MEDLARS

MEDLARS

INTRODUCTION*

The dramatic rise in the number and types of biomedical primary publications, such as books, journals, and technical reports, often makes direct access to this literature difficult and confusing. Increasingly, secondary publications, including indexes, bibliographies, abstracts, and catalogs, serve as important elements in information transfer by directing users to primary sources. Development and use of computerized bibliographic information systems have further facilitated access to the growing body of primary health-related information publications.

The National Library of Medicine (NLM or the Library) has been in the forefront of this field and

———————
*Information in this chapter was obtained primarily from NLM staff and NLM publications.

pioneered the "first large-scale library-based reference retrieval system" for health-related information (34). Since its inception, MEDLARS has become more sophisticated, accessible, and inclusive; and MEDLARS II, the Library's current system, is utilized more than any other system by health communities in the United States and abroad.

This chapter describes the development of MEDLARS, its current operations, and its projected future operations, providing information for the discussion of system issues in chapter 3 and for the analysis of public/private issues in chapter 6. As requested by Congress, the primary focus of the discussion is on MEDLINE, MEDLARS'S major biomedical bibliographic data base.

DEVELOPMENT OF MEDLARS

The development of MEDLARS parallels the evolution in medical bibliography that began in the 1950's with the first attempt to apply computer technologies to information processing. (See app. D for a general description of the development of computerized biomedical bibliographic retrieval systems.) The system's first data base, MEDLINE, was a byproduct of the computerized production of the printed Index *Medicus.* One of the earliest accomplishments of NLM's predecessor, the Library of the Army's Surgeon General, is attributed to John Shaw Billings and Robert Fletcher, who together began the first monthly index of the world's periodical medical literature in 1879 (74).

The early history of *Index Medicus* was marked by severe financial deficits, frequent changes in sponsorship, and the commitment of a few dedicated individuals and organizations. The Army Medical Museum and Library, the successor to the Surgeon General's Library, struggled to keep the index afloat despite financial problems until

1899, when the publication failed. It was then revived for 3 years as *Bibliographic Medica* by a group of French physicians. In 1903, the Carnegie Foundation undertook its financial sponsorship, with Fletcher as chief editor. In 1927, *Index Medicus* was merged with a similar index produced by the American Medical Association, and its circulation quickly increased sixfold, though it continued to lose $25,000 to $50,000 annually. In 1960, NLM assumed responsibility for publishing *Index Medicus* monthly, and the American Medical Association began the annual *Cumulated Index Medicus (CIM).* In 1965, with the advent of computer applications to publishing, the American Medical Association turned *CIM* over to NLM (74).

Index Medicus continues to provide physicians and other health professionals the major access to biomedical literature worldwide (142). Published monthly, *lndex Medicus* cites articles from 2,600 biomedical periodicals published in 36 languages. NLM endeavors to include references to

published articles in periodicals "judged to be of greatest potential use to the international biomedical community" (35). In 1879, Billings' and Fletcher's index cited some 20,000 articles. One hundred years later, Index *Medicus* adds citations of about 220,000 articles annually, and about 6,000 biomedical institutions subscribe to it each year (36).

By the late 1950's, NLM recognized that the exponential growth in biomedical literature required revision of the manual procedures for preparing Index *Medicus* and other publications. With the aid of a $73,800 grant from the Council on Library Resources in 1960, the Library developed a mechanized process that improved the method of preparing citations for publication.

In 1964, a true computerized system, MEDLARS, was put in place. The system was developed between August 1961 and December 1963, by the General Electric Corp., under contract to NLM. MEDLARS cost $3 million, some of which was supplied, at the recommendation of the National Advisory Heart Council, by the National Heart Institute. The new system improved the quality of the printed *Index Medicus,* enlarged its size, and decreased the time required for its printing by using the first computer-driven photocomposing device, Graphic Arts Composing Equipment (GRACE). The research and development of GRACE by the Photon Corp. was sponsored by NLM. The new technology represented a significant advance in the technology of computer typesetting (3). It also allowed for an increase in the number of articles that could be included in *Index Medicus* and in the number of subject terms assigned to each article. A module supporting the production of NLM's published catalog and its catalog cards was added in 1965. MEDLARS was used to produce all of NLM's publications from 1964 to 1975.

It was evident to NLM that the data base prepared for *Index Medicus* could be used for machine searching. In 1964, therefore, the Library began a batch computer operation. Users of MEDLARS would telephone, mail their requests, or make personal visits to the Library and other selected sites, and trained search analysts would access the system for the designated information.

However, there was no direct interaction between the searchers and the computer, the process was expensive, and the time between the submission of a request and receipt of the resulting bibliography ranged from 3 to 6 weeks.

In 1968, NLM began planning a more advanced, on-line automated support system to create new files and allow on-line formulation of searches. The specification and test of the logic of the search was to be done on-line, but the actual searching were to be done later in the batch mode. Meanwhile, in cooperation with the System Development Corp. (SDC), NLM was experimenting with an on-line bibliographic retrieval system. This system, AIM-TWX *(Abridged Index Medicus* via the Teletypewriter Exchange Network), used an abridged data base consisting of 5 years of citations from the 100 most important English-language journals in clinical medicine. The system became operational in May 1970, and was so successful that by November of the same year over 80 institutions were using it. AIM-TWX ran on SDC's IBM 360/67 computer. Telecommunications enabling remote access were provided via the TWX network (then part of the American Telephone & Telegraph Co. but subsequently transferred to the Western Union Corp.).

But the new MEDLARS, MEDLARS II, took more than 6 years to be completed. The original contractor could not fulfill its obligations, and a contract was negotiated with SDC. As an intermediate solution, SDC merged the old MEDLARS I with the already successful on-line retrieval system AIM-TWX and put MEDLINE (MEDLARS On-line) service in place in October 1971. * Because the communications facilities provided via the TWX network were disproportionately high in cost, NLM contracted with Tymshare, Inc., in 1971, to provide data communications services for MEDLINE. The MEDLINE data base included citations from 1,000 to 1,200 journals for the current year and the previous 3 years. The completely new MEDLARS II, which finally became operational in January 1975, is the system now in place. The computer was successfully connected to TYMNET, a telecommunications network, in 1977.

---

*Currently, the term MEDLINE is mainly used to describe the data base and not the service.

CURRENT STATUS OF MEDLARS

NLM is continuously refining MEDLARS II to extend its capabilities. The system has evolved into a complex multiprocessing system that maintains data files, provides on-line retrieval services, and produces computer-photocomposed publications. The main computer facility is located in the NLM building in Bethesda, Md.; a second facility for processing searches is at the State University of New York (SUNY), Albany. In September 1980, NLM enlarged its computer capacity twofold with the transfer of an IBM 370/168 multiprocessor system from the National Institutes of Health's (NIH) Division of Computer Research and Technology.

Commercial firms and the Government cooperate in providing telecommunications to NLM's computerized data bases through a "complex arrangement of lease lines, public dial-up switched circuits, commercial telecommunications networks, and the U.S. Government-sponsored research network (ARPANET)" (134). TELENET and TYMNET are the principal commercial networks and carry 80 percent of the on-line traffic.

Data Bases

General Description

MEDLARS II contains almost 20 data bases. The machine-readable data tape that is the source of data for MEDLINE is used to produce Index *Medicus* and other publications including the *Abridged Index Medicus, Health Science Serials, National Library of Medicine Audiovisuals Catalog,* lists of citations in specialized biomedical fields (termed recurring bibliographies), and individual bibliographic searches considered to be of general interest. Table 1 lists the scope and date of initiation of NLM's current on-line data bases. Most of the data bases, like MEDLINE, are bibliographic, and contain references to the primary journal literature (monographs, serials, etc.); a few contain numeric or representational information.

Some of the data bases–MEDLINE, CATLINE, AVLINE, HISTLINE, SERLINE, and SDILINE– are created and maintained solely by NLM.

Table 1 .—NLM Data Bases, Fiscal Year 1981

Data base and scope	Date initiated
AVLINE (AudioVisuals on-Line) —citations and abstracts to about 10,000 health science audio-visuals	February 1976
BIOETHICS LINE *(Bioethics on- Line)*— 13,000 references to materials on bioethical topics .	March 1978
CANCERLIT *(Cancer* Literature) –285,000 references dealing with various aspects of cancer	July 1974
CANCERPROJ *(Cancer* research PROJects)— 20,000 descriptions of ongoing cancer research projects	November 1975
CATLINE (CATalog on-LINE)-300,000 references to books and serials cataloged at NLM since 1965	September 1973
CHEMLINE (CHEMical dictionary on-LINE)– a file of some 1,000,000 names for chemical substances, representing 500,000 unique compounds	May 1974
CLINPROT (CLINical cancer PROTocols)— summaries of clinical investigations of new anticancer agents and treatment techniques . .	February 1976
EPILEPSY LINE (EPILEP'SY on-LINE)—contains about 25,000 references and abstracts to articles on epilepsy	January 1976
HEALTH *(HEALTH* Planning and Administration) —contains about 200,000 references to literature on health planning, organization, financing, management, manpower, and related subjects	November 1978
HISTLINE (HISTory of Medicine on-LINE)– some 48,000 references to articles, monographs, symposia, and other publications dealing with the history of medicine and related sciences	August 1978
MEDLEARN—computer-assisted instruction program which teaches the novice user how to search the NLM on-line system	November 1976
MEDLINE (MEDLARS on-LINE) –600,000 references to biomedical journal articles published in the current and 2 preceding years. An English abstract, if published with the article, is frequently included	October 1971
Back files that total some 2,700,000 references	
MED 66. ,	November 1975
MED 69.	December 1974
MED 72	May 1975
MED 75.	December 1977
On-line	
MED 77.	December 1979
MED 79. ,	December 1981
MeSH Vocabulary File (Medical Subject Headings Vocabulary File)—an on-line vocabulary file of the 14,000 medical subject headings that are used for subject cataloging, and also approximately 20,000 chemical records	April 1974
NAF (Name Authority file) -an authority list of 125,000 personal names, series names, corporate names and series decision records	September 1973
POPLINE *(POPulation* information *on-LINE)*– about 80,000 citations and abstracts to journal articles, monographs, and technical reports in the field of population.	March 1980

**Table 1.—NLM Data Bases,
Fiscal Year 1981 —continued**

Data base and scope	Date initiated
RTECS (Registry of Toxic Effects of Chemical Substances)—an annual compilation of toxicity data for approximately 50,000 substances	July 1977
SDILINE (Selective Dissemination of Information on-LINE)-references to the most current month of MEDLINE of approximately 20,000 citations.	October 1972
SERLINE (SERials on-LINE')-bibliographic information for about 38,000 serial titles, including all journals which are on order or cataloged for the NLM collection	November 1977
TDB (Toxicology Data Bank) —contains chemical, pharmacological, and toxicological information and data on approximately 3,500 substances	July 1977
TOXLINE (Toxicology Information on-LINE)– a collection of about 600,000 references from the last 6 years on published human and animal toxicity studies, effects of environmental chemicals and pollutants and adverse drug reactions	April 1974
Contains 660,000 references to older materials	
TOXBACK 65 ...., ...	September 1980
TOXBACK 74—on-line	September 1980

SOURCE: **National Library of Medicine.**

MEDLEARN, MeSH and NAF, which are basically support files, are also products of the Library. The other data bases are supported or produced in collaboration with various institutions. BIOETHICSLINE is produced in cooperation with the Kennedy Institute of Ethics, Center for Bioethics at Georgetown University. CANCERLIT, CANCERPROJ, and CLINPROT are sponsored by the National Cancer Institute of NIH. HEALTH is produced in cooperation with the American Hospital Association and the Health Resources Administration. POPLINE is produced in cooperation with the Population Information Program of Johns Hopkins University and the Center for Population and Family Health at Columbia University. And RTECS is an on-line searchable version of a publication prepared by the National Institute for Occupational Safety and Health. CHEMLINE is created by NLM's Toxicology Information Program in collaboration with Chemical Abstracts Service. TOXLINE contains secondary information from Chemical Abstracts Service, BIOSIS (formerly Biosciences Information Service), the American Society of Hospital Pharmacists, the Environmental Protection Agency, Oak Ridge National Laboratory, the now defunct Smithsonian Science Information Exchange (SSIE) data base, the Hayes File on Pesticides, and a Toxicity Bibliography produced by NLM. *

MEDLINE

MEDLINE, with its related back files (MEDLINE from 1966 to 1977), is the largest data base in MEDLARS and is, by far, the most extensively used (see tables 2 and 3). The literature indexed for MEDLINE contains not only information on the science and practice of medicine and public health, but also on bioengineering, bioethics, and

● **At this writing, May 1982, NLM offers EPILEPSYLINE, a bibliographic data base produced by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) of NIH from abstracts generated by the private firm, Excerpta Medica. Excerpta Medica provides NINCDS with computer tapes under a contract, at the cost of the tape. The data base is expected to be available through Bibliographic Retrieval Services (BRS) in the very near future, and thereafter will no longer be accessible through NLM. Use of EPILEPSYLINE is funded by NINCDS, and is rather low (fewer than 40 hours per month on average). Both NINCDS and NLM believe the data base will find a larger audience through BRS (159).**

Table 2.—NLM Data Bases and Updates, Fiscal Year 1981

Data base	Total records at data base[a]	Average number of records added per update	Frequency of updating
MEDLINE	638,374	**22,000**	Monthly
On-line and backfiles.	3,631,463		
TOXLINE	562,130	12,800	Monthly
On-line and backfiles.	1,225,013		
CHEMLINE	506,271	**7,300**	Every 2 months
CATLINE	333,800	260	Weekly
CANCERLIT	269,672	**3,100**	Monthly
HEALTH	186,999	**2,500**	Monthly

a Does not include the 116,620 retrospective records added July to September 1981,

SOURCE: National Library of Medicine.

Table 3.—MEDLARS: Usage of NLM Data Bases During September 1981

Data base	Number of hours
MEDLINE	5,565
MED 77	968
CATLINE	919
TOXLINE	826
CANCERLIT	398
MEDLEARN	389
HEALTH	318
CHEMLINE	285
SERLINE	162
POPLINE	161
TAB	160
SDILINE	149
AVLINE	137
RTECS	94
MeSH VOCABULARY	86
TOXBACK 74	70
NAME AUTHORITY	42
HISTLINE	38
EPILEPSY	34
BIOETHICSLINE	31
CANCERPROJ	31
CLINPROT	17
STORESEARCH	2
Total	10,882

SOURCE: National Library of Medicine

many other health-related disciplines. The intent is to provide references to the most useful biomedical literature from the most significant biomedical journals. Author abstracts are included for about 47 percent of the articles cited. Literature selection, thesaurus maintenance, and indexing are three MEDLINE-related activities that require subject matter knowledge. MEDLINE's ability to provide relevant bibliographic references depends on the performance of these activities.

Literature Selection. —Literature selection is used as a quality filter for the indexed biomedical literature database. About 95 percent of the citations in MEDLINE are those selected for the printed Index Medicus; the remaining 5 percent include citations from the Index *to Dental Literature, the Internatiortal Nursing Index,* and a limited number of nonserial publications. Although NLM chooses which journals are to be indexed for Index Medicus it has no formal influence on journal selection for the other indexes.

The selection process is highly structured and involves a critical review of the literature by a panel of expert consultants. Journals preselected on the basis of scientific merit and relevance to

NLM objectives. The number of serials indexed for Index *Medicus* is low in relation to the Library's collection: only 2,664 of the more than 20,000 serials collectedly the Library were indexed in 1981; the number of articles indexed that same year was 273,750. (See app. E for a comprehensive discussion of literature selection.)

Medical Subject Headings (MeSH).-Central to the development of MEDLINE is its controlled vocabulary, MeSH, a list of subject headings consisting of over 14,000 terms used in the indexing process to characterize an article's content. One way to retrieve citations to articles in *Index Medicus* and MEDLINE is by using these terms. Catalogers also use the MeSH terms to catalog books and other documents in NLM's collection. AVLINE, BIOETHICS, CATLINE, HEALTH, MeSH, POPLINE, and SDILINE are also "MeSH"-searchable data bases.

MeSH is arranged alphabetically and categorically. The 15 categories are subdivided, and arranged in a hierarchical manner, to show relationships between broader and narrower terms. The terms are constantly being updated to reflect changes in knowledge and practice. As of 1981, there were 9,000 major terms and 5,000 minor terms in the MeSH terminology. Both major and minor terms are used to describe articles cited in MEDLINE: only major descriptors are used to describe articles in *Index Medicus.* When an indexer assigns a minor descriptor for on-line searching in MEDLINE, the computer assigns the appropriate major descriptor under which the citation will appear in *Index Medicus.* The MeSH vocabulary is also used by catalogers in libraries throughout the world as subject descriptors for books and monographs.

MEDLINE can be searched not only by the controlled MeSH vocabulary, but also by using any word contained in the title of an article, or in the abstract when included. (For a full description of MeSH, see app. E.)

Indexing. -Subject indexing is a disciplined, intellectually demanding activity that requires consistency and accuracy in assigning subject headings to articles. In part, the effectiveness and reliability of a bibliographic retrieval system are built on the reliability of its data bases, which is deter-

mined by the indexing quality of its records. Indexers assign MeSH headings to describe an article's contents (indexing) on the basis of their orientation, training and judgment of the article's major and minor points and the headings' congruence with the subjects discussed in the article. They may modify the MeSH headings with one or more subheadings.

An article may be indexed exhaustively (indepth) and assigned about 10 MeSH terms, or not exhaustively (nondepth) and assigned approximately 5 MeSH terms, depending on the length and content of the article. All the articles in most of the journals covered by *Index Medicus* are indexed. In some journals, however, only selected articles are indexed; that is, since some journals carry health-related articles only occasionally, indexers are instructed to scan these journals and index only relevant articles.

Articles are indexed not only by NLM staff, but also by commercial contractors and by centers in foreign countries with which NLM has quid pro quo bilateral agreements. (See app. A for a more complete description of NLM's international activities.) Of the 273,000 articles indexed for *Index Medicus* in fiscal year 1981, NLM staff indexed 70,000 (26 percent), four U.S. commercial contractors indexed 52,000 (19 percent), foreign centers indexed 28,000 (10 percent) and U.S. commercial contractors indexed 123,000 (45 percent) for the foreign centers. Initially, all of the foreign centers did their own indexing. Currently, centers in the United Kingdom, France, Sweden, West Germany, South Africa, and the Pan American Health Organization Library in Brazil all have indexers on their staffs.

NLM exercises a considerable degree of quality control over the indexing process. In addition to requiring high educational standards for indexers, it provides them with a formal training course and continuous contact with able and experienced indexers, NLM revisors, who monitor the proficiency of all indexers. The Library also periodically updates training, and employs computerized validation routines and proofreading at a number of stages during the indexing process. (See app. E for a technical description of indexing.)

Leasing

NLM leases 12 of its data bases through the National Technical Information Service (NTIS) of the Department of Commerce. Payments from these leases are turned over to the U.S. Treasury after NTIS deducts a brokerage fee. As of January 1982, Bibliographic Retrieval Services (BRS) and DIALOG Information Services, Inc. (DIALOG) had signed such agreements with the Library. Until January 1982, the MEDLINE tape was leased to a subscriber at $50,000 for the first year and $30,000 for each subsequent year. Other MEDLARS data tapes were leased at different fee levels. NLM has established the leasing charges and arranged for and negotiated the agreements. NTIS, which acts as an accounting broker, receives payments from the leases and returns these revenues to the U.S. Treasury after taking a 10-percent brokerage fee; in the case of some data files such as CHEMLINE, NTIS pays royalties to the organizations that compile some of or all of the data file.

As of January 1982, the MEDLINE data base tape is leased on a fixed fee plus a use fee basis. There is a use fee of $4.00 per on-line connect hour and $0.01 per off-line printed citation for MEDLINE and MEDLINE back files, with a minimum fee of $20,000. After the $20,000 minimum is met, the use fees of $4.00 per on-line connect hour and $0.01 per off-line printed citation continue in effect. NTIS continues to serve as NLM's broker.

On-Line Services Users and Uses

In fiscal year 1981, NLM conducted over 2 million on-line searches, fully one-third of all such searches performed in the United States (165). In fiscal year 1977, 754 domestic institutions had direct access to MEDLARS. By 1981, this figure had grown to almost 1,550 domestic on-line centers. Users now include the major medical school libraries, over 700 hospital libraries, and libraries in pharmaceutical and other commercial firms (see table 4) and centers in 13 foreign countries. In all, there are more than 1,890 domestic and foreign institutions with direct access to NLM's data bases

Table 4.—MEDLARS: Number of Domestic Institutions On-Line by Region and Type, September 1981

					Type of Institution						
Region	Medical school	Allied health	Other university college	Other library	Hospital/ clinic	Research	Foundation Society Association	Commercial	Information agency	Other	Total
01	9	3	2	1	45	3	0	21	0	2	86
02	14	7	7	3	55	13	4	82	4	2	191
03	7	5	1	1	35	8	3	28	3	2	93
04	15	4	4	5	51	57	10	44	11	16	217
05	7	5	4	0	104	7	3	35	1	5	175
06	17	10	9	2	52	8	2	12	1	2	115
07	14	9	8	2	95	4	9	36	4	0	181
08	10	8	4	1	54	4	3	15	0	3	102
09	14	10	5	0	43	7	4	19	3	3	108
10	2	4	3	3	37	5	2	2	0	2	60
11	12	5	10	2	113	11	5	34	10	7	209
Total	125	70	57	20	684	127	45	328	37	44	1,537
					September 1979[a]						
Total	120	NA	91	NA	**400**	**93**	NA	273	NA	34	1,011

NA = not applicable.
[a]During 1980, classifications were expanded

SOURCE: National Library of Medicine.

(see fig. 1). Many other institutions, including 3,000 to 4,000 hospital libraries in the United States, provide indirect access to MEDLARS by referring requests to facilities with on-line terminals for searching.

Although NLM does not collect data on the individual users of its services, some information is available on institutions with direct access to MEDLARS. In fiscal year 1981, the Library ran 2.02 million search requests. Of these, 32 percent came from hospital on-line centers, 21 percent from medical schools, and 12 percent from commercial firms (see fig. 2). Hospitals, more than any other type of institution, have direct access to MEDLARS. In 1981, they were the largest users of MEDLARS as measured by both the number of searches requested (32 percent) and the number of computer connect hours (34 percent). Medical schools ranked second, with 21 percent of searches and 17 percent of computer connect hours. Hospitals and medical schools also ranked first and second, respectively, in their utilization of the MEDLINE and HEALTH data bases (see fig. 3).

There are 328 domestic commercial firms with direct access to MEDLARS. Commercial firms are the third highest user group in the utilization of all MEDLARS data bases, as measured by searches performed (11 percent), and the fourth highest

when measured by number of connect hours utilized (12 percent) (see fig. 2). Ten percent of the MEDLINE connect hours are used by commercial firms (see fig. 3). These firms tend to use CHEMLINE and TOXLINE more often than MEDLINE and utilize an estimated 30 percent of the total connect hours to TOXLINE.

There are no recent nationwide data on the ultimate users (i.e., end users) of MEDLARS data bases. Most studies of MEDLARS users have design limitations, are dated, or have been conducted on an institutional, local, or regional base. They indicate that a variety of health professionals, including students, request searches for a variety of reasons (see app. C). For example, a survey conducted by NLM in 1975 found that 41 percent of MEDLARS end users were physicians, 19 percent were nonphysician scientists, 10 percent were librarians, 10 percent were students, and 20 percent were reported as "other" (85). Preliminary results from a survey of hospital on-line centers, mainly located in New England, indicate that slightly more than one-half of search requests are from nonphysician health providers, including nurses, ancillary service providers, and administrators. Searches are requested to aid in providing patient care, preparing presentations and journal articles, and planning new services and purchasing equipment (50).

Figure I.—MEDLARS: Number of Domestic and Foreign Institutions On-Line, 1971=81
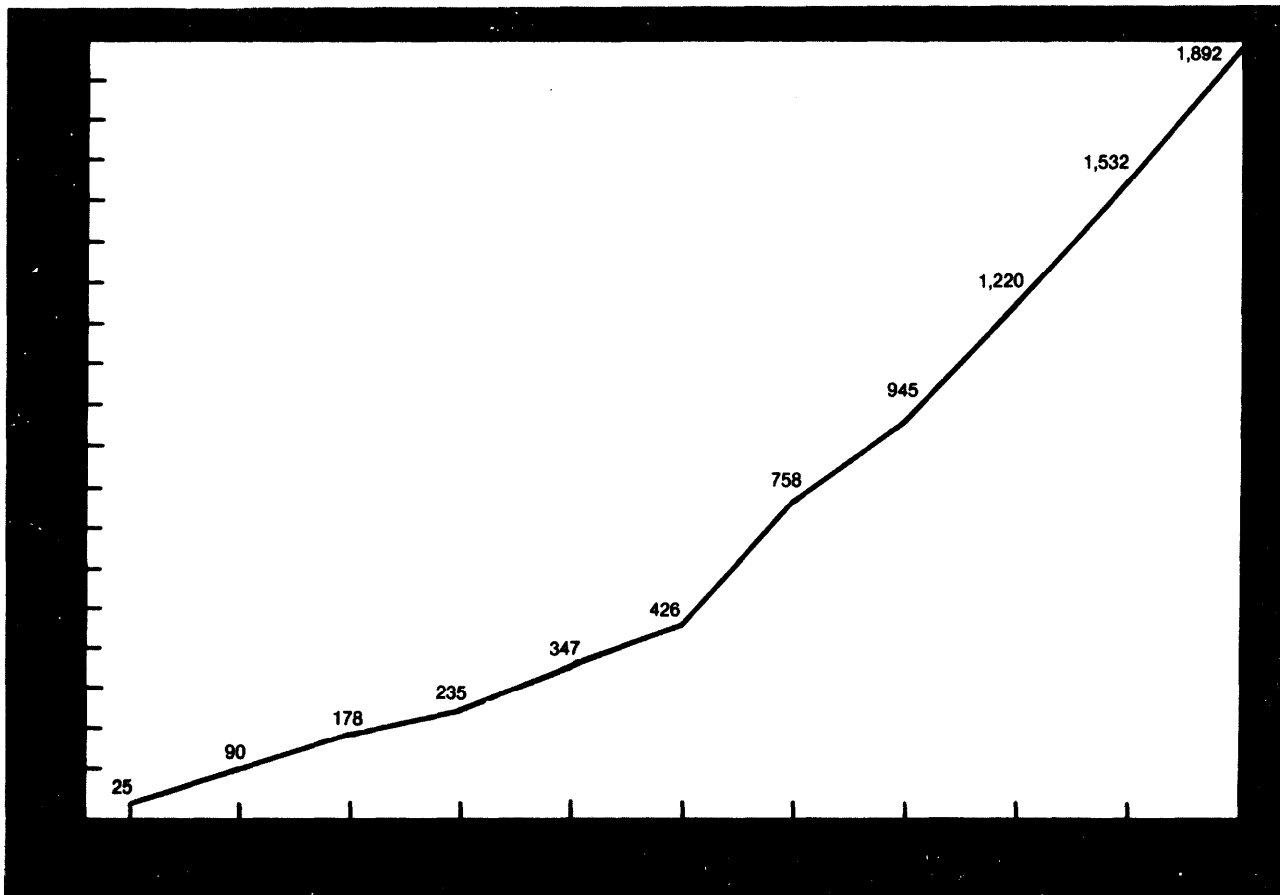
Searching and Retrieval

MEDLARS II allows for direct communication with the computer in on-line, interactive fashion. That is, the user can carry on a "dialog" with the computer, refining the search by typing in successive questions until the needed references are identified. An on-line search usually takes 10 to 15 minutes. (See app. E for a description of searching methodology.)
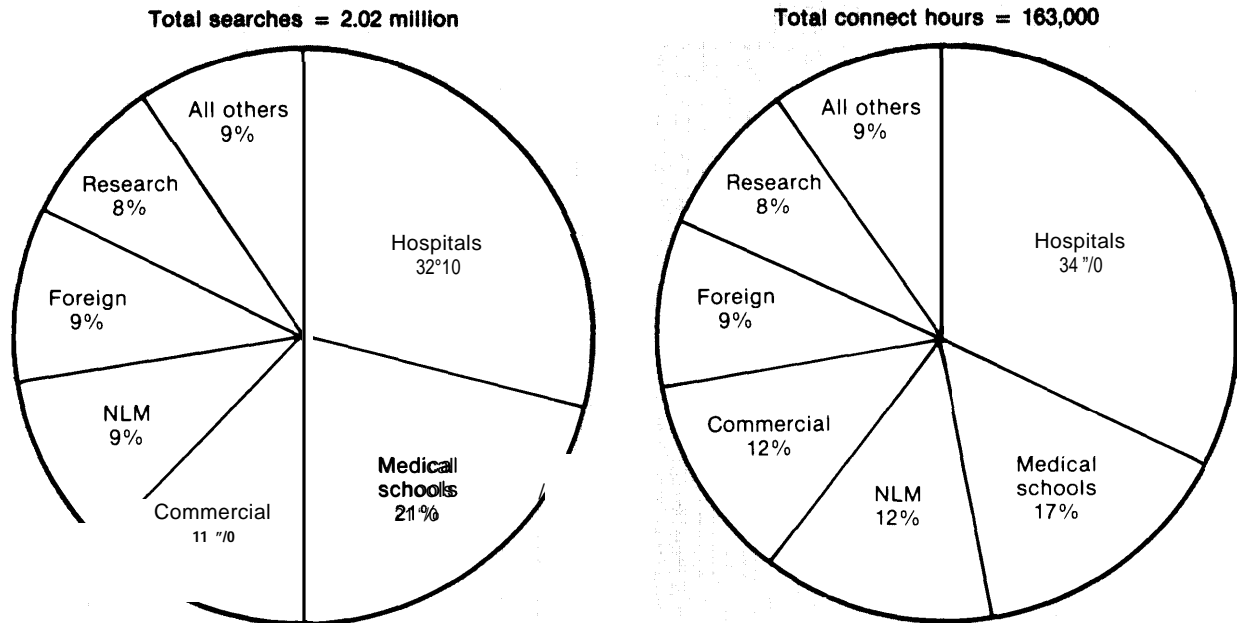
The references can be either printed at the time the search is entered (on-line) or printed during the offpeak hours (off-line) and mailed to the requester from NLM or SUNY the next morning. (Searches can also be formulated on-line and then stored in the computer for reference and later use.)

The number of on-line searches more than doubled from 1977 to 1981 (see fig. 4).

For the most part, trained search analysts, termed intermediary users, perform searches. Although an untrained individual can perform searches, the current system, like most others currently in use, is not designed for it. NLM is investigating methods to make the system more cordial to users without library or information systems training (64).
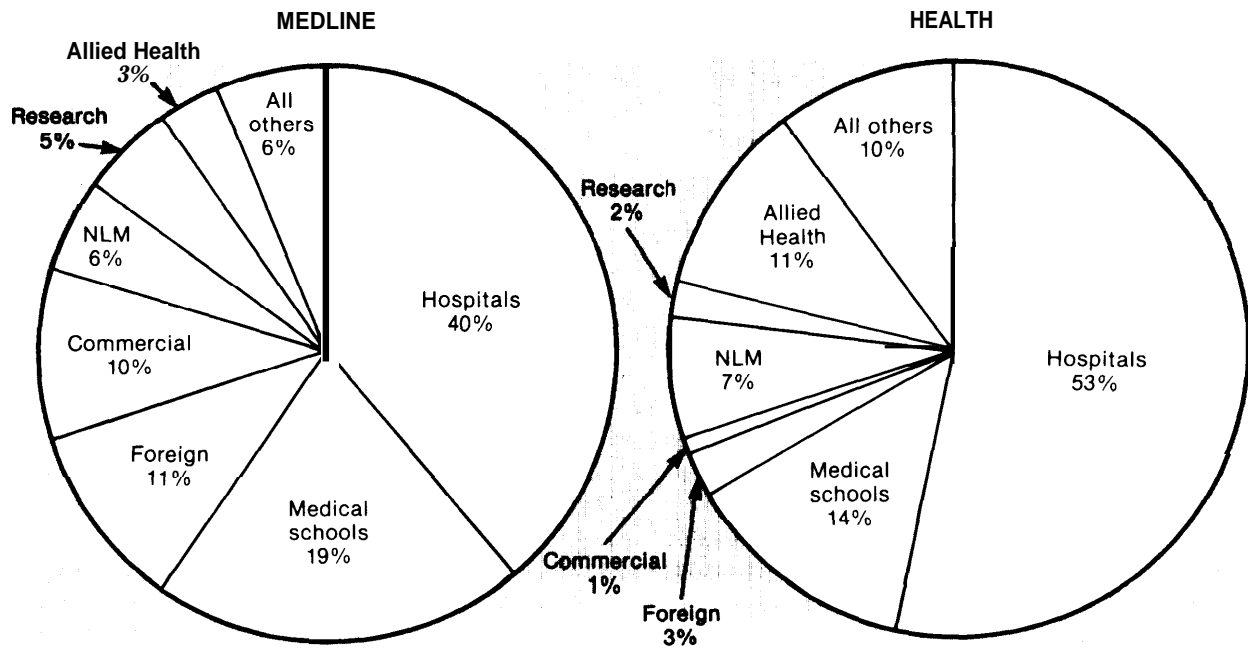
It is generally agreed that an ideal search requires that a trained searcher have a reference interview with the person requesting information

Figure 2.—MEDLARS: Percent of Searches and Connect Hours by User Categories, Fiscal Year 1981

Total searches = 2.02 million

Total connect hours = 163,000

All others 9%
Research 8%
Foreign 9%
NLM 9%
Commercial 11 "/0
Hospitals 32°10
Medical schools 21%

All others 9%
Research 8%
Foreign 9%
Commercial 12%
NLM 12%
Hospitals 34 "/0
Medical schools 17%

SOURCE: National Library of Medicine

Figure 3.—MEDLARS: Use of MEDLiNE and HEALTH On-Line Data Bases by Connect Hours by User Categories, Fiscal Year 1981

MEDLINE

HEALTH

Allied Health 3%
Research 5%
All others 6%
NLM 6%
Commercial 10%
Foreign 11%
Hospitals 40%
Medical schools 19%

Research 2%
All others 10%
Allied Health 11%
NLM 7%
Hospitals 53%
Medical schools 14%
Commercial 1%
Foreign 3%

SOURCE: National Library of Medicine

Figure 4.—MEDLARS: Number of On-Line Searches Performed, 1972-81
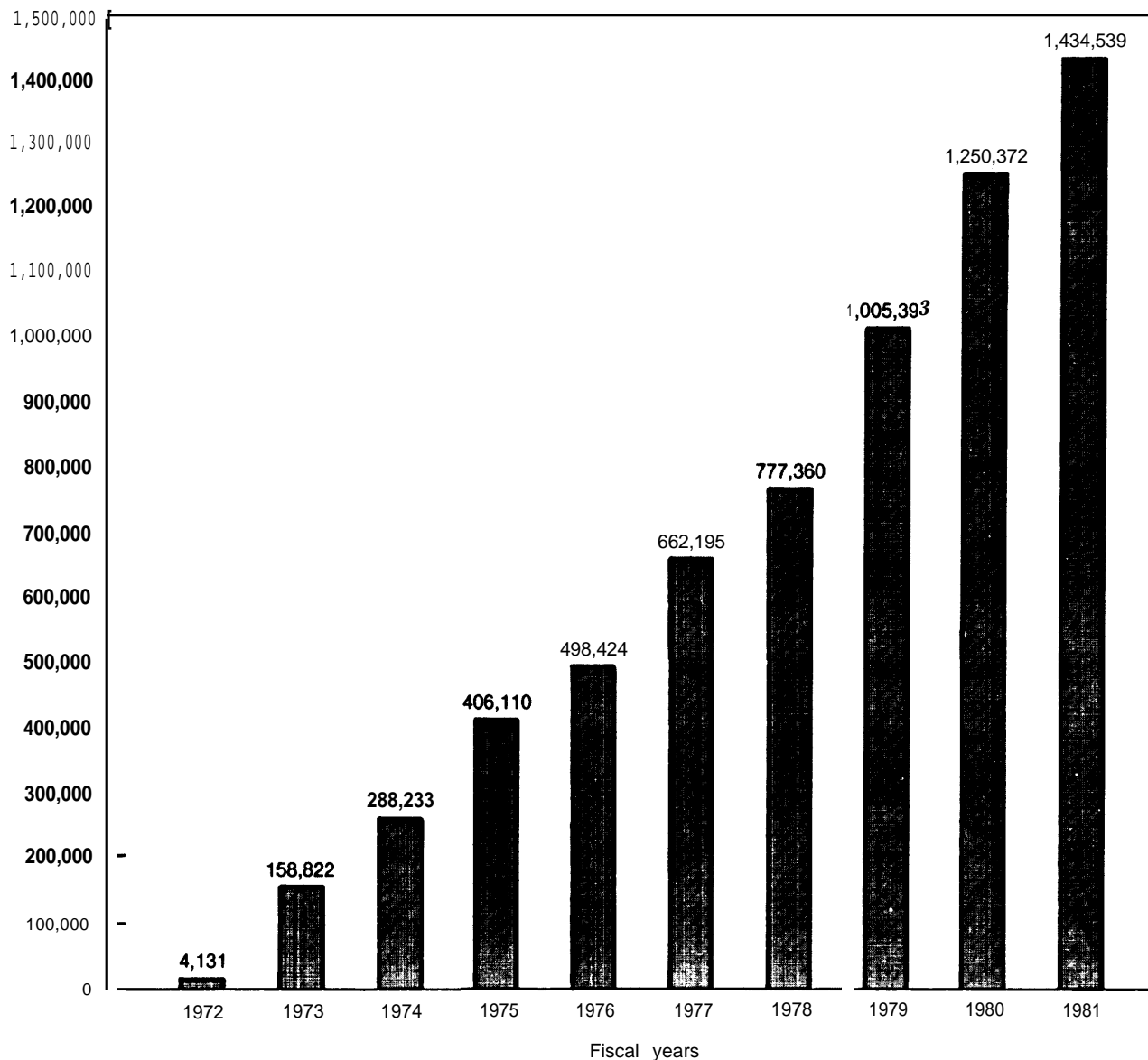
SOURCE. National Library of Medictne

to determine the purpose of the request, the sources the requester has consulted, the facets of the subject the requester wants emphasized or eliminated, new terms in the field the requester can supply, and the character and volume-of the retrieval the requester expects. Requesters are often familiar with such information from their own experience and from previous consultation with written sources. It appears that not all search requests are conducted in this manner, resulting in a wide variation in the quality of searching results.

Aggregate data on how search requests are made are not available, but there appears to be variation from library to library, mainly because of differences in populations served. For exampie, the New York Academy of Medicine Library

estimates that more than 80 percent of its searches are requested by telephone or mail. Most of its other inquiries are by telephone. The reference room of NLM estimates that about sO percent of its research requests are made in person, while the Biomedical Library at the University of California at Los Angeles, an academic health science center library, reports that over 80 percent of its requests are made in person.

User Training

NLM has developed an on-line training course in MEDLARS for information specialists who act as intermediary searchers for end users. The first courses, held in the mid-1960's, initially trained analysts in batch searching techniques during a rigorous 8-month session; this was later decreased to 2 months. With the introduction of interactive on-line searching in October 1971, the need for lengthy training decreased, and the training program was shortened to 3 weeks. The current search training program lasts only 1 week. Before attending the formal l-week course, a searcher must complete MEDLEARN, the Library's computer-assisted instruction program.

The l-week training course, normally held either at NLM or at the Regional Medical Library in Los Angeles or Omaha, is comprehensive and includes didactic instruction in system mechanics, Boolean logic, search formulation, controlled vocabulary searching, free text searching, the scope and content of the MEDLARS data bases, special system capabilities, as well as hands-on experience at the terminal. The training course is set up in a modular fashion, in order to assist searchers who need more complex and more comprehensive training. Additional training is available after the initial class, and many searchers return for a l-week advanced training course. NLM also teaches abridged update/review courses.

During fiscal year 1981, 942 search analysts were trained in 43 initial and advanced training classes at NLM, UCLA, the University of Nebraska, and in the field. This is more than triple the number (254 analysts) training in 1977.

As noted previously, more than 1,890 domestic and foreign institutions now have direct access to

MEDLARS via NLM. New institutions have been admitted to the MEDLARS network, as computer capacity permits, on a priority basis. Priority standing is given to direct patient care facilities, health professional education institutions, organizations primarily engaged in health protection activities, and Federal and State health agencies. As a condition of obtaining access to MEDLARS, an organization is required to send at least one person to attend the formal l-week initial training program. Salaries and other expenses during training must be paid by the sponsoring organization.

In the past few years, more hospitals/clinics and commercial institutions have obtained direct access to MEDLARS. As a result, the character of the student population has become more heterogeneous. The original analysts were experienced librarians from large medical schools, hospitals, research institutions, and Federal agencies, resulting in a fairly homogeneous class. Today, the varying levels of expertise represented in the classes have stimulated NLM to experiment with new training methods and curricula.

Notwithstanding the mandatory searcher training program, there is some opinion that searcher variability may well be the weakest link in the entire system. The MEDLARS system may be too complex to be understood in a l-week course by one who has little background in information systems or medicine and biology. Since there are neither entry nor exit standards for trainees, there is no control over the level of expertise of those who attend the sessions. The trainees' difficulty is mainly in the complexity of the subject, the data base content, and the thesaurus. NLM training provides searchers with highly technical skills, yet the knowledge base grows ever more complex, and thus, so must searchers' skills. In addition, once a terminal is acquired by an institution or organization, anyone can use it.

The interactive process between the searcher and requester is critical to identifying relevant citations in the data base. Inexperienced and poorly trained searchers cannot respond adequately to such a need. Searchers must meet the needs of practicing physicians who tend to use MEDLARS to obtain information related to clinical problems.

These physicians, often not accomplished researchers, are also the ones most in need of guidance to ask the right questions. The critical point is whether the searcher really understands what the user wants. Such issues become more important as NLM extends MEDLARS throughout the hospital community. At many places in the Nation, including the University of Virginia, a concerted effort has been made to upgrade the skills of librarians in community hospitals remote from major medical centers and large resource libraries, so that they may adequately conduct search interviews with physicians and forward information requests to resource libraries.

There are some observers who suggest that problems with retrieving information from the system may not be particularly attributable to the searcher, and they point to an evaluation study of MEDLARS' on-line process conducted in 1978-79 (156). That study concluded that the type of training (formal or informal) had no statistically significant effect on search performance, and suggested that in some areas improving search effectiveness may be more a matter of system redesign than of training or retraining. The study also found the predominant reason searchers elected to use one system over another was that the system they elected offered access to more of the data bases that they needed. In addition, searchers tended to prefer the system they had learned first. Searchers were also influenced by the range of a system's capabilities.

User Services

In addition to providing search training, NLM supports the users of MEDLARS in numerous ways. In order to facilitate interaction between NLM and user institutions, a professional library staff member is assigned to a user service desk at NLM to receive telephone calls and respond to problems which users are having with the system. In addition, information about the system is transmitted to enrolled users by newsletters and other forms of communication. When a new MeSH is published, and when new on-line data bases are added to MEDLARS, the regional medical libraries update the training of the searchers in their regions. Searchers also have access to the *NLM Technical Bulletin* and *Users Manual* and may contact a more experienced searcher for assistance.

On-Line Access Charges

NLM began charging users in the United States for access to its data bases in April 1973. The prices have increased a number of times over the years. As of January 1, 1982, computer connect time for all data bases, except CHEMLINE, TOXLINE, TOXBACK 65, and TOXBACK 74, is $22.00 per hour of prime time (10:00 a.m. to 5:00 p.m. EST) and $15.00 per hour of nonprime time. The price for printing a page of citations off-line is $0.15. The price of CHEMLINE is $101.00 for each hour of prime time and $94.oo for each hour of nonprime time, with a charge of $0.45 for each page printed off-line. TOXLINE and its backfiles are priced at $52.00 per hour of prime computer time and $45.00 per hour of nonprime time, with a charge of $0.35 for each page printed off-line.

CHEMLINE, TOXLINE, TOXBACK 65, and TOXBACK 74 are priced to include the royalty charges paid to the organizations that provide information for them (e.g., Chemical Abstract Service, BIOSIS, and the American Society of Hospital Pharmacists). Foreign centers establish their own rates.

NTIS serves as a collection agent for NLM. In 1980, roughly $2 million in charges were paid by domestic user institutions in on-line access charges. NTIS paid 70 percent of this sum to contractors supplying NLM with telecommunication systems and additional computer capacity; 20 percent of the charges went for billing, collection, mailing, and postage; 10 percent was returned to the Treasury.

THE FUTURE OF MEDLARS

Current Standing

The developmental efforts of NLM that led to MEDLARS II promoted the development of on-line systems by both the public and private sectors. The private sector took the opportunity and studied the NLM system: "Tests and analyses of the system and its use had a significant impact on information science researchers and helped guide their thinking along new lines" (165). NLM had acted in the traditional role of Government by demonstrating the application of novel and potential technologies that require large and long-term investments not typically assumed by private industry,

Many of the on-line information services that were formed in the wake of NLM's developmental efforts have since revised and expanded their search capabilities more rapidly than NLM, and MEDLARS II is no longer a state-of-the-art system.

In May 1980, the Director of NLM established a MEDLARS III Task Force to plan for NLM's future automation needs. The Library's computer experts are troubled that the functional life span of the MEDLARS II software is only 3 or 4 more years. All software eventually becomes "fragile" because of the continuous modifications required. The search services MEDLARS II provides are not as sophisticated as those available on other on-line systems.

More important, many of NLM's internal operations are now carried out through a combination of manual techniques and separate automated systems, and the Library is finding it difficult to continue providing quality services without enlarging its staff. Since such an increase is not likely in the immediate future, NLM hopes that more and improved automation of internal functions will prove compensatory. New techniques for computerizing internal library functions are available and in use in other libraries.

The MEDLARS III Task Force completed the conceptual framework and the functional analysis of MEDLARS III in September 1980. The systems development plan was completed in September 1981. Currently, a systems analysis is under way, with a request for a proposal to design and implement MEDLARS III expected in 1982. The new system is expected to be completed incrementally between 1983 and 1985. The system analysis and systems development plan are being prepared by NLM staff supplemented by specialists from the private sector. Current plans for the design and implementation of MEDLARS III are that both tasks be carried out by a commercial firm under contract.

MEDLARS III is a line item in the NIH/NLM budget and is expected to cost $6 million. The functional analysis component cost $1 million; the design, scheduled for 1982, is expected to cost $2 million; the design and implementation, scheduled for the year after, will be funded at $2 million; and the installation, scheduled for the last year, is funded at $1 million.

The MEDLARS III Task Force was advised by experts and consultants from both the health sciences library community and the general information systems community in both its analysis of the Library's needs and in its formulation of functions to be incorporated into MEDLARS III. The task force expects that MEDLARS III "when implemented [will] represent a significant transition from a partially automated, loosely integrated set of functions [characteristic of MEDLARS II] to a highly automated, tightly integrated system that is responsive to the needs of many" of its potential users (104). This statement reflects the pace at which technological advancements are moving in this field.

Generally, MEDLARS III is expected to improve NLM's internal and external capabilities. It will automate and improve: 1) the acquisition of biomedical literature; 2) the creation, maintenance, and distribution of bibliographic records; 3) the retrieval of bibliographic information; 4) inventory control; and 5) the delivery of primary documents (i. e., interlibrary loans). Some of these processes are now automated, whereas others are essentially manual. MEDLARS III will not perform functions that are purely commercial and available from private vendors, such as

putting the user's own private files on-line for searching.

Improvements in MEDLARS' retrieval capabilities will result from the construction of an integrated data base that permits a single search access to the bibliographic data bases currently in the system. There will be other features in MEDLARS III to assist and simplify a search analyst's efforts. It will adhere to standards for forming search strategies now being established by the Council of Library Resources' Committee on Man-Machine Interface.

MEDLARS III is designed to build on MEDLARS II, permitting the incorporation of new technology and allowing for incremental implementation. For example, NLM has a contract to explore the initial phase of on-line indexing. As soon as this indexing is achieved, it can be incorporated into the old system until MEDLARS III is ready. Much of the new system will not be at the leading edge of computer technology, but will update the current capabilities of MEDLARS II and incorporate newly established or developing technologies. Electronic mass storage and distribution of textbook material is being studied at the Lister Hill National Center for Biomedical Communications, and an experimental retrieval language, CITE (Current Information Transfer in English), which should make the system more accessible to an untrained searcher, is under development at NLM.

The plans do allow some latitude for examining new and undeveloped innovations. One innovation that could be considered for MEDLARS 111 is a natural-language approach to computers (i.e., involving the use of standard English) that was developed at Beth Israel Hospital in Boston, under a research grant from NLM. This approach has been very successful in a limited system, and NLM is considering funding a feasibility study for enlarging the system to include all the hospitals in the Boston area, hoping that the results may be applicable to MEDLARS III. The purpose of using natural language is to make the system easier for untrained people to use. A comparison of features between MEDLARS II and MEDLARS III is presented in table 5.

At the same time that MEDLARS III is evolving, NLM administrators are considering the fu-

Table 5.—Summary Comparison of the Functional Capabilities of MEDLARS II and MEDLARS III

MEDLARS II

Retrieval: Highly automated, oriented towards the technically trained search analyst.

Technical Processing—Level of Automation: Many critical functions such as serials processing, indexing, catalog maintenance, circulation, and an interlibrary loan are essentially under manual control.

Record and File Creation and Maintenance: Currently automated functions have deficiencies which include duplicate redundant record creation, maintenance, and searching; inclusion of nonauthoritative bibliographic data; and incompatibility between NLM files and with other national library data files such as the Library of Congress, Ohio College Library Center.

Card Catalog: Manually controlled; difficult to maintain up-to-date.

Network Access and Interface: On-line access to NLM records is available only for the retrieval. Only limited access to capabilities such as record creation and maintenance. No capabilities to provide for local records creation.

Regional Medical Library Document Delivery System: Manual system for processing requests, maintenance activity and administrative controls for almost 1,000,000 requests per year.

MEDLARS III

Retrieval: Enhanced user aids with natural English-like language queries will benefit not only the trained searcher, but will enable health practitioner users to access the system directly and effectively for many queries.

Technical Processing—Level of Automation: These functions will be more completely automated and integrated.

Record and File Creation and Maintenance: Will be based on the concept of a master machine-readable record for each bibliographic entity and processing unit. Record structure will be based on the Library of Congress MARC record. All bibliographic elements will comply with established standards.

Card Catalog: Fully computerized and available on-line through MEDLINE network to users all over the United States.

Network Access and Interface: Users will have access to national locator information and the retrieval system files. The system will be capable of interfacing on-line with other automated systems to provide data for support of local processing, or to permit distributed processing of NLM records in local systems.

Regional Medical Library Document Delivery System: Fully automated system for generating and routing requests. Fully computerized management control of traffic including fiscal and program control.

SOURCE: National Library of Medicine.

ture of health information transfer. The technology and the times may warrant the introduction of more distributed kinds of arrangements for the provision of data base information (see app. I), and NLM has undertaken a limited experiment with the Toxicology Data Bank to test this hypothesis (33).