

IRWIN H. PIZER
University Librarian and
Professor of Library Administration
University of Illinois at
the Medical Center
Chicago, Illinois

On-Line Technology in a Library Network

ADVANTAGES OF ON-LINE SYSTEMS

The advantages of computers in libraries, although not a supposition which one can afford to accept blindly, are as real as the advantages gained from the other pieces of mechanical equipment which have become everyday tools for accomplishing libraries' objectives. A major difference, however, is that a library's investment in computers, attendant staff, supplies, etc., is so much greater in terms of time, money, and energy, and in general commitment to examine minutely the operations the computer is to perform, that the comparison with other machines seems less valid. It is not a crisis if a system planned around a tape-operated typewriter does not work and one is forced to return to a more traditional method. The situations are similar in that it is not necessarily the technology at fault, but perhaps the technique. One might call the problem "The fault, dear Brutus, . . . syndrome."

The problems which many libraries have had with mechanization become magnified greatly when they begin to work with on-line systems where the stakes involved in success or failure are higher. No one claims that everything ought to be done with machines, and there is no reason to suppose that a combination of manual and mechanical, on-line and off-line systems will not serve the library better than any one type of operation by itself.

On-line technology offers an unparalleled opportunity to accomplish many of the things which libraries have always said they would like to if they had the opportunity. Now there is the chance not only to catch up with the flood

of information pouring into the library, but even to turn some of the flood to advantage and, of course, to the benefit of users.

The library has been viewed as a total system for years, and many have dogmatized that this approach was the only rational one to take in library automation. There is, however, no reason why all of the parts of the system must be developed simultaneously. The valid approaches are many and varied, and more important, there are many approaches which have worked—a success that library automation efforts have not always achieved. The literature is filled with lavish descriptions of planned systems replete with glowing estimates of what they will accomplish, but the search must be long and hard to find reports that these efforts have been successful. One searches in vain for reports that a previously well-publicized system has failed. The expenditure that has gone into abortive system development is scarcely credible and certainly not creditable to the profession. Even the documented reports of failure would have, perhaps, helped someone else.

Why, then, is there an emphasis on on-line systems, which not only cost a great deal more than off-line systems, but are vastly more complicated to organize and place in operation. The answer lies in societal needs and expectations. We expect things to be fast, we equate speed with machines, we believe that quality is inherent, and we assume that speedy machines are efficient (although with the example set by Detroit for the production of defective machines, one does tend to wonder at the credulity of people).

On-line technology will make the development of a national or regional library network a real possibility. On-line systems will enable us to learn more quickly, to alter our behavior patterns accordingly, and thus, to advance the sum of human experience and the quality of life.

The main purpose of on-line technology in a library is to better serve the user, which is after all a library's main purpose for existing. Few libraries today see their major role as custodians of the past for the benefit of future generations. On the contrary, all that we do should improve our service capabilities to the person seeking information now. On-line technology not only allows services to be provided more quickly, but for the first time allows the library to disregard the limitations of its physical structure and interact with users at other sites, and at times of their own choosing. We thus begin to approach the goal of making a library available to users on a twenty-four hour basis with few of the attendant costs of maintaining a physical facility or providing large numbers of personnel. On-line technology allows us exploitation of limited resources of people, money, and information more effectively. On-line technology also frees users from many of the cumbersome restraints which libraries interposed between them and their needs in the past, forcing them to adapt to the internal operations of the library rather than vice versa.

In discussions of on-line technology, most still think in terms of the familiar computers and the adjunct hardware, but it is already evident that

this is too narrow an approach to the subject. Other elements of technology which complement the computer must be considered. Libraries have not really considered the possibility of cable television as a major tool to serve them in their role as important educational elements in their community, whether public or university. Users should be able to call the library from their homes, select programs from a printed catalog, and request that they be sent via the phone line for playback through television equipment. Home videotape cameras for making home movies, recording programs from television broadcasts for later playback, purchasable cassette programs and a modified television receiver to accept them are already on sale in Chicago. If the current test marketing proves successful, it may be assumed that an extensive national sales campaign will follow.

Few libraries have made computers available to users to solve problems of their own. Not until recently has there been the possibility of using the coin-operated, self-service computer terminal at \$0.25 for five minutes of computer time.

There are other possibilities which libraries will need to be aware of and which will radically alter the services and collections of libraries. What holographs and lasers will produce and provide has hardly begun to be explored.

On-line technology therefore provides the library with a powerful tool to enable it to do new things and to perform many of its present operations in new and better ways.

USES OF ON-LINE SYSTEMS

Interlibrary Loan

An example of the possible power of an on-line network involves the interlibrary loan process. The term "possible power" is used because the problems involved in placing it in operation are not technological.

The data base in the State University of New York (SUNY) network (fig. 1) consists of citations to journal articles, citations to books, and a union list of serial titles with holdings statements and location information for those titles which are represented in the journal citation file. Therefore all information regarding citations which is necessary for interlibrary loan is available. The user, after obtaining the output to his search and scanning the retrieved titles for relevance, indicates which citations he wishes to see. If the user is in the library where a citation is located, he is given any necessary information to retrieve the document, i.e., call number, shelf title or special location.

The SUNY system was originally planned to include all circulation records for each member station, and, in that case, the circulation files would also have been checked to determine whether the item was out or otherwise unavailable, in which case a request would have been placed on the record for

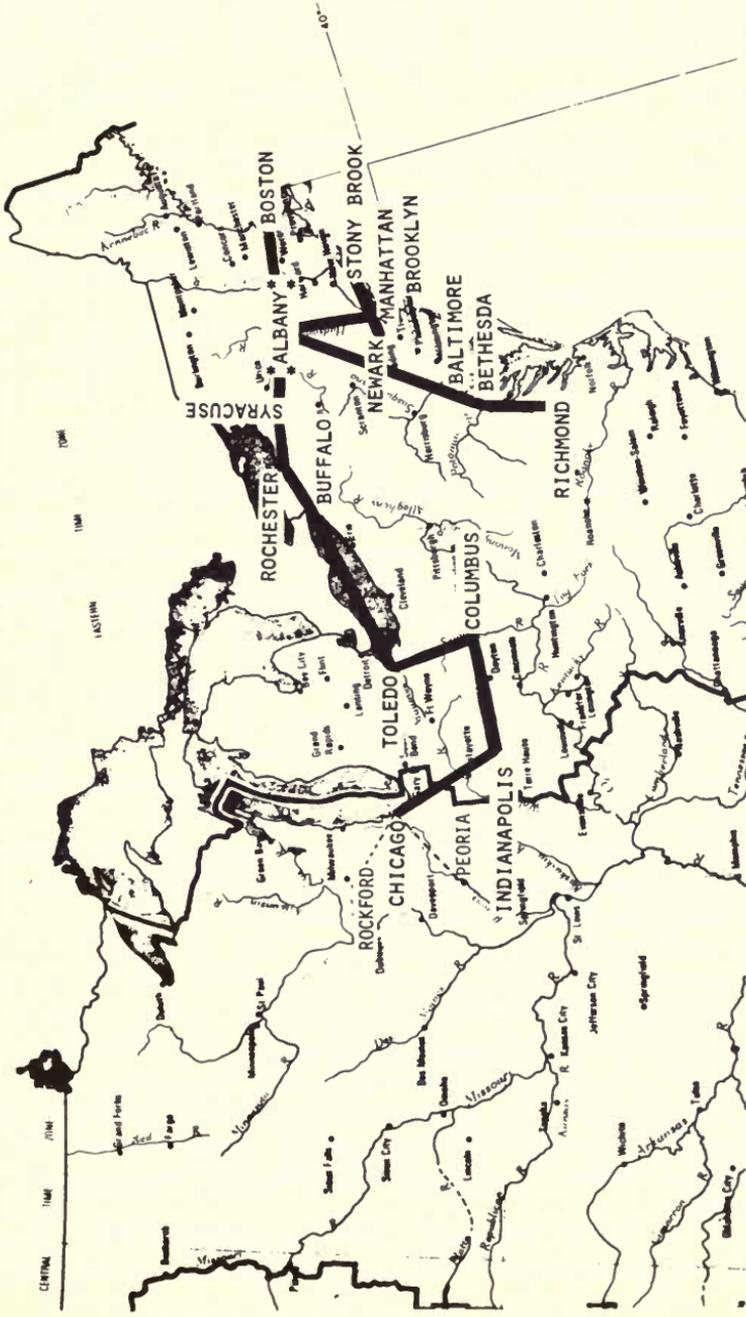


Fig. 1. SUNY Biomedical Communication Network as of March 1972

the user. For items which were not owned by the user's library an interlibrary loan request in standard telecommunications format was to have been sent automatically to the nearest network station which was able to supply the item; a duplicate copy of that request would have been generated in the interlibrary loan office of the user's library. Prior to this the user was to have been given the opportunity to reject the interlibrary loan segment of the service for each selected citation.

This service, therefore, tied together the cataloging aspects of the system, circulation, and serials control, and added the journal article citation data from an external source. This type of service begins to approach the "total system" goal toward which libraries have been striving. It eliminates a number of steps for both the user and the library. Gone is the need for the user to copy down his citation from the secondary source and to recopy the citation on an internal interlibrary loan request form after he has determined that his library does not have the item. The library's need to verify the citation has been eliminated since it has come from a verified source and has not been transcribed which might have introduced errors or diluted information content. The step of retyping the verified citation with the attendant possibility of additional clerical error in this third transcription is also removed. In addition, the delays attendant upon these processes within the library are reduced, together with the major delay caused by mailing the request to a library which may be able to supply. It should be noted that since the computer system has already verified the source of supply, the delays caused by repeated efforts to obtain items from libraries which cannot supply are minimized. The net result of this chain of events is a marked increase in the speed of delivery of the document to the requester, a document which will be even more rapidly supplied if telefacsimile equipment is employed.

This entire process is not a complicated one, and it does not depend on new technology, major reorganization of library functions, or large increases in library budgets. The SUNY network performed all of the necessary programming and testing of the required procedures, and the entire system was declared operable in 1969. At that point, the problems which face a library network, but which are unrelated to its technology, became evident. As the librarians of the member institutions were faced with the reality of accepting a larger number of interlibrary loan requests than they had been accustomed to receiving, they balked. The Network Advisory Council felt that the anticipated avalanche of requests would render normal service in this area unworkable, thus the automatic interlibrary loan procedure was never tried, even on a limited basis.

Another use of the bibliographic data base described above is in the verification of requests for interlibrary loan which are not generated through

computer search. Using search options like "title scan," it is possible to verify citations which may be incomplete or only partially accurate. It is not possible to perform an author search of the MEDLARS file on either the SUNY or MEDLINE systems at the present time, although this disadvantage is not as serious as it would be without the printed *Index Medicus* author index at hand.

Multiple Data Bases

One of the principal benefits which can be justifiably expected from computer systems is that which derives from their ability to do repetitive jobs quickly, accurately, and with a minimum of human intervention. Libraries have looked forward to the day when a number of secondary sources could be searched with a single command or search strategy (single in the sense that it needs input only once). Also anticipated was the ability to progress along levels of information in a sequential fashion, depending on the results of the preceding portion of our search. The SUNY network stated this type of search capability as a phase two goal when it was planned in 1966, and the University of Chicago has recently restated that this type of activity is one of its continuing (although unfunded) goals.

When it is considered what could be achieved for the library user by searching a combination of the data bases available in machine-readable form from *Chemical Abstracts*, *Biological Abstracts*, *Excerpta Medica*, *Science Citation Index*, *Index Medicus*, and a number of other automated publications, it is apparent how far there is to go in on-line or off-line information retrieval, even before the millenium when full texts of documents can be retrieved. If these services were correlated so that the searcher could locate first the bibliographic information, and then the abstracts of articles which he selected, a new level of user service would have definitely been achieved.

Although the SUNY network is not yet approaching this kind of service capability, it is about to begin an experiment which may finally lead to such an on-line data base. For an experimental period of four months, SUNY will load 50,000 citations from the 1971 files of the *Drug Literature Index (DLI)*, published by the Excerpta Medica Foundation. This tool is a hybrid, appearing to be a cross between the bibliographic index and the abstracting service, in that it combines the citation with what might be called a telegraphic abstract comprised of a number of thesaural terms. The *DLI* provides indexes by drug class, generic name, trade name, author, and a separate listing of new drugs. The citations are somewhat different than those of many other indexes in that the foreign-language titles are first given in English translation, followed by the affiliations of the authors. See Appendix at end of chapter for additional information.

The indexing terms used are from a controlled vocabulary called MALIMET (Master List of Medical Indexing Terms), but the *DLI* also uses what it calls secondary indexing terms. MALIMET consists of some 40,000 preferred terms together with *see also* references and up to five class assignments which indicate biomedical fields, subjects areas, or disciplines. In addition to the primary or preferred terms, there are approximately 500,000 synonyms which are also keyed to the primary related term. These secondary terms are not those which would normally be used to look up an article but which provide information as to the nature of a given investigation (e.g., anatomical, demographic, or electrophysiological study), the species of experimental animal used, the type of primary document (e.g., review, textbook, etc.) or even quantitative data on the scope of the study and the value of the results (e.g., nineteen patients, result). The sources of the citations are 3,400 biomedical journals appearing worldwide, and each monthly index contains about 4,000 citations. The price of this service (\$2,000 a year for the twelve monthly issues including two semiannual cumulations) makes it prohibitive for most libraries, but the use of the data in a cooperative network brings the information to a number of locations which could not individually afford to obtain it.

Using the STAIRS operating system developed by IBM, the SUNY network will permit the user to search the file either through the controlled vocabulary or by the use of the secondary, natural language, terms. The internal indexes are constructed as a series of inverted files, and the computer retrieves the citation only after all of the search parameters have been satisfied. This experiment will be monitored closely and evaluated, and the use of the file will be studied by an outside (non-SUNY or network) research team. If this test is successful, the network plans to make available a two-year file of the *Excerpta Medica*, now published in thirty-nine monthly subject sections, which will also be searchable using the controlled vocabulary or natural language. *Excerpta Medica* abstracts are English-language summaries, not telegraphic lists of subject headings. This data base would contain 500,000 citations, and the abstract words would be arranged internally in inverted files.

The recent *UNISIST Study Report on the Feasibility of a World Science Information System* noted the parallel development of MEDLARS and the *Excerpta Medica* systems and then called for the coordination of the two services in this area in which they closely overlap. Both of the parent organizations have agreed to adopt a common communication format to facilitate the exchange of data and the possible interconnection of the two systems.

SUNY			MEDLINE		
	Citations	Dates		Citations	Date
Medlars files on-line	645,378	Oct.1967-June 1972	Medlars files on-line	400,000	1969- June 1972
off-line	1,043,114	1964-1969			
NLM book	48,000	1968-1971			
NETBOOK	15,890	1962-1970			
DLI	50,000	1971			
Total on-line file	759,268				

Table 1. SUNY-MEDLINE Data Bases

Multiple Networks

A valid often-raised question concerns the benefit or necessity of having two networks in existence such as MEDLINE and SUNY. Why should a library pay SUNY \$8,000 to \$10,000 a year when it could participate in the MEDLINE network for the cost of a TWX terminal, which it may already be using? The major reason is that the two networks do not duplicate each other at the present time, and each can do some things that the other cannot. This gives additional power to the library which is able to employ both systems, allows it to provide more services, and, of course, doubles the access of its users to the available data bases. Even as MEDLINE adds additional data bases such as the NLMBOOK files, the networks will not necessarily converge (table 1). Each plans new services and the things that each system is able to do will continue to differ because of the internal operating systems which have been selected. The SUNY Network Advisory Committee has urged the network to select a different type of system for its operation because of the potential that results from testing different methods of doing things, a legitimate function of a university and a capability which might well have been lost were the SUNY network to have turned into a back-up MEDLINE operation. As evidenced by the forthcoming work with the *Excerpta Medica* data bases, the need for a viable alternative system exists and would benefit the profession in many ways.

An unresolved problem of on-line systems which does depend on their technology is that which arises when more and more people attempt to use the system simultaneously. A graph produced by the National Library of Medicine (NLM) shows clearly that an important drop in response time occurs when the number of simultaneous users rises from forty-five to fifty-five (fig. 2). Considering that there are at least 100 users—some of whom are shown with their use figures in table 2—who constitute NLM's primary population

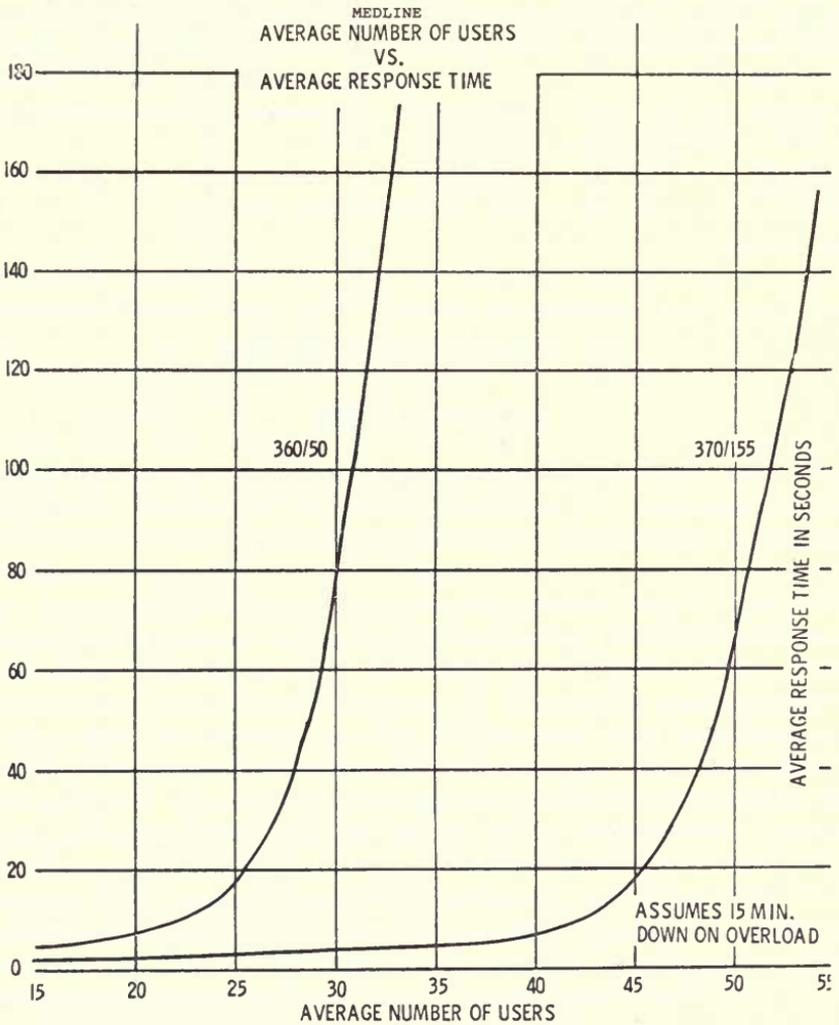


Fig. 2. How Number of Users Affects Response Time

(including the medical schools and the regional medical libraries), and that there are plans for extending the service to secondary sites such as hospitals, this is an important factor to consider. Increases in search activity are shown in figs. 3 and 4.

Although the difference between a response time of 20 seconds and 160 seconds may not appear to be great, if we were to stop and wait for those 160 seconds to pass it would begin to seem interminable. To the user seated

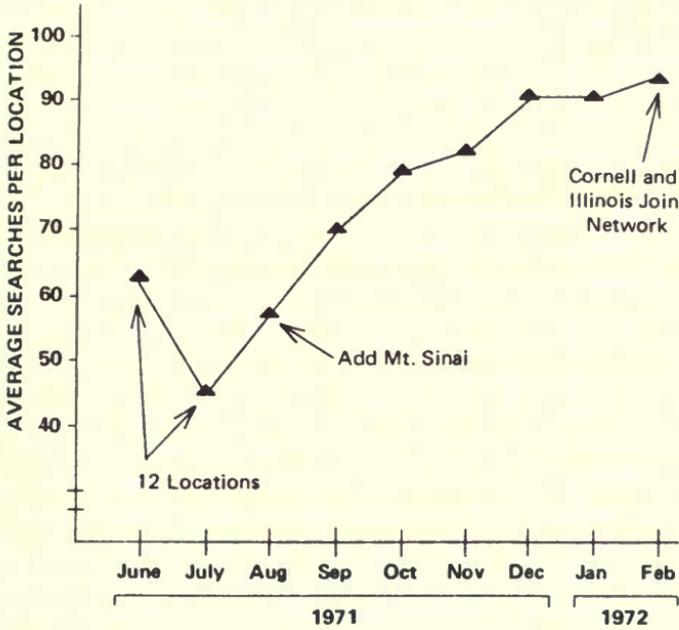


Fig. 3. Average Monthly Searching Activity

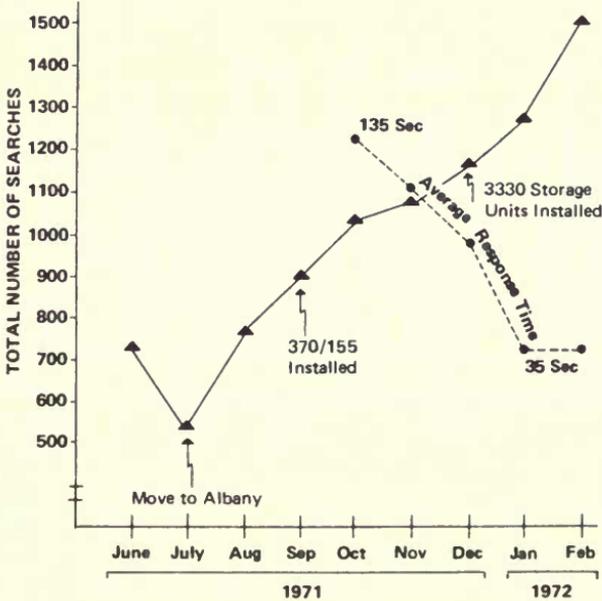


Fig. 4. Total Monthly Searching Activity

Terminal	Completed Searches											
	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.		
*02 Food and Drug Administration	101	70	66	130	112	94	97	60†	60†	28		
03 Upstate Medical Center								140	185	170		
04 "												
*05 Medical College of Ohio, Toledo	50	62	64	54	77	59	97	37	72	97		
06 University of Rochester	39	36	54	85	80	91	92	85	79	77		
07 Indiana University Medical Center	102	72	113	182	227	218	207	94	103	122		
08 SUNY at Buffalo								231	260	255		
09 "												
10 "												
11 Albany Medical College	69	33	65	79	87	76	67	103	100	75		
**12 National Library of Medicine	76+	6+	25+	41+	37+	139+	253+	475+	388+	75		
**13 National Library of Medicine	29	20	31	35	67	65	44	55	65	62		
14 Ohio State University	11	12	5	21	18	23	31	11	23	27		
15 Downstate Medical Center	65	40	73	56	59	126	130	110	92	144		
16 College of Med & Dent of NJ, Newark	25	26	33	36	32	36	50	57	71	83		
18 Countway Library of Medicine	99	52	76	44	66	63	95	87	83	115		
19 Medical College of Virginia												
***20 New York State Medical Library												
***21 SUNY at Stony Brook												
22 Cornell University Medical College	60	36	60	62	76	83	113	121	109	77		
23 Johns Hopkins University	74	81	90	84	96	90	95	99	143	140		
***24 US Army Medical R & D												
25 Mount Sinai Hospital			18	30	32	44	65	45	43	68		
26 Cornell University Life Sciences									60	79		
27 University of Illinois, Chicago									25	66		
Total	724	540	748	898	1029	1068	1183	1275	1513	1760		

† not included in totals

* incomplete

** terminal not used for "searching"

*** no data available

Table 2. Some SUNY Biomedical Communication Network Users and Their Searches

at a computer terminal, who knows that there is a powerful device connected to his station which operates at electronic speeds, a delay of 160 seconds becomes intolerable. We learned very early that the threshold of user impatience when using an on-line system is very low. One can choose to ignore the problem, but one then finds that the system is regarded as unsatisfactory by many, and ignored altogether by others. Driving away the potential user is not the solution to the problem.

NLM postulates that there may be a need for shifts during which groups of terminals are authorized to use the system; this method of expanding the capacity of the SUNY system was also considered. SUNY has not yet, however, had to face the problem because their terminals still number less than thirty. The system designer must also consider that when the system overloads and fails, it may require fifteen minutes to restore the programs and enable searching to resume. If this happens with any frequency, not only is one faced with the problem of user dissatisfaction, but also with what happens to the data which were being transferred at that time. When dealing with a system other than searching from a file, that is, when entering new data to the system, one does not know, for example, which circulation records were lost when the system dropped without going back and redoing a certain number of transactions at each terminal location—a difficult task if the user's ID card is required as well as the book card, and the user has already left the circulation area. When one is dealing with internal library operations, the problem may be tedious, but no more than that; when the user is involved, however, the problem is far more complicated.

CASE and CRIB

The Library of the Health Sciences now has four on-line systems combining a variety of functions and services available for use by the patron. The systems are the SUNY Biomedical Communication Network, the National Library of Medicine's MEDLINE (MEDLARS On-Line) Network, and the University of Illinois at the Medical Center's CASE (Computer Assisted Simulation of the Clinical Encounter) and CRIB (Computer Randomized Item Bank) systems.

CASE is a unique program which enables the medical student or practitioner wishing to review or test a course of action to select a case study in a particular area from the series available (emergency orthopedics, obstetrics and gynecology, pediatrics, psychiatry, and internal medicine) and proceed with analysis of the problem, diagnosis and treatment. The program provides an introductory description of the patient and other related information and the physician is then able to ask the patient (i.e., the computer) questions, order laboratory tests and review the results, obtain physical information and any other data which he deems significant. After this stage is complete and the physician feels that he understands the case and is ready to proceed, he

then enters his own diagnosis and treatment of the case. After this is done, he receives the author's diagnosis and treatment plan, and is told what happened to the patient as a result of the course of treatment he prescribed. The system also supplies a list of critical concepts relating to the case which should have been explored and notes whether the physician covered them or not. The user may terminate his interaction with the system at any time by typing the word "quit."

CRIB is a system designed to allow a student to test himself in a subject to determine his grasp of the material as well as to instruct himself. The subject areas covered by CRIB are: anatomy, dermatology, microbiology, physiology, pathology, histology, pharmacology, medicine and orthopedics. The questions are in multiple choice form and the student is credited with only his first answer, i.e., he cannot change his mind on a particular question after answering it. He can then request the correct answer by typing the word "answer" to see if he was correct. He can skip questions in a series, or terminate his test at any time. The system maintains a record which is available only to the student, so that he can see how he scored on the question group. The system uses a Hazeltine 2000 Cathode Ray Terminal which is connected to the computer using an acoustical coupler and a voice grade telephone line. There are eleven terminals now connected to the system including those in the library's branches in Peoria and Urbana. Terminal access will also be provided in the branch library in Rockford in 1972 when the first class is admitted.

Plans are also underway for the CASE system to be used through the National Library of Medicine's Biomedical Communications Network using the TYMNET circuits (also used by MEDLINE).

Coordinated Staffing

An intriguing area of networking in an on-line mode is that involving the sharing of professional resources. Although this is basically an uncomplicated concept, it is one which has not been applied in an organized program.

By coordinating staffing patterns, the various stations in a network would be able to provide professional service over a broader range of hours of operation than any one of them could provide alone. In this way, weekend and evening hours, which are usually considered disadvantaged shifts by most staff members, could be distributed among a large group of people. It does not, of course, require a computer system to achieve this type of program. A coordinated staff sharing program can be developed using telephone lines, and the Library of the Health Sciences will be testing the effectiveness of this program using both telephone lines for direct calls to Chicago when a professional staff member is unavailable in Peoria, and the SUNY terminal, using the computer as a message switching device. One convenient feature of the

computer message is that if an addressed terminal is busy, the computer stores the message and transmits it as soon as it can capture the terminal. Messages show the time of input as well as the time of delivery.

A major problem with this type of service is the human problem of convincing the user to pick up the phone to make the call for assistance, or getting him to sit down and type his message. The use of Picturephone may well be an added inducement to the user to try such a service, and it has the advantage of making the program more personal. Another important feature of Picturephone service is that it enables the professional to place printed copy on the screen for the user to either read quickly or note and removes the disadvantage of much telephone reference work which results from the reading of sections of a book to answer the question. This combination of graphic display and two-way verbal communication has not yet been exploited by libraries; it will be interesting to see what the impact of such service will be.

STANDARDIZATION

One would be remiss in discussing network technology without at least mentioning the importance of standardization. The small differences in the ways libraries have recorded information and what information they have required have caused innumerable problems in the development of national data bases. The efforts which went into the establishment of the MARC format were monumental, and the agreement of libraries to use MARC as a communication format has been relatively slow in gaining wide acceptance. The fact that all of the major national library associations have endorsed this policy seems to have made little if any difference. The significance of the term "communication format" should not be lost, for libraries are asked to standardize not their internal systems, for which needs differ considerably, but only to provide a common format which will enable the data captured at one location to be used at another. This means, in effect, that each library or computer center need write only one program to convert its data into the standard format and one program to convert MARC format tapes to its internal processing format.

That standardization should be difficult for libraries is a little strange, since many libraries have standardized their subject heading lists and classification systems by basing their work on the most common U.S. systems available. But the implied loss of autonomy which standardization means has caused many libraries to resist, and has hindered progress toward a national network or even toward regional centers.

Nonetheless, networks imply a degree of standardization in both the input to the system, and the techniques used in querying the system. It would seem preferable to be found in the vanguard of such a movement and to assist in

the development of standards, than to be caught in the rear where one may be forced to adopt them for sheer survival.

On-line technology is the tool which will enable libraries to meet the challenge of providing information services to their constituencies in the 1970s, and it is the greatest hope that libraries have for establishing a relevant relationship with their users.

APPENDIX

By mid-summer of 1972 the SUNY network had decided not to use the *DLI* for a test data base, but instead to work with the *Excerpta Medica* files.

Plans were made to prepare two to three months of the entire data base of the thirty-nine sections of *Excerpta Medica* for searching in an on-line mode. Four test centers were selected, John Hopkins Medical Center Library, the Francis A. Countway Library of Medicine (Harvard), the SUNY Upstate Medical Center Library, and the University of Illinois Library of the Health Sciences.

The network will conduct a comparative evaluative test using both free text and controlled vocabulary searching on the *Excerpta Medica* file, and using controlled vocabulary searching on the MEDLARS file. Each test center will initiate approximately forty searches and the resultant 160 searches will be performed in each center. After the evaluation of the test data, it is expected that the entire *Excerpta Medica* file for the most recent two-year period will be loaded into the system and made available to all network stations.