vocabulary subject search leg.

Taken together, tables 3, 4, and 5 address hypotheses 4, 5, 6, and 7. Based on the

Table 4. Average Persistence and Search Output of Second Search Legs in Mixed Access Subject Search Sessions

	Second Legs		CV as Second Leg		UCV as Second Leg	
Month	Total FIND Attempts	No. of Bib. Records Retrieved	Total FIND Attempts	No. of Bib. Records Retrieved	Total FIND Attempts	No. of Bib. Records Retrieved
2/90	52 (2.26) n = 23	1,132 (49.22) n = 23	43 (3.07) n = 14	226 (16.14) n = 14	12 (1.33) n = 9	906 (100.66) n = 9
3/90	48 (1.92) n = 25	363 (14.52) n = 25	21 (1.75) n = 12	46 (3.83) n = 12	27 (2.08) n = 13	317 (24.38) n = 13
Total	100 (2.08) n = 48	1,495 (31.14) n = 48	64 (2.46) n = 26	272 (10.46) n = 26	39 (1.77) n = 22	1,233 (55.59) n = 22

n = the number of search legs in the set. Averages per leg are given in parentheses.

Table 5. Average Persistence and Search Output of Third and Subsequent Search Legs in Mixed Access Subject Search Sessions

	Third and Subsequent Legs		CV as Third and Subsequent Legs		UCV as Third and Subsequent Legs	
Month	Total FIND Attempts	No. of Bib. Records Retrieved	Total FIND Attempts	No. of Bib. Records Retrieved	Total FIND Attempts	No. of Bib. Records Retrieved
2/90	44 (4.89) n = 9	1,747 (194.11) n = 9	29 (7.25) n = 4	258 (64.50) n = 4	15 (3.00) n = 5	1,489 (297.80) n = 5
3/90	15 (1.88) n = 8	641 (80.13) n = 8	9 (1.80) n = 5	420 (84.00) n = 5	6 (2.00) n = 3	221 (73.67) n = 3
Total	59 (3.47) n = 17	2,388 (140.47) n = 17	38 (4.22) n = 9	678 (75.33) n = 9	21 (2.63) n = 8	1,710 (213.75) n = 8

n = the number of search legs in the set. Averages per leg are given in parentheses.

evidence gathered, hypothesis 4 was neither confirmed nor refuted. Overall, there was some indication that persistence, in terms of the number of FIND searches attempted, increased, especially if the search session lasted into a third search leg. The data for February 1990 strongly supported the hypothesis, but the data for March 1990 tended to refute it insofar as this measure of persistence decreased slightly. Hypothesis 5 was refuted. Persistence seemed to be greater during controlled vocabulary legs than during uncontrolled vocabulary legs. Overall, controlled vocabulary legs averaged 2.74 FIND searches, while uncontrolled vocabulary legs averaged 1.93 FIND searches. Concerning hypothesis 6, search output averaged less during the second leg than during the first (which tended to refute the hypothesis), but third and subsequent legs resulted in the highest output (which tended to confirm the hypothesis). Hypothesis 7 was refuted. Controlled vocabulary legs retrieved an average of 19.23 bibliographic records, while uncontrolled vocabulary legs retrieved an average of 105.25 bibliographic records.

The March 1990 subject search sessions were used to test hypothesis 8. The 47 subject search sessions not exhibiting evidence of uncontrolled vocabulary subject searching produced a ratio of unique, data-entry controlled vocabulary search statements to subsequent FIND search statements of 1:1.39. The sessions averaged 2.43 unique, data-entry searches and 3.36 FIND searches. In comparison, the controlled vocabulary first legs of the 13 mixed access subject search sessions in March 1990 had a ratio of 1:1.40, with an average of 1.92 unique, data-entry searches and an average of 2.69 FIND searches. Although the average number of unique, data-entry controlled vocabulary subject search statements and subsequent FIND search statements was less in controlled vocabulary first legs of mixed access subject search sessions, the ratio was almost the same as in subject sessions not exhibiting evidence of uncontrolled vocabulary subject searching, tending to refute hypothesis 8.

DISCUSSION

Without replicating any of the search sessions studied, we were unable to relate the numeric totals and averages presented in the tables to the relative success of the sessions as a whole or the search legs contained in them. It was nevertheless clear from the data that, in the system studied, patrons regularly used the title keyword search for uncontrolled vocabulary subject access. Using transaction logs to describe the contextual circumstances of that use was as problematic as determining its success.

Counting the number of FIND searches in the mixed access subject search sessions showed that requests for the display of bibliographic records did not really increase as the searches progressed. How this outcome relates to the success of those searches is open to question. On average, controlled vocabulary search legs resulted in slightly more FIND searches than uncontrolled vocabulary legs. This is perhaps due to the architecture of the system; FIND searches in uncontrolled vocabulary legs are dependent on the entry of search terms for each FIND search, whereas in controlled vocabulary legs, FIND searches are alphanumeric option selections.

More shifts between vocabulary types tended to result in more records retrieved. Most users, however, shifted between controlled and uncontrolled vocabulary subject access only once. This was true in spite of the fact that the searches entered after the first vocabulary shift tended to retrieve fewer records than those entered previously. Fewer records retrieved during second legs did not often prompt additional vocabulary shifts.

Uncontrolled vocabulary search legs retrieved more records than controlled vocabulary search legs. Whether the greater search output associated with uncontrolled vocabulary search legs also resulted in greater numbers of relevant records retrieved requires further study.

It was impossible, from the logs, to relate the shifts from controlled to uncontrolled vocabulary subject access to the search characteristics and results that were present prior to the shift. Further analysis is again required.

The specific architecture of the system seemed to influence the subject searching behavior of public users, making it difficult or impossible to generalize about subject searching behavior. In fact, although users' desires to find information on a topic may exist outside of information systems, subject searching behavior probably does not, since system architectures seem to influence subject searching behavior. The quest to discover abstract subject searching behavior may be futile, because it may not exist.

One shortcoming of this study was the inability to compare the search output of a title keyword search being used for uncontrolled vocabulary subject access to the search output of a controlled vocabulary keyword subject search. The system studied did not offer a true keyword search of the controlled vocabulary subject fields. The official subject term search treats each main heading and subheading of the assigned subject headings as key phrases. The subject browse and medical subject browse searches drop the user into alphabetized lists of assigned subject headings at the next closest match to the word or phrase entered.

INSTRUCTIONAL RECOMMENDATIONS

The results of this study indicate that online catalog searchers are using title keyword searches for subject access. To maximize the benefits of that use, instructional activities should encourage creative, expansive approaches to subject access that include the use of title keyword searches. Instructional efforts should stress the relationships among the subject headings displayed in online catalog bibliographic records, those present in online authority files, and those contained in printed subject heading lists. Instruction librarians also should teach effective scoping and truncation practices that can facilitate the subject access uses of title keyword searching.

Specific techniques for using title keyword searches to provide subject access can be demonstrated to online catalog users. There are at least four possible methods in which the title keyword search can be used to identify and locate items on a topic. The first method assumes that the user comes to the search session with at least one known item. The user can use the title keyword search to retrieve the bibliographic record for the known item, note the assigned subject headings, and execute a controlled vocabulary subject search on the most promising assigned heading to identify and locate other items that have been assigned the same heading.

Users can also use a known item and the title keyword search to achieve a simultaneous synthesis between known item and subject searching by carefully choosing the keywords from the title of the known item that are likely to retrieve other items of interest. For example, if the known item being sought is the book *Online Catalogs, Online Reference: Converging Trends*, the user could choose to search on the three keywords "online, catalogs, reference" as a tactic for locating not only the known item, but also perhaps other items about the relationship between online catalogs and reference service. Title keywords that are not essential facets of the topic, such as "converging" and "trends" in this example, would not be entered.

The third and fourth uses of title keyword searches in subject retrieval assume that no known items are brought to the search session. In the third method, the user can try the title keyword search as an uncontrolled vocabulary subject search simply to retrieve at least one potentially pertinent item. If such an item is retrieved, the assigned subject headings can be examined. Controlled vocabulary subject searches can then be used to increase the recall of pertinent items.

In the fourth method, title keyword searches can be used for subject access in lieu of the controlled vocabulary subject searching. As Bates has proposed, if the user wants to retrieve only a few relevant items, exclusive use of uncontrolled vocabulary subject access may be an efficient use of the system.[6]

In order to help users better visualize the ways that title keyword searching can be used with controlled vocabulary searching, instruction librarians may want to enable them to view the transaction logs of their search sessions, as Simpson has suggested.[7] This could be done during or after search sessions to encourage users to conceive of their search sessions as more than sequences of unrelated search statements.

SYSTEM DESIGN RECOMMENDATIONS

The present study reflects the persistence of online catalog users in terms of their willingness to view lists of subject terms and then retrieve the bibliographic records attached to them. The same users bring their own subject terms to the system in order to use title keyword searches for subject access. Recent discussions in the library literature have described ways to help users match their own subject vocabularies to the vocabularies used in online catalog systems. Markey has proposed a design for online catalog searching software that links users' search terms to an online catalog's subject authority file, to the machine-readable Library of Congress Subject Headings, and to selected subject-rich fields in bibliographic records, in order to increase the likelihood that users will locate available relevant library materials.[8] Also seeking a better way to match users' search terms with relevant bibliographic records, Bates has proposed searching software that connects a "superthesaurus" of search terms to subject headings in online authority files.[9] Either of these two systems would improve the user's subject searching chances.

Once an online catalog user retrieves a relevant bibliographic record, the catalog system should provide easy methods for retrieving more records for items which cover the same topics. For example, an online catalog system should lead the searcher from successful title keyword searches to the corresponding subject headings or class numbers assigned to a broader range of related materials.[10] The bridge function available in many online catalog systems, which allows users to jump from retrieved bibliographic records to other records containing the same subject headings, would be useful in the context of the types of searches studied here.[11]

CONCLUSION

The results of this study indicate that using title keyword searches for uncontrolled vocabulary subject access is not primarily an option of last resort. Users of the online catalog studied combined uncontrolled vocabulary subject searches with controlled vocabulary subject searches during the same search session.

Actual use of an automated library system is a synthesis of what users want to do and what the system allows or encourages them to do. Although the system studied does not explicitly encourage the use of title keywords for uncontrolled vocabulary subject access, evidence in the transaction logs indicates user need or demand for such activity. If the debate over controlled versus uncontrolled vocabulary subject access is to be grounded in actual user behavior, studies examining the ways that users test the limits of system functionality can improve future online catalog design and instruction.

REFERENCES AND NOTES

1. Thomas H. Martin, John C. Wyman, and Kumud Madhok, *Feedback and Exploratory Mechanisms for Assisting Library Staff Improve On-Line Catalog Searching* (Syracuse, N.Y.: Syracuse University, ERIC Document ED 258 577, 1983), p.4.
2. The catalog is a modified version of software originally purchased from what was then the Washington Library Network (WLN).
3. The DISPLAY LOCAL HOLDINGS function is used by public users to retrieve circulation information about library materials.
4. The ratio between unique, data-entry searches and FIND searches during title keyword searches is always 1:1, but the ratio between the two measures during controlled vocabulary subject searches can vary. For example, a unique data entry subject BROWSE can be followed by zero, five, or five hundred FIND searches.
5. Christine L. Borgman, *End User Behavior on the Ohio State University Libraries' Online Catalog: A Computer Monitoring Study* (Dublin, Ohio: OCLC Office of Research, ERIC Document ED 265 888, 1983), p.21.
6. Marcia J. Bates, "Rethinking Subject Cataloging in the Online Environment," *Library Resources & Technical Services* 33:403 (Oct. 1989).
7. Charles W. Simpson, "OPAC Transaction Log Analysis: The First Decade," in *Advances in Library Automation and Networking*, ed. Joe A. Hewitt (Greenwich, Conn.: JAI Press, 1989), v.3, p.62.
8. Karen Markey, "Integrating the Machine-Readable LCSH into Online Catalogs," *Information Technology & Libraries* 7:303 (Sept. 1988).
9. Bates, "Rethinking Subject Cataloging," p. 409–11.
10. Charles R. Hildreth, "General Introduction; OPAC Research: Laying the Groundwork for Future OPAC Design," in *The Online Catalogue: Developments and Directions*, ed. Charles R. Hildreth (London: The Library Association, 1989), p.18.
11. The software of the system studied includes a bridge-type HEADINGS command, but at the time of the study it had not been made available to menu-mode users. ■■

The Development of the National Research and Education Network

Roberta A. Corbin

The development of a new technology occurs in several phases. First is the replacement of traditional manual functions with automated ones. Next, people see the potential of using the technology, and new uses and ways of doing things are devised. Finally, society itself changes as a result of that technology. These societal changes are occurring with the development of networks on a local, regional, national, and international scale. The purpose of this paper is to trace the development of national networks, describe their current condition, and discuss the future implications of and obstacles to the accomplishment of the vision.

In 1962, the library at the University of California–San Diego claimed that a new serials management system was the first use of a computer in a university library operation.[1]

Early local systems like the one in San Diego, along with the installation of computers at both the National Library of Medicine (1963) and the Library of Congress (1964), and developments such as the MARC (MAchine-Readable Cataloging) record standard (1966), were the beginnings of revolutionary changes in libraries. The momentum continues to accelerate today.

The development of any new technology occurs in several phases. When a new technology (automation) is adopted, the first phase is the replacement of traditional manual functions with automated ones. Shared cataloging through utilities like OCLC, UTLAS, and WLN was one of the first activities to become automated on a major scale. The serials system mentioned above is an example of automation of acquisitions functions. These were followed by automated circulation systems which appeared in the late 1970s.

Once traditional functions have been automated and people begin to see the potential of using the technology, new uses and ways of doing things are devised and organizational change evolves. Dilys E. Morris describes this phase and its impact on libraries:

> The online catalog with its enhanced keyword and Boolean access allows retrieval never before possible. Database search services are moving reference librarians into stage two as they perform tasks very different from traditional duties.[2]

The final phase of a technology occurs when society itself changes as a result of that technology. This is occurring with the development of networks on a local, regional, national, and international scale.

The purpose of this paper is to trace the development of national networks, describe their current condition, and discuss the future implications of and obstacles to the accomplishment of the vision.

Roberta A. Corbin is Head of Bibliographic User Services at the University of California, San Diego Libraries.

NETWORK DEFINED

What is a network? The word has many different definitions in the world of automation, and its meaning has changed as technology has developed. Initially, computing was only done on large mainframe computers that, due to their expense and the specialized environments required for their operation, were centralized operations. During this period, network referred to the distributed system of terminals that were connected to a central mainframe computer. The most well known systems of this type in libraries were the networks established by OCLC and RLG, which provided bibliographic records for cataloging. There were also resource databases such as DIALOG and ORBIT, which became tools for reference librarians at academic libraries in the 1970s.

> Each of these networks had a single purpose: to link terminals to a host computer that provided a particular service (or set of services provided by a single organization).[3]

As computers became more dependable, costs came down, more applications were developed, decentralization occurred, and departmental computers became common on university campuses. In the early 1980s microcomputers brought the advantages of wordprocessing, spreadsheets, and database applications to the desktop. However, there was still a need to share common resources that resided on large computers. This was done using communications software or terminal emulation programs that allowed the microcomputer to act as a terminal on a mainframe.

As these developments took place, a change occurred in the use of networks. Microcomputers could not only communicate with mainframes and each other, but data could be moved to and from other computers and processed on these desktop machines. They could also provide many other capabilities terminals could not. Caroline Arms writes:

> These "peer-to-peer" networks are designed to support more general forms of communication than terminal-to-host networks. The underlying model is no longer of masters and slaves. Computers on the network, whether large mainframes or personal workstations, are attached to the network as equals, each identified by a unique address. Network traffic consists of individual messages from one computer to another, with communication between workstations as easy as communication from a workstation to a large time-sharing computer.[4]

The ability to move from terminal-to-host networks to peer-to-peer networks necessitated significant technological development because this transition required computers and networks with dissimilar hardware and software to communicate with each other. To understand how this occurred requires some discussion of the history of these developments.

ARPANET, THE DARPA INTERNET, AND BITNET

In 1969 the ARPANET was formed as a project of an agency of the Department of Defense variously known as the Advanced Projects Research Agency (ARPA) or the Defense Advanced Projects Research Agency (DARPA). The network's purpose was to connect computers at geographically distributed sites "to allow scientists to share data, to use computer facilities and programs interactively at each other's sites, and to exchange electronic mail messages."[5] At the time of its development, those involved predicted the primary use of the network would be for connecting to resources on specific computers. While electronic mail was mentioned as a possible use, the overwhelming popularity of this function and its support of publication was not anticipated by the developers.

By the mid-1970s the additional need to communicate between networks was recognized and the DARPA Internet, later to become the Internet, was formed to develop the architecture needed to make these network connections possible.

The ARPANET continued to grow and develop until 1984 when it was split into the MILNET for military use and a research component which retained the name ARPANET.

Among the many hurdles overcome, two major technological breakthroughs for networking came from these projects (ARPANET, MILNET, and the DARPA Internet). First, packet-switching was developed as a means of sending messages over networks. Arms states that

> on packet-switching networks, messages [data] are sent as individually addressed packets; long messages are broken into short segments and sent as sequences of separate packets. Packets from different communication sessions can be interspersed on the same line, and if equipment fails, packets can be rerouted individually without affecting the sessions.[6]

This technique increases the speed and efficiency of the transmission, making networking economically feasible and providing better security for messages since they are sent in medium-sized pieces rather than as whole messages or individual bytes (characters).

Second, communications protocols were developed. Protocols are described in a report from the Office of Technology Assessment (OTA).

> The ability of any two systems to interconnect depends on their ability to recognize and deal with the form information flows take in each. These "protocols" are sets of technical standards that, in a sense, are the "languages" of communication systems.[7]

The protocols that were developed allow dissimilar networks to communicate with each other. The Transmission Control Protocol (TCP) and Internet Protocol (IP) are used in conjunction with each other and are jointly referred to as TCP/IP. These protocols work through devices called "gateways" which use the TCP/IP standards to convert messages from the "language" of one system to that of another as the message moves through the networks.

Within the TCP/IP suite, other protocols were developed to support electronic mail (SMTP), file transfer (FTP), and terminal emulation to access remote computers (Telnet—not the same as Telenet, the commercial network).

With the implementation of TCP/IP and the ability to send and receive data between different networks through gateways, internetworking was born. Internetworking refers to the ability to communicate between heterogeneous or homogeneous networks and requires

> a consistent form of internet addressing—to enable communication between users on the various constituent subnetworks; points of interconnection between the networks; a mechanism to route and control the flow of information on the internet; and common protocols at some level in order to enable different constituent networks to communicate with each other [TCP/IP gateway].[8]

The ARPANET and Internet are not open to all educational institutions; only those involved in government-sponsored research can belong. As a result, in 1981 the BITNET (Because It's Time NETwork) was formed. It is an international educational "cooperative network serving more than 2,300 hosts at several hundred sites in 32 countries."[9] While BITNET and the Internet use different protocols within the two networks, electronic mail communication is possible through gateways. However, file transfers between the two networks are not possible, and inconsistencies in addressing continue to cause some problems in communication.

NSFNET

The National Science Foundation (NSF), which is "charged with supporting scientists and maintaining the nation's leadership role in science and engineering,"[10] began its efforts in the network arena with the funding of CSNET in 1981. It was intended for use by engineers and computer scientists at all institutions, not just those with DARPA projects. It has since merged administratively with BITNET.

Recognizing the need for high-speed computing and networking for scientific research, the NSF began plans for the development of six supercomputer centers and a high-speed network in 1984. The centers are located in Princeton, Boulder, Pittsburgh, Urbana-Champaign, Ithaca, and on the campus of University of California–San Diego. The NSFNET is a three-level internet made up of a transcontinental backbone that connects separately operated mid-level networks, the NSF-funded supercomputer centers, and campus networks.[11]

The NSFNET is a general-purpose research network and is part of the Internet. It is jointly administered by the NSF, which determines policy; Merit, Inc. (a network cooperative of eight Michigan universities), which handles operation and management; and IBM and MCI, which provide research and equipment. In November 1990, IBM, MCI, and Merit announced the establishment of Advanced Network and Services (ANS), "a not-for-profit organization that will manage and operate the federally funded National Science Foundation Network (NSFNET) backbone under subcontract to Merit."[12] It is intended that the NSFNET and the Internet will provide the backbone and foundation on which to build a national research and education network.

The growth rate of the NSFNET has been phenomenal. By May 1990 approximately 300 institutions were connected, with nearly 1,000 individual subnetworks. The rate of growth of user networks was approximately 20 percent a month.[13]

In 1989 William A. Wulf, who heads the NSF's Computer and Information Science and Engineering Directorate that operates the National Supercomputer Centers and NSFNET, stated that a 20 percent growth rate "translates to a doubling every four and a half months in the amount of traffic on the network" and predicted that the NSFNET would be saturated in a year and a half.[14]

THE INTERNET

The Internet is a network of networks, and it is the internetwork.

It exists to facilitate sharing of resources at participating organizations (which include government agencies, educational institutions, and private corporations) and collaboration among researchers, as well as to provide a testbed for new developments in networking.[15]

It is international in scope and is composed of 400 to 500 other networks, some of which are in themselves very large and have component networks within them. Estimates of host computers on the Internet range from 40,000 to 500,000 and numbers of users are probably somewhere between 500,000 and a million.[16]

In addition to MILNET, NSFNET, CSNET, and ARPANET (now called DRI), regional networks such as Los Nettos (founded by Caltech, UCLA, USC, Trusted Information Systems, and Information Sciences Institute), CERFnet (California Education and Research Federation network), the Merit Computer Network (founded by a consortium of Michigan universities), and THEnet (the Texas Higher Education Network) are part of the Internet. University level networks such as the UCDLANet of the University of California are also Internet participants.

Through the Internet, librarians and other users can exchange electronic mail messages with colleagues across the United States and in foreign countries. They can subscribe to bulletin boards through which participants exchange views and information on a variety of topics. They can access library catalogs such as the Melvyl catalog of the University of California and CARL (the Colorado Alliance of Research Libraries). In addition, there are specialized resources such as the Online Mendelian Inheritance of Man, a catalog of inherited traits and disorders at Johns Hopkins University; the text of Dante's Divine Comedy and sixty commentaries on the work, developed by the Dartmouth Dante Project; and the Southwest Research Data Display and Analysis System, which provides access to data collected by the Dynamics Explorer satellites.[17]

While the Internet is heavily used by its participants and is of enormous value, in reality it is a rough skeletal structure and provides only the very basics of what is and will be needed in the future to change raw data into the information needed to advance our knowledge of the universe and resolve the many difficult problems facing us today.

THE NATIONAL NEED

The current networks are limited in the amount of data they can transmit by speed (the rate at which data moves across the network) and bandwidth (the volume of data that can be transmitted at one time). Vast increases in both speed and bandwidth are essential for the United States to maintain its competitive edge in the world. Projections are that it will be necessary for a national network to carry at least three gigabits of data per second by 1996 to begin to meet our needs. This translates into

100,000 typed pages or 1,000 satellite photos each second. The Internet and NSFNET currently have bandwidth speeds of up to 1.56 megabits per second, which is the equivalent of fifty typed pages per second and represents a thirty-fold increase in capacity when compared to speeds prior to 1988. The NSFNET plans to expand to forty-five megabits by early 1991, or 1,500 typed pages per second.

At the present time supercomputers can generate data ten times as fast as any digital network can transmit it, which means that to use the full power of our current supercomputers it is necessary to be in the same building as the computer because even our fastest current networks cannot handle the enormous volume of data produced. Senator Albert Gore of Tennessee, who sponsored the national network legislation in the Senate, writes:

> We are now looking . . . at the Mission to Planet Earth program, which will send down to the earth's surface every day from orbit a quantity of information equal to all the bits of data in the entire Library of Congress. We can't even handle the information we now have about planet earth. In fact, someone argued that we already have all the information we need to decipher the operation of the climate system and the changes we're making to that system. One of the critical questions that must be answered . . . is where are the clouds? Another is, what has happened to land use on the surface of the earth? NASA is proposing to send up a new generation of data-collecting satellites and spend seventeen years compiling a lot more information in order to answer those two questions and many others. And yet we have taken, with just one satellite system, Landsat, a complete photograph of the earth surface every two weeks for the past eighteen years. More than 95 percent of those pictures are stored in digital form and have never fired a single neuron in a single human brain. We know where the clouds are, or, more accurately, we have the data to tell us where the clouds were. We don't know where they are yet, because we have not been able to find the needle in that particular haystack. Teams of researchers, using the most complicated models now available, cannot communicate with each other and work together productively to solve that and other problems, because in order to communicate they have to find a way to download their models and their results onto magnetic tapes, take the tapes to another lab, and read the tapes so other researchers can review and talk about them.[18]

Douglas E. Van Houweling, vice-provost for Information Technology at the University of Michigan and chairman of the board of MERIT, discusses our situation in relation to higher education and the foreign challenge when he writes:

> Judging by the demand for its output from students all over the world, the United States has developed the world's leading system of higher education, especially at the graduate level in science and technology. Almost a quarter of the basic science graduate students and almost half of the engineering graduate students enrolled in our universities are foreign nationals.[19]

Van Houweling goes on to predict that if the current trends continue with more and more of our graduate degrees going to foreign nationals, the United States will not be able to produce enough well trained individuals in the areas of science and engineering to maintain our current lead. He also expresses concern over the continued transfer of knowledge developed in the United States to foreign countries and states that

> despite the magnitude of our research and development effort, the proportion of U.S. patents granted to U.S. inventors has been declining for more than a decade. If current trends continue, sometime in the 1990s more U.S. patents will be granted to foreign nationals than to U.S. inventors.[20]

Japan and the soon to be united European countries have seen the benefits of high speed networks and have national efforts in place to develop and implement ones that will reach into all levels of their societies.

While the first step will be to provide communication for the scientist in what Wulf describes as the national colaboratory, "a framework in which scholars across the nation interact almost as if they were across the hall from each other,"[21] it is no less important for the future of the country that children learn to take advantage of the vast resources of information that will be available through this network. The possibilities are unpredictable and defy the imagination.

GOVERNMENTAL ROLE AND LEGISLATIVE STATUS

In the mid-1980s both the Senate and the Office of the President began studying the current network situation to determine what role the government should play. Between 1987 and 1989 several government agencies were charged with determining the need and developing the conceptual framework for a national network. In 1987 the Federal Coordinating Council for Science, Engineering, and Technology (FCCSET), a working group of the White House Office of Science and Technology Policy (OSTP), issued a report that

> called for a balanced program consisting of four parts: high performance computers, software and algorithms, networking, and basic research and human resources.[22]

This was a general plan with no proposed details as to who would do what or how the goals would be accomplished. According to Paul G. Huray and David B. Nelson, members of FCCSET committees working on this project, this report and a later implementation document form the basis for a coordinated federal program to implement a national high speed computing network.[23]

Legislation to establish a national network has been introduced several times throughout the decade of the eighties. The FCCSET committee reports were followed by a series of Congressional hearings that led to the reintroduction of legislation in both the House and the Senate in 1989. These events coincided with the appointment of Dr. Allan Bromley as the President's science advisor and director of OSTP. One of his first acts was to issue *The Federal High Performance Computing Network* report developed by FCCSET and issued by OSTP. In his first Congressional testimony following his appointment, Dr. Bromley stated that

> there is no technological activity in the federal government which will have a greater impact upon our future as a global competitor than support of the high-performance computing program.[24]

The 1989–90 Senate version of the NREN enabling legislation was introduced by Senator Gore, and the House bill was introduced by Representative Doug Walgren of Pennsylvania. The original bill included sections addressing the four parts of the FCCSET report mentioned above and was rewritten in the spring of 1990 to include more language acknowledging the role of libraries.

Despite the importance placed on this legislation in Congress and the White House, the budget problems that occurred at the end of the 101st Congress left no time for negotiations between the House and the Senate to develop an acceptable version for passage in both houses. In addition, the Senate version was so compromised in order to garner the support of the Senate Energy Committee that Paul Evan Peters, Director of the Coalition for Networked Information, reported a number of those representing libraries and higher education came to the conclusion that it was better that no bill had passed than the final version of the Senate bill. Two of the critical changes made were removal of the National Science Foundation as the lead implementing agency and removal of the strong commitment to a uniform architecture and standards. According to Peters,

> the first is critical because the National Science Foundation is unique as a Federal agency in its commitment to basic rather than mission-oriented research and development. The second is critical because the complexity and cost of multiple, parallel, and possibly incompatible Federal networks is a situation that cannot be allowed to continue let alone increase.[25]

At this point new legislation has been reintroduced to the 102nd Congress and is making its way through both houses.

THE COALITION FOR NETWORKED INFORMATION

As the legislation for a national high-speed computing superhighway made its way through Congress, a coalition of interested library and education groups began to develop, culminating in the formation of the Coalition for Networked Information (CNI) in the spring of 1990. It was formed by an alliance of the Association of Research Libraries (ARL), CAUSE (the professional association for the management of information technology in higher education), and EDUCOM (a nonprofit consortium of colleges, universities, and other institutions involved with campus information networking). According to Thomas Kirk and Thomas Michalak, although the creation of the Coalition was stimulated by proposed federal legislation that would create a National Research and Education Network (NREN), its purpose is not solely to promote and influence the shape of NREN. The sponsoring organizations see the Coalition and its related Task Force as a vital force in shaping the merging national networks and the information they transport.[26]

Since the formation of CNI, several hundred colleges and universities have joined to show their support for national networking. Following the failure of passage of the NREN legislation, the Coalition, under the direction of Peters, has been actively developing its strategy for the new session of Congress.

POLICY ISSUES AND TECHNOLOGICAL REQUIREMENTS

While many forces are coming forward to develop the so-called national high-speed computing superhighway, there are still many legal, economic, technological, policy, social, and even behaviorial issues that must be resolved along the way.

Legal issues involving both copyright and patent law in relation to electronic media have been argued without resolution for several years. In the meantime, electronic publishing is an ever-increasing reality and individual institutions are developing relationships that will test the current laws and possibly force the passage of new ones to handle the conflicts that will inevitably arise.

Funding is a complex and difficult issue. There is little disagreement that the costs of research, implementation, management, and maintenance are high, but who pays? As Ed Brownrigg writes:

> One of the basic problems with the issue of funding the NREN is that most of the organizations connected to the Internet currently pay for the leased telecommunications circuits that link them to an Internet gateway. To add to the confusion, some of the long high-speed circuits in the Internet are underwritten by their common carriers. These practices may give rise to the appearance that, in large part, the proposed NREN would start as self-funding and, thus, not be in need of public support.[27]

Public support is critical if this network is truly to become a national network available not only to those in institutions with money to spend but also to institutions and individuals who may not have the funds but could benefit from the access. It is imperative that this network does not further divide our country into information haves and have-nots.

Technological issues are twofold. Hardware advances are needed to meet the computing goals described and the standards are needed to provide the mechanisms for implementing compatible architectures. In order to meet the supercomputing and network needs of researchers, more powerful computer designs, improved technologies for visualizing data results, and higher data transfer speeds need to be achieved. Standards are what make transfer of information across networks and from computer to computer possible. Kibbey and Evans describe four categories of standards that are needed. They include document representation and exchange standards, standards for data structure and indexes, communication and networking standards, and display interface standards.[28]

Policy issues include such areas as who manages the network, how international linkages are controlled, who has access and to what, what services are provided, and how to balance the specialized needs of researchers with the broader needs of more diverse users.

A whole range of social and behavioral issues exist that have generally been ignored in the literature and the planning. According to McClure et al. little research has been done in such areas as how to provide adequate education, training, and documentation for a national network; how to develop network policies that fit the prevailing social structure; do individual organizational cultures support and encourage networking; how does an organization measure the benefits of network participation on productivity; whether networks change behavior or merely facilitate work; and whether "technophoria" leads to the unfounded belief that all technology is good without providing any sound basis for that conclusion.[29]

LIBRARIES AND NREN

Phrases such as "library without walls," "the electronic library," and the "virtual library" have entered our vocabulary recently. These terms have different meanings to different people. To the scholar they may mean the ability to work with other scholars in his field through the network as though they were in the same laboratory, to run experiments with others and see results on computers thousands of miles away, to send documents to colleagues for comment and review, to publish papers or read the latest research in his field on a daily basis through electronic media.

For teaching institutions a national network could provide a means of communicating with a widely distributed population of students and faculty and of allowing access to resources that might otherwise not be possible. Libraries of all types could provide faculty, students, and the community with access to resources that were totally inaccessible in the past.

A lot of thought, discussion, and plain old hard work will be required for these visions to materialize. Librarians can play a unique and important role in the realization of this new environment by developing information organization and indexing strategies, and providing guidance both on user friendly interfaces and on what resources people will need. Through their knowledge of bibliographic instruction, librarians can assist in developing documentation, training and instruction strategies, and materials.

SUMMARY

The development of a national research and education network is of vital importance to our nation. McAdams et al. write that

> research and education networks merit the same kind of support that the federal government has in the past extended to the national mail system, the continental railroad, and the national highway network.[30]

A national information network is also of great importance to all citizens whether or not they currently recognize the role of information and knowledge in this Information Age. It will take leaders of vision from many fields to fulfill the dream.

REFERENCES

1. Caroline Arms, "Libraries and Electronic Information: The Technological Context, Part One," *EDUCOM Review* 24, no.2:41 (Summer 1989).
2. Dilys E. Morris, "Electronic Information and Technology: Impact and Potential for Academic Libraries," *College & Research Libraries* 50, no.1:58 (Jan. 1989).
3. Caroline Arms, "A New Information Infrastructure," *Online* 14, no.5:15 (Sept. 1990).
4. Caroline Arms, "Libraries and Electronic Information: The Technological Context, Part Two," *EDUCOM Review* 24, no.3:39 (Fall 1989).
5. Dennis G. Perry, Steven H. Blumenthal, and Robert M. Hinden, "The ARPANET and the DARPA Internet," *Library Hi Tech* 6, no.2:51 (Summer 1988).
6. Arms, "A New Information Infrastructure," p.17.
7. Office of Technology Assessment, *High Performance Computing & Networking for Science: Background Paper* (Washington, D.C.: Government Printing Office, 1989), p.23.
8. Larry L. Learn, "NETWORKS: A Review of Their Technology, Architecture and Implementation," *Library Hi Tech* 6, no.2:43 (Summer 1988).
9. John S. Quarterman, *The Matrix: Computer Networks and Conferencing Systems Worldwide* (Bedford, Mass.: Digital Press, 1990), p.230.
10. Arms, "A New Information Infrastructure," p.18.
11. Quarterman, *The Matrix*, p.301.
12. "IBM, MCI, Merit Form Organization to Manage NSFNET," *Advanced Technology Libraries* 19, no.11:5 (Nov. 1990).
13. Paul G. Huray and David B. Nelson, "The Federal High-Performance Computing Program," *EDUCOM Review* 25, no.2:23 (Summer 1990).
14. William A. Wulf, "Government's Role in the National Network," *EDUCOM Review* 24, no.2:22 (Summer 1989).

15. Quarterman, *The Matrix*, p.278.
16. Ibid.
17. Caroline Arms, "Using the National Networks: BITNET and the Internet," *Online* 14, no.5:28 (Sept. 1990).
18. Albert Gore, "Remarks on the NREN," *EDUCOM Review* 25, no.2:13 (Summer 1990).
19. Douglas E. Van Houweling, "The National Network: A National Interest," *EDUCOM Review* 24, no.2:15–16 (Summer 1989).
20. Ibid, p.16.
21. Susan Wintsch, "Toward a National Research and Education Network," *Mosaic* 20,no.4:32 (Winter 1989).
22. Wulf, "Government's Role in the National Network," p.24.
23. Huray, "The Federal High-Performance Computing Program," p.17.
24. Ibid., p.20.
25. Paul Evan Peters, Washington, D.C. to PACS-L (Public Access Computer Systems Forum), 1 November 1990, electronic mail message received by R. Bruce Miller at the University of California, San Diego.
26. Thomas Kirk and Thomas Michalak, "The Coalition for Networked Information," *College & Research Libraries News* no.8:751 (Sept. 1990).
27. Edwin Brownrigg, "Developing the Information Superhighway: Issues for Libraries," in *Library Perspectives on NREN: The National Research and Education Network*, ed. Carol Parkhurst (Chicago: Library and Information Technology Association, 1990), p.58.
28. Mark Kibbey and Nancy H. Evans, "The Network is the Library," *EDUCOM Review* 24,no.3:16–17 (Fall 1989).
29. Charles R. McClure and others, "Realizing the Promise of NREN: Social and Behaviorial Considerations—A Status Report on a Study in Progress," in *Library Perspectives on NREN: The National Research and Education Network*, ed. Carol A. Parkhurst (Chicago: Library and Information Technology Association, 1990), p.23–29.
30. Alan K. McAdams and others, "Economic Benefits and Public Support of a National Education and Research Network," *EDUCOM Bulletin* 23,no.3:63–64 (Summer/Fall 1988). ■■

Communications

Computers and Romanization of Chinese Bibliographic Records

Karl K. Lo and R. Bruce Miller

The Library of Congress is faced with pressure from many institutions to change from Wade-Giles romanization to pinyin for Chinese characters. For libraries with substantial East Asian collections, such a change conjures up images of split files, new cataloging rules, retrospective conversion, and public confusion. This paper provides a brief overview of romanization for Chinese characters, discusses two identified major problems, and poses an automated alternative to the proposed change. The appendix contains the text of a computer program that shows that words and phrases romanized according to the Wade-Giles scheme can be mechanically converted to the pinyin format and vice versa. Libraries should not change from the use of Wade-Giles to pinyin romanization, should not aggregate the romanized Chinese words, and should not further complicate bibliographic records with additional fields for romanization. It would be more cost-effective to apply computer resources to the efficient processing of nonroman scripts with the goal of developing a better language interface between the computer and library staff and users that would automatically link varied romanized forms with the original script. Instead of suffering with romanization as unavoidable processing overhead, we could use Wade-Giles, pinyin, and other romanization schemes to enhance access to bibliographic records. Libraries with East Asian materials should take a close look at the proposal to change romanization schemes and should seek an alternate, more productive way to help the Library of Congress respond to its external pressures.

Karl K. Lo is Head, International Relations and Pacific Studies Library, and R. Bruce Miller is Assistant University Librarian-Technical Services, University of California, San Diego.

"For many years it has been a matter of astonishment and regret that it has not been possible to create a system of transcription for Chinese which is scientifically adequate and at the same time simple and practical and internationally accepted." Such were the words of Bernard Karlgren spoken in 1928 to the China Society in London about the romanization of Chinese.[1] In the same speech, Karlgren continued: "It is really no wonder at all that we have not arrived at a definite result so far, for we have never made it quite logically clear to ourselves what it is we want."[2] This last statement and its potentially significant impact for libraries in the near future is the paramount concern of this paper.

Transcription of the sounds of Chinese dialects into the Roman alphabet has a history as long as the history of visits of Westerners to China. Many schemes have been developed. Some were devised by Westerners, many others were designed by Chinese. For Mandarin alone, more than one hundred different schemes are known. At this time, two of the schemes, Wade-Giles and pinyin, are used internationally. The Wade-Giles system "is named after two Englishmen: the system's inventor, Sir Thomas Wade (1818–1895), whose book *Peking Syllables* was published in 1859; and the man who modified it into the form we know today, the Cambridge professor Herbert Allen Giles, whose famous

Chinese-English Dictionary was published in 1912."[3] Advocates of each system proclaim its merits and often downplay any values of the other. While many of the social and philological values espoused by the advocates of both sides may be valid, it is often completely overlooked that the two schemes are completely interconvertible, i.e., the text that results from the conversion of Chinese script into the Roman alphabet using the Wade-Giles rules can be directly converted to pinyin format and vice versa.

This superficially arcane and scholarly issue has the potential to be a significant problem in the near future for the systems and technical services staff of academic and research libraries that collect Chinese material. Although the *Anglo-American Cataloguing Rules* do not specify romanization schemes, most North American libraries use the ALA/Library of Congress tables.[4] For Chinese characters, that scheme is Wade-Giles. The problem stems from the fact that in the early 1980s, with the notable exception of American libraries, the pinyin system began to be widely used throughout the world. Ten years later the Library of Congress is faced with pressure from many institutions to change to pinyin while those libraries that have been using the Wade-Giles system have a strong vested interest in maintaining the status quo.

In March 1990, the Library of Congress asked the American library community to respond to its paper "Pinyin: Possible Approaches for Cataloging and Automation."[5] The responses indicated that "the more general libraries tend to favor the change to pinyin, while libraries with large East Asian collections are more hesitant, being only too well aware of the effort and expense that such a change would entail."[6] Resistance from the East Asia collections notwithstanding, "The Library of Congress believes that a change to pinyin is inevitable, but also understands that two major problems must be resolved before making the change."[7] This paper provides a brief overview of romanization for Chinese characters, discusses the two identified major problems, and poses an automated alternative to the proposed change.

THE PURPOSE OF ROMANIZATION IN AMERICAN LIBRARIES

"Romanization is a method for converting a word that is written in a nonroman script into a word that sounds like the original but is written in the roman alphabet. . . . Phonetic transcription . . . uses the roman alphabet to reproduce the nonroman word in the approximate sound of a target language, e.g., English."[8] This is where the problem begins: what sounds like "Peking" to one ear may sound like "Beijing" to another.

Romanized proper names and book titles have been used in Western catalogs, bibliographies, and other publications for many years, primarily to enable convenient transcription or substitution for nonroman characters. For example, romanization is found on the Harvard-Yenching Library cards that were printed in the 1930s, but for many years filing in that catalog was still based on the Chinese characters. (The cards were filed numerically according to the four-corner system based on the vernacular script and not alphabetically in the romanized Wade-Giles form.) In the late 1950s, when numerous new Chinese collections came into existence partly due to the National Defense Education Acts, the Library of Congress led the way in standardizing the format of Chinese cards. This included the addition of a romanized title at the right bottom corner of each card. By that time, most American libraries had begun filing their Chinese cards according to Wade-Giles, in the word-by-word convention.

In the 1970s, when LC, OCLC, and some other libraries developed online systems, only romanized data could be entered. Non-alphabetic characters were excluded due to the then current computer inability to process script-based languages like Chinese in their original form. Romanization was used in the bibliographic record to provide surrogate characters. To make romanized words look more like English, a few adaptations were made, e.g., hyphens and upper case characters were used to signify proper names and a space was used to separate two words from each other. In the 1980s, the Research Libraries Group did pioneering development to create a CJK (Chinese, Japanese, Korean) workstation for RLIN that allowed the use of Chinese characters in machine readable bibliographic records for the first time in the United States. In the latter part of the decade, OCLC also developed the capability for input and retrieval of CJK bibliographic records. At that time, key fields in the RLIN and OCLC implementation of MARC were already used for romanized data. For expediency, both utilities elected to place the original script characters in subfields and specially coded redundant fields in order to leave the romanized data where it was for the use of most libraries. The original script characters are used only by those few that have CJK capability. The value and need for romanization as a substitute for the original characters is confirmed by the thousands of libraries with CJK materials but with no automated CJK processing capability.

Both romanized and original script CJK data are indexed in the RLIN and OCLC databases. Romanized data are primarily found in title and responsibility statements, but there are numerous name, subject, and series fields that are under authority control that also are romanized forms. Romanized data are found either in these authoritative headings or in the reference structure that guides users to these access points.

This paper later discusses the potential for romanization to be managed in the interface portion of a database instead of requiring manual input, record-by-record. Related to this point, it is important to understand that there are many romanization schemes for Chinese that can be found throughout library information resources. A shift from Wade-Giles to pinyin for catalog bibliographic records merely begs a larger question regarding how we manage and provide information. Consider the possibilities for romanized variations in citations that a user might bring to a library. The Yale scheme, which was invented during the second World War, has been popular in language classes across the country; many students are "fluent" in this scheme. Many published sources use the Mathews scheme from the popular Mathews dictionary,[9] which is basically Wade-Giles with a few spellings that vary from the Library of Congress romanization rules. In recent years, pinyin gained popularity in many circles and has been accepted as an ISO standard which is clear evidence that it is challenging Wade-Giles popularity. (For example, note the shift from "Peking" to "Beijing" in the news media.) While it is difficult to prove which scheme is most popular with American users at the moment, it is safe to say that many individuals and organizations are using more than one scheme simultaneously.

Although the following is an authority control issue and is beyond the scope of this paper, it is interesting to note that a number of subject and name headings that are obviously romanized are not done so according to a systematic scheme. These headings are generally the result of Anglicization and common usage, e.g., Confucius, Chiang Kai-shek, Taoism. Some other proper names, such as Peking University and Tsingtao, which are neither pinyin nor Wade-Giles romanization, are the preferred forms supported by the owners of those names.

Even though the volume of data in any given bibliographic record is not inordinately large, the sheer mass of distinct titles and other romanized information in those databases and the databases or card catalogs of any library with even a moderately sized collection of CJK material is impressive. However, "daunting" is the adjective that comes to mind when faced with the possibility posed by the Library of Congress to change from Wade-Giles to pinyin romanization. At least for libraries with substantial East Asian collections, such a change conjures up images of split files, new cataloging rules, retrospective conversion, and public confusion that are all too painfully reminiscent of the implementation of *AACR2*.

Americans use various schemes for romanizing Chinese for numerous purposes: pronunciation, alphabetizing, indexing, transcription, and even as a surrogate language. Most of these adaptations are practical, as in the case of geographic names. Some applications are ingenious, such as those used for linkage with Chinese characters in computer word processing. Unfortunately, an attempt to achieve a national consensus on one scheme is probably doomed to failure because of the diverse vested interests and the lack of a common purpose. Not unlike Karlgren's contemporaries, we in the United States are not exactly clear on what we want most from a romanization scheme.

Although this paper is not the forum to debate our national political interests, library economy, or linguistic principles, it can provide useful information to be used as we address those issues in the library context. The following discussion focuses on romanization as an indexing tool in library automation and evaluates Wade-Giles, pinyin, word division, and bibliographic record structure in that light. Appendix A contains the text of a computer program prepared by one of the authors, Karl Lo. This program illustrates that words and phrases romanized according to the Wade-Giles scheme can be mechanically converted to the pinyin format and vice versa. It is hoped that this establishes one important technical point: Wade-Giles and pinyin are equals as far as indexing and online systems are concerned and are potentially interchangeable. We conclude that automation can be the solution to the Wade-Giles/pinyin romanization conflict in libraries and that it is not necessary to disrupt existing catalogs or cataloging workflows. We invite developers to refine the logic of this program and to incorporate the results into online library systems in order to let our machines take the burden and problems of romanization away from library staff and the public.

PROBLEM ONE: WORD DIVISION

One obstacle to the adoption of pinyin in libraries is the lack of standards for word division. In general, an individual Chinese character stands for a meaningful syllable. However, the language also includes numerous multi-character compounds that form a single morpheme. Although certainly not the only reason, the foremost use of romanization in American libraries is for filing and indexing. American libraries romanize and file romanized Wade-Giles Chinese word-by-word, more to follow the convention of sinologists than for any other reason. However, the Wade-Giles "word" really equals a single Chinese character, i.e., two or even three Wade-Giles generated "words" may actually represent a single morpheme. Pinyin aggregates these "words" into a single linguistic unit. This aggregation in pinyin romanization became an issue only when pinyin was suggested for adoption within libraries. If you are knowledgeable about filing systems, it should be immediately apparent that a single Chinese morpheme that is composed of two or more characters will file in one location if romanized by Wade-Giles rules and will file in another location if romanized according to pinyin. For example, the Chinese morpheme for flourishing becomes "mao sheng" when romanized according to Wade-Giles and "maosheng" when romanized according to pinyin. This is a problem both in card files and online indexes.

For aggregation of pinyin words, there is no national or international standard, even though pinyin has been designated as the ISO standard for romanization of Chinese. Without standard rules for aggregation, different results can be produced by different people. For example, the title "Gu sheng wu xue shi" can be aggregated in a number of ways that will create different words for indexing. On the other hand, no matter how a string of words (characters) is aggregated, it can always be divided into single syllable words consistently and accurately. This is a very important point that we wish to make in this paper. The computer program that is discussed below demonstrates this quite well.

Divided single syllable words rather than aggregated words are arguably easier for a computer to manage and a user to search if you can provide suitable computer resources to handle the large number of homonyms. For example if the four words of "Zhong guo li shi" are aggregated into two, "Zhongguo lishi," the title can only be searched by zhongguo lishi, zhongguo, or lishi. If the title remains divided as four words, it can be searched by any one or any combination of the four words. Non-aggregated titles thus have a greater possibility of being found.

Large retrievals due to multiple homonyms can be a problem for some computer systems. However, our perception is that this is a limitation of constrained computing power. As technology advances and the price of performance continues to decline, this limitation should become less and less important. In a system that can easily perform Boolean or multiple keyword searching, individual indexing of each Chinese character is transformed from a liability to an asset. The difficulties in establishing rules for aggregation are reminders that aggregation is not obvious and that we will create errors and cause confusion if we require our users to retrieve Chinese records that are filed and indexed according to the results of pinyin romanization.

PROBLEM TWO: INTERCONVERSION OF PINYIN AND WADE-GILES

Most people would agree that pinyin and Wade-Giles are interconvertible. However, most would also think that the process would be complex, particularly in the handling of the aggregation of the pinyin words. The computer program is provided in Appendix A to show that the interconversion, including word division, is not only possible but also very simple. The simplicity of interconversion of the two schemes is pivotal to the argument that these schemes are technically equal.

There are many strategies that can be used for the mechanical interconversion of the two schemes. One simple method is to create a table of about 445 unique word pairs used in the two schemes, mapping one with the other. This is simple in design, but it results in a long table. Most new dictionaries which contain one of the schemes include a printed table to assist readers in conversion.

This program breaks words into components and takes advantage of the fact that most of the words and components need not be converted. In fact, only forty-five component pairs are needed for conversion from Wade-Giles to pinyin. Conversion from pinyin to Wade-Giles requires more pairs in order to break aggregated words into single syllable words.

The first step in shortening the table of 445 pairs is to eliminate all the components and words that require no change. For example, m, n, f, l, and a are equal in both schemes. The words ma, man, la, fa, fan, and some others also need no conversion. The second step is to convert those that have consistent equivalences in the other scheme. For example, t' in Wade-Giles is t in pinyin. Some of the other Wade-Giles to pinyin pairs are: t and d, p and b, and p' and p. The third step is to convert the irregular components that are exceptions to the steps outlined above. Consider the letter o, for example. It is equal in most of the conversions, except that "ong" in pinyin is "ung" in Wade-Giles and the word of a single "o" in pinyin is a single "e" in Wade-Giles.

After these three steps, the conversion is complete. Input a word or words generated according to Wade-Giles and the output is in the pinyin form; input a word or words generated according to pinyin and the output is in the Wade-Giles form. Running the program demonstrates that the interconversion between the two schemes is fast and efficient. The machine language used in the program is TurboBasic. Even on a PC/AT, the conversion of a book title is faster than one can blink an eye. Surely even greater speed could be achieved with more powerful machines and better programming. Machine interconversion is possible, simple, and fast. Large computer systems could readily support multiple romanization schemes to facilitate easy processing by library staff and user friendly retrieval by the public.

OPPORTUNITIES AND UNANSWERED QUESTIONS

The focus of this paper is on romanized data as an indexing tool in order to provide better database indexes using mixed romanization schemes. We strongly feel that neither U.S. librarians nor the ISO will be able to establish a clear demarcation between the pinyin era and the Wade-Giles era with the result that U.S. libraries will have to contend with a mixed environment for a long time to come. Fortunately, the pinyin and Wade-Giles romanization schemes are not only compatible but are completely interconvertible.

In this section we suggest answers to some related questions and offer some possible opportunities to be pursued.

Diacritics

Both schemes use different diacritics that could pose sorting problems in some computer systems. Both use the umlaut-u "ü". Wade-Giles uses the aspiration mark "'" (an apostrophe is often substituted in limited type fonts). Pinyin uses the apostrophe "'" for word division. The circumflex "^" is no longer necessary in Wade-Giles. Most systems can be instructed to sort or ignore diacritics in any manner that meets user needs. Otherwise, the two schemes are technically the same for indexing purposes.

Users

When a book contains romanization, a library should transcribe the romanized data the way it is found, i.e., in pinyin, Wade-Giles, Taiwanese, or any other form. Proper names, when used in subject, added, or main entries, should be in whatever romanized forms the owners of such names prefer. Library users who may not be familiar with various romanization schemes should have easy assistance with conversion. Should this assistance come from multiple versions of the same access point or should the romanization be managed through automated conversion of the form input by the user?

Computer-aided Romanization

Whether information in the database is homogeneously in one romanized scheme or in a mix of several, there should be an internal link to all schemes. The library user's input in one scheme can be machine translated into the internal scheme for searching. When the bibliographic data is retrieved, the front end program can present the data in either pinyin or Wade-Giles, or both, as the user desires. In order to facilitate machine translation, the user could declare the intention to use one of the schemes for input at the start of the program. (Although this paper is focused on Chinese, this concept can be extended to other nonroman alphabets and scripts. See "Nonroman Scripts and Computer Terminal Developments" for additional discussion.)[10]

Synonym List for Searching

This is a variant of the above technique and concept. A number of computerized library systems today can retrieve synonyms or alternative words. When the word "color" is entered, for example, the word "colour" can also be retrieved. Some other examples are Mac and Mc, center and centre, catalog and catalogue. It may be feasible to add dozens of pairs of pinyin-Wade-Giles words to the list, such as "xia" and "hsia," and "zi" and "tzu." Some words may have two synonyms: "pai" may be "p'ai" or "bai," for example.

International Exchange of Data

It is likely that U.S. libraries will import machine-readable bibliographic data from other countries. This simple statement may turn out to be the most important observation of this paper. International sharing of bibliographic data has much untapped potential that could benefit our information services. The ability to convert or manage more than one romanization scheme will be vital to the success of such ventures.

Redundant Fields in Bibliographic Records

Currently the romanized fields are linked to redundant original script fields through the use of a special subfield. To have more than one romanized field would add to this redundancy. This unnecessarily enlarges storage requirements without truly adding new information to the record.

Single-Syllable Word Division

All romanized data should be stored word-by-word (i.e., Chinese character-by-character) for indexing purposes to attain consistency that aggregation cannot attain. For searchers who input aggregate words, the machine can break the aggregations before searching. When the search result is presented, words can be re-aggregated in the same way as the user entered them. Words that have not been entered by the user will remain as separate single syllable words.

Pre-search Conversion of Romanized Strings into Characters

Several Chinese word processors on the market can search for phrases in Chinese characters by using romanized phrases with greater than 90 percent precision.

Tests are needed to see if conversion of bibliographic data employing this technique can match or improve on the impressive results of such word processors.

Search by Characters

Library systems should index nonroman fields as well as romanized fields. Searches with characters are far more precise than searches by romanization. Both RLIN and OCLC allow users to search by characters. Librarians should expedite the coming of character searching capability in local systems.

THE FUTURE

There is one important question still to be raised concerning the future of romanization in bibliographic records. In a computerized library system that can perform Chinese character input and output efficiently, will romanized fields still be needed for books that have no romanization in their bibliographic data? As suggested above, the technical answer is no. Once a Chinese word can be indexed without an intermediate step in romanization, there is no need to manually create and store romanized data in bibliographic records.

One may still argue that libraries that have no multiscript terminals will continue to need the romanized data. This will hold true as long as multiscript hardware remains expensive. As mass-produced workstations become multiscript capable and low priced, multiscript workstations will be a natural choice to replace ASCII terminals with limited capabilities.

To expedite the coming of a popular multiscript library environment independent of human generated romanization, U.S. libraries would do well to begin promoting the development of a multiscript, general purpose workstation now. Today, a $900 microcomputer in the PC/AT class has the potential to be a low end workstation for that purpose. Soft fonts, operating systems, and terminal emulation programs in multibyte modes that can manage nonroman character sets are being developed and sold by U.S. companies. Many more possibilities exist in Asia. Better technologies are rapidly emerging. However, the packaging of these technologies into library workstations implies significant development costs. We can interest granting agencies and vendors in supporting these development costs by opening their eyes to possible applications that go beyond managing Chinese characters and romanization for library bibliographic databases.

CONCLUSION

For now, American libraries should not change from the use of Wade-Giles to pinyin romanization, should not aggregate the romanized Chinese words, and should not further complicate bibliographic records with additional fields for romanization. Instead, it would be more cost-effective to apply computer resources to the efficient processing of nonroman scripts with the goal of developing a better language interface between the computer and the library staff and users that would automatically link varied romanized forms with the original vernacular script. Instead of suffering with romanization as unavoidable processing overhead, we could use Wade-Giles, pinyin, and other romanization schemes to enhance access to bibliographic records.

The authors do not presume to have answered all of the questions associated with this complex issue, but we do hope to provoke enough interest within libraries with East Asian materials to elicit a closer look at the proposal to change romanization schemes, and that we can find an alternate, more productive way to help the Library of Congress respond to its external pressures.

REFERENCES

1. Bernard Karlgren, "The Romanization of Chinese," paper read before the China Society on January 19, 1928 (London: China Society, 1928), p.1.
2. Ibid.
3. Hanyu Tao and Charles Cole, "Wade-Giles or Hanyu Pinyin: Practical Issues in the Transliteration of Chinese Titles and Proper Names," *Cataloging & Classification Quarterly* 12, no.2:106 (1990).
4. *Anglo-American Cataloguing Rules*, 2d ed. 1988 revision (Chicago: American Library Assn., 1988), p.3.
5. Lucia J. Rather, "Pinyin: Possible Approaches for Cataloging and Automation," Library of Congress memorandum, March 30, 1990.

6. Henriette D. Avram, "LC Position on the Use of Pinyin Romanization," Library of Congress memorandum to those interested in Chinese romanization, January 3, 1991. p.1.
7. Ibid.
8. R. Bruce Miller, "Nonroman Scripts and Computer Terminal Developments," *Information Technology and Libraries* 1:144 (June 1982).
9. Robert Henry Mathews, *Chinese-English Dictionary*, rev. American ed. (Cambridge, Mass.: Harvard Univ. Pr., 1963).
10. Miller, "Nonroman Scripts," p.143-48. ■■

APPENDIX A. COMPUTER PROGRAM FOR CONVERSION BETWEEN WADE-GILES AND PINYIN

```
' InterConversion of Pinyin and Wade-Giles
' A Turbo Basic* program written by
' Karl K. Lo, University of California, San Diego, Library
' May 31, 1991
' Copyright reserved by the author
                    INITIAL CONSTANTS AND VARIABLES
longline$ = string$(80,chr$(176))
longdark$=string$(80,chr$(177))
longblank$ = space$(80)
target$ = "Pinyin"
source$ = "Wade-Giles"
mode = 1
lencheck%=0
work$=""
                              MAIN PROGRAM
      Gosub FrontStatement              ' Opening screen
      StartingGate:                     ' To start or repeat program
      Gosub InputScreen                 ' A menu-and-input screen
      Gosub Choices                     ' User to select actions
      work$ = FN Convert$(work$)        ' Machine to convert input
      Gosub PresentResult               ' Result displayed on screen
      END                               ' Exit
                        FUNCTION AND SUBROUTINES
DEF FN Convert$(inString$)
  LOCAL neo$ = "-", work% = 0
  work%=len(inString$)
  color 15,0
                  REM *** Mark and lower-case the caps
  for     i%=1 to work%
          chk$ = mid$(inString$,i%,1)
          if chk$ < "[" and chk$ > "@" then chk$ = "-*-" + chr$(asc(chk$) + 32)
          neo$=neo$+chk$
  next i%
  inString$ = neo$ + "-"
                  REM *** Restore data for conversion
  restore
  dat$ = ""
  if source$ = "Pinyin" then
        while dat$ < > "pinyin" : read dat$ : wend
    end if
                  REM *** Divide and convert
  Data1$ = "" : neo$ = ""
  while Data1$ < > "done"
        read Data1$, Data2$
        if instr(inStrings$,Data1$) < >0 then
```

```
            work%=len(inString$)
            py%=len(Data1$)
            for i% = 1 to work%
                if mid$(inString$,i%,py%)<>Data1$ then
                neo$ = neo$ + mid$(inString$,i%,1)
                else
                neo$ = neo$ + Data2$
                i%=i%+py%-1
            end if
        next i%
        inString$ = neo$
        neo$ = ""
    end if
  wend
                    REM *** Convert "v"
  while instr(inString$,"v")<>0
      z% = instr(inString$,"v")
      if mode=-1 or instr(inString$,"lv")<>0 or instr(inString$,"nv")<>0 then
          inString$ = left$(inString$,z%-1) + chr$(129) + mid$(inString$,z%+1)
        else
          inString$ = left$(inString$,z%-1)+"u"+mid$(inString$,z%+1)
      end if
  wend
                    REM *** Recapitalize
  while instr(inString$,"*")<>0
      z% = instr(inString$,"*")
      chk$ = mid$(inString$,z% + 2)
      inString$ = left$(inString$,z%-1) + chr$(asc(chk$)-32) + mid$(inString$,z% + 3)
  wend
                    REM *** Clean off extra hyphens and apostrophes
  while mid$(inString$,len(inString$),1) = "-"
      inString$ = left$(inString$,len(inString$)-1)
  wend
  while (left$(inString$,1) = "-" or left$(inString$,1) = " ' ")
      inString$ = mid$(inString$,2)
  wend
  FN Convert$ = inString$
END DEF
PresentResult:
  color 15,1
  work%=len(work$)
  margin%=6
  column%= (80-work%)/2 - margin%
  border$ = ""
  for i = 1 to work% +(2*margin%)
     border$=border$+chr$(205)
  next
  locate 12, column%
  print chr$(201);border$;chr$(187)
  locate 13, column%: print chr$(186)
  locate 13, column%+work%+2*margin%+1: print chr$(186)
  locate 14, column%
  print chr$(186);"        ";work$;"        ";chr$(186)
  locate 15, column%: print chr$(186)
  locate 15, column%+work%+2*margin%+1: print chr$(186)
  locate 16, column%
  print chr$(200);border$;chr$(188)
  color 15,0
  locate 12, (80-len(target$))/2
  print " ";target$; " "
```

```
    if lencheck%=1 then
        locate 16, column%+5
        print " Input phrase too long, only "; longend%;
        print " characters were converted "
        lencheck% = 0
    end if
    gosub PauseKey
    goto StartingGate
return
FrontStatement:
    cls
    color 1
    locate 1,1
    print longline$
    dim t$(20)
    for i% =0 to 19: t$(i%) = "": next i%
    t$(3) = "Pinyin and Wade-Giles"
    t$(4) = "A program for interconversion and word division"
    t$(6) = "*"
    t$(8) = "Written by Karl K. Lo"
    t$(9) = "Library, University of California, San Diego"
    t$(10) = "Revised, May 1991"
    t$(12) = "---------------------"
    t$(14) = "The writer, who reserves the copyright of this program and"
    t$(15) = "appreciates any comments users may have, does not warranty"
    t$(16) = "the program's performance nor promise any further technical"
    t$(17) = "support for its use."
    color 15,0
    for i%=1 to 17
        locate i%,(79-len(t$(i%)))/2
        print t$(i%)
    next i%
    gosub frame
    gosub PauseKey
    cls
return
SwitchMode:
        mode = mode*(-1)
        if mode = 1 then
            source$ = "Wade-Giles"
            target$ = "Pinyin"
            else
            source$ = "Pinyin"
            target$ = "Wade-Giles"
    end if
    work$ = ""
    goto StartingGate
return
InputScreen:
    locate 4,1
    color 12
    print longdark$
    color 9
    print longline$
    color 15,0
    locate 5,28
    print " ";source$; " to"; target$;" "
    locate 10,1
    for i=1 to 7: print longblank$; : next
    gosub frame
```

```
locate 21,1
print longblank$
locate 21,4
print " Type words in "; source$; "; OR "; chr$(250); chr$(250);
print " <X> "; target$; " input ";chr$(250); chr$(250);
print " <H> Help "; chr$(250); chr$(250);
print " <Q> Quit "
locate 10,1
for i = 1 to 7: print longblank$; : next
locate 10,6
print source$; ":"
work$ = ""
while work$ = ""
 locate 10,19
 line input , work$
wend
if len(work$) >60 then
      work$ = left$(work$,60)
      lencheck% = 1
      for longend%=60 to 50 step -1
         if right$(work$,longend%) = " " then exit for
      next
end if
return
HelpScreen:
  cls
  locate 2,1
  print longline$
  locate 2,37
 print " HELP "
 Print
 Print "This program follows the conventions used by the Library of"
 Print " Congress for Wade-Giles, and the Han-Ying Cidian for pinyin."
 Print " In addition, please also note that"
 print
 print" (1) Words may be entered in full lower case, such as xie wanying;"
 print "          or capitalized, such as Xie, Wanying; but not full-caps."
 print
 print " (2) Each <ü> should be substituted with a <v> in data input."
 print
 print " Some sample strings in ";source$; " to convert:"
 print
 if mode = -1 then
      print "              Cao, Xueqin."
      print "              Zhongguo Yanjiuyuan."
   else
      print "              Yen-chiu-yvan."
      print "              Kuo min so te."
 end if
  locate 5,28
  color 1,7
  print "Han-Ying cidian"
  color 15,0
  gosub frame
  gosub PauseKey
  cls
  goto StartingGate
return
Choices:
 While left$(work$,1) = " "
```

```
        work$ = right$(work$,len(work$)-1)
    wend
    While right$(work$,1) = " "
        work$ = left$(work$,len(work$)-1)
    wend
    if work$ = "x" or work$ = "X" then gosub SwitchMode
    if work$ = "H" or work$ = "h" then gosub HelpScreen
    IF WORK$ = "Q" OR WORK$ = "q" THEN
        CLS
        END
    END IF
  return
PauseKey:
    locate 21,1
    print longblank$
    LOCATE 21,27
    PRINT " PRESS ANY KEY TO CONTINUE "
    while inkey$ = "": wend
  return
  frame:
      color 9
      locate 20,1
      PRINT longline$;
      print
      color 12
      print longdark$;
      color 15,0
      locate 23,77
      Print "KL"
  return
  data "--*", "!-*", "-*-", "* ", " .", "-.", " ,", "-,", " ", "- -"
  data ch'i, qi, ch'v, qv, ch', CH, ts', C, ts, z
  data j, r, chih, zhi, chi, ji, chv, jv
  data ssu, si, tzu, zi, tz'u, ci, ung, ong, ien, ian
  data hs, x, -ho-, -he-, ch, zh
  data t, d, p, b, k, g, -i, -yi, yen, yan, yu-, you-, yv, yu
  data d', t, b', p, g', k, CH, ch, C, c, qih, chi
  data h-, "-", -o-, -e-, ho-, huo-, co-, cuo-, do-, duo-, go-, ge-
  data ko-, ke-, no-, nuo-, ro-, ruo-, so-, suo-, to-, tuo-, zo-, zuo-
  data n-g, "n'g", "n'gu", ngu, -a, " 'a", -e, " 'e", -o, " 'o", -na, " 'na"
  data -ne, " 'ne", n'gon, ngon, -hu'n, -hun, -he'n, -hen, fu'n, fun
  data wu'n, wun, "ai' ", ai, "ei' ", ei, "ui' ", ui, "ou' ", ou
  data n'n, nn, g'n, gn, a'a, aa, e'e, ee, "o' ", o
  data "-", "", " ' ", " ", "!", "-", done, pinyin
                    REM *** Divide the non-convertibles:
  Data " ", "-", " ' ", "-", " ,", "-,", " .", "-."
  Data ong, ung-, nu, -nu, ni, -ni, nn, n-n, f, -f
  Data fu, fu-, l, -l, m, -m, s, -s, w, -w, wo, wo-, wu, wu-, y, -y, aa, a-a
  Data ae, a-e, ai, ai-, ea, e-a, ee, e-e, ei, ei-, eo, e-o, eru, e-ru
  Data iu, iu-, io, i-o, oa, o-a, oe, o-e, oo, o-o, on, o-n, ou, ou-, ui, ui-
                    REM *** Mask the multiple letter initials and finals:
  Data ch, -CH, sh, -SH, zh, -ZH, ng, NG, er, ER, yin, YIN
  Data qu, qv, ju, jv, xu, xv, yu, yv
                    REM *** Divide the convertibles:
  Data h, -h, t, -t', k, -k', g, -k, p, -p', b, -p, po, po-, pu, pu-, d, -t
  Data x, -hs, c, -ts', z, -ts, tsi, tzu-, ts'i, tz'u-, j, -J, q, -ch'
  Data r, -j, ji, -jih-, ien-, i-en-, ian, ien-, CHi-, ch'ih-
                    REM *** Unmask the multiples:
  Data CH, ch', SH, sh, ZH, ch, NG, ng, ER, er, YIN, yin
                    REM *** Adjust the irregulars:
```

```
Data gn, g-n, gu, -ku, geng, -keng, -ker, -ke-j, eri, e-jih-
Data eru, e-ju, ern, -erh-n, ie-, ieh-, ve-, veh-, hi-, hih-
Data uo-, o-, ko-, kuo-, k'o-, k'uo-, sho-, shuo-, -ho-, -huo-
Data -e-, -o-, ke-, ko-, k'e-, k'o-, -si, -ssu-, -he-, -ho-, er-, -erh-
Data r-, rh-, you, yu, ye, yeh-, yan-, yen-, yi-, i-, -ng, ng-, J, ch
        REM *** Clean up:
Data "- -", " ", " -", " ", "--", "-", "- ", " ", " -". " ". "- ", "-"
Data "-,", ",", "--", "-", "-.", " .", done, done
```

*Turbo Basic is a registered trademark of Borland International, Inc.

From Bibliographic Networks to Information Networks: The Networks for Networkers II Conference

Dennis Reynolds

The Networks for Networkers II Conference met at the Dulles Airport Marriot Hotel just outside Washington, D.C., on December 17–19, 1990. Sponsored primarily as a Library of Congress Network Advisory Committee (NAC) function, there were just over sixty participants and speakers attending what was billed as a networking pre-conference to the Second White House Conference on Library and Information Services (WHCLIS2).

The agenda consisted of six sessions and ten speakers. The first session met on the evening of December 17 and, after a welcome by Henriette Avram from the Library of Congress, featured overview papers by Barbara Markuson of INCOLSA and Kenneth King of EDUCOM.

Each of the remaining four sessions over the next day and a half focused on various aspects of information networking, including the technology, national information policy issues, governance and funding, and users and services. Each session featured two speakers and two reactors, followed by participant discussion periods during which the attendees broke into four groups. The charge to each discussion group was to draft two potential WHCLIS2 resolutions on the theme of networking of the session just completed, and to identify at least three "action items" representing possible activities for NAC or some other group to undertake in support of enhancing networking in American libraries. Throughout the conference, a group of three facilitators compared the various sets of potential resolutions, eliminated duplicates, and combined similar ones until a manageable number of resolutions emerged, which were then discussed by the participants as a whole as the final function of the meeting.

This article is based in part on a report submitted by the author to the LITA Board in January 1991. The opinions expressed in this paper are those of the author.

All in all, the conference was a success. The speakers were obviously selected with a great deal of thought and planning, and their presentations set the tone and direction for the conference. The titles of the papers and their authors were as follows:

Session I—General Welcome

Henriette D. Avram, Library of Congress

Session II—Overview and Critical Issues for the Future

Barbara Markuson, INCOLSA, "Historical Overview of Library Networking"

Kenneth M. King, EDUCOM, "Network Development Toward NREN"

Session III—Networking Technology

Clifford Lynch, University of California, "Strategic Planning for Library and Information Networking Technology: The Next Decade"

Douglas Van Houweling, University of Michigan, "Gigabit Networks: Development of NREN and the Potential of Research, Education, and Library and Information Networking"

Session IV—Network Policy: Goals for National Information Networking

Peter Lyman, University of Southern California, "National Information Policy Issues Overview"

Fred W. Weingarten, Computing Research Associates, "National Information Issues and the Formulation of Federal Information Policies"

Session V—Governance and Funding

Howard R. McGinn, North Carolina State Library, "Governance and Funding: State and Local Infrastructure"

Lewis Branscomb and Brian Kahin, Harvard University, "Governance and Funding: the National Infrastructure"

Session VI—Users and Services

Kenneth E. Dowlin, San Francisco Public Library "Users and Services: Citizens Information for a New Century"

Nina W. Matheson, Johns Hopkins University "Users and Services: Scholarly and Research Goals for a New Century"

Most of the papers were distributed prior to the conference. This was extremely beneficial. It allowed speakers to highlight or expand upon what they felt to be the key points of their papers rather than simply reading them aloud in full. In addition, it provided all attendees an opportunity to read and reflect on the papers ahead of time. The papers were informative and thought-provoking. The eight reactors to the papers also added valuable insight into the themes addressed at each of the sessions. The discussion periods saw spirited conversation and an overall sense of working together to explore some important issues that libraries and other institutions will face in the years ahead.

FROM BIBLIOGRAPHIC NETWORKS TO INFORMATION NETWORKS

This conference was held eleven years after the first Networks for Networkers, which took place in Indianapolis in 1979. When comparing the context leading up to that first conference with the proceedings of the 1990 conference, there is an alluring similarity between the scale of networking on which both focused. Certain exceptions notwithstanding, much of the discussion of networking in librarianship during the 1970s concentrated on developments and activities on a national scale. Likewise, the presentations and discussions at the 1990 conference centered principally on the promises and prospects of networking of national proportion.

On first impression, this continuity in orientation between the 1979 and the 1990 conferences might seem ironic, given that the most important strides in library networking during the years between the two conferences occurred *not* at the national level, but rather at the local, state, and regional levels. When all is said and done, the most exciting and substantive developments in library networking during the 1980s were initiatives undertaken by such organizations as ILLINET in Illinois, the University of California Division of Library Automation, CARL in Colorado, the Florida Center for Library Automation, and others. Cooperative ventures like these defined the leading context for library network development during the 1980s.

What at first glance may seem like a conference agenda out of step with a still prominent phase in library networking was actually a response to an important shift in the environment in which networking developments affecting libraries may occur during the 1990s. In the 1970s and 1980s, library networking was essentially synonymous with bibliographic networking. The most significant thing about Networks for Networkers II is that with an audience composed of people very comfortable in that milieu, the topic of bibliographic networking was conspicuous by its nearly complete absence, both in the papers and in the discussion sessions. Instead, the focus was on potential developments in broader *information* networks and the possible role of libraries therein. Since the conference was held on the heels of legislation introduced in the last session of Congress and reintroduced in the current session to fund a National Research and Education Network (NREN) initiative, it was little surprise that the NREN—either as an entity in itself or as a euphemism for a national information network by some other name or names—emerged as the center of attention.

By focusing on the NREN, it was by no means implied that bibliographic networking is passé as an issue of importance in librarianship. But what was brought home so well at this conference is that there is a very large world out there involved in the communication and exchange, through electronic means, of many of the types of information to which libraries have historically provided access, and that if libraries are to have a stake in an evolving national electronic information infrastructure, their involvement will have to take on a character much broader than just the transmission and display of bibliographic citations and library holdings symbols. The planners of the 1990 conference are to be commended for recognizing this and extending the focus of the conference beyond the traditional bounds of bibliographic networking, placing it in the context of much broader emerging networks for the communication and exchange of scientific, research, and other types of information.

BE IT RESOLVED . . .

Networks for Networkers II provided a preview of what may become some of the more important developments and issues *vis-á-vis* libraries and networks in the years ahead, and this insight may come to be recognized as the most important long-term contribution of the conference. There was also more immediate business at hand, however, in the form of the charge to draft resolutions to be submitted to the Second White House Conference on Library and Information Services office.

The format for drafting potential resolutions—that is, with each of four discussion groups submitting two resolutions after each of four sessions—resulted in a total of thirty-two draft statements. After comparing, consolidating, and editing the draft resolutions submitted by the four groups, the conference facilitators presented seven resolutions for discussion by the attendees meeting as a whole. In summary form, the resolutions addressed the following issues and made recommendations pertaining thereto:

- Draft Resolution 1 urged WHCLIS2 to support the concept of the evolving National Research and Education Network (NREN), and made recommendations regarding breadth of participation, the governance structure, funding, and the applicability of network standards and protocols.
- Draft Resolution 2 urged that NREN policy specifically enable public and school libraries to participate as access points to the network for use by the general public, and that there be equal opportunity of access for all regardless of geographic location or socioeconomic status.
- Draft Resolution 3 urged a reaffirmation by government, in an age of electronic dissemination of information, to the principles of protecting privacy and freedom of expression, and also urging a definition of the extent of the role of the public and private sectors in national information infrastructures.
- Draft Resolution 4 urged that public information developed and maintained by government in electronic formats be made accessible to the public, and that to accommodate the dissemination of information in electronic formats, the functions and services of the Depository Library Program be continued and strengthened.
- Draft Resolution 5 urged recognition of the importance of electronic information literacy among the citizenry, and recommended that librarians play an important role in training and educating users in the effective use of information networks and systems.
- Draft Resolution 6 addressed the importance of research as a means through which to better understand and develop information networks, and recommended funding to support such research undertaken collaboratively by the various disciplines, including library and information science.
- Draft Resolution 7 recommended the formation of partnerships among library and information centers with public and private service providers, and that librarians and other information professionals

take a leadership role in ensuring broad public access to networked information and services.

During the final editing that took place after the conference, part of the original Draft Resolution 7 was removed and formulated into a separate Resolution 8. This resolution urged the WHCLIS2 to reaffirm the historic and current federal roles supporting library networking and resource sharing through federal programs, agencies, and funding. A ninth resolution was also added, which urged adherence to standards and protocols in library networking and recommended further development of applicable standards and protocols.

The process of drafting the resolutions at the conference was an interesting one. The facilitators who worked on the resolutions during the proceedings were essentially aiming at a moving target, trying to capture the commonalities of the various groups' statements while at the same time recognizing subtle but sometimes important distinctions. The resolution editors were also confronted by the onerous task of translating some otherwise fairly lucid observations and recommendations into "resolutionese," that foreboding sphere from which little that enters ever emerges nearly as clearly as it went in.

Fortunately, the conference schedule included some time at the end for attendees to meet as a whole and discuss the preliminary versions of the draft resolutions. Some of the rough spots were ironed out, passages more appropriate as professional action items than as statements for the broader audience of the WHCLIS2 were expurgated, and some of the redundancies were consolidated. The conference adjourned with attendees realizing that some fine-tuning of the resolutions would be necessary. It was clear that their discussion had been marked by the same spirit of enthusiasm and cooperation that characterized the conference as a whole. A list of some sixty potential "action items" that came out of the group discussions was also distributed at the conference, but there was insufficient time to examine the items in any detail. Participants were asked to rate each on a priority scale and return the list to the NAC office for tabulation.

THE FUTURE: "LIBRARIES AND NETWORKS" VERSUS "LIBRARY NETWORKS"

Even by December, when Networks for Networkers II took place, there reportedly had already been more than 2,600 resolutions submitted to the WHCLIS2 office by state pre-conferences, library organizations, and other groups. It is unlikely that any of the nine resolutions submitted by Networks for Networks II will make it to the White House Conference in their exact form. It can be expected, however, that the substance of the resolutions will closely enough mirror, and be mirrored by, resolutions submitted by other groups. This will ensure that crucial initiatives like the NREN and critical issues centering on information disseminated and accessed electronically will, in one form or another, be brought to the attention of the WHCLIS2 delegates with the importance they deserve. For this reason, as well as for purposes of professional discussion among those involved in networking, the drafting of resolutions by the conference was a worthwhile endeavor, and one that should serve as a benchmark against which to measure developments in the decade ahead.

As mentioned earlier, the most significant outcome of Networks for Networkers II will not be the resolutions drafted, but rather the tone set by the conference for defining the context for the relationship between libraries and networks in the 1990s. In addition, more than a semantic distinction, the focus indeed is likely to move from "library networks" to "libraries and networks." The former alludes to activities carried out by libraries among themselves; the latter acknowledges the likelihood that the networks, which may come to be of such importance to libraries in the 1990s, will be broader than that in terms of their scope and the participants involved.

When viewed in retrospect several years hence, Networks for Networkers II will likely be seen as being on one end or the other of a spectrum. If networks on the scale of NREN prove impractical or otherwise do not come to fruition, or if libraries are unable to effectively tap into the mainstream of such developments, this conference may come to be seen as a curious historical blip reflecting some rather naive notions about the promise and prospects of electronic information on a societal scale or about the role of libraries therein. If, on the other hand, events turn out quite differently, the conference could indeed come to be viewed as something of a landmark, helping to redirect a focus on library networking that may have been appropriate for the 1970s and 1980s but that will be far too narrow for the 1990s. Either way, the planners of Networks for Networkers II are to be commended for their foresight in putting together a program that challenged participants to reflect on the changes surrounding them, and to think about how libraries and library networking organizations will fit into the possible scenarios of the information networks of the 1990s. ■■

Recent Publications

Book Reviews

Computers in Libraries '91. Proceedings of the Sixth Annual Computers in Libraries Conference, March 10–13, 1991. Ed. by Nancy Melin Nelson. Westport, Conn: Meckler, 1991. 210p. paper $40 (ISBN 0-88736-753-4).

Getting speakers to prepare executive summaries of their presentations and publishing them ahead of time is no small accomplishment for any conference chairperson. Nancy Melin Nelson, conference chair and editor for these proceedings, is to be commended for managing just such a feat. *Computers in Libraries '91* is a compendium of papers ranging from brief annotations of paper topics to nearly complete texts of presentations given at the sixth annual Computers in Libraries (CIL) Conference, held at the Hyatt Regency, Oakland, California, March 10-13, 1991. Speakers were instructed to prepare executive summaries for publication, and most fall into that category in length and detail.

Papers are available for most presentations which, the editor notes, include a number of topics new to the annual CIL meetings. Among the new sessions covered most comprehensively in the proceedings are integrated online public access catalogs, electronic publishing, and strategic planning, all of which received attention in several conference sessions. Other session topics covered electronic networks, word-processing, artificial intelligence, staff training, and bibliographic instruction using hypercard technology. Preconference publication of these proceedings brings with it some drawbacks in terms of coverage and consistency, and papers are absent for some potentially intriguing sessions. The articles vary widely in writing style and grammatical and typographical consistency. It is apparent that there was minimal time to edit contributions prior to publication.

Papers are organized by author's name and an Index to Papers by Session is provided to give access to the topics of greatest interest. The session topics offer some information on research value and technical level of the intended audience.

Despite the gaps and inconsistencies, *Computers in Libraries '91* serves an important function beyond being a memento of the conference for attendees. Computers in Libraries' annual conferences have grown to be among the largest and best organized forums on library automation, and the proceedings are a "hot off the press" reflection of these meetings.

While papers from academic librarians predominate, there is good representation, in terms of presenters and paper topics, from public, special, network, and vendor sectors. The presentations and their executive summaries vary widely in level of technical detail, many appealing to the beginner in information technology, and many of greater interest to researchers in the state of the art.

In short, there is something for everyone in *Computers in Libraries '91*. These proceedings are recommended for all who attempt to stay current with library technology "state of the art," and for medium to large library information technology collections.—*Elizabeth D. Nichols, Stockton-San Joaquin County Public Library.* ■■

Coping with Information Illiteracy: Bibliographic Instruction for the Information Age. Papers presented at the Seventeenth National LOEX Library Instruction Conference held in Ann Arbor, Michigan, May 4–5, 1989. The Library Orientation Series, no. 20. Ed. by Glenn E. Mensching and Teresa B. Mensching. Ann Arbor, Mich.: Pierian, 1990. 202p. paper, $35 (ISBN 0-87650-267-2).

This volume contains the proceedings of the Seventeenth National LOEX Library Instruction Conference held in Ann Arbor, Michigan, at the Ann Arbor Inn on May 4 and 5, 1989. Six major papers are presented including an inspiring one by the keynote speaker, Patricia Senn Breivik, challenging academic librarians to focus on information literacy as a comprehensive educational issue and to be proactive in bringing about changes in higher education at their institutions.

Other major papers discuss important themes in bibliographic instruction (BI) generated by the rapid expansion of technology accompanied by the information explosion of the past decade. Jan Kennedy Olsen and Bill Coons describe the Mann Library's Information Literacy Program at Cornell University which began in 1986 and serves as a prototype of the integration of information literacy in a selected department's undergraduate curriculum. Harold Tuckett's paper examines the concepts of computer literacy and information literacy and argues for broadening the role of the bibliographic instruction librarian. Trish Ridgeway suggests several methods for integrating active learning techniques into a one-hour BI lecture, and accompanies her paper with useful worksheets. Sandra Yee's work with minority students prompted her paper concerning culturally based learning styles and information literacy skills—how to locate, analyze, evaluate, and apply information. The final major paper is by Rebecca Jackson, who describes the University of Maryland's experience in adapting the ACRL/Bibliographic Instruction Section's *Model Statement of Objectives for Academic Bibliographic Instruction* and in redesigning their library instruction program based on this statement.

The conference included a symposium on the dichotomy surrounding instruction program goal limits, that is, short-term specific goals versus long-term lifelong learning goals. Richard Feinberg took the short-term stance, Carolyn Dusenbury advocated lifelong learning as the ultimate goal, and Carolyn Kirkendall moderated. Panel papers are included.

Short papers from the four instructional methods sessions are presented in their entirety for a change in conference reportage. The editors of the proceedings present this caveat in the preface: "Spoken presentations will always vary in readability." This applies to these papers as well as others in the volume.

Poster session abstracts and discussion group handouts are included. The volume concludes with Hannelore Rader's well-done annotated bibliography of library orientation and instruction materials published in English in 1988, and a list of conference participants.

Although the conference took place two years ago, the papers are as timely and thought-provoking today as then, proving once again the value of making these proceedings available to all. Information literacy has been chosen as the theme for several programs, including the ACRL President's Program at the Atlanta ALA Annual Meeting in June 1991. These papers should provide excellent background reading on this topic.—*Claudette S. Hagle, University of Dallas.* ■■

Dewey, Patrick R. *National Directory of Bulletin Board Systems.* Westport, Conn.: Meckler, 1990. 114p. paper, $39.95 (ISBN 0-8736-554-X).

Patrick R. Dewey is a well-known and knowledgeable writer on electronic bulletin board systems (EBBS) and other technologies. His *Essential Guide to Bulletin Board Systems* (Meckler, 1987) is still in print. This title, *National Directory of Bulletin Board Systems* (1990), is the third edition of a Meckler directory that began in 1986–87. The fourth, updated edition is scheduled to appear later this month.

The main body of this 114-page book is an area code listing of thousands of EBBS from every state and some parts of Canada and abroad. The information is presented in three columns: telephone number, EBBS name, and City/State. Appendices (Chicago area systems, major public systems, UNIX systems, and IBM PC boards) are considerably more descriptive. Additional information on the EBBS software, number of lines, hours of operation, subject orientation, and baud rate are often included.

However, the book is of limited use. The criteria of currency, completeness, and cost are not met. EBBS are notorious for their short lives. Most EBBS system operators of long standing state that the attrition rate in one year's time is around 70 to 80 percent. Dewey's primary sources are the bulletin boards themselves and vendors. Many electronic bulletin board systems are shareware systems operated by teenagers and last only a few months. It appears that many or most of the listings are from 1989, making much of the list out-of-date at publication.

The list is also far from complete. Wichita, Kansas (area code 316) is listed with nine EBBS, yet there are thirty to thirty-five at any one time in that city. Three of the best, long-standing systems are not listed. In addition, the listings are only a fraction of the total available in the United States and may miss the more stable, reliable, and fully developed electronic bulletin board systems in a state or region.

One can often log onto a local CompuServe network, request the number of a desired city's or region's CompuServe service, connect and download current, correct, and complete information of bulletin boards carried on it for cheaper and more reliable listings.

The directory would also be more useful if it consistently listed the software used, the date of establishment, and the number of incoming lines supported. A reader could dial-up with much more confidence if the EBBS has been ongoing for several years, has more than one line, and is known to be using more substantial EBBS software, such as TBBS (which costs between $300 and $1,200 and not the inexpensive or free downloaded EBBS software). The subject orientation (games, sports, women's issues, adult) of the EBBS should also be included in a useful directory.—*Gregory Zuck, Southwestern College, Winfield, Kansas.* ■■

Dewey, Patrick R. *Public Access Microcomputers: A Handbook for Librarians*, 2d ed. Boston: Hall, 1990. 155p. paper, $35 (ISBN 0-81611-896-5).

The second edition of Patrick R. Dewey's book provides clear and useful information for librarians who are either interested in starting a public access microcomputer project or who want to expand, improve, change the focus of, and/or evaluate current projects. The emphasis of the book is on public library programs designed for adults, teenagers, and children.

One of the strengths of this book is the concrete examples of both successful projects and of problems faced by libraries. The author gives examples of microcomputer projects from as far back as 1977. He includes such areas of public access as electronic bulletin boards, circulation of hardware and software, online services, desktop publishing, and computer-assisted instruction. He notes important questions to consider before undertaking a project. Some questions that seem so obvious as to be overlooked are actually extremely important (for example, how many people are likely to use the computers, how many patrons already own home computers). Throughout the book, the author stresses the importance of considering the goals and objectives of the library and how the microcomputer will fit into the library's mission.

In his chapter on selecting hardware and software Mr. Dewey emphasizes that, while hardware is important, the selection of software and the organization of the project are of utmost importance. He reviews purchasing guidelines for PC components and peripheral devices, as well as security and maintenance issues, and general care of the equipment.

Mr. Dewey notes the importance of offering users a variety of different kinds of software to keep the project exciting. Since computers can make learning fun, educational software should be a prime consideration. The author includes a sample survey to be used to find out who is likely to use the equipment, a software evaluation form, and a sample software collection of one hundred programs.

One chapter covers issues related to circulating hardware and software including a brief discussion of copyright, license agreement, shareware, and public domain software with examples of how specific libraries face these issues. The author includes chapters on running a bulletin board service (BBS) and other online services; on library programs using the microcomputer; on managing a microcomputer center, where he points out that training staff and patrons may be biggest challenge; and a chapter giving examples of public access microcomputer projects throughout the country.

The unique considerations of academic and school libraries are briefly touched upon, as are wall charts and desktop publishing. Appendices give model project proposals, software sources, computer journals, and bulletin board software vendors. A comprehensive bibliography is included.

This book's strength lies in the fact that it is put together in such a way that the reader can select the parts that meet his or her needs. *Public Access Microcomputers* can easily serve as a checklist for starting a new project or evaluating an on-going one. It can also be used to strengthen aspects of a current project and improve one that seems unsuccessful. The concrete examples supplied throughout the book give the reader many ideas that can be adapted to different library situations. This book is recommended for libraries that have, or are considering offering public access microcomputers.—*Bonnie Birman, New York Public Library.* ■■

Olson, Nancy B. *Cataloging Microcomputer Software: A Manual to Accompany AACR2 Chapter 9, Computer Files.* Englewood, Colo.: Libraries Unlimited, 1989. 267p. paper, $33 (ISBN 0-87287-513-X).

Olson's manual is a thorough and helpful guide through the often confusing process of cataloging software for microcomputers. The book is large (28cm) with type and spacing conducive to easy reading. The extensive section of 100 examples illustrates the rules, decision-making processes, and standards covered throughout the text. Included in the examples are copies of actual information appearing on disks, labels, containers, and/or accompanying material. Samples of descriptive cataloging according to *AACR2* Chapter 9, including suggested call numbers and subject headings, are also illustrated. Some OCLC MARC tagged examples accompany selected titles.

The practical questions and suggestions in chapter 2, "Decisions To Be Made Before Cataloging," are excellent reminders of issues that may appear obvious but are sometimes overlooked. Will they circulate, will they be available for interlibrary loan, and where are labels placed are some of the questions answered. Suggestions include tips for handling the disks, procedures for damage repair, and a discussion indicating how loss of data may occur when certain theft security systems are used.

Particularly helpful are the chapters dealing with "Descriptive Cataloging," using and interpreting *AACR2*, chapter 9, and "Choice and Form of Access Points." The rule numbers are cited following the arrangement of *AACR2*, making referral from the manual to the book easy. The comments are useful and informative. Slight discrepancies from the draft version of *AACR2*, 1988 revision, and the actual publication were observed. Most changes are minor and will not limit the usefulness of this manual. Also, some new rules are included in the 1988 revision, which this book does not address. One notable printing error occurs in the *AACR2* section 9.6A1 in which the instructions are, "Precede the ISBN of a series or subseries by a comma." "ISBN" should read "ISSN." The sections of *AACR2* dealing with determining access points are covered in the same manner as the descriptive cataloging. Appropriate rules are listed with useful comments. Lists of computer program names, microcomputers, operating systems, and languages are included that will be helpful in choosing specific entries.

Olson encourages use of national standards and recommendations of national committees in making local decisions for subject access and selecting a classification scheme. She provides lists of selected headings from *Library of Congress Subject Headings* and *Sears List of Subject Headings*. To aid in assigning classification numbers, sections of Library of Congress and Dewey Decimal classification schedules are also included. She concludes the instructional part of the book with information concerning the evolution of MARC tagging of bibliographic records. Although interesting, the history of microcomputer software cataloging and bibliography is a distraction from the focus of the manual.

Many other companion manuals to *AACR2* for cataloging various formats have been published. Olson's is one of the most practical available. Updating chapter 3, "Descriptive Cataloging," to coincide with the small changes in *AACR2*, 1988 revision, would enhance its usefulness. Both veteran and beginning catalogers will find this manual helpful.—*Sylvia C. Cornell, Jacksonville Public Libraries, Florida.* ■■

Rewritable Optical Storage Technology. Ed. by Judith Paris Roth. Westport, Conn.: Meckler, 1990. 172p. $45 (ISBN 0-88736-534-5).

This book focuses on magneto-optical (M-O) disk storage systems, which use a combination of magnetism and laser heat to record or erase data. Two other forms of rewritable optical storage technology—dye-polymer and phase-change—are briefly discussed. However, at the time of writing (mid-1990), all rewritable optical disks and drives available commercially used M-O technology. The most well-known example may be the optical drive in the NeXT computer.

The seven chapters in this slim volume present a balance between the present status of the technology and future possibilities, between technical descriptions of the media and practical applications of rewritable optical storage, between industry expectations and actual experiences with system implementation. The various authors express much optimism in the future of the technology and they clearly describe the role of such "direct access secondary storage" devices in handling information needs today. The performance and cost of rewritable optical disks place them between high performance fixed disks (primary storage) and sequential tape systems (secondary storage).

M-O storage systems share several characteristics with fixed magnetic disk drives. Both are mass storage systems, offer random access to their contents, and permit intentional erasure and rewriting of data. Furthermore, digital information is stored in the form of magnetic flux directions, as with standard magnetic storage, rather than the pits and lands of CD-ROMs. Rewritable optical disks not only hold enormous amounts of data, but are also relatively inexpensive, durable, and transportable. They are highly suited for graphic- or image-based tasks such as computer-aided design and engineering in work group environments.

However, barriers remain in the way of wider adoption of this technology and the book does not fail to address these issues. For example, data interchange standards must be developed and adopted so that M-O disks can operate on a variety of hardware platforms, computer operating systems need to take into account the two-step process of erasure and rewriting inherent in M-O technology to ensure proper file management, and prices of rewritable optical drives have to drop considerably before widespread use can be feasible.

This reviewer found that the case studies at the end of the book brought to life some of the performance issues discussed in earlier chapters. Also particularly helpful were the comparisons made between rewritable optical disks and other optical storage systems, write-once (or WORM), and read-only (or CD-ROM). The editor has provided an annotated list of organizations and institutions involved in the development and application of the technology. The appendices include a reading list and a short but useful glossary of terms and acronyms.—*Ka-Neng Au, Dana Library, Rutgers University, Newark, New Jersey.* ■■

Teaching Technologies in Libraries: A Practical Guide. By Linda Brew MacDonald and others. Boston: Hall, 1991. 275p. $34.95 (ISBN 0-8161-1906-6).

Utilizing new technology to teach new technologies in libraries is the theme of this guide, written by four practitioners at the University of Vermont. As the Introduction states, the chapters in the book "discuss various teaching technologies, highlighting their particular strengths and unique applications in libraries."

The authors contend that, while the costs of teaching technologies are high and while there are no provable benefits of training, "designing pedagogically sound instructional programs for library technology is nonetheless extremely important for both the library and the library user."

This book is intended to serve as a "basic guide for librarians working with automated resources in all types of libraries," and emphasizes six technologies which can be used for library instruction: OPAC help screens, optical discs, online tutorial services, video, CAI, and expert systems. Each chapter describes the pros and cons of using the technology for instruction and indicates potential difficulties and costs. In many instances, actual case studies are included. Although arguments for and against training are presented, the conclusion strongly states that the reasons for providing training are more numerous and substantial than the arguments against it. The importance of staff training is also repeatedly emphasized.

New approaches to bibliographic instruction are essential because many library users view computers alternatively as "simple mechanical tools, as mysterious answer machines, or as fearsome monsters." The authors argue persuasively that the only way to effectively teach these new technologies is to use the same technologies as part of the instructional process.

One unique chapter includes brief essays, contributed by other libraries, that deal with specific applications of teaching library technologies to public, school, academic, medical, and business libraries. However, the primary emphasis of the book is on academic library applications.

In many cases the suggestions are fairly traditional. For example, the chapter entitled "The Audiovisual Renaissance" does not really introduce any new uses or techniques for audiovisual technologies. In fact, the use of liquid crystal display units with overhead projectors and other types of computer projection devices are barely mentioned.

Two other chapters include good overviews of CAI and expert systems, but they also are fairly traditional in their approach. One of the best chapters offers practical suggestions for developing the "ideal functionality of onscreen help." Suggestions for designing the most useful local OPAC screens deal with online documentation, online tutorials, and other functional specifications of help screens.

While most of the points raised are intuitively known by reference and BI librarians, this book effectively brings them all together in a coherent whole. Consequently, it is recommended as both an introduction to the topic and as a good checklist for any library undertaking a redesign of its instructional program to accommodate and support new technological tools. Given the arguments and evidence presented in this book, all libraries should be reviewing their methods of instruction as more electronic information sources are necessarily introduced into libraries.—*Greg Byerly, Kent State University School of Library and Information Science.* ■■

Tessmer, Martin, David Jonassen, and David C. Caverly. *A Nonprogrammer's Guide to Designing Instruction for Microcomputers*. Englewood, Colo.: Libraries Unlimited, 1989. 228p. paper, $26.50 U.S.: $32 elsewhere (ISBN 0-87287-680-2).

This guide is aptly titled. It is not about programming, scripting, or creating stacks. It is about teaching and learning. The authors state firmly that "CAI (computer assisted instruction) should be developed by teachers and trainers, not by programmers . . . teachers should function as designers and subject matter experts of the program, not as program coders."

The guide does a good job of presenting critical instructional considerations. The difficult dilemma of presenting verbal or textual material is treated well. Also included are chapters that deal with the basic components of "conceptual" learning and "rule" learning, each of which require a different instructional approach. One of the most helpful features of the work is a set of guidelines for determining which type of information is being presented and suggestions for appropriate drill, practice, and feedback.

The authors have carefully avoided slanting the material toward any specific software or hardware environment. The material presented in the guide could as easily be applied to hypertext applications as to fairly complex authoring systems. In fact, the techniques presented could even be applied to the design of classroom lectures and lessons.

The unique interactions between students and computers are largely ignored by this guide. While the limitations of screen display are noted, such machine limitations are not a major focus in the lesson design. Examples are presented in screen-shaped frames, but serious issues of screen design are not addressed. Presumably, this would be left to the programmer.

The discussion of "problem solving behavior" is fascinating and leads the designer through an analysis of how students learn to solve problems. The role of games and simulations is explored, including guidelines for evaluating the effectiveness of games as instructional devices. Unfortunately, this section is less successful in helping a novice CAI designer integrate problem solving into computer assisted lessons.

It is interesting to note that the authors believe computer assisted instruction should not always be designed to be used independently. They suggest that CAI is more effective when integrated into broader educational activities. Well-designed CAI lessons might serve as a supplement to more traditional classroom activities.

This work demands a careful reading. Since the emphasis is on applying theories of learning, the narrative relies heavily on educational jargon. The reader may struggle with unfamiliar terms like "learning outcomes," "knowledge engagement," "non-examples," and "sub-skills."

If your goal is to create a finished CAI application in a specific hardware and software environment, you will need to seek your technical information elsewhere. This guide is intended to help the CAI lesson designer understand and use a variety of instructional techniques. The focus is on the material to be taught and how that material is learned by the student. As a tool for analyzing and designing lessons, the guide is a good purchase.—*Tamara J. Miller, University of Tennessee, Knoxville.* ■■

Other Recent Receipts

Listed here are books and other publications received for review that are of potential interest to LITA members. Some of these materials may be reviewed in later issues of ITAL.

Advances in Library Resource Sharing, V.2. Ed. by Jennifer Cargill and Diane J. Graves. Westport, Conn: Meckler, 1991. $55 (ISBN 0-88736-739-9).

Advances in Librarianship, V.15. Ed. by Irene P. Godden. San Diego, Calif.: Academic, 1991. 293p. $55 (ISBN 0-12-024615-5).

ALA Survey of Librarian Salaries, 1990. Prep. by Mary Jo Lynch, Margaret Myers, and Jeniece Guy. Chicago: American Library Assn., 1990. 55p. paper, $30 (ISBN 0-8389-3385-8).

Bibliographic Access in Europe: First International Conference. Ed. by Lorcan Dempsey. Brookfield, Vt.: Gower, 1990. 320p. Illus. $69.95 (ISBN 0-566-03644-4).

Brandt, D. Scott. *UNIX & Libraries: Supplement to Computers in Libraries Series no.20.* Westport, Conn.: Meckler, 1991. 200p. $47.50 (ISBN 0-88736-541-8).

Bryson, Jo. *Effective Library & Information Centre Management.* Brookfield, Vt.: Gower, 1990. 450p. $58.95 (ISBN 0-566-05637-2).

Carmen, Michael. *A Manual of Library Network Management.* Brookfield, Vt.: Gower, May 1992. 200p. $56.95 (ISBN 0-566-03559-6).

CD-ROM Librarian's Optical Product Reviews, 1987-1990. Ed. by Norman Desmarais. Westport, Conn.: Meckler, 1991. 100p. $34.95 (ISBN 0-88736-733-X).

CD-ROM Local Area Networks: A User's Guide. Supplement to Computers in Libraries, no.24. Westport, Conn.: Meckler, 1991. 175p. $39.50 (ISBN 0-88736-700-3).

CD-ROM Market Place: An International Directory. Westport, Conn.: Meckler, 1990. 150p. $30 (ISBN 0-88736-680-5).

Cummins, Thompson R. *Planning, Measuring & Evaluating Library Services & Facilities.* New York: Neal-Schuman, 1991. 250p. paper. $39.95. (ISBN 1-55570-070-5).

Ensor, Pat. *CD-ROM Research Collections.* Supplement to Computers in Libraries Series, no.38. Westport, Conn.: Meckler, 1991. 200p. $55 (ISBN 0-88736-779-8).

Evans, John R. *Serials Control Systems for Libraries.* Essential Guide to the Library IBM PC Series, V.12. Westport, Conn.: Meckler, 1991. 165p. spiralbound. $34.95 (ISBN 0-88736-186-2).

Eyre, John. *Computers in Libraries International, 1991.* Proceedings of the Fifth Annual Computers in Libraries Conference, February, 1991. Westport, Conn.: Meckler, 1991. 200p. $45 (ISBN 0-88736-754-2).

Fenichel, Carol, and Thomas H. Hogan. *Online Searching: A Primer, 3d ed.* Medford, N.J.: Learned Information, 1990. $16.95 (ISBN 0-938734-30-X).

Fortson, Judith. *Disaster Planning and Recovery: A How-to-Do-It Manual for Librarians and Archivists.* New York: Neal-Schuman, 1991. 150p. $35 (ISBN 1-55570-059-4).

Greene, Robert J. *The Guide to Application Programs in BASIC.* New York: Neal-Schuman, 1991. 250p. $39.95 (ISBN 1-55570-063-2).

Hernon, Peter, and Charles R. McClure. *Evaluation & Library Decision Making.* Norwood, N.J.: Ablex, 1990. 272p. $47.50 (ISBN 0-89391-640-4); $24.50 paper (ISBN 0-89391-686-2).

Kim, David U., and Douglas M. Kim. *Policies of Education Software Publishers: A Guide for Authors.* The Woodlands, Tex.: New Technology Press (P.O. Box 9154, The Woodlands, TX 77387), 1991. 231p. paper, $25.

Kovacs, Beatrice. *The Decision-Making Process for Library Collections: Case Studies in Four Types of Libraries.* Contributions in Librarianship & Information Science Series, no.65. Westport, Conn.: Greenwood, 1990. 208p. $39.95. (ISBN 0-313-26042-7).

Library Hi-Tech Bibliography, V.5. Ed. by. C. Edward Wall. Ann Arbor, Mich.: Pierian, 1990. 213p. paper, $45 (ISBN 0-87650-262-1).

Library LANS: Case Studies in Practice and Application. Ed. by Marshall Breeding. Westport, Conn.: Meckler, 1991. 200p. $42.50 (ISBN 0-88736-786-0).

Library Technical Services: Operations & Management. 2d ed. Ed. by Irene P. Godden. San Diego, Calif.: Academic, 1991. 285p. (ISBN 0-12-287041-7).

Losee, Robert M., Jr. *The Science of Information: Measurement and Applications.* San Diego, Calif.: Academic, 1990. 293p. $29.95 (ISBN 0-12-455771-6).

Machalow, Robert. *Using Microsoft Excel: A How-to-Do-It Manual for Librarians.* New York: Neal-Schuman, 1991. 175p. $35 (ISBN 1-55570-075-6).

Mates, Barbara T. *Library Technology for Visually & Physically Handicapped Patrons.* Westport, Conn.: Meckler, 1991. 175p. $42.50 (ISBN 0-88736-704-6).

Murphy, Catherine. *Automated Systems for School Library Media Centers.* Englewood, Colo.: Libraries Unlimited, 1991. 200p. $27 (ISBN 0-87287-771-X).

Operations Research for Libraries and Information Agencies: Techniques for the Evaluation of Management Decision Alternatives. Ed. by Donald H. Draft and Bert R. Boyse. San Diego, Calif.: Academic, 1991. 188p. $49.95 (ISBN 0-12-424520-X).

Optical Information Systems Internation, 1991. Proceedings of the Eighth Annual Conference on Optical Information Systems, London, 1991. Westport, Conn.: Meckler, 1991. 250p. paper, $45 (ISBN 0-88736-758-5).

Pitkin, Gary M. *Library Systems Migration: Changing Automated Systems in Libraries & Information Systems.* Westport, Conn.: Meckler, 1991. 142p. $42.50 (ISBN 0-88736-738-0).

Rader, Hennelore B., and Trish Rideway. *Information Literacy: A How-to-Do-It Manual for Librarians.* New York: Neal-Schuman, 1991. 150p. $35 (ISBN 1-55570-109-4).

Ranade, Sanjay. *Optical Storage Jukebox Technology: A Systems Integrator's Perspective.* Westport, Conn.: Meckler, 1991. 275p. $49.50 (ISBN 0-88736-571-X).

Williams, Brian. *A Directory of Computer Conferencing for Librarians.* Westport, Conn.: Meckler, 1991. 250p. paper, $59 (ISBN 0-88736-771-2).

Yerkey, A. Neil. *Using dBase: A How-to-Do-It Manual for Microcomputer Information Management in Libraries.* New York: Neal-Schuman, 1990. 150p. $35 (ISBN 1-55570-079-9).

Ziegler, Kurt Jr. *Distributed Computing and the Mainframe: Leveraging Your Investments.* New York: Wiley, 1991. 387p. $44.95 (ISBN 0-471-51753-4). ■■

Software Reviews

Total Library Computerization. On Point, Inc., 2606 36th St., N.W., Washington, D.C. 20007; (202) 338-8914. Price: first module, $550; additional modules, $500; complete, $2,050.

Hardware requirements: IBM PC, XT, AT, or PS/2 or compatible computer with DOS 3.0 or later version, a hard disk, and 640K RAM. Printing requires an IBM or Epson compatible printer.

Nearly six years ago *PC Magazine* surprised librarians by printing an article about bibliographic database software. At the time, relational databases like Ashton-Tate's dBase III and flat file managers like PFS File were popular. Such products and similar software were about the only database programs available to the PC user for bibliographic management. The review of InMagic, Notebook II, and Finder introduced non-librarians to software that could help them with problems that librarians generally consider as problems of bibliographic control—the recording and retrieval of documents.

The problem with these bibliographic managers and other notable products that followed, like Personal Bibliographic Software's ProCite, is that they only attain partial bibliographic control. Bibliographic control roughly means keeping accurate track of what documents you have and where you put them, and some way, presumably easy, of finding documents by subject, title, author, and other descriptive elements. This would be the most elemental bibliographic control.

In order to run a library of any size, however, more control is required, as well as ways of managing what actually happens to the collection when users use it. For example, you need to keep track of who is using documents, what has been ordered, what has been lost, and how much has been spent performing these and other library tasks.

Unfortunately, this more elaborate type of bibliographic control is not a strength of typical relational databases or commercially available bibliographic managers. InMagic or dBase IV, for example, can be programmed to achieve bibliographic control, but this takes skill, time, and labor. In addition, although many bibliographic managers can be tweaked and finessed into doing bibliographic control, it is at the expense of their basic simplicity and ease of use.

Clearly, a desirable bibliographic manager, especially for small libraries working in non-MARC data format, would be a PC-based database program that is reasonably priced, easily learned and used, and that would provide bibliographic control.

TLC OVERVIEW

On Point Inc.'s Total Library Computerization (TLC) is such software. TLC is a full library service (cataloging, borrowing, interlibrary loan, serials control, ordering) integrated library software package developed to meet the needs of small libraries but at a fraction of the cost of software designed for large institutions and considerably more adaptable to library needs than less expensive bibliographic managers or relational databases.

TLC comprises six independent files (modules) that use individual menus to manage cataloging, borrowing, ILL, serials control, ordering, and memos. The cataloging module supports data input for journal articles, books, reports, chapters within books, reprints, congressional publications, serials, and miscellaneous formats. Cataloging templates contain predetermined fields for essential bibliographic information. For instance, the book template's fields include author, editor, title, edition, series, place of publication, publisher, copyright date, keywords, classification number, LC classification number, location in library, cost of book, and record date. ASCII records from other database sources may be imported.

ILL functions include items lent and borrowed, information on cooperating libraries, and return status. The ILL module can print out standard interlibrary loan forms. The serials control module supports importing of serials lists in ASCII format, printing routing slips and labels, serials check-in, binding schedules, serials cancellations and claims, and information about binderies. Vendor information, as well as information on items ordered, can be recorded in the ordering module. Items lent and returned are tracked in the borrowing module. The memos module allows users to record notes about the organization.

Bibliographic records may be searched by author, title, or the records' full text. The full record may then be displayed or printed. All database modules may be searched globally in two different ways. The chief difference between the two methods is the end result. The first method retrieves the actual data record (the complete template) while the second method produces a finished report listing the records that contain the search terms. TLC supports Boolean logic, word truncation, segment searches, field searches, and proximity searches.

TLC supports printing on standard paper and spine labels. For proper formatting, an Epson or IBM printer is required; TLC does not support other printer drivers. Customized printer programs include those that customize incoming and outgoing ILL forms, routing slips, and claim forms.

TLC also supports numerous database and catalog maintenance functions essential to a library system. All modules contain backup and delete functions. Various sorted lists (for example, ILL loans sorted by title) can be produced. Statistics include monies paid to vendors between any dates, costs within any dates subtotaled per vendor, and so on.

But there is even more. Readers interested in this package should contact the vendor to determine all the features. It seems all encompassing; that is, it contains the programs necessary for a tidy library operation. Of course a cataloger could prove me wrong—some important statistical or spinal element may be missing. But in the main it is all there—to buy, find, and circulate books, journals, and other library items.

TLC works smoothly, too. It can be installed in seconds and the code is written economically, taking just over 400,000 bytes. The search engine is the quick and powerful askSam retrieval software. Help, both online and written, is straightforward. At critical points, functions are passworded for database protection. At the same time, the borrowing function is so easy that, if desired, borrowers can record their own transactions.

SUMMARY

In many ways, TLC allows you to create a do-it-yourself library. Librarians will have no problem using TLC. The logic, purpose, and functions are eminently suited for a small library that wants basic access and bibliographic control but does not use MARC records.

Certainly the major criticism of this package is its non-MARC format. For libraries that want to share their records in a consortial environment or in other cooperative projects this will be a major problem, since records output from the system would need to be converted into a quasi-MARC format through special programming. Other MARC-based PC products do offer a true MARC format and buyers may want to compare features and pricing with these packages before making a final decision.

Users less familiar with library standards, practice, and bibliographic control will have to study not only TLC but the ways of librarians and libraries. There is no way around having to plan for all the contingencies of the recording, retrieval, and management of documents. But TLC will take you through the steps of managing a small database of library records and offers a resilient system with which to provide thorough bibliographic control on a small scale.—*Dennis R. Brunning, Arizona State University, Tempe.* ■■