The SUNY Biomedical Communication Network: Six Years of Progress in On-Line Bibiographic Retrieval

By JANET EGELAND, *Director*

SUNY Biomedical Communication Network
State University of New York
Albany, New York

ABSTRACT

The SUNY Biomedical Communication Network became operational in 1968 as the first on-line bibliographic retrieval service for biomedical literature. Since 1968, the SUNY/BCN has grown in size from nine to thirty-two medical and university libraries and has expanded its data base coverage to include the ERIC and *Psychological Abstracts* data bases in addition to the full ten-year retrospective MEDLARS data base.

Aside from the continuous provision of an on-line searching system, the SUNY experience over the last six years has yielded valuable information in the following areas of: (1) monograph indexing and retrieval, (2) shared cataloging, (3) user interaction and education in on-line systems, and (4) member participation in Network policy-making processes.

The continued success of the SUNY/BCN is evidence that it is possible to provide a high quality on-line bibliographic retrieval system at cost to academic institutions. SUNY's success in this effort is the result of centralized resource sharing and effective regional networking, combined with thoughtful planning by user advisory committees.

THE State University of New York (SUNY) Biomedical Communication Network was the first on-line information retrieval service for biomedical literature. As early as 1966, plans were under way at the Upstate Medical Center Library in Syracuse, New York, for making on-line access possible to the existing MEDLARS (Medical Literature Analysis and Retrieval System) data base produced by the National Library of Medicine (NLM).

Under the direction of Irwin Pizer and with the cooperation of NLM and significant financial support from SUNY, the technical personnel and equipment necessary for the development and eventual operation of an on-line information system were secured and arrangements made for the initial participation of nine medical libraries in the proposed Network. An IBM 360/40 was installed at the Upstate Medical Center Library in 1967,

and programmers began work on system design. Programming was completed and pilot testing performed during 1967 and 1968.

In October 1968, the SUNY/BCN became fully operational, offering on-line searching of the following data bases:

1. MEDLARS (1964–68),
2. NLMbook (the NLM Current Catalog tapes from 1966–68),
3. Netbook (Monograph holdings of the three SUNY medical libraries).

In July 1971, Network operation and administrative headquarters were transferred from Syracuse to the SUNY Central Office Computer Center in Albany, New York, under the direction of Ronald Quake. Consolidating the Network operation within the larger central office computer facility enabled SUNY to maintain its service at a lower cost to the user than would have been possible had operation continued at Syracuse on the smaller, dedicated computer system.

A good review of early Network goals and operations may be found in a 1969 article by Pizer (1).

The SUNY Network has remained in continuous operation since its opening day in October 1968, and the invaluable buildup of experience gained during this six-year period should be of great interest to all those persons involved with biomedical information retrieval or with on-line information retrieval in general. This paper is a discussion of the SUNY experience and its implications for the future of on-line information systems in biomedicine.

THE SUNY/BCN TODAY

Membership Expansion

The overwhelming success of the SUNY Network is evidenced by the fact that it has grown in

Bull. Med. Libr. Assoc. 63(2) April 1975

189

size since 1968 from the original nine medical libraries to a Network of thirty-two medical and university libraries from Massachusetts to Minnesota.

Each of the member libraries has a thirty character per second on-line communication terminal connected via a dedicated telephone line to the IBM 370/158 computer at the SUNY Central Office Computer Center in Albany, providing them with unlimited on-line access to the various data bases at any time during the forty-six-hour operational week. SUNY members have indicated a high level of satisfaction with dedicated line service because it avoids many of the problems such as frequent system overloads ("out of ports") and accidental disconnects that are experienced in dial-up services. Dedicated line service also permits a high volume of searching activity at minimal cost.

Data Base Expansion

Between 1968 and 1973, continued refinements and improvements were made to the on-line SUNY/MEDLARS search service. Then, in 1973, the decision was made to broaden the scope of the Network's data base coverage to other academic areas so that non-medical institutions could also have the benefit of on-line bibliographic services.

The first step in this direction was the selection of a new software package that would permit both natural language (free text) and controlled vocabulary access to multiple data bases. The modified IBM Document Processing program which ran the on-line MEDLARS service from 1968–73 was not flexible in these respects. After an extensive comparative evaluation of available on-line software packages by the Network systems programmers, the IBM STAIRS (Storage And Information Retrieval System) package was selected for this purpose and the MEDLARS data base was reloaded under the STAIRS program in March 1973.

In response to many requests from educational institutions in New York State, the ERIC (Educational Resources Information Center) data base was loaded under STAIRS and made available for on-line searching in October 1973. The *Psychological Abstracts* data base was the third file added to the Network search system in March 1974. A brief description of each of these files follows.

Data Base Descriptions

SUNY Network members now have three different data bases available to them for on-line searching. In any one session at the on-line terminal in their library, these users may elect to search any one or all of these files:

1. *MEDLARS.* The MEDLARS data base, produced by NLM, now contains over two million citations to articles in the more than 2,700 international journals that are indexed for the printed *Index Medicus, International Nursing Index,* and *Index to Dental Literature.*

The SUNY/MEDLARS file is available for retrospective searching back through 1964. The file is updated monthly, and there is an automatic SDI service available to provide users with monthly updates to a search profile.

The SUNY/MEDLARS data base is accessible by the controlled vocabulary *Medical Subject Headings (MeSH), MeSH* tree numbers, and title words. A search may be further refined by scanning for certain authors, journal titles, or subsets. Full bibliographic information is available for each citation, including those *MeSH* terms used to index the article.

2. *ERIC.* The ERIC data base contains citations to the literature of education and its related disciplines. The citations come from two separate sources: (1) *Research in Education (RIE)*— research reports filed by contractors and grantees on the results of funded educational research; (2) *Current Index to Journals in Education (CIJE)*—articles from more than 500 education and education-related journals.

The *RIE* file contains approximately 70,000 citations with abstracts covering the period from 1966 to the present, with approximately 12,000 new citations added each year. The *CIJE* file contains over 90,000 citations from January 1969 to the present and grows at the rate of approximately 18,000 per year. Thus the total ERIC file currently contains over 160,000 citations.

The SUNY/ERIC data base is accessible by subject terms from the *Thesaurus of ERIC Descriptors,* as well as via free text searching on all abstract and title words, author names, and other meaningful bibliographic information such as institution names.

3. *Psychological Abstracts.* The *Psychological Abstracts* data base, produced by the American Psychological Association in Washington, D.C.,

190

Bull. Med. Libr. Assoc. 63(2) April 1975

contains approximately 160,000 citations with abstracts from 1967 to the present.

These citations come from over 800 psychological and behavioral science journals, plus book literature, technical reports, conference proceedings, and other material that is considered relevant to the behavioral and social science community.

The SUNY/PSYCH ABSTRACTS data base is accessible via subject terms listed in the *Thesaurus of Psychological Terms*, as well as via free text searching on all title and abstract words, author names, classification codes, and publication type.

The SUNY/BIOSIS Project

Of special interest to the biomedical community is a cooperative research project still in progress at this writing between SUNY and the BioSciences Information Service, Philadelphia, to determine the applicability of the *BA Previews* data base for on-line searching.

The major goals and the actual design of the experimental project are discussed in two recent issues of the *SUNY/BCN :Network Newsletter* (2) and in a paper by Durkin and Egeland (3). The final results of the experiment and SUNY's plans for future on-line implementation of the *BA Previews* data base will be available at the completion of the project.

SUNY CONTRIBUTIONS TO THE ACCUMULATION OF KNOWLEDGE IN THE AREA OF ON-LINE INFORMATION RETRIEVAL

Aside from the continuous provision of an on-line literature searching service, the SUNY Network experience provides important guidelines for the improvement and expansion of this type ɔf service in the future.

Specifically, the SUNY experience has provided valuable information in the following areas:

1. Monograph indexing and retrieval,
2. Shared cataloging,
3. User interaction and education in on-line systems,
4. Member participation in the Network policy-making process.

Monograph Indexing and Retrieval

One of the original goals of the SUNY/BCN was the provision of on-line access to monographic

as well as journal literature. Two different monograph data bases were made available for on-line searching in 1968: (1) NLMbook and (2) Netbook.

The NLMbook file contained the *Current Catalog* data from the National Library of Medicine from 1966–68. These data were searchable by all the standard MARC bibliographic fields, as well as the *MeSH* descriptors used as subject tracings on the catalog card for each book.

The Netbook file contained two distinct types of information: (1) catalog card information for the monographic holdings of the three SUNY medical libraries with a publication date of 1962 or later; (2) depth-indexed monograph holdings from the Upstate Medical Center Library, 1962 or later.

As a pilot project, the in-depth subject indexing of monographic works for future on-line retrieval via the SUNY/BCN searching system was begun at the Upstate Medical Center Library in 1967. This project involved the chapter-by-chapter subject analysis of all monographs (excluding texts, laboratory manuals, atlases, etc.) in the library with a publication date of 1962 or later. A combination of *MeSH* and LC subject headings was used to describe the subject content of the books selected for indexing. The depth indexing of conference proceedings was considered of particular importance to the project.

At the conclusion of the project in December 1969, approximately 8,000 monographs, including over 800 conferences and symposia, had been depth indexed and entered into the Netbook file for on-line searching. A detailed account of this pilot project is available in an earlier paper by Egeland (4).

The Netbook file remained on-line in the SUNY system from 1968 to July 1971, at which time it was removed to make room for more on-line journal citations. The NLMbook file remained on-line until December 1972.

Although no formal statistical evaluation was made of the actual use of the Netbook and NLMbook files, a brief survey of all Network members brought forth the following useful information about on-line accessibility to monographic works.

1. On-line accessibility to monographic literature was desirable only if the amount of on-line storage necessary for journal citations was not affected. Without exception, all the member libraries surveyed stated that the two book files should be removed from the Network search system if

maintaining them reduced the number of journal citations that could be stored on-line.

2. Subject accessibility to monographic literature via only the subject tracings on the catalog card was judged inadequate for on-line retrieval purposes since, in most cases, only very general *MeSH* headings are used as subject tracings for the overall content of a given book.

3. The continuation of an in-depth subject analysis for conferences and symposia was considered as extremely worthwhile for future on-line information retrieval.

This information may be useful to those persons planning projects dealing with monographic literature retrieval in the future.

Shared Cataloging

In cooperation with NLM, the Francis A. Countway Library of Medicine, and the Upstate Medical Center Library, the SUNY/BCN developed the framework for the first on-line cooperative cataloging system.

In April 1970, SUNY's SCATT (Shared Cataloging Access Through Terminals) system became operational, providing an on-line searching capability to an in-process file of books being cataloged at each of the participating libraries. This file was accessible via the author, title, or a combination of author and title. Additional information about SCATT is available in an article by Onsi and Pelosi (5).

After a year's experience with the experimental SCATT file, a meeting of the cooperative cataloging participants was held in Boston in March 1971. At this meeting, the announcement was made by the NLM representatives that the development of a similar file to be called CATLINE was concurrently under way at NLM. Thus the SUNY project was discontinued at that time, after demonstrating that on-line shared cataloging was a viable solution to the increasing need for cooperative programs in this area of medical library operation.

User Interaction and Education in On-line Systems

Perhaps some of the most useful information resulting from SUNY's six years of experience with on-line bibliographic retrieval pertains to the problems encountered in user-system interaction and user education in the retrieval process.

Early User Interaction Problems

It was the philosophy of the early Network planners that the SUNY system be designed to allow the physician or researcher to perform a search at the input terminal without the aid of an intermediary librarian or search analyst. With this goal in mind, a query language interface was written by Network programmers to guide the user through the process of entering a search without requiring any knowledge of machine language or coding (6).

When the Network first became operational, users were encouraged to sit at the terminals and perform their own searches via the query language. After several months, it became obvious that this arrangement was unsatisfactory for many reasons, the most important of which were:

1. The majority of the users were not able successfully to translate their request into the appropriate *MeSH* terms necessary for search entry. Regardless of their familiarity with their subject areas, patrons were not aware of the specificity and structure of *MeSH* and the process of coordinate searching that is required for retrieving information on specific topics. As a result, the retrieval accomplished by the user himself was frequently not adequate to meet his needs and he was disappointed in what he felt was the inadequacy of the system itself.

2. Potential users were discouraged because the access terminals were constantly occupied by other patrons performing searches. The time schedules of most of these patrons did not permit them to wait in line or to return at regular intervals until a terminal was free. An appointment schedule was suggested, but it met with negative reaction from clinicians, in particular, who maintained that their needs for information were not predictable in this sense.

3. Even when terminals were available, many patrons indicated that they were not interested in taking their own time to perform searches and asked for assistance from a library staff member. The results of a questionnaire distributed at random to requesters substantiated this observation. Of the 241 users who returned the questionnaire, only forty-four (18%) said that they preferred personally to operate the terminal and perform their own searches, while 187 (78%) specified that they preferred to have their searches processed by a trained assistant.

192

In response to these problems, all the Network member libraries either placed an existing staff person (usually a reference librarian) in charge of the SUNY search service or hired a "search analyst" or "information specialist" to be trained to perform this function. In either case, patrons were no longer encouraged to perform their own searches but were asked to seek the help of an intermediary so that the quality of the retrieval could be controlled. The emphasis, then, switched from the education of the patron to the education and training of the intermediary.

This change in philosophy from the idea of a totally user-oriented system to the current intermediary system was based on actual experience with thousands of health professional users.

The Education and Training of the Intermediary

Throughout the years, the SUNY/BCN staff has provided assistance in the training of the intermediary search personnel in each of its member libraries.

It had become quickly apparent from retrieval statistics and patron feedback during the first year of operation that these intermediary searchers needed at least a minimal knowledge of the basic indexing procedures governing the input of the citations into the MEDLARS data base to accomplish even adequate success in the retrieval process. The majority of search failures identified in early on-line operation were the result of an incomplete or inaccurate search strategy and not of internal system processing errors or, certainly, of the inadequacy of the MEDLARS data base. These early observations of the SUNY on-line MEDLARS service were substantiated by Lancaster in his evaluation of the MEDLARS demand (batch process) service at the National Library of Medicine. In his report (7), he states that when recall and precision failures are analyzed together, the searching subsystem is the greatest contributor to all the MEDLARS failures. Specific problems he identified in the evaluation included the use of inappropriate vocabulary terms, the level of specificity or exhaustivity used in search strategies, and the failure of searchers to cover all reasonable approaches to the retrieval of relevant articles.

To help avoid some of these retrieval problems, the Office of Director of Educational Programs was established in 1969 to provide SUNY Network members with assistance in search profiling

for successful retrieval. Through this office, user training programs were developed to include intensive training in vocabulary usage and search strategy formulation based on MEDLARS indexing principles. All new users who joined the Network after this time were given this instruction in addition to the usual instructions in terminal operation and software functions. To provide continuing education in this area, the monthly *SUNY/BCN Network Newsletter* was published containing hints for efficient search strategies, problem searches, and vocabulary clarifications. Annual workshops were held to discuss specific aspects of the on-line searching process. In addition, the Director of Educational Programs was available at all times via either terminal or telephone for consultation on searching problems.

It is important to point out here that these educational programs were developed at the request of the users themselves in response to their own feelings of need for help with retrieval strategies. Even those medical librarians who had used the printed *Index Medicus* for years expressed dissatisfaction with the quality of their retrieval efforts from the MEDLARS data base.

Most of those same users are still part of the Network today and are now being confronted with the same problems in searching the new data bases available to them on-line. It has not been possible for them to make a direct transfer of training from the MEDLARS data base to these new files because of the many differences in the organizational structure of these files.

There are many special problems encountered in learning to use multi-data base on-line searching systems, and more attention will have to be directed to the educational programs for Network members in the future (8).

Member Participation in Network Policy-making Processes

One of the unique and most important aspects of the SUNY/BCN is that all the major Network policy has, since its inception, been determined by advisory councils of Network members themselves.

In 1968, the Network Advisory Council was composed of the directors of the original nine medical libraries. The current council has changed its name and its composition. The advisory group is now called the User's Task Force Committee,

and it consists of eleven elected representatives from the thirty-two member institutions. This group meets semiannually and is responsible for all major policy decisions affecting the operation and the future of the SUNY Network.

In 1971, a special Subcommittee on Network Operations was appointed by the User's Task Force Committee to serve as the liaison between the Computer Center staff and the Network users and to make recommendations to the larger group regarding necessary changes or modifications in system operations. The Subcommittee meets at regular intervals in Albany and has been instrumental in affecting several major operational changes to date.

These two committees work closely together to see that the best interests of the user group as a whole are represented when policy is determined. This type of user participation in the decision-making process has proved to be an extremely effective way to administer a large information network of this nature so that its service reflects the real needs of the people for whom it is intended.

SUMMARY STATEMENT

The success of the SUNY Biomedical Communication Network demonstrates that it is, indeed, possible to provide a high-quality on-line bibliographic retrieval service at cost to the academic community. This success has been the result of centralized resource sharing and effective regional networking, combined with thoughtful planning by user advisory committees.

This successful combination of factors should be considered in the planning of other on-line retrieval services in the future, not only for biomedicine but for all other scientific information systems.

REFERENCES

1. PIZER, IRWIN H. A regional medical library network. Bull. Med. Libr. Assoc. 57: 105-15, Apr. 1969.
2. SUNY/BCN Network Newsletter. v. 7, Nos. 1-3, 1974.
3. DURKIN, KAY, and EGELAND, JANET. An experiment to evaluate the use of *BA PREVIEWS* in an on-line interactive mode. In: Zunde, Pranas, ed. Information Utilities: Proceedings of the 37th ASIS Annual Meeting. Washington, American Society for Information Science, 1974. p. 141-44.
4. EGELAND, JANET S. In-depth indexing of monograph literature for an on-line retrieval system: a pilot project. Bull. Med. Libr. Assoc. 60: 432-8, July 1972.
5. ONSI, PATRICIA W., and PELOSI, SHEILA J. Cooperative cataloging from the participants' point of view: a record of SUNY's and Countway's experience. Bull. Med. Libr. Assoc. 58: 126-33, Apr. 1970.
6. EGELAND, JANET S. User-interaction in the State University of New York (SUNY) Biomedical Communication Network. In: Walker, Donald E., ed. Interactive Bibliographic Search: The User/Computer Interface. Montvale, New Jersey, AFIPS Press, 1971. p. 105-20.
7. LANCASTER, F. W. Evaluation of the MEDLARS Demand Search Service. Washington U. S. Department of Health, Education and Welfare, 1968.
8. EGELAND, JANET. The importance of user education and training in a multi-data base on-line information network. In: Zunde, Pranas, ed. Information Utilities: Proceedings of the 37th ASIS Annual Meeting. Washington, American Society for Information Science, 1974. p. 137-40.

194

Bull. Med. Libr. Assoc. 63(2) April 1975