volume 12
number 4

ISSN 0022-2240

journal of library automation

december, 1979

jlauay 12(4) 297–408 (1979); j libr automat

journal of library automation

Volume 12, Number 4: December 1979

CONTENTS

JLAUAY 12(4) 297–408; J Libr Automat

Editor: **William D. Mathews**

journal of library automation

Communications Editor: **Mary A. Madden**

Book Review Editor: **Katherine King**

Advertising Editor: **Judith G. Schmidt**

Journal of Library Automation 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 Secretary:* Donald P. Hammer. The journal is issued quarterly in March, June, September, and December.

Journal of Library Automation 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. The *Journal* welcomes unsolicited manuscripts. Submissions should follow the guidelines stated under "Instructions to Authors" on page 408 of this volume.

Manuscripts of articles should be addressed to William D. Mathews, Editor, *Journal of Library Automation,* 73 E. Linden Ave., Englewood, NJ 07631. Technical communications and news items should be addressed to Mary A. Madden, *JOLA Communications,* 1605 SW Upland Drive, Portland, OR 97221. Copies of books submitted for review should be addressed to LITA Office, ALA Headquarters, 50 E. Huron St., Chicago, IL 60611. Advertising arrangements should be made with Judith Schmidt, 1408 D St., SE, Washington, DC 20003.

Journal of Library Automation is a perquisite of membership in the Library and Information Technology Association. Subscription price, $7.50, is included in membership dues. Nonmembers may subscribe for $15 per year. Single copies, $4.

Circulation and Production: American Library Association, 50 E. Huron St., Chicago, IL 60611. Please allow six weeks for change of address.

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 *Current Contents, Current Index to Journals in Education, Education, Library Literature,* and *Quarterly Bibliography of Computers and Data Processing.* Microfilm copies available to subscribers from University Microfilms, Ann Arbor, Michigan.

Second-class postage paid at Chicago, Illinois, and at additional mailing offices.

A Question of Quality

Depending on your point of view, library automation is entering a period either of diminishing returns or of unparalleled opportunity. Some people believe that, as a matter of survival, it is time to curtail the growth of automation in the library and to emphasize more visible priorities. Others feel we must not ignore the promise of electronic publishing, the paperless office, and the wired nation. They believe automation should be an implement of fundamental change, transforming the library into a new kind of information entity. The truth probably resides between these two extremes. Quite likely there will be sustained growth of automation activities in the decade ahead, but realistically we should recognize that new applications must be carefully justified on the basis of tangible benefits. We will have to adopt a conscious discipline of measuring and tailoring each application to identifiable needs.

At first, this discipline may be difficult to learn. The dominant climate of networking and cooperation has conditioned us to accommodate, to compromise, and to think primarily in terms of the general good. But there is no end of *generally* good ideas. The challenge is to make a value judgment on specific alternatives. Having accepted, for instance, the general premise that bibliographic control is desirable, it still remains for us to determine whether a particular project to translate between two subject heading schemes has practical relevance. In short, we must stop installing systems merely because others have done so, stop joining networks simply because they exist, and start making deliberate choices that suit the individual case. What this really amounts to is a process of relearning the importance of quality.

It may be difficult to imagine that a silver lining exists in the clouds of inflation, recession, and eroding tax support that darken our horizon. Yet, if these fiscal constraints encourage us to develop consistent ways of judging quality and assessing the worth of each new automation activity, there may be a ray of hope. For once, if a system is expected to bring operating efficiencies, to organize information in a better way, or to improve the user interface, we will possess the means to demonstrate the relevance of these effects. We will be able to show an improvement in quality for every service we support, and the systems we build will withstand the severest scrutiny.

POSTSCRIPT

This is my last issue as editor of *JOLA*. I wish to express my appreciation to all those who have contributed to the success of the journal during the past two years. In particular, I wish to thank the members of the Editorial Board for their advice and encouragement, and the staff of the Central Production Unit at ALA for their unfailing patience and diligence. WILLIAM D. MATHEWS

DOBIS/LIBIS: An Integrated, On-Line Library Management System

Caryl McALLISTER and A. Stratton McALLISTER: IBM Deutschland, DV Service Zentrum, Stuttgart, West Germany.

DOBIS/LIBIS is an on-line, integrated, interactive system that includes the major library functions of searching, cataloging, circulation, and acquisitions processing for a network of libraries. The integration of all library functions in a network environment permits sharing of computer resources and costs while enhancing the value of the shared catalog. All files are updated in real-time. Full authority file control is supported for all indexes to the data base. Attention has been paid to the interface between the system and the user in an effort to remove all computerese from the dialogue language.

INTRODUCTION

There are strong library and data processing arguments in favor of integrated, on-line library management systems. Since holding, circulation, and acquisition records are all related to bibliographic records, it makes economic and procedural sense to store the catalog records just once, but make them available throughout the system when needed.

Similarly, authority file entries can best be controlled when an entry is stored only once no matter how many bibliographic records are indexed by that entry and how many copies of these records are held. When provision is made for choosing authority file entries from the files instead of reentering them each time a document is ordered or cataloged, authority file control and accuracy are increased.

The library staff should be able to use an on-line system simply and directly. Various techniques can be used to simplify operation, to reduce keying, and to make the system approachable by people who are not data processing experts.

DOBIS/LIBIS was therefore designed with these three goals in mind:

- An integrated system with
- strong authority file control that can be
- used directly by the library staff and its borrowers.

System Organization

The system is designed not only for an individual library with its branches but also for a network of libraries.

Figure 1 is a diagram of the DOBIS/LIBIS library network structure. The system catalog is the pool of bibliographic records shared by all member libraries. Attached to it is the system holdings file giving the names of those libraries holding each document in the system catalog. The system catalog has author, title, subject, classification, publisher, ISBN, ISSN, LC card number, and other number indexes. These indexes serve not just as indexes to the bibliographic files but, more important, as the system's authority files as well. Figure 2 is a summary of the indexes available.

A file of local copies, held separately for each library in the network, tells how many copies of the document a particular library holds and the copy information for each of them. In addition to the local holding files, each library has the local files needed for circulation and acquisitions.

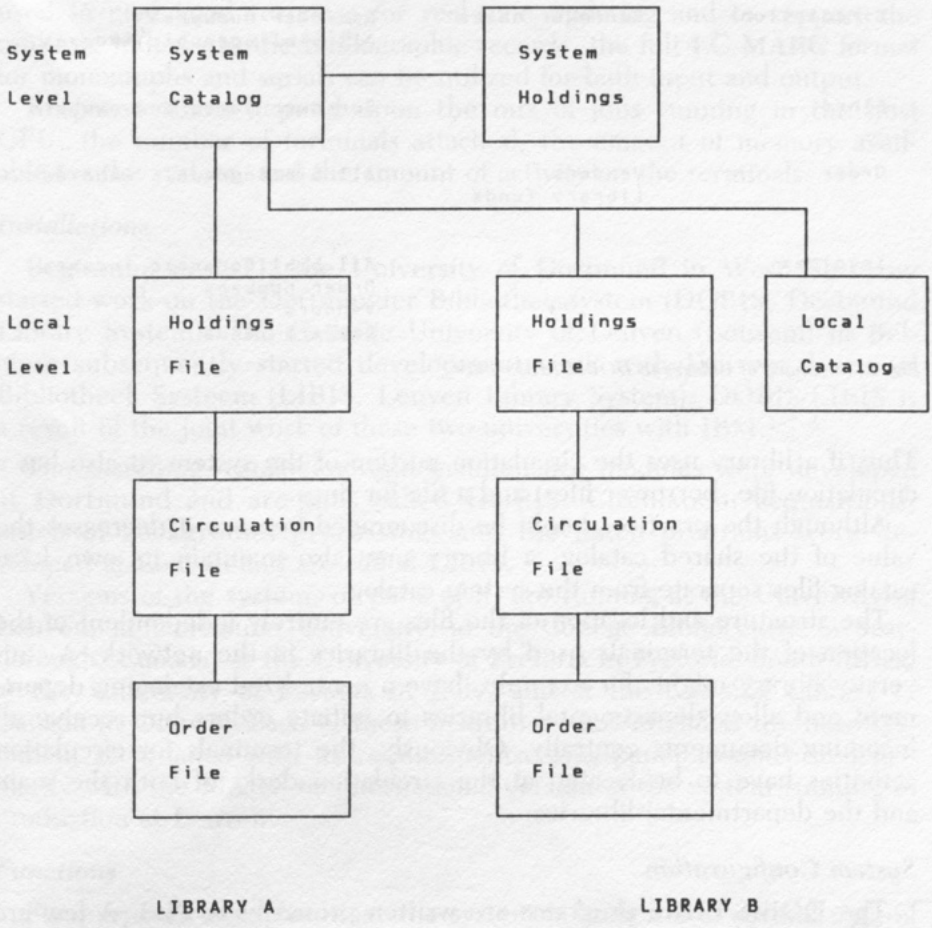


Fig. 1. Structure of the DOBIS/LIBIS Data Base.

Document File	Primary Indexes	Secondary Indexes
Bibliographic	Names	Vendors
	Titles	Library funds
	Subjects	Borrower names and numbers
	Classification	Order numbers
	Publishers	
	ISBNs and ISSNs	
	LC card numbers	
	Other numbers	
Copies	Call numbers	All bibliographic indexes
		Borrower names and numbers
		Order numbers
Borrower numbers	Borrower names	All bibliographic indexes
Circulation	Borrower names	Borrower numbers
		All bibliographic indexes
Fines		Borrower names and numbers
Order	Vendors	All bibliographic indexes
	Library funds	
Invoices		All bibliographic indexes
		Order numbers
		Vendors
		Library funds

Fig. 2. Summary of Indexes to DOBIS/LIBIS Files.

Thus if a library uses the circulation portion of the system, it also has a circulation file, borrower files, and a file for fines.

Although the practice should be discouraged because it decreases the value of the shared catalog, a library may also maintain its own local catalog files separate from the system catalog.

The structure and location of the files are entirely independent of the location of the terminals used by the libraries in the network. A university library might, for example, have a centralized cataloging department and allow departmental libraries to initiate orders but receive all incoming documents centrally. Obviously, the terminals for circulation activities have to be located at the circulation desks of both the main and the departmental libraries.

System Configuration

The DOBIS/LIBIS programs are written primarily in PL/I. A few are in assembly language for performance reasons. The real-time monitor is

CICS/OS/VS (Customer Information Control System), which in turn runs under any IBM OS operating system.

The system has run on computers ranging in size from an IBM System 370/Model 135 to a Model 168. It does not require a dedicated machine, but runs in a partition just as any other job. It uses direct access devices for storing all records. With the exception of batch printing operations, which are done at a high-speed printer, and the printing of single documents done at a typewriter terminal, all searching and updating are done from IBM 3270 display terminals. A scanner attached to the terminal is used to read bar-coded labels identifying borrowers and documents for circulation.

The maximum number of bibliographic records and copies for one installation is 2 billion. Up to 2 billion borrowers may be registered with the system. The maximum indexable length of all types of alphanumeric keys in the system is 255 bytes. Although a special internal format is used to give good response for real-time updating and to support the multiple indexes to the bibliographic records, the full LC MARC format for monographs and serials can be utilized for both input and output.

Response times depend upon the mix of jobs running in the host CPU, the number of terminals attached, the amount of memory available for the system, and the amount of activity on the terminals.

Installations

Beginning in 1971, the University of Dortmund in West Germany started work on the Dortmunder Bibliothekssystem (DOBIS, Dortmund Library System). The Catholic University of Leuven (Louvain) in Belgium subsequently started development work with Leuvens Integraal Bibliotheek Systeem (LIBIS, Leuven Library System). DOBIS/LIBIS is a result of the joint work of these two universities with IBM.[1-7]

The searching, cataloging, and maintenance modules were developed at Dortmund and are thus called DOBIS. Circulation, acquisitions, much of background processing, and the batch programs were developed at Leuven and are called LIBIS.

Versions of the system, or parts of it, are running at the University of Leuven, at Dortmund University, at the College Bibliocentre in Scarborough, Canada, at the University of Pretoria in Pretoria, South Africa, at the University of Perugia in Italy, and at the National Library of Canada in Ottawa. Each of these institutions has modified the base system in accordance with its requirements. This paper is concerned primarily with the MARC or international version of the system running in production at Leuven.

Functions

For purposes of explanation, the system is broken into several natural subdivisions, corresponding to a large extent with the way librarians

think about the departments in a library. Although it is impossible to describe each of these functions in detail, some of the most important features are summarized here to give a feeling for how they operate.

Searching

Searching is an important system function, both in its own right and because this display sequence is used when any file in the system is searched. In circulation and acquisitions, the choice of files is different but the sequence of displays is the same. Searching also allows us to illustrate the dialogue language of DOBIS/LIBIS.

Searching begins with the choice of a file. Figure 3 shows the list of the files available for searching the catalog as they are displayed at the terminal. Those listed on the left are the authority files and indexes for the system catalog. Those on the right are indexes for the local catalog and the local copies files. At the top of the display are one or more headings that tell the user exactly what he or she is doing: in this case, "Searching." Subsequent displays add to this heading to show the file being searched, and so forth.

Near the bottom of the display is an instruction: 'Enter number.' Instructions for displays always appear at this place on the screen. The user has chosen to enter a '1,' indicating that a search is to be made of the system names file. Only the single character "1" has been entered. The remainder of the display was created by the system. User input is brighter than system-supplied information when it is displayed at the terminal. We will indicate this difference by using boldface type for user input and normal type for system-supplied information.

The system's dialogue language can easily be translated from English

```
Searching


System files                        Local files
 1 Names                            10 Names
 2 Titles                           11 Titles
 3 Subjects                         12 Subjects
 4 Publishers                       13 Shelf list numbers
 5 Classification                   14 Copies, volumes
 6 ISBN / ISSN                      15 Document numbers
 7 LC card numbers
 8 Other numbers
 9 Document numbers



Enter number
1_
```

Fig. 3. Selection of Files in Searching.

into other languages. Additional facilities make it possible to run two or more dialogue languages simultaneously: For example, the system can "talk" to one user in English, while another user might have Dutch as a dialogue language and a third still another language. If the dialogue language were German all of the characters not in boldface type would appear in German.

After an index has been chosen, the system requests entry of a search term (figure 4). This term must be at least one character long but need not be an exact match with any entry in the indexes. The system ignores capitalization when searching, and so it is easiest just to enter lowercase characters.

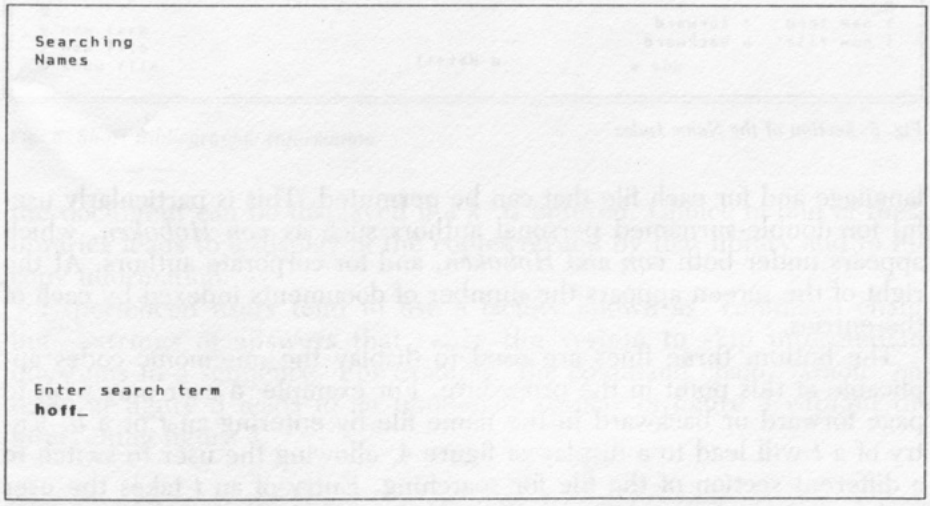


Fig. 4. Entry of the Search Term.

The system uses the search term only to display the portion of the appropriate file immediately surrounding the search term in alphabetical order (figure 5). Fourteen lines from that file are displayed: one immediately before the search term and then the succeeding thirteen entries in alphabetical order. At the left of the display are the numbers from 1 to 14. These numbers are arbitrary numerical designations for the entries from the name file that are listed on the display. The symbols immediately following the line numbers give additional information about the entries listed. An asterisk indicates that diacritical characters are included in the entry; an *x* indicates that there are cross-references for that entry; the slash on line 7 indicates that the entry had to be truncated for display purposes.

Corporate and conference authors in the name file and all titles in the title file are permuted so that an entry may be found under each word that is not a stop word. There is a separate stop word list for each

```
Searching
Names

 1                          van  Hoboken, Anthony, 1887-
 2*                              Hofferee, Arthur Charles Ernest, 1897-          2
 3                               Hoffnung, Gerard, 1900-1967                     6
 4                               Holst, Gustav Theodore, 1874-1934               1
 5                               Holst, Henry, 1899-                             2
 6 x                             Holywood, John                                  0
 7*x/              Verein fur    Holzblasinstrumentbau. Arbeitsgruppe fur Blockfl
 8                          van  Hoogstraaten, Willem, 1884-
 9                               Humperdinck, Engelbert, 1854-1921               4
10                               Hungarian String Quartet                        1
11                               Husa, Karel, 1921-                              3
12                               Hutschenruijter, Wouter, 1796-1878              1
13                               Hutschenruijter, Wouter, 1859-1943              1
14                               Huybrechts, Albert, 1938-                       2

Enter number or code
3_
t new term    f forward
i new file    b backward
                              d detail
```

Fig. 5. Section of the Name Index.

language and for each file that can be permuted. This is particularly useful for double-surnamed personal authors such as *von Hoboken*, which appears under both *von* and *Hoboken*, and for corporate authors. At the right of the screen appears the number of documents indexed by each of the entries.

The bottom three lines are used to display the mnemonic codes applicable at this point in the procedure. For example, a user may wish to page forward or backward in the name file by entering an *f* or a *b*. Entry of a *t* will lead to a display of figure 4, allowing the user to switch to a different section of the file for searching. Entry of an *i* takes the user to figure 3 so that a different file may be searched. The authority file detail (subfield codes, MARC indicators, notes, and cross-references) for a particular entry is displayed when a *d* is entered.

If the line number of an entry is chosen, the system displays a short summary of each document associated with that entry (figure 6). Here the first author and title from the system catalog as well as the publication date are displayed.

Entry of a *w* asks the system to display the file again (figure 5). Entry of an *e* indicates that searching should be terminated.

Entry of the number of the appropriate short description leads to display of the complete bibliographic information (figure 7). The standard descriptive section of a bibliographic entry is shown at the top: title, remainder of title, publisher, place, date, collation. This is followed by a summary of all the elements in the bibliographic record. Entry of the mnemonic code *s* causes the system to display short bibliographic information again (figure 6).

From the bibliographic information, those libraries holding copies of

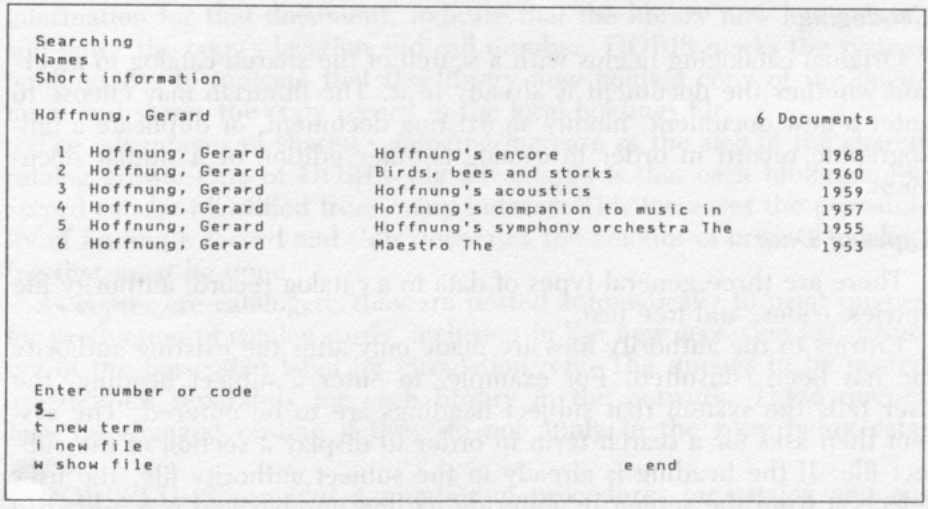


Fig. 6. Short Bibliographic Information

the document can be displayed if a *k* is entered. Choice of one of these libraries leads to a display of the copies owned by that library and to full copy information.

Experienced users tend to use a facility known as "command chaining"—strings of answers that cause the system to skip intermediate questions in a sequence. For example, the command chain "/1/hoff" entered at figure 3 leads to an immediate display of figure 5 without the intervening figure 4.

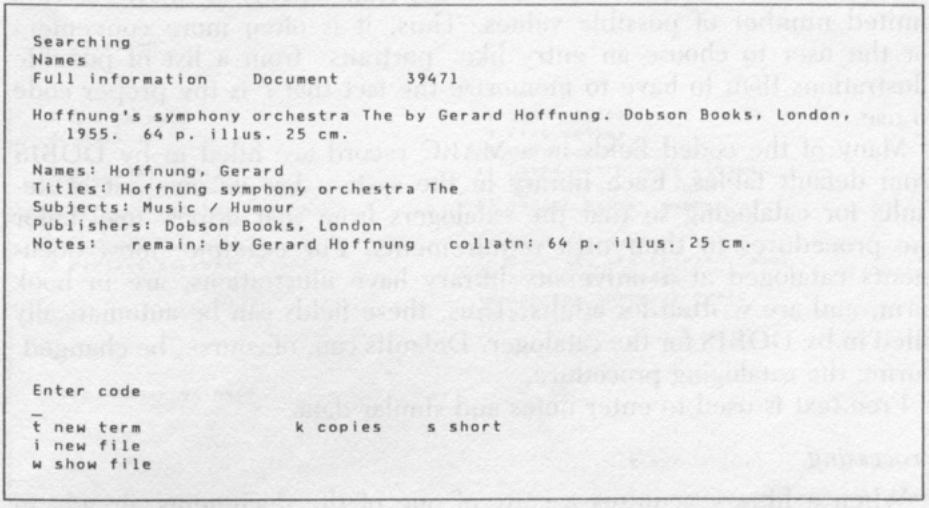


Fig. 7. Full Bibliographic Information.

Cataloging

Original cataloging begins with a search of the shared catalog to ascertain whether the document is already in it. The librarian may choose to enter a new document, modify an existing document, or duplicate a bibliographic record in order to catalog another edition or a similar document.

Types of Data

There are three general types of data in a catalog record: authority file entries, codes, and free text.

Entries to the authority files are made only after the existing authority file has been consulted. For example, to enter a subject heading, the user tells the system that subject headings are to be entered. The system then asks for a search term in order to display a section of the subject file. If the heading is already in the subject authority file, the user selects it from the screen by entering its line number and it is added to the document. The user is not required to reenter the subject heading. If, however, the subject is not already present, it can be added to the file; the user selects the MARC indicators for type of heading (topical, geographic, etc.) and the source (LC, NLM, etc.) from code tables and then types in the heading itself.

Entries may be made in all the files without the necessity of having a document in the catalog with that entry on it. Thus it is possible for the user to enter subject headings or descriptors from tape, thereby decreasing the typing required of the librarian and increasing the accuracy and consistency of subject heading assignments.

Tables are used throughout the system to standardize and simplify entry of coded data. Many of the fixed and leader fields in MARC have a limited number of possible values. Thus, it is often more convenient for the user to choose an entry like "portraits" from a list of possible illustrations than to have to memorize the fact that *c* is the proper code to use.

Many of the coded fields in a MARC record are filled in by DOBIS from default tables. Each library in the system has its own set of defaults for cataloging so that the catalogers from that library may tailor the procedures to their own requirements. For example, most documents cataloged at a university library have illustrations, are in book form, and are written for adults. Thus, these fields can be automatically filled in by DOBIS for the cataloger. Defaults can, of course, be changed during the cataloging procedure.

Free text is used to enter notes and similar data.

Processing

When a library acquires a copy of one of the documents already in the system catalog, the librarian needs only to locate the bibliographic

information for that document, indicate that the library now has a copy, and enter the copy's location and call number. DOBIS marks the system holdings file to indicate that the library now holds a copy of the document and stores the copy record in the local holdings file.

The advantages of shared cataloging increase as the size of the shared catalog grows. One of DOBIS' big advantages is that each bibliographic record can be identified from many indexes. This increases the probability of finding a record and thus decreases the amount of original cataloging that must be done.

As copies are cataloged, they are posted automatically to print queues for production of catalog cards, inclusion in the new accession list, printing of the bar-coded label for circulation, etc. The queues to be posted are defined separately for each library in the network. These queues may be changed on-line if they do not apply to the copy being cataloged.

DOBIS/LIBIS provides a number of procedures for catalog and authority file maintenance. For example, an author's name might be entered incorrectly in the catalog because of a mistake by the cataloger or perhaps because of a change in the cataloging rules. The cataloging for *all* documents in the system by that author may be corrected at once with one of the catalog maintenance functions.

Circulation

The subfunctions available under the area of circulation are listed in the circulation summary screen shown in figure 8. The most important of them are described briefly in the following paragraphs.

Circulation policies control the length of time a document is loaned to

```
Circulation


 1 Charge out                          8 Loan policy
 2 Check in                            9 Fine policy
                                      10 Overdue policy
 3 Document status                    11 Dates library is closed
 4 Borrower status - name             12 Semi-permanent loan
 5 Borrower status - number           13 Change master number

 6 Replace document label             14 Send overdues
 7 Replace borrower label             15 Generate pre-printed label
                                      16 Delete borrowers
                                      17 Correct borrower names
                                      18 Delete paid fines



Enter number or code
_

                                                  e end
```

Fig. 8. Circulation Subfunctions.

a borrower, the amount of the fine to be charged if it becomes overdue, and at what interval overdue notices are to be sent. These are different for each library in the network and can be displayed and changed on-line.

Charge out and check in are straightforward. The borrower number and the numbers on the labels are read with a bar-code scanner. The borrower's record in the file is checked to verify his or her right to borrow. Optionally, the document records can be read and the copy's call number displayed. The loan period is calculated automatically from the loan policy. When a document is returned, the fine is likewise calculated automatically from the fine policy.

Status information for a borrower includes the documents on loan, fines due, fines paid, and holds placed, as well as the usual fields needed for borrower records such as addresses, telephone numbers, etc. Fines may be recorded as paid, holds may be placed, and the values for various codes and other fields may be changed on-line.

The circulation status of all copies of a title or of a single copy may be displayed. This information includes the borrower, the date borrowed and the date due, renewals made, whether the document has been recalled and when, and whether there are holds on the title or on the copy.

A damaged borrower label or document label is replaced by requesting that a new label be printed at a typewriter terminal near the circulation desk.

Documents may be loaned from one departmental library to another on a semipermanent basis. The documents are treated, for circulation purposes, as if they belonged to the borrowing library.

Acquisitions

The acquisitions function includes placing orders, receiving copies and invoices, controlling library funds, claiming, and renewing subscriptions. Both monographs and serials are handled. The acquisitions subfunctions are listed in figure 9. Some of these are discussed briefly below.

When a document is to be ordered, the vendor for the document and the library fund to be debited are first identified. Then the bibliographic record is found in the system catalog. If the desired document is not already in the catalog, it can be added. LIBIS then displays an order summary, filling in many fields with the order defaults tailored for each library and vendor. Individual fields may be changed as desired before the order is placed. If the document is to be circulated to a borrower when it is received, the borrower file is searched for the name of that person and the hold recorded in the files.

Documents and invoices may be received individually or together. In both cases, the document ordered is identified with either the biblio-

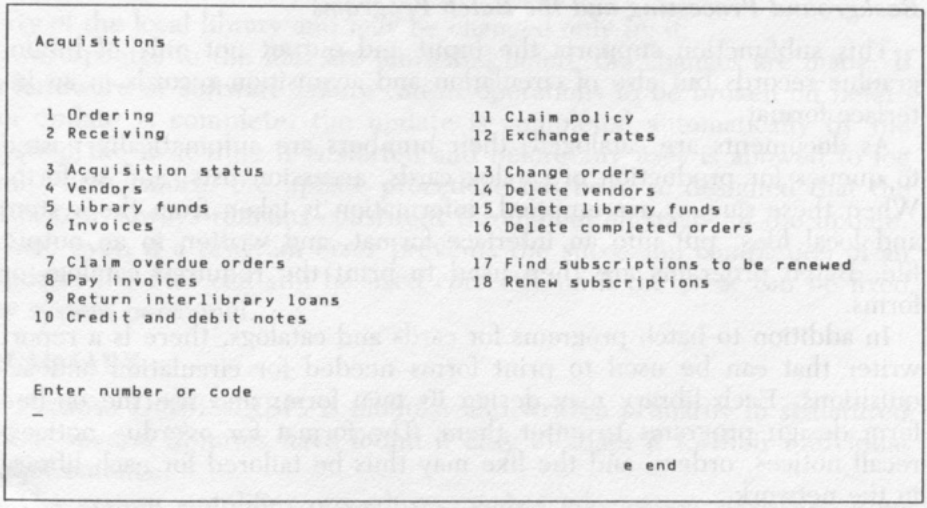


Fig. 9. Acquisition Subfunctions.

graphic indexes or the order number. The number of copies received and/or the invoice information is noted.

Periodic searches of the order file by programs locate those orders that are complete and where the invoices can be paid. The order numbers of these invoices are posted to a queue so that LIBIS can prepare financial information for the accounting department.

The acquisitions status for all documents in the order file can be found with any of the bibliographic indexes and with the order number. Statistical information about vendors and library funds is available online.

Each library in the network has a claim policy that defines when the various types of documents ordered are to be claimed. The claim forms are printed with the report writer (see "Background Processing," below).

Interlibrary loans are handled much as normal purchase orders are handled. Provision is made for the document to be treated as a copy temporarily assigned to the borrowing library. A special function is used to return the loans to the lending library.

Maintenance and the Utilities

Maintenance is a function used primarily by programmers responsible for taking care of the system. It includes programs that display statistical information about the condition of the files and that allow direct changes to records in the file.

The utilities provide unload, reload, and reorganization facilities for the files. Files are reorganized when large numbers of records added to the files degrade system performance.

Background Processing and the Batch Programs

This subfunction supports the input and output not only of bibliographic records but also of circulation and acquisition records in an interface format.

As documents are cataloged, their numbers are automatically posted to queues for production of catalog cards, accession lists, and so forth. When these queues are emptied, information is taken from the system and local files, put into an interface format, and written in an output file. Batch programs are then used to print the required catalogs or forms.

In addition to batch programs for cards and catalogs, there is a report writer that can be used to print forms needed for circulation and acquisitions. Each library may design its own forms and use the on-line form design programs to enter them. The format for overdue notices, recall notices, orders, and the like may thus be tailored for each library in the network.

Records from the system are written on an output file when a queue is ready for printing. The report writer then processes these records, printing the information in them according to the format given in the print specification. It can sort the records before printing, number pages, print fields conditionally, underline, print vertically, sum fields, print totals, and print fixed information.

Security

Access to the system is restricted to those persons registered as users. (Borrowers are those persons who may borrow documents, and thus information about them is found in the borrower files. There are, of course, individuals found in both file categories who may borrow documents as well as catalog them.) Each user has a name used to log onto the system and a password. Both must be entered correctly before the system will allow access. The password is entered and changed by the user and is stored in encoded form. Knowledge of the encoded form does not permit reconstruction of the password.

With each user is associated a set of authorization levels that control the amount of access he or she is allowed in each of the system functions (searching, cataloging, circulation, maintenance, etc.). If the user is not authorized to use a particular function, the system will not allow him or her to access it. Authorization levels are controlled by the one or two people at each installation who perform system maintenance.

Although searches in the catalog are allowed without authorization or password, changes to the data base require authorization and a password. With the exception of the system catalogs, to which any cataloger in the system with the appropriate authorization level may make additions, a system user may update *only* the files for his or her local library. The system thus ensures that the local data are wholly the prop-

erty of the local library and may be changed only by it.

All updates to the files are journaled before the changes are made. If a hardware or software failure causes operations to be broken off before an update is complete, the update is completed automatically by the system the next time it is started and before any user is allowed to log on. Furthermore, the update procedure has been so designed that the data base always remains consistent throughout all stages of the update. Thus, even if a program error prevents the successful completion of an update, the files can still be used correctly until the error can be fixed by system personnel.

SUMMARY

Because DOBIS/LIBIS is modular and written primarily in structured PL/I, several libraries have found it easy to adapt it to their individual requirements.

The system simplifies procedures, makes information accessible from many locations, permits effective intervention at critical points in procedures, makes all data available immediately, and can save money for libraries.

The authors of this paper designed and programmed DOBIS/LIBIS while working at Dortmund and Leuven. The functional design of the system has been influenced by many people at Dortmund, at Leuven, and in Canada and the United States.

REFERENCES

1. IBM Corporation, *DOBIS Librarian's Guide* (SB11-5019; Copenhagen: Oct. 1978).
2. IBM Corporation, *DOBIS Installation and Operations Guide* (SB11-5024; Copenhagen: Oct. 1978).
3. IBM Corporation, *LIBIS—Online Librarian's Guide* (SB11-5311; Copenhagen: Sept. 1979).
4. IBM Corporation, *LIBIS—Online Installation and Operations Guide* (SB11-5312; Copenhagen: Sept. 1979).
5. IBM Corporation, *LIBIS—Batch Program Description and Operations Guide* (SB11-5356; Copenhagen: Oct. 1979).
6. IBM Corporation, *Dortmund Library System—DOBIS: Executive Guide*. Guide for Application Programming (GE15-6068; Copenhagen: Oct. 1975).
7. IBM Corporation, *Dortmund Library System—DOBIS: Application Guide*. Guide for Application Programming (GE15-6069; Copenhagen: Oct. 1975).

Caryl and Stratton McAllister have been with IBM Germany since 1971 and have been involved with DOBIS-LIBIS since its beginning. When they are not writing, enhancing, and teaching DOBIS-LIBIS, they like to play music on medieval and renaissance instruments.

Summary Statistics for Five Years of the MARC Data Base

Martha E. WILLIAMS, Stephen W. BARTH, and Scott E. PREECE: Information Retrieval Research Lab (IRRL), University of Illinois at Urbana-Champaign.

MARC data base statistics are presented for the benefit of processors of the MARC file who may use those data for purposes of planning file structures, for selecting subsets of MARC for local processing, for estimating processing time based on record sizes, and for estimating future file sizes based on growth rates. Statistics pertaining to record lengths, field tag occurrence, data element length per field tag, and distribution of records by Dewey Decimal Division and Library of Congress class code are given. Library of Congress categories versus record length and main entry length versus record length distributions are also provided. The statistics can also be valuable in the preparation of user profiles and search strategies for subject searching of the MARC data base online.

INTRODUCTION

In June 1976 Williams and Shefner published a paper entitled "Data Element Statistics for the MARC II Data Base" in the *Journal of Library Automation*.[1] Responses to this paper indicated an interest in, and request for, the same type of statistics on a wider scale plus additional breakdowns and multiyear comparisons. In this paper we have responded to those requests. We have provided statistics for five years of the MARC data base. As in the 1976 paper we have included data for the 1969–70 year from the report by Gerald L. Swanson.[2] We did not process MARC tapes for that year. We did process tapes for the other four years.

The years covered are 1969–70, 1974–75, 1975–76, 1976–77, and 1977–78. The first series of MARC* tapes began with volume 1 in 1969 and continued through volume 4, which covered the year 1972–73. The next series, MARC Books All, began again with volume 1 and covered the year 1973–74. In our study we processed tapes for volumes 2 through 5 of the MARC Books All series and compared data with Swan-

*There was a previous volume, volume 0, which included pre-1968 records and is referred to as *Recon*.

son's volume 1 data from the initial series. Our volume 2 data were processed by us in 1975 and we did not rerun those tapes. Some of the analyses in this paper provide breakdowns and statistics that were not included in Swanson's volume 1 series 1 data nor in our earlier study of volume 2. Where possible, data are provided for all five years but in some cases they are for four years or three years. All tables are appropriately labeled to indicate coverage.

In all cases the term *record* refers to one bibliographic record that constitutes one logical record. Documentation consulted for the MARC records, tag numbers, field definition, etc., was provided by OCLC and the Library of Congress.[3,4] Records excluded from our processing were records with the following encoding level codes (in byte position 17 of the leader where the first position is position 0):

E—OCLC system identified an error
J—MARC deleted records
K—OCLC member input sublevel record
O—Order records

We included new, corrected (corrected records include all fields—not just corrected data), and previously CIP records as designated in byte position 5 of the leader.

In most years the MARC tapes are issued fifty-two times but in some years there are fifty-three or fifty-four issues because of changes in the date for subscriptions or partial weeks being included. Table 1 provides a summary of the statistics that characterize MARC tapes. Numbers of issues, numbers of records, average number of records per issue, aver-

Table 1. Summary of Statistics

	*1969–70 V.1 Ser. 1**	*1974–75 V.2*	*1975–76 V.3*	*1976–77 V.4*	*1977–78 V.5†*
Total Records	52,294	124,355	146,199	176,127	173,978
Number of Issues	54	52	53	52	(52)
Avg. No. Records per Issue		2,391	2,758	3,387	3,346†
Standard Deviation			799	745	
Max.		4,104	5,406	4,997	
Min.		1,076	1,155	1,501	
Avg. No. Fields/Record	10.70	11.45	13.53	13.63	13.50
Avg. No. Unique Fields/Record	10.03	10.59	10.63	10.71	10.60
No. Fields‡	48	49	49	50	50
Field Tags Added		043 + 086		830	
Field Tags Deleted		652			
Mean Rec. Length	636.00	666.42	681.93	689.67	685.21
Standard Deviation		192	196	196	194
Max.	1,987	2,032	2,030	2,047	2,050
Min.	280	263	248	251	247

*The 1969–70 MARC tapes are from the first series of MARC tapes. Numbering volumes began again with volume 1 in 1973.
†Based on a set of annual tapes ÷ 52.
‡Excluding fields 1 and 8.

age number of unique fields per record, mean record length, and number of fields per volume are provided for the five MARC years in summary form in table 1. Further specification is given in later sections of this paper, as are field lengths (excluding tag), distribution of records by Dewey Decimal Division (DDC numbers), distribution of records by Library of Congress (LC) categories, LC categories versus record length distribution, and main entry length versus record length distribution.

Numbers of records in table 1 may vary slightly with numbers counted by other investigators. The reason is that bad records (parity errors, etc.) were not counted or processed. The number of bad physical records per volume ranged from less than .01 percent to slightly more than .02 percent.

FIELD TAG OCCURRENCES

Field tag occurrence statistics were generated on a volume basis for the years 1974–75, 1975–76, 1976–77, and 1977–78. Slightly more than half the field tags occur once per record in those records in which they occur: 59 percent in volume 2, 51 percent in volume 3, and 56 percent in volumes 4 and 5. Although half occur once per record only 10 percent occur more than 1.2 times per record. For volumes 2, 3, 4, and 5, these are tags 500, 650, 651, 700 and 710. Tag 650 is consistently the one with the highest frequency of occurrence. Tags 1 (control number) and 8 (fixed length data elements) occur in all records. These are not included in our analyses, as they do not vary. Additional statistics are presented in tables 2(a)–2(d). The column marked "Field Tag" indicates the MARC tag numbers. The "Total Records" column indicates the total number of records that contain the particular element at least once. "Percent of Records" indicates the percentage of the records in which the element occurs. These numbers have been truncated to one decimal place. The "Total Occurrence" column indicates, for each tag number, the total number of times that data element (including multiple occurrences per record) occurs in the data base.

In table 3 comparisons of field tag occurrences were made, using volume 2 data as the base. Thus data for volumes 3, 4, and 5 are compared with volume 2. Field tag percent occurrences and changes in occurrences data are provided in terms of percentage change relative to volume 2. In the interest of saving space only changes of some magnitude are provided in this paper, and they are given in the text below.

Some tag occurrences changed significantly from volume to volume. For example, tag 15 (National Bibliography number) occurred in 25.57 percent of the records in volume 2. It increased in each year, going from 25.5 percent to 29.9 percent in volume 3, to 36.3 percent in volume 4, to 39.9 percent in volume 5, for a total percentage increase of 14.3 percent. Tag 350 (Bibliographic Price) decreased in occurrence from 25.7 percent in volume 2 to 5.6 percent in volume 5, for a de-

crease of 20.1 percent. The reason for the decrease in use of this tag is that a decision was made by the Library of Congress to remove price data for ISBD(M) records from the 350 field and include it in a subfield of 020. Consequently, the 020 field (ISBN) increased in length (in volume 2 the average length was eighteen characters and in volume 5 it was twenty-one characters). Tag 830 (Series Added Entry for Uniform Title Heading) was introduced in volume 4 (occurring in only one record in volume 4 and in seven records in volume 5) and could not be compared to the base year.

Other tags with significant changes include: 043 (Geographic Area Code), which increased in use by 4.4 percent from volume 2 to volume 5; tag 082 (DDC number), which decreased from 96.6 percent use in records in volume 2 to 69.3 percent in volume 5, for an overall decrease of 27.2 percent; tag 100 (Main Entry—Personal Name) decreased from 79.4 percent to 74.2 percent in volume 5, for a change of −5.2 percent; and tag 500 (General Note) increased from 41.7 percent, to 51.9 percent for a change of +10.2 percent. There were no great fluctuations; the largest decrease was 27.2 percent and the largest increase was 14.3 percent.

These DDC percent occurrence data must not be considered apart from the absolute numbers, which show an increase in assignment of DDCs; i.e., volume 2 included 120,218 DDCs, volume 3, 132,314, and volume 4, 147,016, and there was a decrease in volume 5 to 120,734. The decline in the percentage of MARC records with DDCs does not reflect the assignment of fewer DDCs. Rather, it reflects the fact that the number of MARC records produced is growing faster than the number of DDCs assigned.

In volume 3 for the twenty tags that occurred in at least 10 percent of the records the average magnitude of change, regardless of direction, is 3.26 (65 ÷ 20 = 3.26). Of the twenty, nine varied less than 1. Of the eleven that varied more than 1 the average magnitude of change, regardless of direction, is 5.70. Some decreased, some increased but the net change for the eleven was an increase of 4.43. In volume 4 the average magnitude of change, regardless of direction, is 4.63. Of the 20, nine varied less than 1; the remaining eleven showed an average magnitude of change of 8.05. The net change for the eleven was an increase of 9.49. For volume 5 the average magnitude of change is 5.35 for the twenty fields that occurred in at least 10 percent of the records. Of the twenty, eight varied less than 1; the remaining twelve showed an average magnitude of change of 8.85. All the above changes are relative to volume 2, which was taken as the base volume.

The average number of fields per record—including multiple occurrences per record—increased, as is shown in table 1, from volume 1 through volume 4 (from 10.70 to 11.45 to 13.53 to 13.63) and dropped slightly in volume 5 (to 13.50).

Table 2(a). Field Tag Occurrence Statistics: Volume 2, 1974–75

Field Tag	*Total Records*	*Percent of Records*	*Total Occurrence*
15	31,803	25.5	31,805
20	68,499	55.0	68,505
25	307	0.2	307
40	100	0.0	100
41	9,271	7.4	9,273
43	50,533	40.6	50,545
50	124,272	99.9	124,288
51	415	0.3	418
60	3,820	3.0	3,820
70	19	0.0	19
82	120,207	96.6	120,218
86	1,109	0.8	1,109
100	98,801	79.4	98,802
110	14,883	11.5	14,385
111	2,705	2.1	2,705
130	471	0.3	471
240	6,385	5.1	6,387
241	8	0.0	8
245	124,355	100.0	124,357
250	17,813	14.3	17,813
260	124,354	99.9	124,356
300	124,132	99.8	124,134
350	32,005	25.7	32,007
400	119	0.0	119
410	2,419	1.9	2,515
411	20	0.0	20
440	10,050	8.0	10,279
490	27,534	22.1	28,715
500	51,882	41.7	64,530
501	68	0.0	68
502	489	0.3	492
504	59,385	47.7	59,632
505	4,822	3.8	4,822
520	7,020	5.6	7,024
600	14,362	11.5	17,280
610	6,414	5.1	7,029
611	130	0.1	140
630	1,048	0.8	1,186
650	90,953	73.1	158,686
651	20,276	16.3	27,936
700	34,193	27.4	43,696
710	12,571	10.1	15,480
711	344	0.2	349
730	1,314	1.0	1,420
740	5,269	4.2	5,634
800	86	0.0	87
810	8,293	6.6	8,706
811	28	0.0	28
840	2,424	1.9	2,499

Table 2(b). Field Tag Occurrence Statistics: Volume 3, 1975–76

Field Tag	*Total Records*	*Percent of Records*	*Total Occurrence*
15	43,802	29.9	43,808
20	94,076	64.3	94,106
25	88	.0	88
40	60	.0	60
41	9,695	6.6	9,700
43	63,510	43.4	63,537
50	146,105	99.9	146,116
51	292	.2	293
60	4,070	2.7	4,070
70	28	.0	28
82	132,307	90.4	132,314
86	986	.6	987
100	110,005	75.2	110,007
110	17,472	11.9	17,475
111	3,281	2.2	3,282
130	435	.2	435
240	6,413	4.3	6,413
241	8	.0	8
245	146,199	100.0	146,205
250	20,447	13.9	20,449
260	146,190	99.9	146,190
300	146,019	99.8	146,026
350	10,200	6.9	10,200
400	74	.0	74
410	1,441	.9	1,479
411	11	.0	11
440	12,631	8.6	12,901
490	34,774	23.7	36,610
500	78,143	53.4	94,075
501	129	.0	131
502	987	.6	992
504	72,325	49.4	72,852
505	5,711	3.9	5,712
520	5,085	3.4	5,089
600	18,345	12.5	21,741
610	7,728	5.2	8,375
611	195	.1	204
630	1,361	.9	1,548
650	108,356	74.1	190,044
651	24,319	16.6	32,670
700	41,668	28.5	54,888
710	15,533	10.6	19,346
711	441	.3	455
730	1,547	1.0	1,644
740	7,624	5.2	8,136
800	46	.0	46
810	11,472	7.8	11,912
811	56	.0	56
840	2,590	1.7	2,611

Table 2(c). Field Tag Occurrence Statistics: Volume 4, 1976–77

Field Tag	*Total Records*	*Percent of Records*	*Total Occurrenc*
15	64,027	36.3	64,046
20	118,261	67.1	118,307
25	102	.0	102
40	45	.0	45
41	11,972	6.7	11,972
43	79,852	45.3	79,882
50	175,965	99.9	175,981
51	394	.2	395
60	3,829	2.1	3,829
70	7	.0	7
82	147,012	83.4	147,016
86	799	.4	799
100	129,287	73.4	129,289
110	21,553	12.2	21,559
111	4,083	2.3	4,084
130	500	.2	500
240	7,159	4.0	7,165
241	12	.0	12
245	176,127	100.0	176,134
250	26,586	15.0	26,587
260	176,115	99.9	176,116
300	175,954	99.9	175,964
350	9,594	5.4	9,594
400	186	.1	188
410	1,354	.7	1,375
411	21	.0	21
440	15,132	8.5	15,355
490	45,342	25.7	47,811
500	93,680	53.1	113,192
501	199	.1	199
502	1,207	.6	1,218
504	86,773	49.2	87,456
505	7,019	3.9	7,019
520	6,303	3.5	6,304
600	21,596	12.2	25,284
610	10,102	5.7	10,924
611	268	.1	286
630	1,524	.8	1,680
650	133,351	75.7	234,473
651	30,215	17.1	39,742
700	52,479	29.7	69,975
710	19,344	10.9	24,113
711	550	.3	571
730	1,656	.9	1,751
740	9,745	5.5	10,440
800	246	.1	246
810	15,099	8.5	15,631
811	45	.0	45
830	1	.0	1
840	3,106	1.7	3,138

Table 2(d). Field Tag Occurrence Statistics: Volume 5, 1977–78

Field Tag	*Total Records*	*Percent of Records*	*Total Occurrence*
15	69,491	39.9	69,513
20	114,512	65.8	114,541
25	341	.1	341
40	47	.0	47
41	12,981	7.4	12,984
43	78,392	45.0	78,401
50	173,725	99.8	173,751
51	313	.1	316
60	2,895	1.6	2,896
70	13	.0	13
82	120,728	69.3	120,734
86	845	.4	845
100	129,131	74.2	129,134
110	20,083	11.5	20,084
111	3,496	2.0	3,496
130	482	.2	482
240	6,847	3.9	6,850
241	12	.0	12
245	173,974	99.9	173,977
250	28,512	16.3	28,515
260	173,963	99.9	173,965
300	173,890	99.9	173,896
350	9,772	5.6	9,772
400	267	.1	267
410	653	.3	670
411	2	.0	2
440	15,560	8.9	15,785
490	44,789	25.7	47,196
500	90,383	51.9	109,646
501	272	.1	273
502	1,348	.7	1,367
504	83,088	47.7	83,812
505	7,222	4.1	7,222
520	9,366	5.3	9,370
600	23,634	13.5	27,680
610	10,158	5.8	10,952
611	247	.1	258
630	1,623	.9	1,833
650	133,097	76.5	229,890
651	28,661	16.4	37,758
700	52,580	30.2	69,482
710	18,352	10.5	22,788
711	466	.2	481
730	1,548	.8	1,633
740	9,402	5.4	10,065
800	259	.1	260
810	14,284	8.2	14,792
811	57	.0	57
830	7	.0	7
840	3,096	1.7	3,123

Table 3. Field Tag Percent Occurrence and Change in Occurrence: Volumes 2, 3, 4, and 5

Field Tag	*V.2 Percent Occurrence*	*V.3 Percent Occurrence*	*Change in Percent versus V.2*	*V.4 Percent Occurrence*	*Change in Percent versus V.2*	*V.5 Percent Occurrence*	*Change in Percent versus V.2*
15	25.57	29.96	+4.39	36.35	+10.78	39.94	+14.37
20	55.08	64.35	+9.27	67.15	+12.07	65.82	+10.74
25		0.06		0.6		0.20	
40		0.04		0.03		0.03	
41		6.63		6.80		7.46	
43	40.64	43.44	+2.80	45.34	+4.70	45.06	+4.42
50	99.93	99.94	+0.01	99.91	−0.02	99.86	−0.07
51		0.20		0.22		0.18	
60		2.78		2.17		1.66	
70		0.02		0.004		0.01	
82	96.66	90.50	−6.16	83.47	−13.19	69.39	−27.27
86		0.67		0.45		0.49	
100	79.45	75.24	−4.21	73.41	−6.04	74.22	−5.23
110	11.57	11.95	+0.38	12.24	+0.67	11.54	−0.03
111		2.24		2.32		2.01	
130		0.30		0.28		0.28	
240		4.39		4.07		3.94	
241		0.01		0.01		0.01	
245	100.00	100.00	0.00	100.00	0.00	100.00	0.00
250	14.32	13.99	−0.33	15.10	+0.78	16.39	+2.07
260	99.99	99.99	0.00	99.99	0.00	99.99	0.00
300	99.82	99.88	+0.06	99.90	+0.08	99.95	+0.13
350	25.74	6.98	−18.76	5.45	−20.29	5.62	−20.12
400		0.05		0.10		0.15	
410		0.99		0.77		0.38	
411		0.01		0.01		0.001	
440		8.64		8.59		8.94	
490	22.14	23.79	+1.65	25.74	+3.60	25.74	+3.60
500	41.72	53.45	+11.73	53.19	+11.74	51.95	+10.23
501		0.09		0.11		0.16	
502		0.68		0.69		0.78	
504	47.75	49.47	+1.72	49.27	+1.52	47.76	+0.01
505		3.91		3.99		4.15	
520		3.48		3.58		5.38	
600	11.55	12.55	+1.00	12.26	+0.71	13.58	+2.03
610		5.29		5.74		5.84	
611		0.13		0.15		0.14	
630		0.93		0.87		0.93	
650	73.14	74.12	+0.98	75.71	+2.57	76.50	+3.36
651	16.31	16.63	+0.32	17.16	+0.85	16.47	+0.16
700	27.50	28.50	+1.00	29.80	+2.30	30.22	+2.72
710	10.11	10.63	+0.52	10.98	+0.87	10.55	+0.44
711		0.30		0.31		0.27	
730		1.06		0.94		0.89	
740		5.22		5.53		5.40	
800		0.03		0.14		0.15	
810		7.85		8.57		8.21	
811		0.04		0.03		0.03	
830				0.001		0.004	
840		1.77		1.76		1.78	

The stability of the occurrences in volumes 3, 4, and 5 probably indicates the number of fields stabilized, with only one new field being added in volume 4.

As indicated in table 1 the average number of unique fields (ignoring multiple occurrences) increased from volume 1 through volume 4 (from 10.03 to 10.59 to 10.63 to 10.71) and dropped back in volume 5 (to 10.60).

RECORD AND FIELD LENGTHS

Record Lengths

Record lengths have varied slightly over the five volumes, as is shown in table 4.

The mean length increased for three years and then decreased in volume 5. From volume 1 (Ser. 1) to volume 2 the mean length increased 4.8 percent; from volume 2 to volume 3 it increased 2.3 percent; from volume 3 to volume 4 it increased only 1.1 percent; and from volume 4 to volume 5 it decreased .6 percent.

Table 4. Record Lengths for Five Volumes of MARC

Volume No.	*Mean Length*	*Standard Deviation*	*Maximum Length*	*Minimum Length*
V.1 Ser. 1	636.00		1,987	280
V.2	666.42	192.0	2,032	263
V.3	681.93	196.1	2,030	248
V.4	689.67	196.3	2,047	251
V.5	685.21	194.6	2,050	247

Field Lengths

Field lengths (mean, maximum, minimum, and standard deviation) are provided for volumes 2, 3, 4, and 5 in tables 5(a)–5(d). In each table, column 1 specifies the MARC field tag number, and column 2 indicates the total number of occurrences of that field tag for the volume. The adjacent columns specify the mean length, the maximum length, the minimum length, and the standard deviation for the data in the field. In general the trend is toward stabilization of field lengths, as can be seen in tables 6 and 7. Using the data presented in table 6, one can calculate the percentage change in the mean length of fields for each volume relative to the prior volume. This was done for the twenty tags that occur in more than 10 percent of the MARC records in volume 2 (nineteen of the twenty occurred in more than 10 percent of the records in volumes 3, 4, and 5 also; the nineteen are the only ones with more than 10 percent occurrence in volumes 3, 4, and 5). When volume 3 was compared to volume 2 only four of the fields had mean lengths that remained unchanged, while eleven remained unchanged from volume 3 to volume 4 and volume 4 to volume 5. Looking at the relative changes over volumes 2 to 5 one can observe that by the time of volume 5

Table 5(a). Field Length Statistics: Volume 2, 1974–75

Field Tag	*Total Occurrence*	*Mean Length*	*Maximum Length*	*Minimum Length*	*Standard Deviation of Length*
1	———	———	———	———	———
8	———	———	———	———	———
15	31,805	12.0	42	7	2.7
20	68,505	18.3	91	8	7.2
25	307	18.8	22	12	1.3
40	100	30.9	69	23	9.3
41	9,273	11.3	31	5	1.4
43	50,545	13.1	39	7	3.3
50	124,288	20.1	56	7	4.7
51	418	30.0	152	12	11.6
60	3,820	20.1	48	9	6.1
70	19	19.1	25	15	3.0
82	120,218	14.2	46	6	4.3
86	1,109	18.2	48	13	3.0
100	98,802	27.1	104	10	7.5
110	14,385	58.8	226	5	27.8
111	2,705	81.1	189	14	22.8
130	471	44.3	232	12	29.1
240	6,387	40.0	307	9	18.4
241	8	26.3	74	14	18.2
245	124,357	89.8	842	9	54.9
250	17,813	17.6	281	7	14.5
260	124,356	47.6	283	8	20.0
300	124,134	26.2	184	7	11.2
350	32,007	11.0	69	7	3.2
400	119	49.0	99	21	15.1
410	2,515	59.3	194	16	24.0
411	20	45.4	76	23	15.2
440	10,279	43.9	171	11	14.9
490	28,715	43.7	296	7	22.8
500	64,530	60.6	1,208	8	54.9
501	68	99.4	445	50	61.2
502	492	40.6	455	17	26.7
504	59,632	33.2	351	17	9.4
505	4,822	261.0	1,470	14	75.4
520	7,024	138.2	609	13	46.2
600	17,280	41.6	228	10	16.0
610	7,029	44.1	248	10	17.6
611	140	51.4	133	14	21.8
630	1,186	39.1	95	9	15.6
650	158,686	33.1	112	5	13.9
651	27,936	43.7	119	9	15.8
700	43,696	35.4	220	9	13.1
710	15,480	54.5	349	10	24.2
711	349	79.0	248	14	31.1
730	1,420	33.1	268	9	18.9
740	5,634	40.4	199	7	19.9
800	87	69.1	115	40	17.0
810	8,706	85.4	259	25	26.2
811	28	87.3	132	44	22.1
840	2,499	47.2	161	13	16.4

Table 5(b). Field Length Statistics: Volume 3, 1975–76

Field Tag	*Total Occurrence*	*Mean Length*	*Maximum Length*	*Minimum Length*	*Standard Deviation of Length*
1	146,199	13.3	23	13	1.2
8	146,199	41.0	41	41	.0
15	43,808	11.9	56	7	2.7
20	94,106	20.0	182	8	8.6
25	88	18.9	21	12	1.7
40	60	31.1	73	23	8.9
41	9,700	11.5	25	7	1.5
43	63,537	13.0	39	9	3.4
50	146,116	20.3	71	6	5.0
51	293	34.7	168	12	25.0
60	4,070	19.3	48	9	6.5
70	28	19.0	24	12	3.4
82	132,314	14.5	44	7	4.3
86	987	17.7	43	13	2.6
100	110,007	26.6	171	9	7.4
110	17,475	57.8	323	9	27.7
111	3,282	81.0	197	27	22.2
130	435	42.0	210	12	21.9
240	6,413	40.1	249	7	18.8
241	8	27.1	41	14	10.1
245	146,205	97.3	763	9	58.0
250	20,449	17.3	329	7	14.9
260	146,190	50.1	396	9	21.8
300	146,026	27.6	191	7	12.3
350	10,200	11.2	65	7	3.9
400	74	50.4	104	22	16.6
410	1,479	57.5	178	17	22.1
411	11	58.3	122	22	28.2
440	12,901	45.6	195	9	17.4
490	36,610	44.4	324	7	23.0
500	94,075	51.2	1,208	5	49.8
501	131	122.0	497	26	72.2
502	992	37.1	455	18	17.9
504	72,852	34.3	258	17	10.1
505	5,712	240.6	1,418	11	252.6
520	5,089	130.3	339	20	41.3
600	21,741	41.9	167	8	17.1
610	8,375	45.0	278	10	17.9
611	204	52.5	173	14	23.0
630	1,548	39.9	232	10	17.0
650	190,044	34.1	123	8	14.3
651	32,670	41.9	112	5	14.3
700	54,888	32.8	232	7	13.3
710	19,346	54.3	281	10	23.9
711	455	75.1	195	14	25.0
730	1,644	32.2	150	8	16.9
740	8,136	45.3	183	9	21.4
800	46	72.0	122	41	16.7
810	11,912	87.3	327	29	29.1
811	56	97.0	198	44	30.1
840	2,611	47.2	142	5	17.4

Table 5(c). Field Length Statistics: Volume 4, 1976–77

Field Tag	*Total Occurrence*	*Mean Length*	*Maximum Length*	*Minimum Length*	*Standard Deviation of Length*
1	176,127	13.3	23	13	1.1
8	176,127	41.0	41	41	.0
15	64,046	11.9	69	5	2.5
20	118,307	20.3	107	8	8.8
25	102	19.0	21	12	1.4
40	45	31.7	58	26	7.4
41	11,972	11.7	32	8	1.8
43	79,882	12.9	39	6	3.1
50	175,981	20.2	61	8	5.1
51	395	31.7	343	12	20.8
60	3,829	19.8	49	10	7.0
70	7	21.1	22	20	.6
82	147,016	14.5	46	6	4.3
86	799	17.9	27	13	2.6
100	129,289	26.3	102	8	7.2
110	21,559	57.9	322	8	28.6
111	4,084	81.3	224	14	23.0
130	500	42.1	232	12	21.8
240	7,165	39.2	201	7	18.2
241	12	28.0	44	13	9.5
245	176,134	100.1	750	8	59.2
250	26,587	17.6	298	8	15.0
260	176,116	50.7	387	9	22.2
300	175,964	27.9	195	8	12.6
350	9,594	11.1	63	7	3.8
400	188	44.1	87	20	16.3
410	1,375	61.5	212	16	25.8
411	21	47.5	93	25	17.7
440	15,355	46.1	173	10	17.6
490	47,811	44.8	328	7	24.1
500	113,192	49.0	1,012	8	46.2
501	199	129.6	755	24	98.1
502	1,218	35.7	192	17	17.2
504	87,456	34.5	314	11	10.0
505	7,019	223.1	1,448	5	240.4
520	6,304	126.3	378	20	39.2
600	25,284	41.9	194	10	17.4
610	10,924	45.3	195	9	17.4
611	286	52.3	261	14	25.5
630	1,680	40.4	119	10	16.3
650	234,473	35.0	138	5	14.3
651	39,742	41.6	117	9	14.3
700	69,975	31.7	178	7	12.8
710	24,113	54.4	251	9	23.9
711	571	76.0	298	19	26.2
730	1,751	32.5	268	9	18.1
740	10,440	46.2	180	8	21.5
800	246	73.8	154	35	17.3
810	15,631	90.7	321	21	31.6
811	45	84.9	166	42	25.1
830	1	86.0	86	86	.0
840	3,138	48.6	185	16	17.9

Table 5(d). Field Length Statistics: Volume 5, 1977–78

Field Tag	*Total Occurrence*	*Mean Length*	*Maximum Length*	*Minimum Length*	*Standard Deviation of Length*
1	173,978	13.4	23	13	1.3
8	173,978	41.0	41	41	.0
15	69,513	12.0	55	6	2.6
20	114,541	20.5	145	8	9.0
25	341	17.4	24	10	1.2
40	47	27.8	49	19	4.3
41	12,984	12.0	521	8	4.9
43	78,401	12.8	31	8	3.0
50	173,751	20.0	57	8	5.1
51	316	30.8	168	10	21.1
60	2,896	19.6	49	8	6.8
70	13	22.3	25	20	1.5
82	120,734	14.3	50	6	4.3
86	845	19.0	48	10	3.4
100	129,134	26.2	133	9	7.2
110	20,084	56.6	224	10	26.6
111	3,496	82.2	210	14	23.0
130	482	42.9	210	10	25.9
240	6,850	39.0	170	8	17.9
241	12	21.3	31	11	5.7
245	173,977	99.4	974	9	59.2
250	28,515	17.3	292	7	14.1
260	173,965	50.5	385	7	22.7
300	173,896	28.1	159	8	12.2
350	9,772	10.7	62	7	3.3
400	267	38.5	92	15	13.9
410	670	60.6	153	13	24.4
411	2	46.0	62	30	16.0
440	15,785	46.1	257	10	17.9
490	47,196	44.0	307	6	24.2
500	109,646	48.0	864	7	44.1
501	273	132.4	755	20	109.4
502	1,367	36.1	178	10	21.1
504	83,812	34.7	438	8	10.2
505	7,222	204.2	1,410	11	216.8
520	9,370	128.2	507	20	42.3
600	27,680	43.8	149	10	18.0
610	10,952	45.3	190	9	17.5
611	258	51.8	149	13	22.6
630	1,833	40.4	255	9	17.6
650	229,890	35.4	146	8	14.5
651	37,758	41.6	130	9	14.4
700	69,482	31.3	198	7	12.7
710	22,788	54.4	248	9	24.0
711	481	75.9	177	14	26.5
730	1,633	33.4	168	8	17.7
740	10,065	46.5	248	8	22.3
800	260	69.7	160	33	17.5
810	14,792	92.5	296	5	32.8
811	57	101.8	212	42	34.3
830	7	46.7	67	36	9.6
840	3,123	47.6	203	13	18.0

Table 6. Field Mean Lengths: Volumes 2, 3, 4, and 5

Field Tag	*V.2*	*V.3*	*V.4*	*V.5*
1		13	13	13
8		41	41	41
15	12	12	12	12
20	18	20	20	21
25	19	19	19	17
40	31	31	32	28
41	11	12	12	12
43	13	13	13	13
50	20	20	20	20
51	30	35	32	31
60	20	19	20	20
70	19	19	21	22
82	14	15	15	14
86	18	18	18	19
100	27	27	26	26
110	59	58	58	57
111	81	81	81	82
130	44	42	42	43
240	40	40	39	39
241	26	27	28	21
245	90	97	100	99
250	18	17	18	17
260	48	50	51	51
300	26	28	28	28
350	11	11	11	11
400	49	59	44	39
410	59	58	62	61
411	45	58	48	46
440	44	46	46	46
490	44	44	45	44
500	61	51	49	48
501	99	122	130	132
502	41	37	36	36
504	33	34	35	35
505	261	241	223	204
520	138	130	126	128
600	42	42	42	44
610	44	45	45	45
611	51	53	52	52
630	39	40	40	40
650	33	34	35	35
651	44	42	42	42
700	35	33	32	31
710	55	54	54	54
711	79	75	76	76
730	33	32	33	33
740	40	45	46	47
800	69	72	74	70
810	85	87	91	93
811	87	97	85	102
830	———	———	86	47
840	47	47	49	48

Table 7. Number of Field Tag Length Changes

Volume No.	*No. Field Tags*	*No. Field Tags That Increased versus Prior Volume*	*No. Field Tags the Same Versus Prior Volume*	*No. Field Tags That Decreased versus Prior Volume*	*New Field Tags*	*Deleted Field Tags*
1 Ser. 1	48	—	—	—	—	—
2	49	44	0	3	+2	−1
3	49	21	14	14	0	0
4	50	18	19	12	+1	0
5	50	11	22	17	0	0

eleven fields have unchanged lengths, and three have a decreasing magnitude of change. The zero change and decreasing magnitude of change indicate stabilization. Exceptions to the stabilization trend are the mean field lengths for tags 110 and 700, which decreased steadily; for tag 600, which increased in volume 5; and for tags 82, 130, and 490, which fluctuated.

Although three fields changed length by twenty or more characters, they are fields that occur in a relatively small percentage of records. Tag 400 had a mean field length difference of twenty characters between volumes 3 and 5; yet the 400 tag occurred in only 0.2 percent of the records. Tag 505 had a field length difference of fifty-seven characters (the largest character change for any field) between volumes 2 and 5, and yet it occurred in only 4.2 percent of the records in volume 5. Tag 830 had a mean field length difference of thirty-nine characters between volumes 4 and 5, but it occurred in only .004 percent (seven) of the records in volume 5, and it occurred only once in volume 4. In table 7 one can see that the numbers of tags for which the mean length remains unchanged relative to the prior volume is zero for volume 2. However, one can see the trend toward stabilization by calculating the percentage of tags for which the mean length remains unchanged over the next three volumes; it increased from 28 percent in volume 3 to 38 percent in volume 4 to 44 percent in volume 5.

Table 6 provides mean lengths for MARC field tags for volumes 2, 3, 4, and 5. These are presented to permit easy visual comparisons. If these data are used for estimating file sizes, etc., they must be considered in relation to the field tag occurrence data (tables 2[a]–2[d] and table 3). A significant increase in the length of a field that occurs seldom may have less impact than a smaller increase in a field that occurs in a high percentage of all records. Table 7 indicates the changes in number of field tags that have occurred from volume to volume. It indicates the number of field tags per year and the number that have increased in length, stayed the same, or decreased relative to the prior year. We also compared the number that increased in length, stayed the same, or decreased relative to volume 2. In volume 3, twenty-one increased in

length; in volumes 4 and 5, twenty-four increased. The number that stayed the same in volumes 3, 4, and 5 relative to volume 2 were fourteen, eleven, and eight. The number that decreased versus volume 2 for volumes 3, 4, and 5 were fourteen, fourteen, and seventeen.

DISTRIBUTION OF RECORDS BY DEWEY DECIMAL DIVISION

Distribution data for MARC records by Dewey Decimal Divisions are presented in tables 8(a) and 8(b). The division numbers in the "DDD" column represent the digits preceding the decimal point in the Dewey numbers. They are grouped in increments of 100 ranging from 0 to 990. All nonnumeric Deweys are listed as "others." The "count" column indicates the total occurrences in all records for each range of 100. The "Percentage" column indicates the percent of occurrence in records that contain Dewey numbers. As can be seen in tables 2(a)–2(d), in volume 2 Deweys appear in 96.6 percent of the records; in volume 3, in 90.4 percent of the records; in volume 4, in 83.4 percent of the records; and in volume 5, in 69.3 percent.

Table 8(a). Dewey Decimal Division Distribution: Volume 2, 1974–75 (Six Months), and Volume 3, 1975–76

Volume 2, 1974–75 (6 Months)			*Volume 3, 1975–76*		
DDD	*Count*	*Percent*	*DDD*	*Count*	*Percent*
0– 99	2,254	3.494	0– 99	5,506	4.161
100–199	1,573	2.438	100–199	3,660	2.766
200–299	2,915	4.518	200–299	6,073	4.590
300–399	18,281	28.337	300–399	40,779	30.820
400–499	812	1.259	400–499	1,707	1.290
500–599	4,934	7.648	500–599	9,846	7.441
600–699	8,208	12.723	600–699	18,631	14.081
700–799	6,714	10.407	700–799	10,548	7.972
800–899	8,114	12.577	800–899	18,643	14.090
900–999	9,155	14.191	900–999	15,035	11.363
Others	1,553	2.407	Others	1,886	1.425

Table 8(b). Dewey Decimal Division Distribution: Volume 4, 1976–77, and Volume 5, 1977–78

Volume 4, 1976–77			*Volume 5, 1977–78*		
DDD	*Count*	*Percent*	*DDD*	*Count*	*Percent*
0– 99	6,110	4.156	0– 99	5,063	4.194
100–199	3,929	2.672	100–199	2,739	2.269
200–299	6,718	4.570	200–299	5,291	4.382
300–399	45,866	31.198	300–399	35,041	29.023
400–499	2,111	1.436	400–499	1,656	1.372
500–599	10,641	7.238	500–599	8,356	6.921
600–699	21,421	14.571	600–699	16,827	13.937
700–799	12,055	8.200	700–799	10,967	9.084
800–899	20,813	14.157	800–899	18,380	15.224
900–999	15,088	10.263	900–999	11,466	9.497
Others	2,264	1.540	Others	4,948	4.098

The Dewey Division groups that ranked in the top ten for at least one of the four volumes are listed in table 9. The position of a group is indicated by a number from 1 to 10 (first to tenth). Eight of the Dewey groups ranked in the top ten in all four volumes, one in three volumes, two in two volumes, and two in one volume.

Table 9. Top Dewey Decimal Division Groups

DDD Group	*V.2*	*V.3*	*V.4*	*V.5*	*Times in Top Ten*
300	6	5	4	5	4
320	8	7			2
330	1	1	1	1	4
340	2	2	2	3	4
360		10	8	9	3
370	9	9	9	10	4
610	5	4	3	4	4
620	10	8	7	8	4
790				7	1
810	7	6	6	6	4
820	4	3	5	2	4
910	3				2
970			10		1

LC CLASS NUMBER OCCURRENCES AND LENGTHS

Our analysis of LC classes is based on the 2-alpha character portions of LC class numbers, which are then collapsed to the 1-alpha character portion of the LC class number. The 2-alpha tables require five times as much space as the single-character tables; therefore, we have included only the single-character tables in this paper. Interested parties may write for copies of the 2-alpha character tables. Because many libraries do not maintain juvenile collections or treat them differently, we have provided separate and combined data for juvenile and nonjuvenile records for volumes 3, 4, and 5. The data for volume 2 were generated at an earlier date and do not contain a juvenile/nonjuvenile distinction.

Tables 10(a)–(c) provide a breakdown by LC single-character class giving absolute numbers of records and percentage of records for nonjuvenile and juvenile records as well as the combined figures.

No single-character category accounts for more than 10 percent of the records except for H (social sciences) and P (language and literature). Four of the categories account for fewer than 1 percent of the records in volumes 2, 3, 4, and 5. The remaining sixteen categories range from 1.0 percent to 8.5 percent occurrence. Table 11 summarizes the four-volume LC class occurrence statistics.

Table 10(a). LC Class Occurrence Distribution: Volume 3, 1975–76

LC Class	Nonjuvenile		Juvenile		Combined	
	Count	Percent	Count	Percent	Count	Percent
A–AZ	667	.4	19	.0	686	.4
B–BZ	10,184	6.9	193	.1	10,377	7.0
C–CZ	1,715	1.1	50	.0	1,765	1.2
D–DZ	10,751	7.3	225	.1	10,976	7.5
E–FZ	6,181	4.2	376	.2	6,557	4.4
G–GZ	4,083	2.7	532	.3	4,615	3.1
H–HZ	22,543	15.4	261	.1	22,804	15.5
J–JZ	3,853	2.6	26	.0	3,879	2.6
K–KZ	5,663	3.8	14	.0	5,677	3.8
L–LZ	7,423	5.0	13	.0	7,436	5.0
M–MZ	1,476	1.0	65	.0	1,541	1.0
N–NZ	5,208	3.5	63	.0	5,271	3.6
P–PZ	24,830	16.9	3,311	2.2	28,141	19.2
Q–QZ	10,828	7.4	518	.3	11,346	7.7
R–RZ	6,947	4.7	64	.0	7,011	4.7
S–SZ	2,678	1.8	109	.0	2,787	1.9
T–TZ	9,578	6.5	343	.2	9,921	6.7
U–UZ	882	.6	4	.0	886	.6
V–VZ	452	.3	11	.0	463	.3
W–ZZ	3,950	2.7	25	.0	3,975	2.7
TOTAL	139,892	95.6	6,222	4.2	146,114	99.9

Table 10(b). LC Class Occurrence Distribution: Volume 4, 1976–77

LC Class	Nonjuvenile		Juvenile		Combined	
	Count	Percent	Count	Percent	Count	Percent
A–AZ	722	.4	27	.0	799	.4
B–BZ	11,622	6.5	249	.1	11,871	6.7
C–CZ	1,791	1.0	35	.0	1,826	1.0
D–DZ	11,372	6.4	240	.1	11,612	6.5
E–FZ	6,911	3.9	367	.2	7,278	4.1
G–GZ	5,291	3.0	626	.3	5,917	3.3
H–HZ	28,534	16.2	238	.1	28,772	16.3
J–JZ	4,784	2.7	21	.0	4,805	2.7
K–KZ	6,214	3.5	11	.0	6,225	3.5
L–LZ	11,154	6.3	22	.0	11,176	6.3
M–MZ	1,994	1.1	65	.0	2,059	1.1
N–NZ	6,181	3.5	63	.0	6,244	3.5
P–PZ	31,361	17.8	3,596	2.0	34,957	19.8
Q–QZ	11,704	6.6	618	.3	12,322	6.9
R–RZ	8,446	4.7	85	.0	8,531	4.8
S–SZ	3,210	1.8	195	.1	3,405	1.9
T–TZ	11,498	6.5	437	.2	11,935	6.7
U–UZ	1,007	.5	19	.0	1,026	.5
V–VZ	606	.3	16	.0	622	.3
W–ZZ	4,559	2.5	30	.0	4,589	2.6
TOTAL	169,011	95.9	6,960	3.9	175,971	99.9

Table 10(c). LC Class Occurrence Distribution: Volume 5, 1977–78

LC Class	*Nonjuvenile*		*Juvenile*		*Combined*	
	Count	*Percent*	*Count*	*Percent*	*Count*	*Percent*
A–AZ	631	.3	27	.0	658	.3
B–BZ	11,525	6.6	268	.1	11,793	6.7
C–CZ	1,457	.8	30	.0	1,487	.8
D–DZ	11,438	6.5	241	.1	11,679	6.7
E–FZ	5,553	3.1	521	.2	6,074	3.4
G–GZ	5,405	3.1	734	.4	6,139	3.5
H–HZ	24,446	14.0	252	.1	24,698	14.1
J–JZ	4,217	2.4	24	.0	4,241	2.4
K–KZ	5,739	3.2	14	.0	5,753	3.3
L–LZ	11,756	6.7	13	.0	11,769	6.7
M–MZ	1,961	1.1	68	.0	2,029	1.1
N–NZ	7,183	4.1	99	.0	7,282	4.1
P–PZ	34,562	19.8	6,406	3.6	40,968	23.5
Q–QZ	10,281	5.9	611	.3	10,892	6.2
R–RZ	7,400	4.2	83	.0	7,483	4.3
S–SZ	3,730	2.1	144	.0	3,874	2.2
T–TZ	10,199	5.8	375	.2	10,574	6.0
U–UZ	971	.5	23	.0	994	.5
V–VZ	505	.2	20	.0	525	.3
W–ZZ	4,795	2.7	29	.0	4,824	2.7
TOTAL	163,754	94.1	9,982	5.7	173,736	99.8

Table 11. LC Classifications Percent Occurrences of Volumes 2, 3, 4, and 5

LC Class	*V.2*	*V.3*	*V.4*	*V.5*
A–AZ	0.3	0.4	0.4	0.3
B–BZ	6.8	7.0	6.7	6.7
C–CZ	1.0	1.2	1.0	0.8
D–DZ	6.3	7.5	6.5	6.7
E–FZ	6.2	4.4	4.1	3.4
G–GZ	3.3	3.1	3.3	3.5
H–HZ	13.4	15.5	16.3	14.1
J–JZ	2.6	2.6	2.7	2.4
K–KZ	4.2	3.8	3.5	3.3
L–LZ	4.3	5.0	6.3	6.7
M–MZ	1.1	1.0	1.1	1.1
N–NZ	3.2	3.6	3.5	4.1
P–PZ	20.2	19.2	19.8	23.5
Q–QZ	8.5	7.7	6.9	6.2
R–RZ	4.6	4.7	4.8	4.3
S–SZ	1.9	1.9	1.9	2.2
T–TZ	6.7	6.7	6.7	6.0
U–UZ	0.5	0.6	0.5	0.5
V–VZ	0.3	0.3	0.3	0.3
W–ZZ	2.6	2.7	2.6	2.7

Record lengths within each of the single-character LC classes are provided for volumes 3, 4, and 5 in tables 12(a)–12(c). In each table, column 1 indicates LC class; column 2 indicates the number of occurrence of records within the class; column 3 indicates the mean length of records in that class; column 4, standard deviation; column 5, maximum length; and column 6, minimum length.

The overall average length of a MARC record is 681.9 for volume 3, 689.7 for volume 4, and 685.2 for volume 5. The records for scientific monographs (Q, R, S, and T) are somewhat longer than average. The mean length for these records in volume 3 was 713, in volume 4 it was 719, and in volume 5 it was 727. On the other hand, Language and Literature records (P) are shorter than average (585 for volume 3, 595 in volume 4, and 588 in volume 5). Table 13 provides a summary of record lengths for LC classes in volumes 3, 4, and 5.

CONCLUSION

The basic statistics and analyses presented in this paper should prove useful for organizations working with MARC records and should help them in estimating file sizes, deciding on file subsets, etc. The number of MARC records per year increased between volume 1 and volume 4 from 52,294 to 176,127 and decreased slightly in volume 5 to 173,978. Mean record lengths increased over the three volumes and leveled off with volume 5. There are now fifty field tags (excluding tags 1 and 8,

Table 12(a). LC Class versus Record Length Distribution: Volume 3, 1975–76

LC Class	*Number of Occurrences*	*Mean Length*	*Standard Deviation of Length*	*Maximum Length*	*Minimum Length*
A–AZ	686	733.3	201.9	1,785	267
B–BZ	10,377	644.4	187.9	2,035	311
C–CZ	1,765	641.5	183.3	1,825	279
D–DZ	10,976	691.5	174.3	2,011	333
E–FZ	6,557	723.3	190.9	1,915	369
G–GZ	4,615	677.6	204.1	2,017	338
H–HZ	22,804	719.2	190.2	2,008	305
J–JZ	3,879	733.6	188.5	2,006	356
K–KZ	5,677	752.6	174.5	1,911	361
L–LZ	7,436	708.0	177.0	2,029	362
M–MZ	1,541	648.2	155.9	1,911	322
N–NZ	5,271	676.3	162.1	1,693	319
P–PZ	28,141	585.2	186.1	2,012	248
Q–QZ	11,346	733.4	201.3	2,009	335
R–RZ	7,011	727.5	205.8	2,030	346
S–SZ	2,787	710.1	203.5	1,797	320
T–TZ	9,921	680.5	200.1	2,022	344
U–UZ	886	720.0	190.5	2,024	388
V–VZ	463	680.2	189.3	1,791	387
W–ZZ	3,975	738.0	182.7	1,818	381
TOTAL	146,114	681.9	196.1	2,030	248

*Table 12(b). LC Class versus Record Length Distribution: Volume 4, 1976–77**

LC Class	*Number of Occurrences*	*Mean Length*	*Standard Deviation of Length*	*Maximum Length*	*Minimum Length*
A–AZ	91	787.9	263.1	1,770	347
B–BZ	1,797	659.3	185.3	1,919	341
C–CZ	216	674.8	183.9	1,591	360
D–DZ	1,603	708.6	172.6	1,836	359
E–FZ	1,142	751.3	206.2	1,935	363
G–GZ	932	690.3	182.0	1,696	365
H–HZ	3,680	728.0	183.1	1,950	367
J–JZ	656	730.2	173.4	1,966	403
K–KZ	963	774.5	187.0	1,961	404
L–LZ	1,646	717.7	180.7	1,971	400
M–MZ	257	662.1	141.3	1,520	403
N–NZ	813	702.2	176.1	1,648	387
P–PZ	4,779	595.0	178.9	1,626	296
Q–QZ	1,608	747.4	199.6	1,881	409
R–RZ	1,173	740.3	206.4	1,905	401
S–SZ	418	701.8	204.6	1,993	396
T–TZ	1,560	685.4	201.2	1,741	338
U–UZ	116	707.9	181.5	1,886	379
V–VZ	79	691.4	159.0	1,158	401
W–ZZ	540	758.8	174.4	1,523	389
TOTAL	24,069	693.4	194.4	1,993	296

*Based on a sample consisting of the last eight issues of Volume 4.

Table 12(c). LC Class versus Record Length Distribution: Volume 5, 1977–78

LC Class	*Number of Occurrences*	*Mean Length*	*Standard Deviation of Length*	*Maximum Length*	*Minimum Length*
A–AZ	658	756.9	215.4	1,504	322
B–BZ	11,793	653.7	178.6	2,010	300
C–CZ	1,487	649.4	163.3	1,764	284
D–DZ	11,679	696.0	175.6	2,031	303
E–FZ	6,074	751.3	199.9	1,984	355
G–GZ	6,139	697.9	188.5	2,004	311
H–HZ	24,698	727.6	185.8	2,050	313
J–JZ	4,241	728.2	180.4	1,970	334
K–KZ	5,753	764.8	178.0	1,992	356
L–LZ	11,769	712.0	177.9	2,019	312
M–MZ	2,029	664.6	153.1	1,600	335
N–NZ	7,282	702.5	173.7	1,748	322
P–PZ	40,968	588.3	179.2	2,034	247
Q–QZ	10,892	745.6	203.8	2,027	302
R–RZ	7,483	736.7	209.2	1,988	347
S–SZ	3,874	731.7	209.7	1,764	325
T–TZ	10,574	690.9	195.6	1,966	336
U–UZ	994	723.8	175.4	1,761	345
V–VZ	525	720.5	190.8	1,622	403
W–ZZ	4,824	749.9	180.9	1,738	360
TOTAL	173,736	685.2	194.6	2,050	247

Table 13. LC Class Mean Lengths: Volumes 3, 4, and 5

LC Class	*V.3*	*V.4*	*V.5*
A–AZ	733.3	787.9	756.9
B–BZ	644.4	659.3	653.7
C–CZ	641.5	674.8	649.4
D–DZ	691.5	708.6	696.0
E–FZ	723.3	751.3	751.3
G–GZ	677.6	690.3	697.9
H–HZ	719.2	728.0	727.6
J–JZ	733.6	730.2	728.2
K–KZ	752.6	774.5	764.8
L–LZ	708.0	717.7	712.0
M–MZ	648.2	662.1	664.6
N–NZ	676.3	702.2	702.5
P–PZ	585.2	595.0	588.3
Q–QZ	733.4	747.4	745.6
R–RZ	727.5	740.3	736.7
S–SZ	710.1	701.8	731.7
T–TZ	680.5	685.4	690.9
U–UZ	720.0	707.9	723.8
V–VZ	680.2	691.4	720.5
W–ZZ	738.0	758.8	749.9
TOTAL	681.9	693.4	685.2

which are invariant) in MARC tapes. Of the fifty, only five occur more than 1.2 times per record, and slightly more than half occur once per record. The length of MARC records has increased from a mean of 636 in volume 1 to 690 in volume 4 and decreased slightly to 685 in volume 5. The mean lengths of fields appears to be stabilizing by volume 5. Only six fields appeared to be unstable. Dewey Decimal Classification numbers are being applied to a smaller *percentage* of records, although this represents a larger *number* of records because the total number of records has increased. Looking at LC categories in terms of the twenty-three major divisions, we found that no single category accounts for more than 10 percent of the records except for H and P, that five categories accounted for fewer than 1 percent, and that the remaining categories each accounts for from 1 percent to 9 percent. It remains to be seen whether the observations and trends related to the five volumes analyzed will continue over future volumes of MARC.

REFERENCES

1. Martha E. Williams and Gordon J. Shefner, "Data Element Statistics for the MARC II Data Base," *Journal of Library Automation* 9, no.2:89–100 (June 1976).
2. Gerald L. Swanson, *Selected Statistics Based on L.C. MARC Bibliographic Records* (Technical Note no.2 [Systems Office, Columbia Univ. Libraries, Sept. 1970]).
3. R. K. Primmer, comp., *OCLC-MARC Distribution Service* (version 1.0) *Specifications for Magnetic Tapes Containing Catalog Records for Monographs and Serials* (Columbus, Ohio: OCLC, Inc., 1 May 1974).

4. Library of Congress, MARC Development Office, *Books, A MARC Format: Specifications for Magnetic Tapes Containing Catalog Records for Books* (Washington, D.C.: The Library, 1972).
5. Gerald L. Swanson, *Dewey to L C Conversion Tables* (New York: CCM Information Corp., 1972).

Martha E. Williams is director of the Information Retrieval Research Laboratory (IRRL) within the Coordinated Science Laboratory at the University of Illinois. Among her current projects are a Council on Library Resources (CLR) grant to continue developing statistics for the MARC data base. **Scott E. Preece** is a Ph.D. candidate in computer science at the University of Illinois and a senior programmer in the IRRL. He has been active in information retrieval research since 1968 and is responsible for coordinating the programming activities of IRRL. **Steven W. Barth** is a candidate for the M.S. degree in computer science at the university and has worked as a graduate research assistant in the IRRL for the past two years.

Expanded Subject Access to Reference Collection Materials*

William H. MISCHO: Iowa State University, Ames.

This paper reports a computer-assisted printed index emphasizing expanded subject access to the more than 6,000 titles in the Iowa State University Library reference collection. The index displays abbreviated-length records and is designed to complement existing catalogs with a minimum of duplication of information. The limitations of subject access to reference materials via Library of Congress Subject Headings is discussed and a system for assigning subject descriptors to reference titles is described. Central to the project are custom-designed software routines that expedite data entry, minimize keyboarding, and use a combinatorial-based method to produce multiple entry points from single input strings.

INTRODUCTION

Library automation activities directed to reference and information department services have principally involved on-line systems designed to enhance retrieval of bibliographic citations contained in secondary abstracting and indexing (A & I) services. Little attempt has been made to expand access to information contained in other types of reference collection materials—such as dictionaries, handbooks, bibliographies, etc. Several studies have investigated the feasibility of automating search procedures to biographical reference books or other discrete subsets of a reference collection.[1-4] It should be noted that these studies were performed in laboratory environments and did not involve an operational system in a library setting.

Catalog use studies have shown that different library users have different needs.[5] Reference collection materials are approached by users who are, in most cases, interested in items that can fill a specific informational need, rather than with locating known-item author or title citations. Patrons typically will be seeking a reference tool that can provide either an on-the-spot piece of information, such as an address, definition, or statistic, or will be interested in a bibliography of primary source

*This research was performed under a Council on Library Resources grant (CLR 677).

citations in a specific subject area. These user needs demand a system capable of providing in-depth access to specific types of reference tools such as handbooks, A & I services, or dictionaries, associated with specific subject areas. This is information retrieval one step removed from the typical machine-assisted reference service in that what is desired is a list of candidate reference titles for meeting the information need. In the case of A & I services, for example, the patron is interested in identifying from among the myriad A & I services those containing citations in a specific area—for example, archaeology or criminology.

However, the standard source of subject access to reference materials in college and research libraries—Library of Congress Subject Headings (LCSH)—is unable to meet these user needs completely. The inadequacies of subject access through LCSH, that is, outmoded descriptor terms, inconsistent forms of headings, lack of term specificity, and lack of indexing exhaustivity, are accentuated by the paramount importance of the subject approach to reference collection materials.[6-11]

Enhanced access using terms taken from the titles of reference works can be easily shown to be of limited effectiveness. The title-derivative or KWIC/KWOC approach has serious retrieval limitations.[12] Thus, the information contained in the machine-readable network records or MARC-related products is unable to provide the in-depth subject access terms required for reference titles, and the need must be met by local augmentation.

Although augmentation of monographic titles for retrieval in an online environment has been successfully demonstrated by Atherton, the nature of reference requests—that is, for a type of tool covering a subject area—suggests that a Boolean search capability in the retrieval of reference titles is not necessary.[13] Malinconico has suggested that increasing the number of precoordinated subject access points in a printed index will tend to duplicate the document retrieval recall performance of an interactive system.[14] This paper reports on a project whose purpose was the design and construction of a computer-produced printed index to provide the needed expanded subject access to titles in the Iowa State University Library Reference Collection. The resulting printed format index, featuring exhaustive subject analysis and allowing the identification of the form of the reference tool, has proved effective in providing the needed access to reference materials. Particularly emphasized was the development of custom-designed mechanized indexing procedures that expedite data input and product multiple entry points by software manipulation of indexer-assigned subject descriptor strings.

OBJECTIVES

This project focused on the development of mechanized indexing procedures that can be applied in the production of printed indexes de-

signed to improve access to materials in academic library reference collections. Particular emphasis was placed on:

1. The identification of optimum access points for reference titles and the design of an indexing language—specifically testing the efficacy of an uncontrolled vocabulary index;
2. The development of streamlined data preparation and data entry procedures;
3. Testing the feasibility of a printed format index and attempting to determine the number of data elements needed in index entries;
4. In conjunction with 1 and 2, the development of custom-designed software utilizing the computer's facility in manipulating word strings to produce index entries.

PROCEDURES

The Iowa State University Library Reference Collection is a multidisciplinary, centralized collection of more than 6,000 titles. Initially, titles from the reference collection were examined and prototype indexes produced and studied for form and type of entry points and index format. It was decided that the printed index would display an abbreviated bibliographic record consisting of lead term, title, publication date, and call number. Palmer has suggested that a short-entry catalog with the abovementioned fields plus author's name will function effectively in an academic library setting.[15] However, in this application the author's name is a less frequently used access point and a less important element for identifying the work. The index functions as a finding aid, as opposed to a source of full bibliographic information, and complements the library's other bibliographic files (card catalog and *Serials Catalog)* with a minimum of duplication of information.

A hybrid indexing language was devised, with entry points for individual titles to include: (1) specific descriptor terms taken from the table of contents and body of the work, (2) title entries, (3) significant words and phrases appearing in the title or subtitle, (4) author, compiler, or editor on an infrequent basis, and (5) form and content designations, identifying the type of reference material, e.g., encyclopedia, atlas, biographical information, etc. Title and personal name access points, although secondary in importance to exhaustive subject analysis, were judged to be approached frequently enough for inclusion as index entry points.

Computer-assisted indexing techniques were developed to facilitate data input and to allow the generation of multiple access points from single input phrases by rearranging the order of the elements within multiword descriptors. These string manipulation procedures rely on data input format recognition software that minimizes keyboarding of duplicate data elements. The abbreviated record length in conjunction with the mechanized phrase manipulation software facilitated the execu-

tion of the retrospective conversion of collection titles and reduces the time dedicated to data preparation, data entry, and editing activities in the ongoing processing of new titles for cumulative supplements.

While the index is intended to be used primarily by reference librarians, it has proven to be easily approachable by patrons and other staff. For the sake of speed and convenience, the index uses a one-step approach; that is, an attempt is made to allow complete searches with one look-up in the index. The incorporation of multiple access routes for individual titles by providing entries under all significant elements of composite subjects and, in particular, the form designator(s) addresses the problem of lack of user perseverance in catalog searches. In one study of an academic library, more than 50 percent of catalog users searched at only one entry point and then discontinued the search.[16] Another study found that multiple access greatly increases the chances of success in non-known-item searches.[17]

INDEX FORMAT AND RECORD STRUCTURE

The index entries, illustrated in figure 1, consist of four fixed-length fields—lead term (designated as "subject"), title, date of publication, and call number—arranged in a title-a-line 132-character KWOC-style format. These elements are normally sufficient to locate a title on the shelf and can also serve as a link-key to the card catalog or *Serials Catalog* at which a patron can obtain more complete bibliographic information concerning a title. The short-entry format is consistent with the fact that the primary focus of the index is to provide expanded subject access to reference materials.The KWOC-style format offers index readability, and the single-line entries proved necessary to keep the size of the index to a minimum. The display line presently allows for 38 characters for subject descriptor and 57 characters for title. Approximately 5 percent of the subject descriptors and 9 percent of the titles are truncated at the output stage by the field length parameters. In the very few cases where index entries are not readily identifiable from the visible terms, abbreviations have been employed to overcome the problem.

The master records are of variable length and contain title, descriptor terms, call numbers, date information, and LC card number fields, all in varying-length format. The abbreviated record format, of course, greatly expedited the retrospective conversion of reference titles into a machine-readable form. So that the short entry records could be linked to their MARC format analogs, the LC card number was included in those records for which the card number was readily available. This would allow augmenting the MARC record by adding descriptor strings to the 690 (local subject headings) field. The average length of the added descriptors field is approximately 115 characters. The compact data entry format greatly reduces duplication of data elements at the keyboarding stage and decreases the length of the input string. How-

REFERENCE COLLECTION INDEX 21

SUBJECT HEADING	TITLE	DATE.CALL NUMBER
AGRONOMY INDEX/ABSTRACT SERVICE	**WORLD AGRICULTURAL ECONOMICS AND RURAL SOCIOLOGY ABSTRACT	HD1405 W893 V.1. 1959+
AGRONOMY INDEX/ABSTRACT SERVICE	TROPICAL ABSTRACTS	S8 T753 V.8-29.1953-1974
AGRONOMY INDEX/ABSTRACT SERVICE	**SOILS AND FERTILIZERS	S591.7 IM7S V.1.1938+
AGRONOMY INDEX/ABSTRACT SERVICE	**PLANT BREEDING ABSTRACTS	S8123 A1P69 V.1. 1930+
AGRONOMY INDEX/ABSTRACT SERVICE	**FIELD CROP ABSTRACTS	S8183 F459 V.1. 1948+
AGRONOMY INDEX/ABSTRACT SERVICE	GEO ABSTRACTS A: LANDFORMS AND THE QUATERNARY	QE601 G2921 1966+
AGRONOMY INDEX/ABSTRACT SERVICE	FERTILIZER ABSTRACTS	S631.5 F418 V.0.1967+
AGRONOMY INDEX/ABSTRACT SERVICE	**HORTICULTURAL ABSTRACTS	S8355 A11H7 V.1.1931+
AGRONOMY INDEX/ABSTRACT SERVICE	**HERBAGE ABSTRACTS	S8193 A1H41 V.1. 1931+
AGRONOMY INDEX/ABSTRACT SERVICE	**BIBLIOGRAPHY OF AGRICULTURE	TABLE 11 V.1. 1942+
AGRONOMY INDEX/ABSTRACT SERVICE	BIOLOGICAL AND AGRICULTURAL INDEX	TABLE 11 V.1. 1916+
AGRONOMY INDEX/ABSTRACT SERVICE	BIODETERIORATION RESEARCH TITLES	QP517 B5 B5 V.12. 1976+
AGRONOMY PERIODICALS	BOTANICO-PERIODICUM-HUNTIANUM	1968 QK1 .A1186X
AGRONOMY PERIODICALS	SERIALS CURRENTLY RECEIVED BY THE U.S. NATIONAL AGRICULTU	UPDATED S493 .A11 U55
AGRONOMY SERIALS CATALOG	SERIALS CURRENTLY RECEIVED BY THE U.S. NATIONAL AGRICULTU	UPDATED S493 .A11 U55
AGRONOMY STATS INFO	MCGRAW-HILL ENCYCLOPEDIA OF FOOD,AGRICULTURE AND NUTRITI	1977 TX349 .M2
AGRONOMY STATS INFO	AGRICULTURAL STATISTICS	ANNUAL REF DESK HD1751 .UN
AGRONOMY- -USE ALSO 'SOIL SCIENCE','CROP SCIENCE'		
AGRONOMY-CORN	BIBLIOGRAPHY OF CORN	1959-1968 SB191 .C8 A11 G43X
AGRONOMY-MEETINGS INDEX/ABSTRACT SERVI	AGRONOMY ABSTRACTS	S591.7 AM35A 1955+
AGRONOMY-MILLET	THE MILLETS; A BIBLIOGRAPHY OF WORLD LITERATURE COVERING	1930-1969 SB191.M5 .A11G4
AGRONOMY-SORGHUM	SORGHUM; A BIBLIOGRAPHY OF THE WORLD LITERATURE COVERING	1930-1969 SB235 .A11G4
AGRONOMY-WHEAT	INTERNATIONAL BIBLIOGRAPHY OF RICE RESEARCH 1951-1961	1951-1961 SB191.R5 .A11L6
AGRONOMY-WHEAT	BIBLIOGRAPHY OF WHEAT	1959-1968 SB191 .W5 A11 I52X
AID SOCIETIES	NATIONAL DIRECTORY OF ADDRESSES AND TELEPHONE NUMBERS	UPDATED REF DESK E154.5 .G
AIR POLLUTION	INDUSTRIAL POLLUTION CONTROL HANDBOOK	1971 TD897 .I42
AIR POLLUTION	POLLUTION ENGINEERING PRACTICE HANDBOOK	1975 TD151 .P64
AIR POLLUTION ABSTRACTS	**AIR POLLUTION ABSTRACTS	TD883 A11 A38 1970-1976
AIR POLLUTION ASSOCIATIONS	WORLD ENVIRONMENTAL DIRECTORY	UPDATED TD12 .W65
AIR POLLUTION BIBLIOGRAPHY	TOXIC SUBSTANCE SOURCEBOOK	UPDATED RA1190 .T76X
AIR POLLUTION BIOLOGY HANDBOOK	HANDBOOK ON ENVIRONMENTAL DATA ON ORGANIC CHEMICALS	1977 TD196 .073 V47
AIR POLLUTION COMPANIES	WORLD ENVIRONMENTAL DIRECTORY	UPDATED TD12 .W65
AIR POLLUTION COMPANIES	NATIONAL DIRECTORY: ENVIRONMENTAL IMPACT EXPERTS, CONSULT	1974 TD169.6 .C76
AIR POLLUTION CONSULTANTS	NATIONAL DIRECTORY: ENVIRONMENTAL IMPACT EXPERTS, CONSULT	1974 TD169.6 .C76
AIR POLLUTION CONSULTANTS	WORLD ENVIRONMENTAL DIRECTORY	UPDATED TD12 .W65
AIR POLLUTION FORMULAS	HANDBOOK ON ENVIRONMENTAL DATA ON ORGANIC CHEMICALS	1977 TD196 .073 V47
AIR POLLUTION GOVERNMENT AGENCIES	WORLD ENVIRONMENTAL DIRECTORY	UPDATED TD12 .W65
AIR POLLUTION GOVERNMENT AGENCIES	DIRECTORY OF GOVERNMENTAL AIR POLLUTION AGENCIES	UPDATED TD882 .D5
AIR POLLUTION INDEX/ABSTRACT SERVICE	TOXIC SUBSTANCE SOURCEBOOK	UPDATED RA1190 .T76X
AIR POLLUTION INDEX/ABSTRACT SERVICE	**AIR POLLUTION ABSTRACTS	TD883 A11 A38 1970-1976
AIR POLLUTION INDEX/ABSTRACT SERVICE	AIR POLLUTION TITLES	TD881 A4 V.4. 1968+
AIR POLLUTION INDEX/ABSTRACT SERVICE	**ENVIRONMENT INDEX	TD172 E52 V.1 1971+
AIR POLLUTION INDEX/ABSTRACT SERVICE	**POLLUTION ABSTRACTS	TD180 A1P6 V.1. 1970+
AIR POLLUTION INSTITUTES-RESEARCH	WORLD ENVIRONMENTAL DIRECTORY	UPDATED TD12 .W65

** THIS ABSTRACTING OR INDEXING SERVICE CAN BE SEARCHED BY COMPUTER. INQUIRE AT THE REFERENCE DESK.

Fig. 1. KWOC Format Index Entries Showing Subject Heading, Title, Date of Publication, and Call Number.

ever, linkage between the reference collection index and MARC standard machine-readable bibliographic records is not envisioned at this time. The entire indexing project, including data preparation, data entry, and production runs, is within the purview of the Reference Department.

Periodicals display varying information in the publication date field. Serials such as A & I services and book reviewing journals display the beginning holdings date and a continuation symbol, such as "1970+," in the publication date field. Items purchased on a yearly basis are assigned the publication date heading "ANNUAL," while those publications that are frequently issued in new editions or are annual publications purchased by the Library on a staggered basis are given the heading "UPDATED." The traditional edition statement is of little value for reference titles. Loose-leaf publications contain "LOOSE-LEAF" in the date of publication field. The use of these terms, in lieu of dates, minimizes the problems of updating the index for new editions and eliminates from the index any record-keeping function.

INDEXING VOCABULARY

As mentioned previously, the problem of access to reference materials was addressed in this project by providing (1) exhaustive subject analysis based on examination of the work, (2) title entries, (3) entry by significant words or phrases from the title, (4) author, compiler and/or editor, personal name entries on an infrequent basis, and (5) form or content designators as subdivisions under subjects and as lead terms in descriptor strings.

The sample indexes revealed that, in most cases, individual reference titles can be categorized by various form or content designations, such as DICTIONARY, A & I SERVICE, BIBLIOGRAPHY, HANDBOOK, etc. However, the form designator DIRECTORY was judged to be insufficient in specificity and not valuable as either a subdivision or a lead term. Rather, one or more content descriptors (such as TRADE NAMES, BIOGRAPHICAL INFORMATION, etc.) or subject terms found in the work (such as COMPANIES, ASSOCIATIONS, FACULTY, etc.) are substituted. The form noun HANDBOOK is not used as a lead term and, while helpful as a subdivision, must be augmented with other content designators or subject terms. Consistent with recent trends in indexing, an uncontrolled vocabulary scheme using author's words taken directly from the work serves as the basis for the subject indexing system.[18] However, it was felt that retrieval was facilitated when subject or geographical terms were combined with their appropriate form noun(s) to produce a precoordinated descriptor string.

The prescribed method of constructing subject descriptors is to identify the significant subject or geographical descriptors and combine them with the title's form/content indicator(s). These descriptors are in a pas-

sive order, natural language string, e.g., POLITICAL SCIENCE DICTIONARY or SIERRA LEONE STATISTICAL INFORMATION. An attempt was made to avoid the pitfalls of LCSH's multiplicity of punctuation symbols; entries in natural language contain no punctuation symbols, and others contain hyphens at the point(s) where natural language order breaks, e.g., DICTIONARY—GERMAN-ENGLISH PHYSICS, DICTIONARY—PHYSICS—GERMAN-ENGLISH. Experience revealed that access by form/content terms was useful and outweighed any disadvantages of added size to the printed index. These form/content noun lead term entries are produced by phrase manipulation and will be discussed in the data entry section.

A problem of all uncontrolled vocabulary indexes is their scatter of variant forms (e.g., the terms *Plan, Planned, Planning*) and synonyms (e.g., *Firms, Businesses, Business Enterprises, Companies, Corporations,* etc.). The problem of scatter due to indexer inconsistency is especially acute in applications such as this one in which a number of indexers are working separately without a thesaurus. For the retrospective conversion, a term normalization procedure operating on both single words and multiword phrases was introduced to address these problems. This procedure changes variant forms and synonyms into a single preferred term at the output stage. Additionally, syndetic structure routines tied to the normalization programs produce both *see* and *see also* cross-references and connect variant forms and related terms. The amount of normalization needed proved to be extensive and covered both subject terms and form/content nouns. This consumed a large amount of time in human editing of the raw output in order to build the authority file of terms coupled to their preferred forms. For the ongoing coding and entry of new titles, the index itself serves as a thesaurus of reference titles subject terminology, and the normalization routines are less extensively employed.

Another shortcoming of natural language indexing vocabularies is the lack of a hierarchical structure between descriptors (such as the BT, or broader term, and NT, or narrower term, designations in thesauri). In this index, the problem is overcome by abandoning the principle of specific entry in favor of exhaustive indexing at both the generic and specific levels. This is consistent with the policy of producing a one-step index. For example, a cell science dictionary is assigned both CELL SCIENCE and BIOLOGY as subject terms. A directory of midwestern manufacturers receives MIDWESTERN, WISCONSIN, ILLINOIS, MINNESOTA, etc., as geographic name headings. The term normalization procedures are used in many cases to accomplish this coassignment of entry terms. In addition, the capability of providing cross-references between terms helps to alleviate the impact of any subjective decisions about index term relationships.

DATA ENTRY SOFTWARE

To expedite the retrospective conversion of titles and to minimize coding errors, keypunching was done from reference shelflist cards using data elements already present (title, call number, date of publication) along with strings and title word tags added, in pencil, to the cards by indexers.

The data entry software admits two methods of indicating desired access points during input. The vast majority of lead terms are generated from data entered in the subject descriptors field. This variable-length field can presently contain up to 2,000 characters, in the form of either single words or multiword strings. These descriptor strings are stream-coded and separated by semicolons.

The phrase manipulation software rearranges the elements of multiword strings in order to produce additional lead terms for the alphabetical index. The multiple entry points retain the whole of the compound subject within each of the lead terms. The rearrangement follows a prescribed manner and is dependent upon the positioning by an indexer of phrase manipulation symbols at desired points within the input string.

Keen has summarized the multiple entry generation schemes, and this report will use Keen's definitions.[19,20] The phrase manipulation procedures used in this project incorporate several of the mechanized indexing routines reported in the literature.

Three optional phrase manipulation symbols are available to the indexer. The symbol "#" produces a cycling operation in which the phrase elements following the "#" are brought to the lead position; a dash is then inserted and the elements preceding the "#" are added. For example, an input string of the form A#BC will yield ABC and BC—A as entries. Any number of manipulation symbols may be coded. The string A#B#C first yields ABC, BC—A, and C—AB. In addition, any string with two or more of the "#" symbols, such as in the above example, will produce two additional articulated string entries generated by (1) combining the elements preceding the first "#" with those following the final "#" and subdividing with the remainder of the string, and (2) bringing the string elements following the final "#" to the lead, inserting the elements between the first and last "#" symbols, and following with the elements preceding the initial "#" symbol, with the three sections separated by hyphens. Thus, in the case of A#B#C, the two additional entries AC—B and C—B—A are generated by articulation. None of the phrase manipulation symbols will appear at the output stage.

As an example, the input string

GERMAN-ENGLISH#PHYSICS#DICTIONARY

generates the five lead terms:

GERMAN-ENGLISH PHYSICS DICTIONARY
PHYSICS DICTIONARY—GERMAN-ENGLISH
DICTIONARY—GERMAN-ENGLISH PHYSICS
GERMAN-ENGLISH DICTIONARY—PHYSICS
DICTIONARY—PHYSICS—GERMAN-ENGLISH.

The use of the manipulation symbol "+" causes a similar cycling operation, but a hyphen is inserted in the original string in place of the space. For example, the string

WOMEN+LITERATURE

produces the two lead terms:

WOMEN—LITERATURE
LITERATURE—WOMEN.

This symbol is used in the rare instances when the original input string cannot be constructed in a natural language, passive order phrase.

The cycling operation rearranges the order of the input string elements and brings significant elements of the string to the lead position, while preserving the contextual relationships of the term elements. The purpose of the articulation operation is to collocate those string element pairs that have become separated because of the length and form of the input string. These term pair coordinations are desirable in instances where other lead terms, either because of fewer constituent elements or a different arrangement, do display the particular pair combination. Without the articulation headings, an effective search would require the scanning of the entire body of entries under the first significant element in the lead term in order to ensure that all the combinations of the two elements would be found. For example, the input string PHYSICS#DICTIONARY generates the lead term DICTIONARY—PHYSICS. In the preceding example of GERMAN-ENGLISH#PHYSICS#DICTIONARY, the cycling operation produced the lead term DICTIONARY—GERMAN-ENGLISH PHYSICS, in which the elements DICTIONARY and PHYSICS are not directly coordinated. If the articulation operation were not performed, a search for all physics dictionaries approached via the form noun DICTIONARY would require reading all the lead terms under DICTIONARY to be certain of satisfying the search. However, the articulation operation produces DICTIONARY—PHYSICS—GERMAN-ENGLISH, and this term pair coordination collocates the two titles. Similarly, the articulation-produced lead entry GERMAN-ENGLISH DICTIONARY—PHYSICS ensures coordination of all German-English dictionaries under the element GERMAN-ENGLISH.

It should be noted that for $N \geq 2$, a string containing N # symbols will generate N+1 lead terms by cycling (including the original string with

the order unchanged), but only two additional leads by articulation for a total of N + 3 entries. This serves to keep the number of index entries to a manageable number. For the same reason, it was decided that because of the passive order input string presentation, lead terms of the form B—A—C (corresponding to the input string A#B#C) were normally not valuable. For situations where nonstandard types of coordinations are desired a truncation routine has been devised. This routine requires that a pair of brackets be placed around a subset of the entry input string, e.g., A#B[#C]. After the complete string is processed intact—that is, as if the brackets were not present—the elements and symbols within the brackets are eliminated, and the string is reprocessed. Thus, in the example above, the five cycling and articulation entries would be produced and, in addition, the two entries AB and B—A. This component truncation routine is most commonly used in conjunction with countries or generic-specific terms. For example, the string [CANADA] CHEMISTRY#FACULTY produces the lead terms

CANADA CHEMISTRY FACULTY
FACULTY—CANADA CHEMISTRY
CHEMISTRY FACULTY
FACULTY—CHEMISTRY

and might be used for a directory of U.S. and Canadian university chemists. Another example of the use of truncation is in the subject term [SOLAR] ENERGY to generate both SOLAR ENERGY and ENERGY.

The second method for placing desired access points in the input stream is by tagging significant elements in the title field for removal and inclusion as subject heading terms. The software, after ignoring initial articles, also includes the full title as a lead term entry. For example, the coded title

WHO'S WHO IN#AMERICAN#ART

produces the lead terms

WHO'S WHO IN AMERICAN ART
AMERICAN ART—WHO'S WHO IN
ART—WHO'S WHO IN AMERICAN
ART—AMERICAN—WHO'S WHO IN
WHO'S WHO IN ART—AMERICAN.

The data preparation procedures are designed to minimize keyboarding strokes (and, hence, the incidence of keyboard error) by greatly reducing duplication of entry string elements. This is possible because examination of sample indexes revealed a word order pattern occurring frequently among groups of subject lead terms assigned to an individual title. A combinatorial-based data input format was designed and the

appropriate format recognition software written to accept input data and convert it to a form amenable to processing by the phrase manipulation algorithms. The result is a data input format that is much more compact than a traditional linear input format, which would require a large amount of duplicate coding of phrase elements.

The data input format is predicated on the fact that a large number of the desired subject descriptors for reference titles are of the form subject/geographical indentifier plus form noun(s). When applicable, two partitioned sets of descriptor elements are entered in the subject descriptors field in a data input string of the type $(A_1; A_2; \ldots; A_p)*(B_1; B_2; \ldots; B_q)$, where the A_i and B_i are single words or multiword phrases and "*" is a phrase manipulation symbol. Elements within a set are separated by semicolons. The partitioned sets input string is then expanded by the format recognition routine in a distributive manner to yield the pq multiword strings A_1*B_1, A_1*B_2, . . . , A_1*B_q, A_2*B_1, A_2*B_2, . . . , A_2*B_q, . . . , A_p*B_1, A_p*B_2, . . . , A_p*B_q. These discrete strings are then passed individually to the phrase manipulation software to generate lead terms for index entries. For example, the input string (PHYSICS; MATHEMATICS)#(DICTIONARY; BIOGRAPHICAL INFORMATION) sends the input strings

PHYSICS#DICTIONARY
PHYSICS#BIOGRAPHICAL INFORMATION
MATHEMATICS#DICTIONARY
MATHEMATICS#BIOGRAPHICAL INFORMATION

to the phrase manipulation routine to produce eight lead terms for the index. Individual elements within the descriptor sets may contain embedded articulation symbols. These are passed intact to the phrase manipulation software.

The third manipulation symbol "-" is most often used in conjunction with the partitioned sets format. This operator's function is to signal the program that no manipulation should be performed. For example, the string (UNITED NATIONS; INTERNATIONAL ORGANIZATIONS)-(OFFICIALS) produces the two lead terms

UNITED NATIONS OFFICIALS
INTERNATIONAL ORGANIZATIONS OFFICIALS

The symbol "-" can be used to shorten input strings under certain conditions. For example, the input form (GERMAN-; FRENCH-; RUSSIAN-)-(ENGLISH#PHYSICS#DICTIONARY) generates the same lead terms as (GERMAN-ENGLISH; FRENCH-ENGLISH; RUSSIAN-ENGLISH)#(PHYSICS#DICTIONARY) but does not require the repeated entry of the term ENGLISH.

Several specialized software routines have been introduced to further expedite multiple entry generation and provide a more compact data in-

put format. The symbol "≠," when placed before a term element, indicates to the phrase manipulation routine that the term should not be moved to the lead position. For example, the string (CIVIL ENGINEERING)#(TABLES; ≠ HANDBOOK) ultimately produces three lead terms:

CIVIL ENGINEERING TABLES
CIVIL ENGINEERING HANDBOOK
TABLES—CIVIL ENGINEERING

The symbol "≠" is effectively used in front of individual term elements in situations where some, but not all, of the elements in either descriptor set merit lead element status.

The term normalization routines are employed to streamline data input in two ways. Heavily used terms, such as the form nouns, can be normalized from abbreviated or mnemonic forms. For example, the descriptor authority file contains the word DY as a variant form of the term DICTIONARY. Thus, from an earlier example, we can write the input string (PHYSICS; MATHEMATICS)#(DY;BI), where BI is normalized to BIOGRAPHICAL INFORMATION. The normalization software is also utilized to replace an individual term with itself and another coupled term, usually a term hierarchically related. for example, the term RAILROADS cues the normalization program to substitute RAILROADS; TRANSPORTATION in its stead. In this way, the broader heading will also appear as a separate lead term element. However, the use of the term TRANSPORTATION will not automatically produce the above substitution.

The number of lead terms produced by a partitioned sets string can be determined, but the formula depends on the particular manipulation symbol included in the string. In instances where the symbol "#" separates the two sets, the number N of lead terms produced can be determined from the formula $N = 2pq + S_Bp + S_Aq + R$, where p is, as before, the number of elements in the first set $A = (A_1; A_2; \ldots A_p)$, q the number of elements in the second set $B = (B_1; B_2; \ldots B_p)$, S_B the total number of "#" symbols within the elements of the set B, S_A the total number of "#" symbols within the elements of the set A, and R the total number of lead terms produced by articulation. The number R can be determined only by inspection of the partitioned sets. However, in situations where each element of B contains at least one "#," or each "#" or each element of A contains at least one embedded "#," $R = 2p$. If the symbol "-" separates the two partitioned sets, the formula $N = pq + S_Bp + S_Aq + R$ holds.

The software for this project was written in PL/I F level and optimizer compiler. The main procedure contains four modules: (1) a format recognition routine to translate input strings for processing, (2) a word normalization routine, (3) a cross-reference production procedure, and (4) the

phrase manipulation and entry production module.

The data entry procedures and/or phrase manipulation algorithm could be used in any indexing application desiring a combinatorial-based data entry format or requiring a multiple entry generation procedure capable of allowing controlled combinations of input string elements.

ACCESS POINTS

The need for certain types of access points was identified primarily from anecdotal evidence compiled by reference staff. Likewise, much of the terminology used in the subject descriptors was derived from experiences using LCSH and other subject heading schemes. In order to better illustrate the raison d'être for this index, selected titles that typify these kinds of access points will be presented and any differences with LCSH examined.

Title access allows easy identification of the call number in those instances where known-item title searches are performed, usually by reference librarians. Access by significant title words is helpful when the complete title is unknown or an incorrect citation has been provided. For example, a search for a "Who's Who in Music" under that title will locate the heading WHO'S WHO IN MUSIC—INTERNATIONAL, a retrieval key for the title *International Who's Who in Music*. Likewise, the reference collection index contains an entry—BOOK OF LISTS—for the title whose main entry is *People's Almanac Presents the Book of Lists*.

Form/content descriptors prove useful in situations where there may not be an exact match for a search request. For example, a search for a plant science dictionary yields nothing under that heading, but searching under the subdivisions at the form descriptor DICTIONARY yields DICTIONARY—FORESTRY and DICTIONARY—BOTANY with associated relevant titles. This type of form descriptor access is not possible in LCSH.

Personal or corporate name entries are indicated for works closely associated with a person or body, such as HAZLITT for the *Bibliography of Early English Literature 1867–1903*, or CHICAGO—UNIVERSITY OF and UNIVERSITY OF CHICAGO for the title *A Manual of Style*, which is associated with that institution.

The real value of the index lies in its exhaustive subject access to reference titles. Specific features of titles are brought out, e.g., ENGINEERING AWARDS for *Who's Who in Engineering*, AGRICULTURAL ENGINEERING STANDARDS in *Agricultural Engineers Yearbook*, and DANCE UNIVERSITIES from *The National Directory for the Performing Arts–Education*. The lack of indexing exhaustivity is a major problem in LCSH. To illustrate, the title *Dictionary of Scandinavian Biography* (LCCN 73-189270) was assigned the two subject headings SCANDINAVIA—BIOGRAPHY—DICTIONARIES and FIN-

LAND—BIOGRAPHY—DICTIONARIES from LCSH. However, the title also includes biographical information for Norway, Denmark, Sweden, and Iceland. The reference collection index contains the twelve retrieval points generated from the data entry string (SCANDINAVIA; NORWAY; FINLAND; DENMARK; ICELAND; SWEDEN)#(BIOGRAPHICAL INFORMATION) and two additional access points from the additional input phrase SCANDINAVIA COOPERATION# TREATY, for a total of fourteen subject entries. Examination of some of the traditional reference tools, such as *Europa Yearbook* (ISSN 71-2302), reveals the wide discrepancies in coverage. For the reference collection index, this title has been assigned the following string in the subject descriptors field:

> (INTERNATIONAL ORGANIZATIONS; ECONOMIC#COMMISSIONS; UNITED NATIONS)-(OFFICIALS); (ALBANIA; ANDORRA; AUSTRIA; etc., the European countries through Yugoslavia; EUROPE)#([ECONOMIC#]STATISTICAL INFORMATION; ALMANAC; AGRICULTURE STATISTICAL INFORMATION; CONSTITUTIONS; GOVERNMENT OFFICIALS; EMBASSIES; NEWSPAPERS; PERIODICALS; PUBLISHERS; BANKS; ASSOCIATIONS; LABOR UNIONS; UNIVERSITIES).

From this some 300 subject entries are produced. Contrast this with the two LCSHs assigned to the title:

> EUROPE—POLITICS—YEARBOOK
> EUROPEAN FEDERATION—YEARBOOKS.

The scatter of titles under synonyms and related terms is another major problem area in the use of LCSH to retrieve reference titles. For example, *National Faculty Directory* (ISSN 77-4472) has been given the heading COLLEGE TEACHERS—UNITED STATES—DIRECTORIES; *Directory of American Scholars* (ISSN 70-5101) has the heading SCHOLARS, AMERICAN—DIRECTORIES; *American Men and Women of Science* (ISSN 192-8576) has the heading SCIENTISTS—UNITED STATES; *AIBS Directory of Bioscience Departments and Faculty* (LCCN 75-33761) has been assigned BIOLOGY—STUDY AND TEACHING—U.S.—DIRECTORIES; and *American Men and Women of Science. Biology* (ISSN 146-0048), in contrast to the previous title, was assigned BIOLOGISTS—U.S.—BIOGRAPHY from LCSH. All of these titles will be found under FACULTY, in addition to other access points, in the reference collection index.

As another example, consider several comprehensive business directories. *Consumer's Register of American Business* (ISSN 98-7344) bears the heading BUSINESS ENTERPRISES—U.S.—DIRECTORIES from LCSH. However, *Moody's Industrial Manual* (ISSN 545-0217) has been assigned CORPORATIONS—U.S.—DIRECTORIES; the *Standard and*

Poor's Register of Corporations, Directors, and Executives (LCCN 28-7849) has been given only DIRECTORS OF CORPORATIONS—U.S.—DIRECTORIES and the identical heading with CANADA substituted for U.S.; and the Dun and Bradstreet *Million Dollar Directory* (LCCN 59-3033) bears the LC subject heading U.S.—INDUSTRIES—DIRECTORIES. These titles appear under the term COMPANIES in the reference collection index.

LCSH's use of outmoded terminology also complicates retrieval. *Energy Technology Handbook* (LCCN 76-17653) has the headings POWER(MECHANICS)—HANDBOOKS and POWER RESOURCES—HANDBOOKS from LCSH as opposed to the fifty headings produced by the input string (ENERGY; COAL; PETROLEUM; NUCLEAR ENERGY; SOLAR ENERGY; GEOTHERMAL ENERGY; GASOLINE; HYDROPOWER; POWER ENGINEERING; INDUSTRIAL ENGINEERING)#(PHYSICAL CONSTANTS;≠HANDBOOK; DICTIONARY) for the reference collection index.

These examples clearly demonstrate the significant problems of effectively retrieving reference titles by subject using LCSH. These problems have been addressed in this project by providing expanded subject analysis in a printed format display. An average of approximately 11 access points have been assigned per title; for titles processed during the last year the average has been more than 16 access points, compared to an average 1.8 subject headings from subject cataloging done at the Library of Congress.[21]

CONCLUSION

This project demonstrated the feasibility of a short entry printed index emphasizing expanded subject access to academic library reference collection materials. In terms of bibliographic role, this index is designed to complement, not replace, other library catalogs. The index provides access points not present in the other catalogs, indicates special locations within the reference collection (e.g., index tables), covers bibliographies (e.g., *Reader's Advisory Service*) and series (Midwest Plan Service publications) not analyzed in the other catalogs, and identifies the A & I services that can be searched through an on-line vendor. While the index is most heavily used by reference librarians, it has proved to be easily approachable by patrons and other staff. It has been a valuable tool for training new reference staff and assisting paraprofessionals. The index is produced in both hardcopy and microfiche, and presently totals more than 1,000 pages and 70,000 entries. The index guide's portability allows it to be placed in branch library reading rooms and other locations in the library system. A separate index covering only the A & I services is produced as a spin-off. As for its adaptability by other institutions, while it may be helpful in reference service, the differences in holdings, call numbers, and location designations between the Iowa

State University Library and other academic libraries would limit its usefulness. However, the procedures utilized in this project could be transferred to other settings and successfully applied to either a reference collection or any other component of a library collection.

A copy of the most recent complete printed index is included in the final report of the project submitted to the Council on Library Resources and is available from ERIC.

REFERENCES

1. Cherie Weil, "Automatic Retrieval of Biographical Reference Tools," *Journal of Library Automation* 1:239–49 (Dec. 1968).
2. N. La Var Harline, *A Computerized Reference Retrieval System*, ED 102 974 (Provo, Utah: Brigham Young Univ., Graduate Department of Library and Information Sciences, 1974).
3. Ching-Chih Chen and E. Robert Kingham, "Subject Reference Lists Produced by Computer," *Journal of Library Automation* 1:178–97 (Sept. 1968).
4. J. C. Meredith, "Machine-Assisted Approach to General Reference Materials," *Journal of the American Society for Information Science* 22:171–86 (May–June 1971).
5. James Krikelas, "Catalog Use Studies and Their Implications," *Advances in Librarianship* 3:195–220 (1972).
6. Hans Wellisch, "Subject Retrieval in the Seventies," in Hans Wellisch and Thomas D. Wilson, eds., *Subject Retrieval in the Seventies* (Westport, Conn.: Greenwood, 1972), p.5–15.
7. Sanford Berman, "The Cataloging Shtik," *Library Journal* 102:1251–53 (1 June 1977).
8. Marcia J. Bates, "Factors Affecting Subject Catalog Search Success," *Journal of the American Society for Information Science* 28:161–69 (May 1977).
9. Lois Mai Chan, "Alphabetical Arrangement and Subject Collection in Library of Congress Subject Headings," *Library Resources and Technical Services* 21:156–69 (Spring 1977).
10. William H. Mischo, "An Augmented Keyword Index to Abstracting and Indexing Services," *Journal of Academic Librarianship* 3:141–45 (July 1977).
11. William H. Mischo, "Library of Congress Subject Headings: A Review of the Problems, and Prospects for Improved Subject Access" (manuscript out for review).
12. Kieth C. Wright, "Trends in Modern Subject Analysis with Reference to Text Derivative Indexing and Abstracting Methods; The State of the Art," *Information Part 2* 1:6 (Sept.–Oct. 1972).
13. Pauline Atherton, *Books Are for Use: Final Report of the Subject Access Project to the Council on Library Resources* (Syracuse, N.Y.: Syracuse Univ. School of Information Studies, 1978 [ED 156 181]).
14. S. Michael Malinconico, "The Display Medium," *Library Journal* 101:2144–49 (1976).
15. R. P. Palmer, *Computerizing the Card Catalog in the University Library: A Survey of User Requirements* (Littleton, Colo.: Libraries Unlimited, 1972).
16. R. Tagliacozzo and M. Kochen, "Information-Seeking Behavior of Catalog Users," *Information Storage and Retrieval* 6:375 (Dec. 1970).
17. D. R. Swanson, "Requirements Study for Future Catalogs," *Library Quarterly* 42:302–15 (1972).
18. F. W. Lancaster, "Vocabulary Control and Information Retrieval Systems," *Advances in Librarianship* 8:2–40 (1977).
19. E. Michael Keen, *On the Performance of Nine Printed Subject Index Entry Types*

(Aberystwyth, Wales: College of Librarianship, Wales, 1978).
20. E. Michael Keen, "On the Generation and Searching of Entries in Printed Subject Indexes," *Journal of Documentation* 33:15–45 (March 1977).
21. Martha E. Williams and Gordon J. Shefner, "Data Element Statistics for the MARC II Data Base," *Journal of Library Automation* 9:89–100 (June 1976).

William Mischo has an abiding interest in computer-assisted indexing techniques. He is presently working on a grant to study authority control of subject headings.

The Impact of Technology on Legislation Affecting Libraries*

Henriette D. AVRAM, Library of Congress, Washington, D.C.

Issues raised by the technology that may affect the operation and services of libraries are discussed. Emphasis is on international problems, but many of the problems have their counterparts in the national scene, i.e., the legal, political, and economic implications of the transfer of data via information networks. Included is a brief description of the problems of ownership and copyright that have developed in the International MARC Network and the efforts to date to resolve the problems.

INTRODUCTION

I have chosen to concentrate on the legislative aspects of the conference theme, "Library Legislation and Management," by discussing some of the issues raised by the technology that may result in legislation that, in turn, may affect the operations and services of libraries.

I shall address aspects of the problem that are international in nature, that is, issues with which both the developed and developing countries are concerned in varying degrees. The orientation of IFLA toward Universal Bibliographic Control (UBC), Universal Availability of Publications (UAP), and the International MARC Program (IMP) indicates a deep international involvement in information and its technology on the part of IFLA's membership.

We are now challenged by an information environment of massive volume and complexity characterized by an exponential increase in information flow, decreasing communications time and distance constraints, and greater dependence on information and communication services. The result is a rapidly diminishing time buffer between technical development and social change.

The globe has shrunk and the information problem has gone international. The marriage of the computer with the technology of communication has created a period in our history that rightly can be called a revolution and marks the beginning of the information age.

*This paper was prepared for the International Federation of Library Associations and Institutions (IFLA) Division of Managment and Technology Program meeting held at the 1979 IFLA conference in Copenhagen, Denmark, and also was to be published in the *IFLA Journal*, V.6 (1980), no.1. Since the topic is of concern to libraries and other information agencies in the United States, it appeared useful to include it in the *Journal of Library Automation*.

Information is now recognized as a commodity and a national resource which must be protected and cultivated. Thus, both nationally and internationally, considerations concerning the sale, exchange, lease, and/or giving away of data are receiving increased attention. National laws are being enacted that will govern the flow of information made possible through the technology. As major processors and distributors of information, libraries could be seriously affected by legislation that would regulate the free flow of information or that would impose tariffs, thus influencing the economics of information transfer. It is, therefore, appropriate in this setting to review some of the major problems caused by the technology.

TRANSNATIONAL ASPECTS OF DATA FLOW

Until very recently, international information flow was principally in printed form and, in this form, was subject to both review and control. Naturally, there always existed ways to carry information surreptitiously from one place to another, bypassing scrutiny. In general, however, a book, a letter, a business record, was human readable, and, although subject to photocopy, the photocopy was also human readable and therefore subject to fairly straightforward review and control if required. All this has changed.

It is no longer science fiction to portray someone at a terminal in Paris dialing a computer in the United States to request information, or a multinational corporation with staff dispersed worldwide in continuous communication with one another and with their respective computers. With the merging of the two forms of technology—computers and communications—it is feasible and cost-effective to rapidly and efficiently move large amounts of data across national boundaries. The terms *transnational* and *trans-border data flow* reflect this international setting. Because the data is machine readable as opposed to human readable, it is no longer readily subject to review and/or control.

The data being transmitted vary widely in information content: airline information, which includes seat and hotel reservations; freight shipping arrangements; banking and credit information; individual tax information; corporate data for production, management control, marketing, personnel, capital expenses and investment; bibliographic data, including the description of the publishing output of many countries and the citations for information in particular disciplines; statistical data such as national census and production information; full text data such as LEXIS files, which contain statutory and judicial data; and a variety of other data spanning current social and economic activities.

BASIC OBJECTIVES OF INTERNATIONAL COMMUNICATIONS

A recent study of European data networks distinguished four basic objectives for the establishment of international communications as:

1. Indirect financial benefits gained by multinational corporations achieving economies through rapid international coordination of production, marketing, and financial management.
2. Direct financial benefits from sharing expensive resources such as computers.
3. Management of inherently international activities such as airline reservations and international money exchange.
4. Coordination and cooperation such as multinationals' assisting subsidiaries in different countries to coordinate their activities.[1]

The above obviously relate to international trade. However, the IFLA-sponsored programs of Universal Bibliographic Control, Universal Availability of Publications, and International MARC can readily fulfill similar objectives in the area of sharing of bibliographic information and bibliographic items.

The above four points could be stated as:

1. Indirect financial benefits gained as the bibliographic agencies achieve economies in research through increased access to information.
2. Direct financial benefits from shared cataloging and materials lending.
3. Management of inherently international activities such as the International Serials Data System and the MARC International Program.
4. Coordination and cooperation such as in the International Standard Book Number program.

LEGAL AND MANAGERIAL CONCERNS

At present, the principal concerns of transnational data flow are:

1. Individual privacy rights—personal information that is protected in one country is transmitted to other countries where laws are less stringent.
2. Data security—data transmitted to one or more countries and subject to different security and communications standards and different licensing agreements are vulnerable to additional risks of unauthorized use and physical damage.
3. National sovereignty—data emanating from a country with a lack of technical capability are processed in another country, with the possibility that (a) the data may not be available to the originating country, (b) sensitive information may be made available outside the originating country, and (c) the originating country lacks control over manipulation of the data and is therefore dependent to a degree upon the other country.
4. Exportation of highly skilled positions—processing of data outside the originating country may result in the involuntary export of skilled positions that would otherwise be established in the originating country, thus seriously affecting its economy.

Although the above considerations may appear to be remote from the

present concerns of libraries, this is in fact not so. Libraries are not merely storehouses of information, but are organizations attempting to provide a vast array of services to their highly diversified clientele, such as access to news services, census information, legislative information, and research of all kinds, as well as the more traditional bibliographic and indexing services. These services are available through telecommunications and will be affected by the problems of transnational data flow.

Eighteen countries either have adopted or have under consideration laws that will restrict in varying degrees the flow of data across national borders. To date, most of the laws attempt to limit the free flow of data to provide economic or civil rights protection for the citizens.

NEED FOR INTERNATIONAL AGREEMENT

Few would contest the validity of protecting the individual person—but one problem is that some of the national laws use the term *legal person* as opposed to *natural person*, and that may include, in addition to people, a corporate entity or a national entity and may encompass information services used by libraries. These differences in legislation clearly point toward the need for international agreement to resolve present conflicts and to prevent further limitation of the free flow of information through barriers imposed by national laws. The Council of Europe and the Organization for Economic Cooperation and Development have been actively involved in furthering efforts toward such international agreement. Individual countries may have to pass enabling legislation to permit them to participate in such international agreements.

QUESTIONS OF OWNERSHIP AND COPYRIGHT

Transnational communications are also affected by the question of ownership of information in shared computer-based systems. At present, copyright laws protect authors and publishers from plagiarism and unauthorized copying. However, the law is still unclear with respect to the rights of ownership in shared computer-based systems, and the question of ownership arises independent of copyright. For example, a data base such as a machine-readable national bibliography may be subject to copyright, while the individual bibliographic records making up the bibliography are not. In both cases, problems of ownership exist and these problems are ultimately economic ones.

In some computer-based library systems, where cataloging data are created in a single system by many organizations and the cataloging records constituting the data base are shared by all the participating members, the problem of ownership exists even though there have been no claims to copyright. Does an individual record belong to the organiza-

tion that created the record? Does a record created by one organization belong to another organization that pays for and uses the same record from a shared system? Or does the record belong to the organization that controls the system?

In other systems the design permits a member library to limit access to its file to its own staff and to staff of other libraries that are parties to an interorganization arrangement. Files are organized either on an individual library basis or as a single union file for a group of libraries. In either case, certain data, such as the MARC files from national bibliographic agencies, are considered resource data to be used by any organization and to be incorporated into its private file.

In both of the above cases, what exactly is meant by *belong*? Does ownership mean the license to use the record in any way, including resale or giveaway, or is the concept of ownership subject to certain restrictions?

If, for example, an organization chose to discontinue its use of one system and join another system, would it be permitted to copy its machine-readable records and have them stored in the new system? The implications are that the users of the second system, in paying a fee for the use of such records to the second system, would cause loss of revenue to the first system. Such problems are now surfacing as the result of national and international networking. There are as yet no solutions.

RELATIONSHIP TO EXCHANGE OF MARC DATA

A combination of the problems of ownership and copyright has developed internationally during the past several years. Since 1969 we have witnessed the implementation of many national MARC systems. It was quickly recognized that national MARC records representing cataloging data for the publishing output of each country could usefully be exchanged as one principal component of Universal Bibliographic Control. National agencies began to make bilateral arrangements covering the conditions for the exchange of MARC data via magnetic tape. Because of differences in copyright laws and practices in the various bibliographic agencies, the exchange agreements vary.

For example, the Library of Congress MARC records are not covered by copyright laws, and the Library is not concerned about any exclusive arrangements for its tape distribution services; the National Library of Canada's publication *Canadiana* is covered by copyright law, but the National Library chose not to exercise copyright protection in the distribution of these records in machine-readable form in its agreement with the Library of Congress. On the other hand, the British National Bibliography (now incorporated into the British Library) is concerned with maintaining some control over the distribution of its MARC records. The result of these differences is that each exchange agreement may be unique. Obviously, the Library of Congress' agreement with the

National Library of Canada differs from its agreement with the British Library. The basic problem appears to be that some of the national agencies involved in the production and distribution through sale of national bibliographies are concerned that, since the MARC records represent records in the national bibliography, the bibliography could be published by another organization. In addition, since many of the national bibliographic agencies plan to provide an on-line service for the distribution of MARC records (and in some cases they have already implemented it), the agencies are concerned whether an on-line service would be economically viable if the records were available to their competition. The Conference of Directors of National Libraries, recognizing the worldwide impact of this evolving international MARC network, decided in 1975 to give high priority to the consideration of an efficient and effective network.* A Steering Committee was appointed and recommended a study to include both bibliographic and technical problems of an international system. The report of that study included a recommendation to establish an international exchange policy that would include a universally acceptable agreement among national bibliographic agencies to cover the distribution and use of MARC records.[2] The Steering Committee, with contractual support, formulated such an agreement. The significant provisions of the agreement are:

1. The exchange of bibliographic records should involve no financial charge imposed by either organization.
2. Each organization may use another's MARC records as they exist or may modify records as needed.
3. The originating organization, if it is a copyright holder, agrees that any of its records that have been modified to meet national bibliographic standards of the receiving organization are not subject to copyright.
4. Records not modified to accord with national bibliographic standards of the receiving organization may be distributed within the country of receipt to not-for-profit organizations; supply of these records to profit-making organizations is permitted but is subject to contractual arrangements made between the originating organization and the profit-making organization concerned.
5. The receiving organization agrees not to produce the national bibliography of the originating organization.

During the Steering Committee meeting in London in May 1979, the decision was made to forward the agreement to the Conference of Directors of National Libraries for its approval during that IFLA session and to recommend the wide distribution of a document that includes

*The Conference of Directors of National Libraries, originally called the International Association of National Librarians, was organized in 1975 to discuss the problems of national libraries with an aim of providing solutions. This body is now incorporated within the IFLA structure.

the background of the agreement and the agreement itself.† This agreement is a landmark accomplishment toward the goal of the free flow of information.

CONCLUDING REMARKS

The increased cost of resources and the general lack of funding to meet the increased costs, all-too-familiar topics of the past decade, have forced bibliographic agencies to acknowledge their need for sharing or networking, and the technology has provided a possible solution. At the same time that the technical feasibility of resource sharing has been established, with the capability of providing wider access to a greater number of information resources for a larger number of uses, laws are being passed to inhibit the free flow of this information. I recommend that IFLA, through the Section on Mechanization of the Division of Management and Technology,‡ monitor developments in transnational data flow and copyright issues concerning machine-readable data files and participate in international conferences concerned with the free flow of information from country to country. Otherwise we will continue to work together toward the sharing of information through UBC and UAP only to find that we are not permitted to do so—laws governing the free flow of information will have been established without any voice from this community.

REFERENCES

1. T. Johnson, "Data Networks and International Data Traffic in Europe, at 162" (Feb. 1978) (Report by Senior Consultant to Logica, Ltd., U.K., presented at the Conference on Transnational Data Regulation in Brussels, Belgium), cited in John M. Eger, "Emerging Restrictions on Transnational Data Flows: Privacy Protection or Non-Tariff Trade Barriers?" *Law and Policy in International Business* 10, no.4:1058 (1978).
2. A. J. Wells, "The International MARC Network: A Study for an International Bibliographic Data Network" (Prepared for the International MARC Network Study Steering Committee), Jan. 1977.

Henriette D. Avram, director of the Network Development Office, Library of Congress, has been at the library since 1965. Avram directed the MARC Project, both the pilot and the operational systems. As chief of the MARC Development Office she was responsible for automation of the bibliographic processes of the Library of Congress and in her present position is principally responsible for the LC interface with the evolving nationwide and international library networks.

†At the Conference of Directors of National Libraries meeting held during the IFLA conference (August 1979), after some discussion among the national librarians (or their delegates) the decision was made to ask for written comments on the agreement before any further distribution of the document was made. Sir Harry Hookway, British Library, chairman of the conference, asked that comments be submitted to him by December 1979.

‡At the meeting of the IFLA Professional Board in November 1979, the board approved a name change from the Section on Mechanization to the Section on Information Technology.

The Library and the Computer Center

Five brief papers from a program conducted by the TESLA Committee are presented here. The subtitle of the original program was "A Marriage Made in . . . ," giving rise to a number of allusions to courtship, marriage, and divorce. The papers touch on the appropriate role of the computer center, sources of conflict, and basic incompatibilities in the relationship with the library. Also discussed are ways of dealing with developmental operating difficulties, the importance of negotiating a basis for a continuing relationship, and, finally, the need for tolerance and mutual respect.

This material was edited by William D. Mathews from a transcript provided by Arlene Schwartz, who chaired the session at the Dallas Conference of the ALA.

What the Computer Center Should Do for a Library

Hugh STANDIFER: Data Systems Administration, City of Austin, Austin, Texas.

INTRODUCTION

If reality offends you, then some of you may not want to hear what I am about to say. The primary question I hope to address is, Can the library use its local computer center for any or all phases of an automation project? Before I respond to that question, let me say that there are poorly managed computer centers and, under the doctrine of equal damnation, there also exist poorly managed libraries. Poorly run entities are excluded from my response for two reasons: first, of necessity my comments must be general, and second, I believe that no board of scholars, nor any rational body of law or custom, exists to instruct directors in their duties. I cannot hope to educate the director to direct or the manager to manage.

JUSTIFYING COSTS AND SERVICES

To return to the question: Yes, a library can use its local computer center for any or all phases of an automation project. Yes, but if your library is a governmental entity, if you are supported by taxes, you should not automate anything that cannot be cost- and/or service-justified.

I will stick my neck out and venture a guess that 75 percent of the libraries that have been automated did so without the aid of a properly conducted feasibility study. Of course, some sort of study was done—one that was directed at supporting a decision to automate that had already been made. The study may have been accomplished by someone in the library at the direction of a division head or perhaps the library director. It was more than likely not done objectively. It almost certainly was not done with the idea in mind that it might be possible simply to streamline the manual process, and forget all about "computerizing."

Probably a lot of "facts" and figures were gathered. Now, even if the study were not objective, for the first time somebody had a pretty good idea of how many patrons they were serving, how many titles they had, what the circulation figures looked like, etc. Let me say something here about statistics. Samuel Langhorne Clemens once remarked: "Statistics are like ladies of the evening. Once you get them down, you can do anything with them." Usually, once the statistics are gathered, someone concludes, "We have such a tremendous volume of activity we obviously have to automate."

Typically, the study that was completed did not itself contain a problem definition at all. And yet, based on an incomplete amalgamation of facts, the decision that had already been made—which was to automate—is now confirmed: get a computer to solve the problem. I ask, What problem? No problem had yet been defined and already somebody had decided to solve it! And all this at the expense of the taxpayers—that's you and me, folks. Of course, if your library happens to be eleemosynary in nature, then it really doesn't matter what you do with sugar daddy's money as far as the taxpayer is concerned. But as a taxpayer, you should demand value for your tax dollar.

I say again, Yes, the library can use the computer center for any or all phases of an automation project. But the library is not going to find a computer center willing to do so unless it can meet justification criteria of cost and/or services.

What, then, should the library expect from the computer center? The computer center is, or should be, a service-oriented operation. So the library should expect a number of services from the computer center, just as the using public expects a number of services from a library. So you see, the computer center and the library have two things very much in common: they are both organized to provide service to users.

FEASIBILITY STUDY AND PROBLEM DEFINITION

One service the computer center can provide is a *feasibility study* by personnel who are trained in conducting one properly. They can define for you what are you doing now—not what you think you are doing, but what you are actually doing in some measurable terms. They can assist you in determining whether what you are now doing actually needs to be done. This need-to-be-done thing is most often discovered by asking the question, Why does it have to be done? Following that, the next question is, How much of it is there to do? It is remarkable at times how much less intimidating a project becomes when a cold count is made: when it is actually quantified. This is especially true when it is contemplated along with the next question, How often does it have to be done?

The purpose of a feasibility study is to determine whether it is feasible to automate, so the computer center, after gathering all the available facts, and after having asked why until you are ready to climb the wall, can determine whether automation is justified. Computer center staff can tell you, after having received the manual tasks, whether a computer can perform those tasks for less money than presently, whether the improvement in service will offset the additional expenditure, or whether a combination of service improvement and less money will justify automation. Of course, sometimes they will tell you that automation can't be justified.

GETTING ADVICE

At this point the library occasionally turns to a hardware/software vendor, who will *advise* with firm conviction that they can certainly justify automation. Just a brief aside here about vendors, if I may, and about private industry in general. Private industry is dedicated to providing the *least* service for the *most* money. You as a governmental employee are dedicated, or should be dedicated, to providing the *most* service for the *least* money. So immediately, you are at opposite ends of the pole. The vendors will *advise* you with great vigor how to show proof of need. They will *advise* you how you can prove it even if they must change the library's problem to fit their prepackaged solution.

Let me also tell you this about advisers. I wish I had said it, but I didn't; John F. Kennedy did. He said: "Your advisers are frequently divided. If you take the wrong course, and occasionally you may do that, then you bear the burden of responsibility while the advisers may move on to new advice."

DESIGN AND IMPLEMENTATION

If automation can be cost-justified, the local computer center can, with your assistance and approval, design a system to meet your needs. They can explain, with the aid of something called a flowchart, the flow of data from the time it enters the system until it is discharged. They can graphically describe for you at what points along the way various things will occur. They will explain to you how to control the use of the system and all of the data—good or bad. They will revise and revise again and continue to revise the design until it is the way you want it.

At that point, they will begin the application programming required to make the system something more than a design. They will convert the system design into a usable, workable tool—a tool that can be modified. Another word to the wise at this juncture. If a computer system cannot be modified in a reasonable amount of time, do not accept the design. The user must be sure that any system being implemented can be changed with relative ease by the people who designed it. The only time that statement can be disregarded is in the unlikely event that you—the user or the librarian—can be certain that *your* activities will *not* change.

DOCUMENTATION AND TRAINING

Another service a good computer center can offer is that it will write training manuals and provide the initial training in the use of the automated system. As with every other part of system development, your assistance will also be re-

quired in development of the training manuals. You must be certain that the manual is written not so that it can merely be understood—you must be certain that it is written so that it cannot possibly be *mis*understood.

Training in the use of a computerized system is one of the most important aspects of automation. I'm sure you know that. But I'm not sure you recognize just how critical it really is. Those of us who are in the data processing business know that nearly 70 percent of all computerized system failures are operational failures. They are operational failures because the user has not been properly trained in the system's use. And I know further that if a human being cannot adequately use an automated system, he or she, whether directed otherwise or not, will resort to a nonautomated, more comfortable method, sooner or later.

AN INTERFACE TO OUTSIDE VENDORS

If the computer center does not have the resources to develop a system for you, and the decision is made to go outside for one, you can still get lots of help from your computer center. They can write specifications for you. If the computer center people have been in business very long, they are probably fairly good at writing specs. They can also lend you technical assistance in the bid evaluation process. And they can separate the wheat from the chaff at contract negotiation time. They can also assist you in dealing with the vendor once a system is selected, by wading through the technical trash thrown at you by the vendor every time he or she misses a deadline.

REALISTIC TIME FRAME

Computer centers exist to provide service, but it should be pointed out that they do not sit on their collective duff waiting for someone to request service. Systems development is usually scheduled twelve to eighteen months in advance. It is important, therefore, to get in the queue long before you need the system to actually work. You probably think that is too long to wait. Maybe it is, but I doubt it. Many times libraries and others turn to hardware vendors when they are told that it will be eighteen months before the computer center can begin work on their system. What they usually find out, however, is that the eighteen months that they were to have waited have come and gone and the vendor, the purchasing department, the budget, and other obstacles have extended the time of the implementation far beyond the original eighteen months anyway. Or, the vendors who promised installation in six months did install, but two full years have elapsed since that time and they still have not made the system perform as promised. They have taken your money, however, right on schedule!

There is one other important thing I feel I should mention about vendors before finishing. Some will tell you that their system is so simple that any fool can use it. But I contend that if a system is built that any fool can use, then only fools will want to use it!

CONCLUDING REMARKS

I can recommend trying your computer center. You don't have to stay with them, but give them an opportunity. They can assist you. After all, you prob-

ably both work for the same agency and you both have something to lose or to gain.

One final remark. This is in regard to the purchase or lease of an automated system or, for that matter, the purchase or lease of anything. I don't know who said this but I believe it is very true: "It is unwise to pay too much, but it is worse to pay too little." When you pay too much, you lose a little money—that is all. When you pay too little, you sometimes lose everything, because the thing you bought was incapable of doing the thing it was bought to do. The common law of business balance prohibits paying little and getting a lot—it can't be done. If you deal with the lowest bidder, it's well to add something for the risk you run, and if you do that you will have enough to pay for something better.

The Sources of Disharmony*

Micki Jo YOUNG: Resources on Request, Reston, Virginia; and
Walt CRAWFORD: The Research Libraries Group Inc., Stanford, California.

FUNDAMENTAL INCOMPATIBILITIES

Librarians and data processing people live in two different worlds. Both of them are professionals, both are extremely serious and dedicated about their work, and both are service oriented. Moreover, they both are oriented toward data and detail. In spite of all these parallels there is frequent discord. Now, what is true of many professionals is true of both the librarian and the data processing professional: both are guilty of tending to think they know more than they actually do. Each one sees reality in personal terms. On one side we have the world according to Dewey, or life on a three-by-five card. On the other side we have the world according to Hollerith, or life on the punched card. Each has a language or jargon of its own, and this professional jargon becomes a serious barrier to communication. No matter what the profession, when two professionals try to communicate with their own jargon there has to be a translation. And this is particularly troublesome in the field of library automation because both speak in homonyms. For instance, the words *record, list,* and *file* mean one thing to the librarian but mean quite another thing to data processing personnel. In the two environments these are entirely different things. There is a fundamental disharmony. We hope to outline some of these sources of disharmony in more detail.

When a library relies on a computer center, typically it is either an academic

*In its original presentation, this paper consisted in part of a role-playing dramatization of the misunderstandings that can arise in the dialogue between the computer center and the library. It would be impossible to adequately convey the wit and charm of that dramatization in a written form, and therefore it has been omitted here. We hope we have nonetheless preserved much of the wisdom contained in the rest of the material that accompanied it.

or an administrative computer center we are dealing with. Now, if it is an academic computer center, there may be an insurmountable incompatibility right at the start. Usually, the campus computer center is running thousands of little jobs with not much attention to file security. They are interested in getting people onto the computer, not in keeping people away. If the system crashes, it is no big deal. They can just ask the students to rerun their jobs. "Priority" jobs are often enough run at midnight because during the day, of course, the machine has to be available for those 6,000 students running their 12,000 jobs. We aren't even going to consider that kind of computer center. That is much too difficult a challenge, and an almost certain cause for divorce.

The business or administrative computer center is more like it. Here the programmer-analyst is at least familiar with problems that look pretty much like the library's. Circulation looks like inventory control, more or less. Acquisitions? Every computer center deals with order writing and fund accounting that should be quite straightforward. Union lists? Cataloging? Naturally, everyone has some experience doing data entry sorting and listing, and so we merely have to put the package together. It isn't clear whether this familiarity with similar problems is an asset or a liability. Sometimes, following this path and with the best of intentions, the library comes up short. Obviously, a shelflist is not just another parts list; a circulation system is not just another incarnation of inventory control. Likewise, acquisitions differs from standard purchasing in a great many ways.

SPECIAL REQUIREMENTS

In the field of automation, the library's needs are really unusual. For one thing, the need is ubiquitous. Automation has been introduced into acquisitions, cataloging, circulation, serials control, interlibrary loan, and administration. And the state of the art these days calls for a sophisticated networking environment. We have libraries sharing information resources through large data bases. And these data bases can be very dynamic at times. For example, if you try to keep track of a title that is claimed and then is back ordered, and then the price changes and you have to go to another publisher, all of these data elements are in constant change.

The list of requirements seems very long indeed. We have heavy demands for sorting by specialized collating sequences. We have textual data in uppercase and lowercase. We have accents and diacritics; we have special characters. We simply have to have brackets. Brackets mean something different to us from what parentheses mean. We have multiple access points for every file. We are not willing to expect our users to know any one single element to look an item up. We have to have immediate access. It wouldn't do for a user to put in a request for titles by John Steinbeck and have to come back the next morning to see what the list looks like. We have to see our output in a variety of different forms: COM, cards, multipart paper. These are things we deal with in the library every day.

In the area of data structure, we have the MARC format, which is very comprehensive and complex. Very often we have a storage capacity problem; we have to have so many tape drives, or extra disks. We might have peripherals that are strange; we need to have light pens or an OCR badge reader. We have

to have a printer that uses the ALA print train; we must be able to print equally well on cards or paper stock. Sometimes we may need to accommodate stand-alone systems with floppy disks. And finally, we have to have the right always to change our minds.

In the data processing environment, these requirements translate into things like variable-length character strings, inverted file structures, special sort key constructions, on-line interaction, telecommunications, special input/output devices, data base management systems, and highly sophisticated operating systems. Altogether, the library's data processing requirements are considerably more complex than those of the average department the computer center deals with.

CAREFUL COMMUNICATION

Immediately, it should be apparent that there are a lot of details to be worked out. The data processing person must listen to the librarian, must learn of these requirements, and must learn the what and the why. There is also a great need for coordination and integration, and this, too, is an area that requires careful communication between the librarian and data processing personnel. It is not unimaginable that a library could be using a vendor's terminal for on-line acquisitions, be cataloging through a bibliographic utility, be using a turn-key circulation system, and have an accounting system operated by the parent agency. Without an integrated systems approach, each part of the automation package might end up serving only isolated and parochial needs.

The interaction between reliability, simplicity, and redundancy is also worth watching carefully. How many libraries have found that they must maintain a redundant manual system or, even worse, introduce more complexities into the manual system simply to ensure the reliability of the automated system? One of the things we all expect will happen once we automate is that routine things will become easier. Yet it seems that after automation more than a few libraries have a brand-new manual file that does nothing more than control the control numbers.

ROOM FOR COMPROMISE

Unfortunately, not only is it true that the analyst must listen to the librarian; the librarian must also listen to the analyst and consider some of the realities of this other world. There are genuine trade-offs. Not all your products necessarily require diacritics, and the extra cost of providing them might not be warranted. When you ask the computer center to mount the ALA print train and to provide half a million characters of core and 600 million characters of disk for twelve prime hours every day, you may want to consider that the line printer can print at only a fraction of its usual speed when the ALA print train is mounted. There is room for compromise in all these things, and the precise areas of compromise can be pinpointed through analysis.

A careful job of analysis, of course, is the backbone of any successful automation effort, and given some library situations, the analysis may in fact be more valuable than the automation that follows it. By asking the right questions, a good analyst will frequently discover precisely what is going on in the library. Moreover, an insightful analysis will cut through historical precedents. The fact

that the library has been doing some particular thing for seventy years does not validate its continuing to do so for eternity. Those historical functions are not to be treated as sacred cows. Often enough, a good analysis will result in the recommendation that certain procedural simplifications be made but that no computer processing need be done. Your best analysts will be willing to say so.

Exception routines need also to be considered. In a large system there may be hundreds of exceptions. It is worth looking at how often these exceptions occur in real practice, and making a judgment on how much effort you are willing to put into programming to handle a situation that might cost only $20 a year to deal with manually. Some exceptions, however rarely they occur, must nonetheless be handled by machine. It is also worthwhile to look at the data elements and see which ones really need to be displayed. Do you need every MARC data element displayed in every product? In fact, would your users even be willing to use such a display? Again, there is room for compromise.

SENSE OF DIRECTION

What it finally comes down to is the recognition that much of the interaction between the librarian and the analyst is based on a sense of overall direction. Do you really know where you are going, and what you want the system to do? Once you know your aim, once you have made that aim clear to the analyst, then there is reason to believe that progress can be made. The two sides will never speak exactly the same language, but the computer people are there to help, and sometimes their contributions can be quite significant. Unfortunately, it is often easy to get started on a project without really much caring where you are going to end up. But you should care. And if you do care, you should also be willing to recognize that you may not know your library as well as you think you do.

The Horror Story

Ken BIERMAN: Public Library, Tucson, Arizona.

PROMISES, PROMISES

Since we are talking about marriages, I can't help but be reminded of the story about a woman who had been married three times and was still a virgin. When asked how this situation came about, she explained that her first husband died on their wedding night, her second husband was a little bit flaky and wasn't interested in her, and her third husband worked for the computer center and all he did was sit by the edge of the bed and tell her how good it was going to be. All of us who have worked with a computer center have at one time or

another had the experience of feeling that the computer center is sitting on the edge of the bed telling how good it is going to be. We have already heard what one should expect from a well-managed computing center.

I am going to talk briefly about what you can expect to get from a *poorly* managed computing center. I hope by this to encourage or frighten you if need be into considering very carefully the choices and decisions that you have to make.

First of all remember that in dealing with a computer center, just as in dealing with a library, there are many different organizational parts with different functions, different goals, different purposes, different people, so you are not dealing in general with just one thing, you are dealing with those many different parts. For my presentation I would like to explain about the computer center as it functions in the two specific areas of development and operations and briefly mention some of the problems that can arise in both these areas.

It can't be done. With respect to development, the attitude of the computer center will be on a continuum and ranging between two extremes. At one extreme will be the response, "Just tell us what you want. Anything you want will be yours by next week." This is the "sitting on the edge" response. At the other extreme is the response, "Oh well we just couldn't possibly do that; we can't hardly do anything at all." In reality of course, the response will most often be somewhere in between the two extremes. But you need first to discern where the computer center is coming from along this continuum. The closer it is to either extreme the closer your library may be to being in trouble.

After the system study and after you have decided to automate something, the computer center has designed something. Presumably some programs have been written, and you are beginning to get some results. You may discover that the system doesn't quite do what you thought it was going to do. It may be that certain minor things have been forgotten. One acquisition system was designed and implemented that neglected to take care of deencumbering items after they had been received. I know a circulation system that checked out books fine, but it didn't quite handle the check-in right. So maybe some minor problems develop that you need to work out. Indeed you may soon learn that the other end of the continuum was indeed correct; "the computing center can't hardly do anything at all."

Always remember that any time estimates which the poorly managed computer center gives you should be at least doubled and indeed possibly tripled. The same is even more true of costs. Most important of all, remember that the size and amount of equipment is to a computer person what the size of the book collection is to a university librarian. Like the university librarian will do almost anything to get more books, a computer center person will do almost anything to get more equipment. So do not let yourself be put into a situation of being used as a pawn to help the computer center justify getting more, perhaps unneeded, equipment. I know, for example, of one circulation system decision that was made strictly on the basis of which vendor bid submitted the largest minicomputer. It made no difference what the specifications were, no difference what the system did; the decision was made solely on the argument, "This is the biggest minicomputer; this is the vendor we will go with."

A QUICK HONEYMOON FOR ONE

So much for a brief overview of some of the things that can and will go wrong during development. Let's move into operations. Assume the computer center is operating your system, either on-line during the hours you are open, or as a batch system, perhaps performing a weekly batch run that is done on one night of the week or on the weekend. The first thing you will discover in operations, if indeed you didn't discover this in development, is that on the computer center's list of priorities if the library isn't at the very bottom it is hard to imagine what *is*. It always seems to work this way, even if at the outset the computing center is being very attentive to your needs. (Since we are talking about marriage, you are in the honeymoon stage.) You will discover that like most honeymoons this one has a very short duration. You may discover that before you realize it the library has gone to the bottom of the list. The computer center has moved on to the next department, the next hot topic of the day. I can't help but think of the cartoon which I am sure most of you have seen with all these people sitting around just laughing uproariously and the caption is You Want It—When?! In terms of scheduling and time, you may soon discover that operations work the same way. I am aware of one library, for example, that went eight weeks without ordering a single book because it was impossible to get the weekly acquisitions system to run. The computer center was involved with the end of fiscal year processing for other departments—with payroll and so on and so forth—and somehow those things, for some reason, seem to be more important than the work of the library.

YOUR ORDERS ARE OVERDUE

You may discover that your on-line system suddenly isn't on-line and no one at the computing center seems to be concerned about it. More common, you begin to have garden variety operational problems. The operator has mounted the wrong tape, the wrong disk, or perhaps the wrong forms. Purchase orders were printed on first overdue notice forms or first overdues were printed on purchase order forms or some listing was printed on label stock or a series of mailing labels was printed on a listing stock. These are the little annoyances that you have to learn to deal with and ultimately accept when you are working with a poorly managed computing center.

SERIOUSLY NOW

Well, enough of the levity. I would like to offer four suggestions that can help overcome the problems to which I have alluded:

Understanding Comes First

Always the key to solving any problems, this includes understanding on the part of the library of the computing center's problems and priorities and an understanding by the library of its own priorities as well. This is an extremely important area.

Communication

One person within the library should be responsible for coordinating the

communication and working with the computer center. I emphasize both formal and informal communication. Take your analyst to lunch now and then, particularly if he or she ever does a favor for you (of course this never happens so this makes it difficult). Some of these considerations can go a long way toward furthering smooth operations between the library and the computer center.

Documentation

I am not referring to documentation in the narrow sense that the computer people use the term. I am referring to the whole area of documenting what you want—what decisions have been made, why they were made—documenting them all in writing.

Planning

The library and the computing center together plan the development of the operation over the short- and long-range future.

I am sure you have all heard of Murphy's law, that is, what can go wrong at the worst possible moment, etc. It may be that after dealing with a computer center you may conclude that Murphy was indeed an optimist.

Negotiating a Workable Relationship

Barbara MARKUSON: INCOLSA, Indianapolis, Indiana.

A QUESTION OF CONTROL

In one of the first jobs I ever had in a library, word came over that the chairman of the home economics department wanted her own library and what she had in mind was that her secretary would run it. And the reason she wanted her secretary to run it was so that she could have control. Now you can imagine how that was greeted by the library's administration, rather alarmed with the idea that the home economics secretary might be running a library. And that is just about the way the computer center director will respond initially to the word that the library is thinking about establishing its own computer center. So we are going to start with the assumption that for various reasons you are not going to be allowed to compete with the computer center but indeed will be working with them.

Now I have several disadvantages in giving this speech; for one thing, I have graduated from the Yale School of Drama, and I left my southern compendium of humor at home. And I also unfortunately left my reading glasses back at the hotel. This is going to be a struggle for all of us. You have an advantage, you can leave and I can't.

LEARNING TO NEGOTIATE

What we are talking about is being involved in a continuing series of negotia-

tions. All negotiations involve risk. You are going to negotiate the best possible position and the best possible system from a computer center and they are going to negotiate as much as they can to protect their resources, their time, and so forth. So we need to think and learn a little bit about the process of negotiation and the characteristics of good negotiation, and there is a lot on this in the literature of business management—checklists of things to go through. You are all good information specialists; I assume you would begin by doing a considerable amount of reading. When two parties start to negotiate, one has to have a clear understanding of exactly what it is you are negotiating for—where you will give in, where you won't give in. The other party has the same situation. Occasionally in negotiation you bring in an outside arbitrator. We frequently call them consultants.

The first thing you might want to do is to learn all you can about the other party, all their strengths and weaknesses. You might want to gather just some basic facts; any good reference librarian should be able to do that. Find out something about the equipment and the services. What kind of staff turnover do they have? What kind of programming languages do they support? Do they have annual reports? Are there long-range plans? Do they have any happy customers? Can you nail down some of their unhappy clients and find out what their problems were? What is the background of the key staff and so forth? Maybe you will be lucky, and the computer center won't think to find out anything about the management of the library.

BASIC DIFFERENCES

In return you might also want to contemplate the differences in these two things that are both called service organizations—both providing service to a central administration. They are both overhead items. They are both agency resources. Then things begin to diverge. The library is theoretically an infinite resource, and generally users are barred from using that resource by such considerations as the hours of service. We rarely have users in contention for equipment because we have very little equipment that prevents a user from having access to the service that we provide. The library usually expects its users to develop their own programming—learn how to use the system, how to make their own schedules, and so forth. So from that point of view the library is a relatively flexible access organization. Also the library *does* want to increase the use of its resource so it can get more money. And it wants to reach new user groups.

The computer center manager has a different problem because he is generally dealing with a finite resource that users can't come in and use very openly. This is particularly true when we are talking about the administrative data processing center which I think you have learned is probably your best bet. The major barrier to a wide provision of that resource and a truly service-oriented computer center is that there is a major investment in staff and equipment before the user can even come in the door. Maybe eventually we will reach that golden day when everything is instantly programmable through truly generalized programs that do everything for anyone at any time. But as it stands now there is a heavy degree of staff involvement to allow the user to use this resource: systems analysts, programmers, operators, and so forth. The situation

that results would be roughly equivalent if the librarian had to decide whether students from the business school or students from the psychology department were to have access to the library between 8 and 10 a.m. As far as the computer system director is concerned, the staff is a finite resource, and so an important barrier is the availability of that staff to solve your problem within the time frame that you want.

Another key characteristic that you might find in some administrative data processing centers is that they are providing some basic administrative service such as payroll for the overall agency, and so they tend to have a lot more clout in convincing the power structure that their decisions are right as opposed to those of the library. The library may not be providing an important basic service to the top administrative structure in an organization, whether this be a business, a university, or a public agency. So in the contention for available resources you may not be as strong as you would wish.

It is also important to recognize that most computer center directors are concerned with maximum effective use of the computer resource and as they evaluate prospective users, they want to maximize their control over those users so they can maintain efficient scheduling. They want to maximize their own ability to establish schedules so that they can get themselves out of difficulty when the need arises. And they want to set the user's behavioral standards as well. That's the goal they are going to try to negotiate for. They have as a primary objective the running of an efficient center. So that is where we start with this negotiation.

Now you can simply accept the computer center's verdict on all of these things and live with it, but I think you heard from what has preceded that this is rarely going to lead to an ideal situation. So you want to negotiate what you are going to get for the library from this available resource and also negotiate as carefully as you can the different roles and responsibilities that you will all play.

In this negotiation you may feel about the computer center staff the way some people feel about consultants. These are the people who borrowed your watch to tell you what time it is. Now in some organizations where the computer center has tremendous clout, any of your attempts to use any computer system may be modified by the computer center director. In other words, this director has a kind of authoritative control over any agency expenditure for data processing and may get involved in whether a library can join a network or buy a turnkey system. I know a library that had to go into a detailed thirty-nine-page justification to prove to the computer center that it could not put up a bibliographic utility locally just for that library alone. So an exhaustive justification is probably going to have to be made in a tightly run organization whichever way you go, and it should go a long way in convincing everyone satisfactorily, we hope, that the right way has been selected.

MORE THINGS TO NEGOTIATE

Documentation is also something that you will want to negotiate. Who has access to this data? Is the library the only one who has access? Will they send duplicate sets of all documentation and programming to the library? How are charges to be calculated? Are there any planned changes in charging algorithms? What about long-range viability? You don't want to be halfway into a sys-

tem (and there are many instances of this) where the computer center decides to change equipment but doesn't even inform the library of this plan. You would want to negotiate some arrangement where you will be regularly notified about changes that the computer center contemplates and what action they are going to take to ensure that your system is not impeded by this. Identify the staff that will be working with you on your project.

Recently I talked to people at a library with a long history of use of a computer center. They have just gone through one of these negotiations so I thought I would bring you some practical types of things that they filled out in their negotiations. This was a circulation system. The library will have a small in-house staff—they have a systems analyst and a programmer in their shop to develop the system specifications, to monitor the programming, to review the project, to specify needed modifications, and to help put the system up after a crash. Now the computer center is not too keen about the library having this kind of staff, but that has been negotiated over time. The library analyst, programmer, and key operational staff at the computer center are all working on a single team so that everybody knows right from the beginning what this system is supposed to do.

In earlier times, the library's programming had been farmed out to people in bits and pieces, and there was no accountability. Nobody really could ever be found who would accept responsibility for any of this. So they have negotiated a team approach. They have lived with the result of the negotiations, that the center will indeed do the bulk of the programming and that the computer center is retaining this right. They have negotiated an agreement that if the center can't bring the system up within an agreed-upon schedule the library has the option of doing its own programming. The library still thinks that the computer center would fight this if the situation ever arose, but at least it is down in writing. At the same time, the library specifications must be developed within the general requirements of the computer center; in other words the library can't ask for something so far out that there is no way that the computer center could reasonably respond.

They have negotiated that 90 percent of the time the library will have one-day turnaround. The library recognizes that occasionally there are situations where the computer center simply can't give them that, but that kind of thing is very important for a circulation system. They have negotiated that the library will sign off after it performs all testing. Moreover, the library has the right to say whether the test runs are doing what the library thought they would. They have negotiated that the documentation will be thorough and consistent, have identified the documentation at all levels, and established that complete sets will be available in the library systems office and in the computer systems office as well. They have negotiated that the library knows the key staff assigned to its systems at all levels and that it can communicate directly if a problem arises. This is something that the computer center didn't like at all; they had preferred to have all communications channeled through one person, but a lot of difficulty developed particularly as the computer center had assigned eight different analysts in four years.

So this library has learned from experience over a period of time. By living within the constraints of the computer center, it has been able to automate cir-

culation at a reasonable cost both to the agency and to the library, and is gradually upgrading it. But with every renegotiation the library has become a little bit more sophisticated and has felt that it cannot mount such a complex system for a university library that relies entirely on the computer center. They feel quite confident that with this joint approach right from the beginning and with a strong commitment both from the computer center and the library director they will put up a workable system within a scheduled period of time. While this is not exactly a bed of roses it is something much rosier than they had in the past.

HOMEWORK TO BE DONE

In sum, I think the library needs to do a considerable amount of homework to establish a framework on both the library's side and the computer center's side and to consider this a negotiation that will go from major negotiations in the beginning to an incredible number of minor negotiations as a stage in this long-term system development process. I would encourage you to study as much as you can in the management literature about negotiation and follow a checklist of essential points for good systems development. On the library's side, see that you are doing your job well because from this you will derive some of the bargaining power you will need.

A Final Word

William D. MATHEWS: NCLIS, Washington, D.C.

ARMIES IN THE NIGHT

An aspect of the exercise we have conducted here amounts to holding a mirror to ourselves and laughing at our own image. Judging from some of the horror stories we should be thankful no doubt that we still have such a finely honed sense of humor, and still derive enjoyment from what we see.

I have some familiarity with the topic under discussion by virtue of the fact that for many years I have been a messenger between the library and the computer "camps." More than once I have been struck with the awesome sense that I was dealing with two armies in the night, waiting for that terrible moment when the first light of dawn brings the features of the landscape into prominence, and the machinery of destruction starts to stir.

It is in the university setting, in particular, that the conflict—this malevolent desire of libraries and computer centers to tear each other to shreads—seems most pronounced. This is peculiar, in a way, since universities should be such tranquil places and, after all, we are talking about two information agencies. There ought to be some mutual respect.

Nonetheless it appears that there isn't much respect, and a central point

about which this conflict revolves has to do with differences of temperament. Reflecting back to Isaac Bashevis Singer's remarks on Sunday night, it strikes me that this tension between computer technologists and librarians is yet another manifestation of the cleavage between the two cultures. The librarian is often conceived of as preserving our culture; bringing order and coherence to our information universe; helping us to sail on untroubled waters. The computer technologist, on the other hand, seems to thrive on change, disorder, chaos, and disarray; tolerant of the noise, clutter, and unanticipated twists of life; willing to skate near the edge of madness and total entropy.

SOME GENERALIZATIONS

I know that generalizations are always dangerous, and I sense that I have just made one. But I think I'll risk being wrong and make a few more.

- In the computer field it seems that everyone is an instant expert. One year of FORTRAN and you're a systems analyst. In the library field, on the other hand, you can have four advanced degrees, two foreign languages, and ten years on the job and still not be considered an expert on anything.
- In the computer field there is a tendency to start solving the problem before it has even been structured or defined. By contrast, the librarian seems to feel that structuring the problem is as good as solving it—forever intrigued by the preliminaries, disinterested in the final results.
- The emphasis in the computer world is on relevance and actionable items in contrast with the more speculative, contemplative, perhaps pensive world of the librarian.
- Then there is the antiseptic, almost barbaric dispensation of user services one finds in a computer center (an abrupt dictum to move a comma three spaces to the left, or to balance parentheses) contrasted with the cordial reception one often finds in the library. Yet, of course, balancing parentheses might turn out to be precisely the right device even if it is abrupt, while I can often recall partial answers leading me down a labyrinth of stacks in the library. Sometimes precision helps.

MUTUAL RESPECT

But these comments only serve to heighten and perpetuate the differences, as did much of our discussion here today. Let me return again to mutual respect. I think it holds the key. To have mutual respect one must deal from positions that seem to have parity. There must be some equivalence in strength. Lately it seems that the bargaining chips are all lining one side of the board. The computer center seems to have the strength.

To be sure, the computer center's magic is awesome to behold. When a computer system truly works it produces something quite marvelous—new data, new knowledge, and sometimes a profound new appreciation of the universe. Without the computer we would never have begun to guess at the ultimate complexity of that greatest of all carriers of information: the DNA molecule. Without very sophisticated data processing techniques we would hardly begin to understand the fine structure of the atom, and without some very intricate navigational computation and advanced telemetry, it would have been impossible even to appreciate the aesthetic qualities of the clouds on Jupiter.

WITHOUT THE LIBRARY

Well, what of the library? Without it the entire climate of our civilization would be altered. In a humorous vein: "Without the library we would have had to invent the TV set centuries earlier." But seriously, the library is the memory for all mankind. Mystical things—the record of religious experience is embodied there. Diaphanous, gossamer things that flee with the wink of an eye abound. And there be dragons too: fulsome creatures drawn 'round the fire with their feet propped up on potbellied stoves, stuffed with their dinners of pheasant and squash. The heights of imagination, illusion, drama, and lyric poetry can all be found. Without the library the classical world would be the merest flickering shadow of what we now perceive. Our technology would be in disarray. Some of the simplest tasks of modern society would become impossible. Much of our research would be fruitless and redundant and our industrial productivity would be a fraction of what it is today.

THE STRENGTHS OF BOTH

And what if we combine the two; fuse the power of electronic information processing with the library's memory? Then do we not have the implements to decipher the first footprints of humanity at Olduvai Gorge, to measure and date the rock, and to piece together that finding with the larger record of anthropology? And can we not decode the faint but pulsing heartbeat of objects in distant galaxies and combine that information with the body of astrophysical research?

It seems to me that both sides can deal from strength, can assert a belief in their inherent values and self-worth. Computers and libraries need each other and deserve mutual respect. As a messenger to both your camps I wish you peace.

Run down the following partial list of ALA/LITA benefits:

Their value alone is considerably higher than the $15 it costs to join LITA if you're already an ALA member, or the $65 if you're not ($15 for LITA plus $50 for ALA).

- *LITA Newsletter*
- *Journal of Library Automation*
- *American Libraries*
- *A member's discount on all ALA and LITA publications. Some free*
- *Free information on consultants and activities in the field*
- *Reduced registration rates at LITA institutes*
- *Excellent group insurance plans*
- *A voice in legislation. LITA representatives testify before congressional committees and regulatory agencies.*

Follow the LITA by supporting the Library and Information Technology Association.

Call or write: Library and Information Technology Association, American Library Association, 50 E. Huron St., Chicago, IL 60611 (312) 944-6780

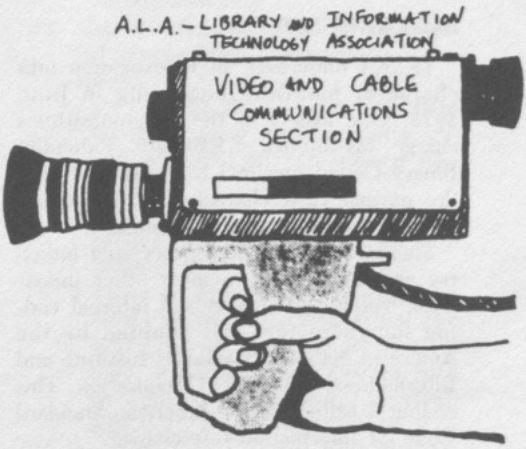
VIDEO/CABLE INFORMATION PACKET

Covering all phases of video and cable in libraries

Compiled by Marilyn Rehnberg

Includes:

- Federal Communications Commission Bulletin on Cable TV 28p
- CATV Bibliography by Marilyn Rehnberg 4p
- A Guide to Video Resources by Arlene Farber Sirkin 10p
- VCCS 1977 National Survey of Library Video Resources 29p
- Listing/Description of Regional Advisors of National Network of Libraries using video and cable
- Sample issues—LITA NEWSLETTER and CABLELIBRARIES



Packet is free to LITA/ALA members; $1.00 prepaid for postage and handling to others. Order from LITA/VCCS, American Library Association, 50 E. Huron St., Chicago IL. 60611.

Communications

EBCDIC Bibliographic Character Sets—Sources and Uses: A Brief Report

Walt CRAWFORD, The Research Libraries Group, Inc.,* for the Technical Standards for Library Automation Committee (TESLA), Information Sciences and Automation Section (ISAS), Library and Information Technology Association (LITA).

SUMMARY

The use of extended EBCDIC for bibliographic processing was studied as part of a TESLA investigation. Results show that extended EBCDIC is heavily used, that there is little support for standardization, and that extended EBCDIC sets typically derive from one of two sources.

No standards activity is recommended. A table showing two extended EBCDIC sets is provided, along with a table of differences for several sets actually in use.

Guidelines are suggested for transmission of bibliographic data in EBCDIC form, with the caution that only ASCII8 transmission can be considered "standard."

BACKGROUND

TESLA undertook an investigation into character handling beginning in June 1978. One aspect of the subcommittee's charge was the use of EBCDIC (Extended Binary Coded Decimal Interchange Code, the normal IBM character representation method) for bibliographic processing.

Standards in the data processing industry are, like those in most other industries, voluntary. A standard internal coding for characters was adopted by the American National Standards Institute and International Standards Organization. This coding is called ASCII: American Standard Code for Information Interchange.

*Work done at Library Systems Office, University of California, Berkeley. This report prepared at RLG, using the SCRIPT and SYSPUB facilities of SCIP.

ASCII exists in two forms, one using six binary digits (bits) per character, the other using seven bits. Neither the six-bit nor seven-bit ASCII is useful for full multilanguage bibliographic processing. A six-bit character can have no more than 64 values. Six-bit ASCII thus does not allow for such niceties as lowercase letters, except through escape codes. Seven-bit ASCII allows 128 values, of which the first 32 are control characters. This leaves 96 possible printable characters, enough for uppercase and lowercase but not an extended set of altered characters (e.g., slash O) or diacritics (e.g., haček). For this reason, a third ASCII set was defined and adopted as a standard for bibliographic processing.

This character set, which is commonly abbreviated "ASCII8," uses eight bits per character, allows 256 values in all, and provides for all altered characters and diacritics appearing on the ALA Print Train, using escape sequences only for superscript and subscript numerals and for Greek letters.

While ASCII8 is usually used for transmission of bibliographic data on tape, most actual processing uses EBCDIC. EBCDIC uses an eight-bit character, but there are only standard definitions for 96 of those characters. There is no "standard" extended EBCDIC, because EBCDIC is not considered a standard. The subcommittee wished to determine the extent of use of EBCDIC for bibliographic processing, whether there was a de facto standard extended EBCDIC, and whether it would be feasible to prepare a pseudostandard.

INVESTIGATION

A draft questionnaire on EBCDIC and display devices was distributed to thirty interested parties at ALA Midwinter. Hastily designed and distributed without planned follow-up, the questionnaire yielded little useful information.

Personal contact and letters did yield the internal character sets of OCLC, RLIN, and WLN and that of the Library

of Congress. The University of California at Berkeley also uses EBCDIC, using the same character set used by UC in general and a number of other institutions (those using services or software provided by Blackwell/North America).

At Midwinter, a representative of Rutgers University suggested that there was, in fact, a standard extended EBCDIC character set. This character set was provided: it is IBM publication 1403-03, and is IBM's suggested codes for the ALA print chain.

These character sets were compared. In total, they probably represent the bulk of EBCDIC use of bibliographic processing. They do not represent the universe: undoubtedly, other institutions are using different sets.*

RESULTS

There are two primary extended EBCDIC character sets: that designed by the Library of Congress early in the history of MARC processing, and that suggested by IBM for users of the ALA train. These two sets are shown with ASCII8 equivalencies in table 1, excluding control characters and unassigned ASCII8 values. (Only those characters *not* in basic EBCDIC are shown, to conserve space.)

Thirty-one additional ALA print train characters are not part of the "regular" ASCII8 set, but are defined as escape sequence characters. Subscript 0, +, −, (,), and 1–9 are, respectively, hex 65–69 and 71–79 for LC, and hex 80, ca, cb, bc, cc, and 51–59 for IBM. Superscript 0, +, −, (,), and 1–9 are hex 55–59 and 41–49 for LC, and hex b0, 8e, 8f, 8d, 9d, and b1–b9 for IBM. Alpha, beta, and gamma are fd–ff for LC and aa–ac for IBM.

OCLC, UC, and B/NA libraries use EBCDIC sets derived from the Library of Congress set. RLIN, WLN, and Rutgers (at least) use sets derived from the IBM publication.* There are minor differences within IBM-based and LC-based character sets. Table 2 gives those cases where LC, OCLC, and UC differ from one another; table 3, the larger set of cases where IBM, RLIN, and WLN differ (generally because IBM does not suggest codes for all cases).

Note that some essential control characters are omitted from these lists, particularly the end of field and end of record characters. These *do* differ among character sets, more widely than other characters, and should be specified in any EBCDIC transmission.

CONCLUSIONS

1. Standardization efforts for EBCDIC do not seem worthwhile:
   - EBCDIC usage tends to be deeply embedded in computer software at institutions using EBCDIC for bibliographic processing.
   - Responding institutions almost universally expressed distaste for the idea of changing their own software to conform to a standard, though some said "it would have been nice if. . . ."
   - Most major users are already accustomed to EBCDIC/ASCII8 translation.
   - ASCII8 does constitute a recognized standard, and there is no recognized method of generating a second standard.
2. Use of extended EBCDIC for internal processing is surprisingly widespread.
   - OCLC, which does not use IBM computers, nonetheless uses a version of EBCDIC for internal processing.
   - Most users contacted used extended EBCDIC for internal processing.
   - There are known cases in which records were translated from one EBCDIC to ASCII8, written to tape, translated from ASCII8 back to another (virtually identical) EBCDIC, then retransmitted.
3. Any effort to transmit bibliographic records in EBCDIC *must* be accom-

*Since the draft version of this note was prepared, Northwestern has submitted an EBCDIC set that differs from those shown, placing all diacritics (not special characters) below hex 40 for ease of sorting and identification.

*IBM Publication 1403–03, Order #GA 24-3073-7, p.36–37.

Table 1. LC and IBM Extensions.

ASCII8	*Name*	*LC*	*IBM*
1f	Delimiter	fa	9a
5b	Left square bracket	3e	ad
5c	Reverse slash	fc	e0
5d	Right square bracket	3f	bd
a1	Polish L—slash L	ce	42
a2	Slash O	db	43
a3	Crossbar D	cb	44
a4	Upper thorn	dc	45
a5	AE digraph	ca	46
a6	OE digraph	cf	47
a7	Miagkiy Znak	51	——
a8	Middle dot	64	49
a9	Musical flat	3d	41
aa	Patent mark	3c	48
ab	Plus/minus	ab	9e
ac	Hook O	da	——
ad	Hook U	dd	——
ae	Alif	53	ed
b0	Ayn	54	ce
b1	Slash l	8e	62
b2	Slash o	9b	63
b3	Crossbar d	8b	64
b4	Lower thorn	9c	65
b5	ae digraph	8a	66
b6	oe digraph	8f	67
b7	Tverdyi Znak	52	——
b8	Undotted i	8d	68
b9	British pound	6a	db
ba	Eth	8c	——
bc	Hook o	9a	——
bd	Hook u	9d	——
e0	Pseudoquestion	23	8a
e1	Grave	2b	79
e2	Acute	0f	74
e3	Circumflex	08	71
e4	Tilde	29	a1
e5	Macron	19	a0
e6	Breve	2c	77
e7	Superior dot	0a	75
e8	Umlaut/Diaeresis	11	72
e9	Haček	0e	78
ea	Angstrom	03	70
eb	Ligature, first half	30	fa
ec	Ligature, second half	31	fb
ed	High comma	1f	——
ee	Double acute	10	69
ef	Candrabindu	18	ef
f0	Cedilla	09	76
f1	Right hook	0c	cf
f2	Dot below	1c	de
f3	Double dot below	1b	df
f4	Circle below	1d	ea
f5	Double underscore	02	eb
f7	Left hook	0b	ee
f8	Right cedilla	0d	——
f9	Upadmaniya	28	——
fa	Double tilde left half	2d	——
fb	Double tilde right half	2e	——
fe	High comma centered	1e	——

Table 2. LC, OCLC, UC Differences.

Name	*LC*	*OCLC*	*UC*
Left square bracket	3e	ad	3e
Reverse slash	fc	64	fc
Right square bracket	3f	bd	3f
Plus/minus	ab	66	ab
High comma	1f	1e	1f
Double dot below	1b	1f	1b
High comma centered	1e	65	1e

Table 3. IBM, RLIN, WLN Differences.

Name	*IBM*	*RLIN*	*WLN*
Miagkiy Znak	——	ae	fc
Hook O	——	c0	9b
Hook U	——	d0	bf
Tverdyi Znak	——	af	fd
Eth	——	ba	da
Hook o	——	8c	8b
Hook u	——	9c	af
High comma	——	be	ee
Right cedilla	——	da	ba
Upadmaniya	——	cd	e1
Double tilde left half	——	dc	dc
Double tilde right half	——	dd	dd
High comma centered	——	bf	ec

panied by documentation:

- A full listing of the sending institution's print set and hexadecimal equivalents is nearly a minimum requirement.
- Separately, the critical MARC control characters (Field terminator, Record terminator, Subfield delimiter) must be listed with the sending institution's values.
- If possible, the sending institution's EBCDIC/ASCII8 translation table, in ASCII8 order, should be sent.
- As a shorthand for telephone work or other occasions, an EBCDIC set can be characterized as LC-based or IBM-based, within the set of exceptions (see tables 2 and 3) given.

4. With increasing use of mini- and microcomputers in libraries, there is also increasing availability of small computers able to write tapes in EBCDIC form. Users should be aware that such machines will probably *not* translate all ASCII characters correctly. Specifically, there are no proper EBCDIC equivalents for ASCII square and curved brackets, reverse slash, tilde, circumflex, or grave.

A Systems Approach to Label Production through the OCLC System

Christina BOLGIANO: James Madison University, Harrisonburg, Virginia.

The implementation of OCLC's label production capability is reviewed from a total systems point of view. Cost/time comparisons with one popular manual labeling system are made. Various factors influencing automated label production, such as the physical characteristics of the printer and format of the label stock, are considered, and the impact of automated label production on the technical services work flow is defined.

The emerging beauty of the OCLC system lies in its capacity for integration of those library operations that in recent tradition have been disparate parts of the work flow but that are essentially related in nature. Particularly within the technical services operation, such integration—provided that it does not lead to abject dependence on any one tool—offers a significant increase in accuracy and efficiency. These benefits, however, while inherent in the OCLC system, may be fully realized only through a systems analysis approach to the implementation of OCLC and its various capabilities. This article will examine in detail from a systems viewpoint the integration of one currently available capability of OCLC, automated label production, into the technical services operation.

Little information is available to assist technical services managers in planning for such integrated use of s14/s16, the commands that result in screen displays showing the call number in spine label format and the call number, author, and title in a horizontal format suitable for cards and pockets. One of the only discussions of the use of a printer with OCLC terminals was written and accepted for publication before OCLC made the s14/s16 capability generally available in early 1976.[1] Scattered references in various numbers of the *OCLC Newsletter* give bits of information and two OCLC documents give technical information on at-terminal production, but these items are buried within the mass of OCLC publications and are especially difficult for newcomers to the system to find.[2] Nowhere is there an overview of the impact of automated label production upon the technical services work flow.

As in any situation where technology offers the opportunity for change, the first and foremost task in considering automated label production must be a careful evaluation of the advantages of the new procedure based on a comparison of equivalent elements in both old and new methods of operation. Cost/time analysis is a traditional means of obtaining data for such a purpose. A cost analysis performed at James Madison University Library of one of the most widely used spine label systems, Se-Lin tape, revealed that the average cost of the spine label in a sample of 1,230 books was $0.0201 per book. Prices of several available OCLC-compatible label stocks, with 10 percent wastage figured in, range from $0.0074 to $0.0089 per label when purchased in quantity (the shelf life factor is negligible compared to Se-Lin). Since full utilization of label production capability will undoubtedly involve more than a single spine label per book, label cost per book must be based on the total number of labels used. Thus, total label cost for a book for which spine, card, pocket, and title page verso labels have been produced, plus two labels for other purposes, may range from $0.0444 to $0.0534 per book.

With regard to clerical time, data collected at James Madison University Library showed that for a sample of 1,230 books an average of 16.8 seconds/book was spent in typing Se-Lin spine labels; for a sample of 595 card and pocket sets, an average of 1.8 minutes/set was spent in typing the call number on the pocket and author, title, and call number on the circulation card. To these figures must be added the time necessary to write or type the call number and/or author and title in the various other locations (title page verso, requestor's form, etc.) designated to receive such information, as well as time spent to revise this work. Problems caused by errors in the recording of author, title, or call number cannot be reliably calculated.

In contrast, the average time necessary to print out the s16 display twice on a 120 character/second printer, based on fifty-four printings spread over three days and measured from the time a reformat command was entered after card production to the cessation of printing, was slightly more than twenty-eight seconds. This figure applies to single-copy, single-volume English-language titles for which the call number, author, and title fields required no editing after card production. For books not in this category, time must be added to edit the bibliographic record in

accordance with the characteristics of label display programming: automatic stamps, which do not display, must be added to the call number or 049 field; copy and volume information must be supplied; diacritics must be deleted for those printers that do not have foreign-language characters; punctuation and tagging, such as #b in 245 field, must be checked and changed as necessary. Response time on any given day will, of course, affect the speed with which such editing can be accomplished. Time necessary for revision is greatly reduced, since information from the bibliographic record is carried through virtually unchanged; unless errors are detected when the catalog cards arrive, only a check to see that the correct labels have been placed in the appropriate locations should be necessary. While it is impossible to measure in detailed fashion all aspects of the labeling procedure, automated label production does appear to offer increased accuracy and efficiency in the technical services operation.

In planning for the use of s14/s16 the physical characteristics of the printer to be used must be reviewed. Some printers are not suited for label production because specially treated paper must be used (e.g., thermal printers); others may have print trains that are too "computerese." If a printer is purchased with label production in mind from the outset, the quality of the final product in terms of clarity for the user should be carefully examined. The standard print train that is normally provided may sometimes be exchanged for one more traditional in appearance; the General Electric Terminet 120, for example, has several print trains, including one with large capitals, from which a potential buyer may choose at no extra cost. In some printers, letters and numbers in a print train that may not be considered acceptable for spine label use may be easily replaced. Thus "ϕ," a standard computer symbol which may be quite alien to a user expecting LC or Dewey call numbers, may be replaced by "O." Speed is another important consideration; although most printers are significantly faster than a human typist, automated label production at the terminal must to some extent prolong the use of the terminal and thus reduce the number of titles cataloged per hour.

Given a printer ready to produce clear and legible labels, the next step is to determine the most effective way of utilizing these labels. Traditional spine, card, and pocket labels are not the only uses for the OCLC label production capability; a much broader application, tailored to an individual library's need for call number and author/title information in its own procedures, is possible. To achieve the fullest possible application of the s14/s16 capability, the technical services work flow should be analyzed to identify every point at which call number and/or author/title information is recorded. In each case the purpose of recording such information should be explored and evaluated. Such analysis often has the added advantage of uncovering duplicative or otherwise wasteful efforts. Most libraries print labels after the bibliographic record has been edited according to local practices and cards have been produced, although it is possible to perform label production at any time that an item is searched and found in the data base. In an analysis of the work flow it should be kept in mind that automated label production for some groups of materials—e.g., reference works, special collection items, added volumes—may not prove practical because of the small number or special nature of labels needed or because cards will not be produced for the title; the total work flow should be planned to accommodate these materials smoothly. When a comprehensive overview of the need for call number, author, and title information has been defined, the format of the label paper may be selected.

Finding a label paper suited to the needs of an individual library may be difficult. Even though a large variety of styles are now available, few seem to have been reasonably designed for library use. Most are too large, impossible to keep properly spaced in the printer, and wasteful of paper in the spine label area. There are some alternatives to label paper that strictly reflects the lopsided spacing on the

screen display: Larlin Corporation (Marietta, Georgia) sells an Avery label that places spine and card labels in an evenly spaced vertical arrangement, thus avoiding waste of the area below the spine label by allowing two or four rather than three labels to be produced (accomplished by positioning the cursor at the beginning of the pocket label before printing); University Products (Holyoke, Massachusetts) and possibly others offer custom-designed labels for those who have the funds. With a little imagination and a good deal of practice with free samples at the terminal, an acceptable printing procedure can be devised.

Several other factors should influence the choice of a label paper. The adhesive must be reliable and should not allow a bleed-through reaction with the book binding. The paper should accept ink without permitting it to spread, smear, or fade. Some libraries coat the labels with a clear plastic spray or glue, but if the paper is of good quality this step may be unnecessary.

The actual mechanics of automated label production will affect terminal use in several ways. Printers, like typewriters, can accommodate only one paper stock at a time; therefore, paper must be changed each time printer use changes. Paper changing is time-consuming and may be wasteful, depending upon the kind of printer used (although use of a leader can eliminate the waste involved in inserting paper in the printer). The ideal situation would be two (or more) terminals, each with its own printer. The most restrictive situation is one terminal, one printer. In this case terminal time for staff to make printouts of bibliographic records and staff concerned with processing should be scheduled in such a way as to facilitate both needs smoothly, yet keep the changing of paper to a minimum.

As mentioned previously, use of the terminal to produce labels will necessarily increase the terminal time devoted to each title cataloged and reduce the number of titles that can be handled within a given amount of time. Libraries with enough terminals to accommodate the cataloging load comfortably with some margin of extra terminal time will not find this to be a problem; those libraries, however, that fully utilize terminal time during their normal working day may occasionally find it necessary to schedule longer days for additional system use.

Follow-up evaluation of any newly implemented procedures and materials is a step that should not be omitted. If a temporary file of records is kept for volumes that have received automated labels, the labels themselves can be checked for durability after a period of time; discussion with the personnel responsible for label production may reveal unanticipated problem areas or may provide insights for more efficient procedural methods.

Last but not least is the matter of personnel training. Once familiar with terminal operation, staff need little additional training for label production. Judgment must be exercised in deciding when to print the subtitle as well as the main title, and a trial period will probably be necessary to determine the most comfortable way to add copy, volume, and other necessary information to the screen display before printing.

The use of the OCLC system to produce labels may not be practical for every library. The point is that any library with an OCLC terminal and a printer is shortchanging itself if it does not investigate the feasibility of automated label production. Only through a willingness on the part of libraries to reexamine traditional methods and formats in light of the total systems-design capabilities of computer technology can that technology deliver real improvements in library operations.

REFERENCES

1. Christina Landram, "Cataloging: OCLC Terminal Plus Printer," *Library Resources & Technical Services* 21:147–55 (Spring 1977).
2. The most recent and comprehensive document is OCLC Technical Bulletin Number 48 (Rev.), *Printing Capabilities with OCLC Terminals* (Feb. 1980). Prior to this, OCLC miscellaneous documents *Operation of the Tycom/Selectric Printing Device from OCLC 100 CRTS* (Feb. 1975) and *Review of Printer Attachments for Use with OCLC Model 100 Terminals* (Feb. 1976) gave information on label production and printer selection.

Computing the Effective Length of a MARC Tag

William R. PRINGLE: Carnegie-Mellon University, Pittsburgh, Pa.

This article explains how to compute the effective length of a MARC tag from the OCLC tapes, which shows how including a particular tag in bibliographic records will affect the average length of those records. How this measurement was developed at Carnegie-Mellon University is demonstrated, and some computer programs available from the university that allow a library to read and manipulate these tapes are described.

This article describes some of the things that Carnegie-Mellon University (C-MU) has done with OCLC tapes. It demonstrates that a library does not have to have large computer resources or a programming staff available before the tapes can be used.

The C-MU Library has been receiving the OCLC-MARC Subscription Service tapes from OCLC since they were first available. This was done with the knowledge that at some time they could be used to build a local machine-readable data base. Unfortunately, these tapes are in such a format that a program has to be written in order to retrieve the information and put it into a form that might be useful to the library staff. No programming staff being available, the tapes were simply stored with the intent of doing something with them at some future date.

The "something" was first realized when the library began to consider how automation could serve both the patrons and the staff. The first thing was to make sure that the tapes were readable, since the library had not had the facilities to determine this in the past. Next, a random sample of the tapes was taken and the average length of an OCLC record was computed. However, once C-MU began to consider what fields were to be included in the bibliographic records, the university had to face the question that many other libraries have tried to answer: "If these tags are used, how long will the records be?"

At first, it would seem that the answer is quite simple: just add up the average lengths of each of the fields that were chosen, and that would give the average length of the record. Unfortunately, it is not that simple. There are several reasons for this, but basically it is because not every record will require every tag. To cite some obvious examples, all the 100 tags deal with main entry, which would certainly be desired in an automated data base, but no record would have more than one of the 100 tags. What, then, is considered as the average length of the main entry field? Is the average length of a personal main entry the average length of a corporate main entry? If not, then the question still remains, How does one compute the average length of a field?

The opposite problem occurs with the 600 tags that are the added entries. In this case, one record might have many added entries; so what is the average length of an added entry? To make matters worse, the average length of all the added entries for a given record depends greatly on the cataloging department of each library. This means that the average length of the added entries at C-MU is not going to be the same as at another library. On top of this, for a given library, the average length of the added entries is probably very much a function of the scope and emphasis of the library. University libraries would tend to have extended added entries only in those areas in which the university concentrated its curriculum, while small public libraries may have few added entries, or none. To solve this problem, a program was written to compute the effective length of a field. The effective length of a tag is a measure of how much that tag will affect the average length of the record if it were included in the data base. The OCLC tapes are used to gather data about the lengths and occurrences of all the tag fields. For this to be significant, the records that entered into OCLC should meet certain requirements:

- They should be representative of the total collection. It does no good to do a

study of OCLC tapes to determine the characteristics of the holdings if those records are only a certain portion of the new items cataloged. At C-MU, 98 percent of the items coming into the library are cataloged onto OCLC, and most of these are cataloged in the same place.

• The records should be in the proper format. Unfortunately, many libraries are using OCLC only to produce cards and are not particular what tag numbers are given to certain pieces of data. In many cases, the record is given an incorrect tag number just so it will be on the card where that department feels it should be. Thus, certain tags may have data that are different from what MARC has defined for them. This, in itself, does not mean that such a library cannot benefit from this study; it simply means that it must be clear on the meaning of each of the tags. A computer cannot look at the data within a given tag and "know" what they are supposed to be. At C-MU the decision was made from the beginning that the record was to be correct, and that the format of the cards was not as critical.

• The record on the OCLC tape should be of high enough quality that the library is willing to accept that record as "The Record." Unfortunately, some libraries have simply "dumped" their retrospective holdings onto OCLC without checking the quality of the data, figuring that they could go back and fix them up later. It makes no sense to compute the average length of a record based on one set of records and then to go through those records and modify them for the actual data base. It is also unreasonable to consider having someone review each record on the OCLC tape and then edit or "correct" it. Such a process would take entirely too much time. The university is very conscientious about making sure that its records accurately describe its collection.

If your OCLC tapes meet those criteria, then the procedure described in this article will be meaningful to your library. If not, then perhaps a presentation should be made to all those involved, so that they are aware of the consequences of such a position.

Consider each tag as a separate and distinct piece of data. Then the average length of the tag (AL) could be computed by taking the total length of all the occurrences of that tag (TL) and dividing it by the total number of times that the tag appears (NA). Thus, for tag "i," this could be expressed by the formula

$$AL_i = TL_i / NA_i. \quad (1)$$

This, however, does not help determine what effect that tag would have on the average length of a bibliographic record, as discussed. To determine this, compute the effective length, which could be expressed in the formula

$$EL_i = TL_i / TR, \quad (2)$$

where TR represents the total number of records in the given sample of the data base.

While this figure is much more meaningful, it is more difficult to compute, probably because one is working with a fairly large number of tapes. This means that such a computation would require reading the tapes more than once or creating a large data file with all the totals that would be required. In either case, this type of process would require fairly extensive computer resources and so is often not practicable for a library that does not have a computer. To simplify this process, modify equation 2 as follows:

$$EL_i = TL_i / NA_i * NA_i / TR, \quad (3)$$

which is identical to

$$EL_i = AL_i * PO_i, \quad (4)$$

where PO represents the percentage of records in which tag "i" occurs.

The attractive thing about equation 4 is that AL and PO may be computed for each tape separately, and then the effective length of each tag may be computed by taking the average of the averages on the tapes. A program was written to compute the average length and percentage of occurrence for each tag on a tape. Another program was written that read the reports and averaged those figures, which it then used to compute the effective length of each tag using equation 4. This program should also record TNA, which is the total number of occurrences of tags.

It was found that the effective length figures were quite helpful in determining

which fields C-MU could afford to include. Even though a field might have been somewhat long, if it appeared infrequently enough an argument could be made for keeping it if the overall impact on the record size was small enough.

Once one has decided which tags to include within a local data base, one may then compute the average length of the record by adding the effective lengths of all the desired tags. One would then have the average amount of storage space required per record to store those data. Thus, the average length of the record would be the total effective length of all the tags—expressed in formula 5:

$$TEL = EL_1 + EL_2 + \ldots \quad (5)$$

for all tags that have been chosen.

It should be noted, however, that this calculation will tell only how much storage space the information itself would take. In order to make the data useful, one would need to have some sort of system that would work with the data. Every system will require a certain amount of overhead space for each tag or record. This could prove to be significant depending on the number of tags and records that will be involved. To get a full picture of how much storage will be required, one would have to allow for this overhead. It would be hoped, however, that the bulk of the storage required would be used for storing the data, and so this calculation would be useful to allow one to compute roughly how much storage would be required. Once one knows what type of system will be used, one should then be able to determine the precise amount of storage that will be needed. This could be done by using the formula

$$TS = (OR + TEL) * TR + OT * TT, \quad (6)$$

where

- TS is the total space required for the data base,
- OR is the overhead required for each record,
- TEL is the total effective length as computed in equation 5,
- TR is the total number of records (or items) that should be allowed for in the entire collection,
- OT is the overhead required for each tag, and
- TT is the total number of tags (or fields) that should be allowed for in the entire collection.

There is a general-purpose program that will read an OCLC tape and perform various operations on any or all records. Rather than several different programs that all read the tapes and do slightly different things with them, the features of Structured Basic have been used to have this program call external routines to do certain functions that will affect the performance of the program. These routines, in turn, have a number of routines available that will do various common manipulations on the record. Only three small, and relatively simple, procedures to do the specific task were written.

The program most generally employed uses these routines to print out a sample portion of the tape. This program will ask for the starting and ending record to read, and it will print those records out into the printer in a format similar to the way they appear on the screen of the OCLC terminal.

As mentioned before, what makes this program so nice is the general-purpose routines that may be written to do the specific job desired. No great programming expertise is required to write these programs, and so a library with access to a PDP-11 under RSTS could write these routines without the need of a programming staff, provided someone in the library has some knowledge of Basic-Plus, or programming languages in general. No knowledge of the OCLC tape format is required. The external routines that the user must write are as follows:

SELECT.OPEN

This routine is called after the tape has completed initial processing and may be used to set up any flags or counters needed before the processing of the tape begins. It may also be used to position the tape to begin processing at any portion of the tape desired.

SELECT.REC

This routine is called after each record is

read. It may do anything with the record. In order to facilitate this, several routines are available to this routine, which it may call if and when it chooses to.

SELECT.CLOSE

This routine is called after all the desired records have been read. It may be used to print a summary report or to perform any special processing required.

The following routines are available to the above routines and may be called by any or all of them.

READ.REC

This routine will read a record from the OCLC tape.

PARSE.REC

This routine will take the current record and parse it into individual tags, placing this information in an array easy for the program to use.

PRINT.REC

This routine will print the current record out to the output file in a format similar to that in which it would appear on the OCLC terminal.

DUMP.REC

This routine will print the record as it appears on the tape. This is useful for determining exactly what data are on the tape.

The above programs are available from Carnegie-Mellon University and are written in Structured Basic, a language that was written at C-MU. This language will run on a PDP-11 computer under RSTS, or any other operating system that will support BASIC-PLUS. Structured Basic is a preprocessor of the Basic-Plus compiler, and the output of the Structured Basic compiler is Basic Plus code, which may be run the same as any other Basic-Plus program. These programs are all available from C-MU in compiled form, Basic-Plus source, or Structured Basic source along with the Structured Basic Language compiler.

Statement of Ownership and Management

Journal of Library Automation is published quarterly by the American Library Association, 50 E. Huron St., Chicago, IL 60611. Annual subscription price, $15. American Library Association, owner; William D. Mathews, editor. Second class postage paid at Chicago, Illinois. Printed in U.S.A. As a nonprofit organization authorized to mail at special rates (Section 132.121, Postal Service Manual), the purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes have not changed during the preceding twelve months.

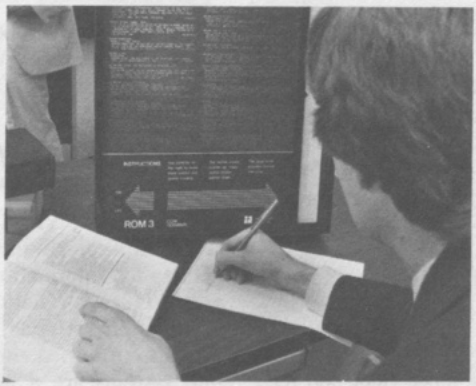
Extent and Nature of Circulation

("Average" figures denote the number of copies printed each issue during the preceding twelve months; "Actual" figures denote number of copies of single issue published nearest to filing date—the June 1979 issue.) Total number of copies printed: Average, 5,638; Actual, 6,044. Paid circulation: not applicable (i.e., no sales through dealers, carriers, street vendors, and counter sales). Mail subscriptions: Average, 4,924; Actual, 5,455. Total paid circulation: Average, 4,924; Actual, 5,455. Free distribution by mail, carrier, or other means, samples, complimentary, and other free copies: Average, 219; Actual, 219. Total distribution: Average, 5,143; Actual, 5,674. Copies not distributed: Office use, left over, unaccounted, spoiled after printing: Average, 495; Actual 370. Returns from news agents: not applicable. Total (sum previous three entries): Average, 5,638; Actual, 6,044.

Statement of Ownership, Management and the Circulation (PS form 3526, Aug. 1979) for 1979 filed with the United States Post Office Postmaster in Chicago, September 28, 1979.

LIBRIS CATALOG SERVICE

Baker & Taylor's new COM technology for libraries.



The LIBRIS Catalog Service offers immediate benefit to libraries. It is a fully supported and field-proven system designed to meet the changing catalog needs of today's libraries.

Library Needs — Rising costs and shrinking budgets have forced many libraries to close branches, reduce hours, limit community programs, and eliminate staff. Yet, the information proliferation creates an ever-expanding need for improved access to library collections.

While cataloging and filing backlogs build up, interlibrary loan programs and networking activities continue to utilize increasing shares of already limited staff time. The situation becomes even more critical as funds available to libraries fail to keep pace with the rate of inflation.

The LIBRIS Response — Baker & Taylor is responding to this challenge with an extensive commitment to automation and the LIBRIS Catalog Service—a fully supported bibliographic database management system.

The LIBRIS Catalog Service provides a uniform, authoritative bibliographic database and a system for readily managing it.

DATABASE MANAGEMENT SYSTEM — Our bibliographic database contains nearly two million titles in MARC-II communications format. It can be easily used by libraries in establishing a computer based resource from which information meeting cataloging and patron demands can be obtained. The time needed for typing and filing catalog cards, updating holdings, and maintaining authority control files and activities is greatly decreased—and, the costly errors inherent in traditional card catalogs are minimized.

In addition to catalog maintenance, applications of computer technology, which include circulation control and interlibrary loan, can be supported through the LIBRIS Catalog Service.

COM — For the library patron, the COM catalog is one of the most apparent and appreciated benefits of computerization. Because of their small size, COM catalogs are highly mobile and may be placed at several locations within the library and other community sites. Catalogs are faster and easier to use than traditional card catalogs because they display multiple entries. And, the database management system assures that the entries are always up-to-date.

LIBRIS Is Ready — Baker & Taylor's team of library and computer professionals are ready to help you determine the appropriate application of computer technology to your library. Every library's needs are special. That's why the LIBRIS Catalog Service was developed with flexibility in mind. Your catalog can be published in any format you need, whether it is microfilm, microfiche, or hard copy. And we can publish on any schedule you desire. In addition, we can produce special catalogs such as: large print catalogs, juvenile catalogs, audio-visual catalogs—virtually any subcatalog fulfilling library and community needs.

The Next Step Is Yours — Contact us today to learn more about our LIBRIS Catalog Service or about the other members of the LIBRIS family of integrated automation services including: MRC (Machine Readable Cataloging); On-Line Ordering Service; and BATAB Automated Buying Service.

LIBRIS
BAKER & TAYLOR'S
Library Information Service
6 Kirby Avenue
Somerville, New Jersey 08876
(201) 526-8000

- ☐ Please send me your new LIBRIS Catalog Service brochure
- ☐ Please send me information on your other LIBRIS services.
- ☐ Please have a representative contact me.

Name_____
Library_____
Address_____
City_____ State_____ Zip_____

Western Division
Reno, NV 89564
380 Edison Way
(702) 786-6700

Midwestern Division
Momence, IL 60954
Gladiola Avenue
(815) 472-2444

Eastern Division
Somerville, NJ 08876
50 Kirby Avenue
(201) 722-8000

Southern Division
Commerce, GA 30529
(404) 335-5000

News and Announcements

PUBLICATIONS

Computer-Readable Data Bases Available

Computer-Readable Data Bases: A Directory and Data Sourcebook, 1979 edition, edited by Martha E. Williams, is by far the most up-to-date, comprehensive directory of machine-readable data bases ever published. It is a must for users of data bases and for reference librarians, for data base producers and information suppliers, and for the many students and teachers working in this dynamic field. The 1979 edition lists more than 500 data bases worldwide, each listing containing: name and producer of the base, its coverage, year of origin, number of items in the base, availability in either batch or on-line mode, pricing, and other pertinent information. The directory will run more than 1,000 pages and be available in a single easy-to-use volume, bound in a durable, flexible softcover material.

Users of the first edition, originally offered in 1976, will note that this edition lists nearly twice as many data bases and has some 50 percent more pages. Several indexes speed location of desired information: a two-level subject index, producer index, processor index, and data base name index. Another invaluable aid will be a listing of data bases that are now outmoded, that have been replaced by more current data bases, or whose existence cannot be verified.

The directory is compiled from the data base on data bases maintained at the University of Illinois's Information Retrieval Research Laboratory, under the direction of Professor Martha E. Williams. Chief editor of *Computer-Readable Data Bases*, she is a world authority on the subject.

The directory is published by the American Society for Information Science in conjunction with Knowledge Industry Publications, Inc., 2 Corporate Park Dr., White Plains, NY 10604, the exclusive distributor. The price is $76 (softcover) to ASIS members and $95 to nonmembers.

The Annual Review of Information Science and Technology, Volume 14 (1979)

Martha E. Williams has edited this latest annual review for the American Society for Information Science. Computer architecture for natural language and for information retrieval, information systems in Latin America, and education for on-line systems are a few of the topics covered in this year's edition. Each chapter reviews current developments and the recent literature, providing an invaluable overview of a dynamic field.

The contents include: "Cost Analysis of Systems and Services" by Colin K. Mick, "Systems Design—Principles and Techniques" by Ronald Wyllys, "Empirical Foundations of Information Science" by Pranos Zunde and John Gehl, "Empirical Techniques of Information Retrieval" by Michael J. McGill and Jennifer Huitfeldt, "Unconventional Computer Architectures for Information Retrieval" by Lee A. Hollaar, "Database Management Systems" by Michael A. Huffenberger and Ronald L. Wigington, "Library Automation" by Mary Jane Pobst Reed, "Education and Training for Online Systems" by Judith Wanger, and "Information Systems in Latin America" by Tefko Saracevic, Gilda Maria Braga, and Alvaro Quijano Solis.

To place an order contact Knowledge Industry Publications, Inc., 2 Corporate Park Drive, White Plains, NY 10604. The price is $28 for ASIS members and $35 for nonmembers.

Communications Directory

Communications and Information Handling Equipment and Services: A Semi-

annual Directory/Index of Advertisements and Evaluations; A Multidisciplinary Approach, edited by C. I. Park, Department of Library, Malcolm X College, City College of Chicago, 1979, is a directory/digest/index of the advertisements and evaluations of the equipment and services, printed in the selected professional and trade journals.

The *Directory/Index* has two unique points. Firstly, information sources are based on the advertisements and evaluative articles. In this media-bombarded world, advertising provides live and persuasive communication between producers and vendors and buyers. The advertisements are fast and active sources to identify new products and developments, new or improved services. To complement the value of the advertisements, nonadvertising sources are covered, such as "evaluation articles, guides, surveys, user reports, and overview articles that are helpful to get objective view of products and services" (Preface, V.1, no.2). It is planned to include publicity notes or new product columns, beginning with volume 2.

Secondly, it provides an integrated approach to the equipment and services in the areas of computer data processing, data communications, audiovisual communications, micrographics/reprographics, data bases/on-line services, word processing, office systems, and library. The information technology covers all across modes, regardless of how information is generated, stored, communicated, or delivered.

This *Directory/Index* is divided into two parts. The first part is "Products and Services," which are listed alphabetically by subjects and keywords. This is designed to help readers identify what products or services are newly introduced, how they are evaluated or perceived in overview articles, or how they are rated by users. The second part is "Advertisers and Evaluators." This provides directory information: full addresses and telephone, TWX, and Telex numbers and facilitates quick identification of who advertises what, where, and how often in half a year (first issue of a volume) and in a given year (second issue of a volume).

The first issue (V.1, no.1) was published in March 1979 under the title *Advertisements Digest: Library and Information Services*. The second issue, published in October, is available from the Advertisements Digest, P.O. Box 165, Morton Grove, IL 60053. Subscription price is $18 per year.

Guide to Information Science

An exciting new interdisciplinary field has begun to deal with the incredible implications of a revolution—the field of information science. This *Guide to Information Science* is the first up-to-date book to combine the technical sophistication necessary to explore this complex field with a clarity and readability that make the *Guide* a perfect introduction for the nonspecialist. Virtually every important facet of information science is discussed: retrieval systems, indexing and abstracting, search techniques, the history and fundamentals of computing, and the organization of data systems. The evaluation of information systems is a particularly difficult task, and there is an entire chapter devoted to various techniques—including descriptive and inferential statistics—that are appropriate for research and administrative decision-making. The all-important human factor, often lost in the maze of technology, is considered from the standpoint of both the user and the information scientist.

The *Guide* became available late in 1979. It was written by Charles H. Davis, former dean of the Faculty of Library Science, University of Alberta, and presently dean of the Graduate School of Library Science at the University of Illinois at Urbana/Champaign, and by James E. Rush, presently director of research and development at OCLC, Inc. Greenwood Press has published the *Guide*, which is priced at $25.

User Market Study Published

The National Micrographics Association (NMA) has published a comprehensive survey of users of micrographics technology. The exercise was conceived to measure the use of micrographics and automated office equipment in the office en-

vironment. The study was conducted for NMA by International Data Corporation, an independent market research company.

Projecting a dramatically increased use of micrographics in conjunction with other forms of information-handling technology within two years, the survey results are encouraging for the micrographics industry. The survey indicates an 87 percent growth rate in the area of computer-assisted micrographic retrieval, an 85 percent in use of word processors outputting to COM, and a 74 percent growth rate in the methodology of merging text and graphics on micrographics—all within two years.

The report is divided into three major sections: "Ways Micrographics Can Be Used by Interfacing with Other Technologies," "Equipment Configurations," and "User Analysis." Several significant and informative case studies are included in the report, illustrating how organizations have successfully married various forms of information-handling technology in the broader context of automating their offices. This eighty-three-page market survey sells for $400. To place an order or request additional information, contact the NMA publications sales office, 8719 Colesville Rd., Silver Spring, MD 20910; (301) 587-8202.

New Library Administration Journal

The Haworth Press is pleased to announce the forthcoming publication of the *Journal of Library Administration*. This is the only journal for library administrators and managers. It is dedicated to providing thoughtful, provocative, and useful articles prepared specifically from the management viewpoint.

Unlike other library journals and periodicals, which must appeal as they do to a generalized library audience, the *Journal of Library Administration* represents the viewpoints, concerns, and perspectives of top administration and middle management. Covered will be rigorous analyses dealing with: the library administrator as a labor relations expert; methods of leading the decision-making group; budget preparation and presentation; methods for measuring effectiveness and efficiency; marketing library services; methods for delegating authority; forecasting, planning, and objectives management; influencing employee motivation and satisfaction; and advancing up the career ladder.

An outstanding editorial board, comprised of national leaders in library management and administration, are associated with this new journal. Each issue will aim at providing new models for administrative practice, important tests of long-standing approaches, and case studies demonstrating how other libraries handle their most critical managerial problems on both a day-to-day and a long-range basis.

Order from Haworth Press, Inc., 149 Fifth Ave., New York, NY 10010. Volume 1, number 1 appears in January 1980. The subscription price is $36.

NEW PRODUCTS, SYSTEMS AND SERVICES

Houston U. Awards Contract to SSI for Book Circulation Control Wands

Software Sciences Incorporated, Atlanta, Georgia, was awarded a contract by Sam Houston State University in Huntsville, Texas, to supply specialized OCR wands for book circulation control in the 350,000-volume campus library.

Customized versions of the Caere model 712, the wands will enable libraries to facilitate book checkouts by recording book item and patron identification numbers in one simple and fast operation. The wands recognize and transmit data on alphanumeric documents using OCR-A size 1 characters specified by the National Retail Merchants Association. Dual guide beams assist the operator to accurately scan OCR-A data at a rate of 130 characters per second. Scanning accuracy is 99.9 percent. The wands will be used for other on-campus data collection applications.

The Sam Houston State University contract is an outgrowth of work that SSI has done in the European and American book trade. A teleordering system, currently operating in Great Britain, was designed by Software Sciences to meet the special needs

of English bookshops. SSI is currently marketing a completely computerized, "stand-alone" point-of-sale and inventory management system—called Duet—to American retail booksellers.

For additional information on the custom OCR wands or the new Duet system, contact David Ellis, President, Software Sciences Incorporated, Suite 200, 2814 New Spring Rd., Atlanta, GA 30339.

On-Line Data Base Management Maintains Copy On-Line

In the fall of 1979, Inforonics, Inc., announced TREEDEX, an on-line data base management system particularly suited to in-house search and processing of private data bases. This module can store, search, and edit text records. TREEDEX was the newest module in Inforonics' Text Processing Service, a system used by publishers for data base compilation, processing, and typesetting.

Publishers can use TREEDEX to store, access, and modify text data files in production. This capability vastly simplifies the modification of manuscript for books with rapidly changing content, such as directories, catalogs, handbooks, manuals, abstract journals, and bibliographies. It also facilitates the incorporation of last-minute changes made to galley proofs in ordinary book composition.

The primary feature of the TREEDEX module is its ability to create a multiple-aspect, on-line searchable index. New entries may be added, a record at a time, in an on-line mode. Access points to data added on-line are immediately integrated into the search index, resulting in a real-time system. Data storage accessed by the search system can accommodate variable-length records, an important capability in text processing applications.

For information contact the Marketing Department at Inforonics, 550 Newtown Rd., Littleton, MA 01460; (617) 486-8976.

News of the World Indexed Daily, via Computer

News of the world, indexed daily, is now instantly available through Lockheed's Dialog computerized information retrieval service. Newsearch, a data base exclusive to the Dialog service, provides cover-to-cover indexing of all news items and articles in the *New York Times, Wall Street Journal,* and *Christian Science Monitor* as well as major features in 375 other popular periodicals including *Business Week, Time, Newsweek, Psychology Today,* and *Science.*

Newsearch acts as a daily news alert for corporate officers, enabling them to identify articles on finance, mergers, stocks, products, trends, and regulatory issues in a quick, simple, and timely fashion. City officials and other government bodies can keep up with the latest on legislation, regulatory matters, and solutions to environmental and zoning problems. Professionals in advertising, public relations, and marketing benefit from information on product reviews, corporate profiles, and changes in management as well as current issues.

To locate articles of interest, users type their requests on a keyboard terminal connected by telephone to Lockheed's computers in Palo Alto, California. The computer responds with a list of articles on the topic, displayed on a TV-type screen or paper printout. Newsearch covers the current month's publications only. References to back issues may be obtained from the National Newspaper Index and the Magazine Index data bases in the Dialog service.

The new file is one of 100 data bases now available through Lockheed's Dialog service, offering 25 million references in virtually all fields of study. It will be offered on a subscription basis with one fee covering connect time, royalties to the data base supplier, off-line prints, and telecommunications within the United States (i.e., Tymnet and Telenet).

Additional information may be obtained from Betty A. Davis, Lockheed Information Systems, Lockheed Missiles & Space Company, 3251 Hanover St., Palo Alto, CA 94304.

New Audiovisual System Brings Sound to Print

In 1979 Microsonics Corporation introduced a totally new dimension to publishing: a compact audiovisual system that brings sound to the printed page. The system uses a hand-held microphonograph

that plays a transparent audiodisc mounted on printed material to give up to ninety seconds of audio information.

The applications for the system are virtually limitless: a history text can come alive with the actual sound of former President John F. Kennedy delivering his famous inauguration speech; a language student can hear proper pronunciation of the words he is learning; a new employee can accelerate her training with audio support of the training manual; or at-home chefs can listen to the wit and culinary wisdom of Julia Child from the pages of her latest cookbook.

The system is designed to inform, teach, and train people with methodology that complements the way they learn. With the Microsonics system, the reader both sees and hears the material, which is presented in short, attention-holding segments. The system can even be adapted to already existing printed or graphic material, which makes it one of the least expensive audiovisual systems available.

Completely compact and portable, the system consists of a hand-held player, the "Microphonograph," and a two-inch, transparent record that can be applied directly to the page of a book or to a specifically designed card to give up to ninety seconds of audio information.

The cost of the system—hardware and software package—is priced from $20 up, depending upon choice of hardware and quality and quantity of software. Since new cards and microrecords can easily be added to the package, the cost of updating the materials and adding additional subject information is minimal.

Additional information can be obtained from Bradley Scott at Microsonics Corporation, 2049 Century Park East, Los Angeles, CA 90067; (213) 553-9105.

Contract Awarded for On-Line Union List of Serials

The Pittsburgh Regional Library Center (PRLC) was awarded a contract in 1979 by the State Library of Pennsylvania for the development and production of an on-line union list of serials for Pennsylvania.

PRLC is a multistate library network with administrative offices on the Chatham College Campus in Pittsburgh. Its membership includes fifty-one public, academic, and special libraries in Pennsylvania, West Virginia, and Maryland. As an organization, PRLC has a basic role in resource sharing and cost savings for libraries.

The $99,304 in Library Services and Construction Act Title III funds for the first year was to be used for system design and for the selection and production of a core list of approximately two thousand serial titles held by all types of libraries throughout Pennsylvania.

Ruth C. Carter, head of the Serials Unit at the University of Pittsburgh's Hillman Library, has been named the project's technical director. She will be assisted in this work by a special full-time staff of four.

The project staff will be advised by a special council of librarians, representing all types of libraries and library cooperatives in Pennsylvania, and consultants from national groups. This Advisory Council will provide guidance to the project staff so that they fill the needs of Pennsylvania while adhering to national standards and plans.

The union list, which will show detailed holdings for each library, will be made a part of the OCLC data base and will be accessible to any OCLC participant in Pennsylvania. The project also calls for the future capability of production in computer-output-microform.

PRLC has had extensive experience with serials on OCLC beginning in March 1974 when the center sold its original serials data base to OCLC, Inc. Since that time, selected holdings of several PRLC libraries have been maintained and made available on-line to various cooperative groups.

CONTENTS

MARCH

JUNE

SEPTEMBER

DECEMBER

INSTRUCTIONS TO AUTHORS

The *Journal of Library Automation* welcomes manuscripts related to all aspects of library and information technology. Some specific topics of interest are mentioned on the masthead page. Feature Articles, Communications, Letters to the Editor, and News Items are all considered for inclusion in the journal. Feature Articles are refereed, other items generally are not. All material is edited as necessary for clarity or length.

Manuscripts must be typewritten; a single copy is sufficient. All text must be double spaced, *including footnotes and references*. Manuscripts should conform to *A Manual of Style*, 12th ed., rev. (Chicago: University of Chicago Press, 1969). Illustrations should be prepared carefully as camera-ready copy, neatly drawn in a professional manner on separate sheets of paper. Manuscript pages, bibliographic references, tables, and figures should all be numbered consecutively.

Feature Articles consist of original research, state-of-the-art reviews, or comprehensive and in-depth analyses. They may be from ten to twenty-five pages in length. An abstract of 100 words or less should accompany the article on a separate sheet. Headings should be used to identify major sections. Authors are encouraged to relate their work to other research in the field and to the larger context of economic, organizational or management issues surrounding the development, implementation, and use of particular technologies.

Communications consist of brief research reports, technical findings, and application notes. These may be up to ten pages in length; an abstract need not be included.

Letters to the Editor may offer corrections, clarifications, and additions to previously published material, or may be independent expressions of opinion or fact related to current matters of concern in the interest area of the journal. A letter commenting on an article in the journal is shared with the author, and a response from the author may appear with the letter. Letters should be no more than three pages in length.

News Items may announce publications, conferences, meetings, products, services, or other items of note. These should be limited to two pages in length.

Book Reviews are assigned by the book review editor. Readers wishing to review books for the journal are invited to contact the book review editor, indicating their special areas of interest and expertise.

Names and addresses of the journal editors may be found in paragraph three on the masthead page. In all correspondence please include your own name, institutional affiliation, mailing address, and phone number.

BLACKWELL

A NAME THAT SPEAKS WITH AUTHO

Right now!

To be more specific—fully automated Subject Authority Control according to LCSH/8.

By the time you read this advertisement, every COM and book catalogue produced for B/NA-customer-libraries will have been fully recatalogued from LCSH/7 to LCSH/8. And all the required cross-references will have been interfiled and verified.

All this will have been done at B/NA with B/NA software, the LCSH/8 database, and B/NA's experienced staff of librarians and editors.

The Subject Authority Control System will process computer cataloguing from such sources as: LC-MARC records, OCLC user records, CAN-MARC records, B/NA-MARC records, and other commercial vendor MARC-like records (which we first convert to MARC).

B/NA is the only vendor which can deliver this service to your library now.

Right now! Just as we del
to LCSH/8) with LCSI
proven software. The sa
staff. The most up-to-date

B/NA technical servi
include colleg esearcl
and sp arou

To sun wou
with you about some, or all of the following:

- shelflist conversion to MARC;
- database management of OCLC (or other network) records;
- book, (COM) fiche, or film catalogues;

and, of course,

- Subject Authority Control by LCSH/8.

Please call (503) 643-8423. Or write to Blackwell North America, 10300 S.W. Allen Blvd., Beaverton, Oregon 97005. Ask for Michael Moen, one of the twelve librarians on our technical services staff, all of whom are knowledgeable about MARC structure, COM fiche and film catalogues and, of course, fully automated Subject Authority Control by LCSH/8.