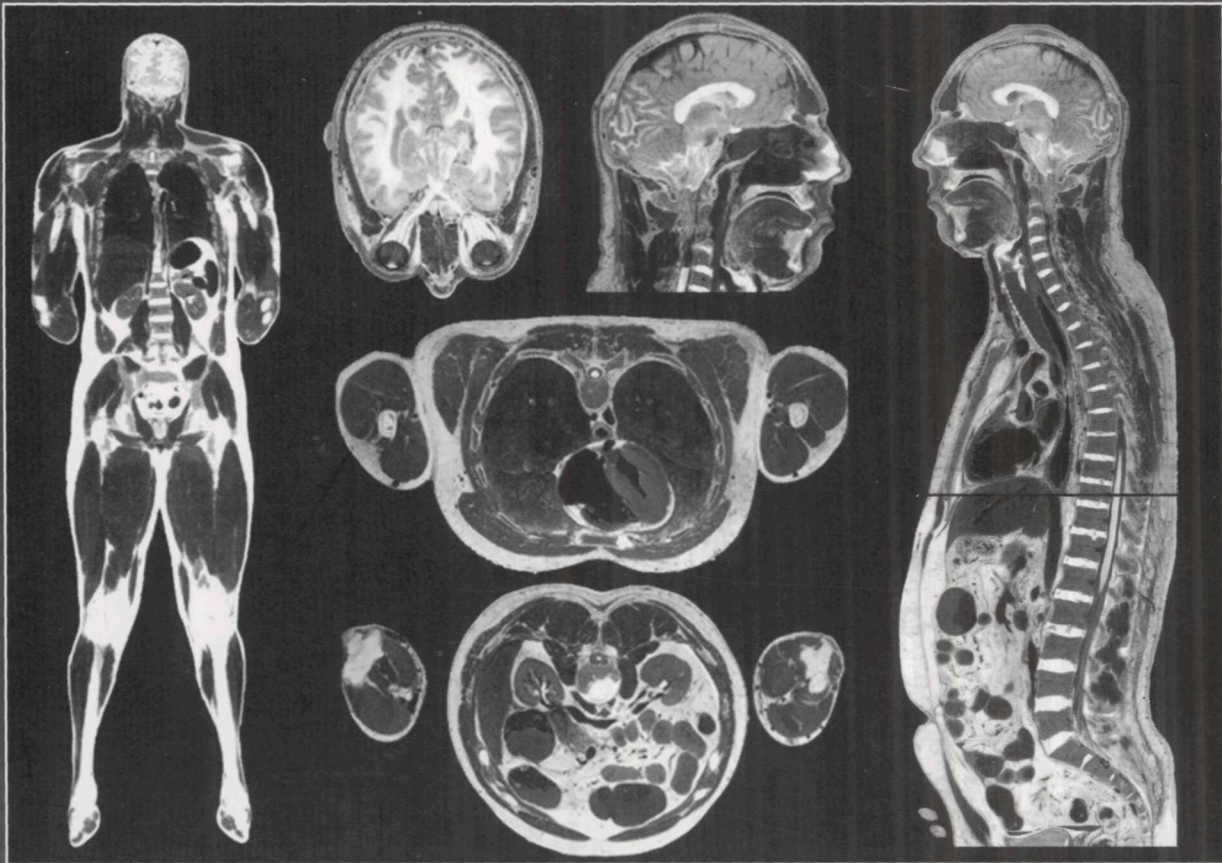


NATIONAL
INSTITUTES
OF HEALTH

NATIONAL
LIBRARY OF
MEDICINE

PROGRAMS & SERVICES • FISCAL YEAR 1994



PROPERTY OF THE
NATIONAL
LIBRARY OF
MEDICINE

DEPARTMENT OF HEALTH AND HUMAN SERVICES • PUBLIC HEALTH SERVICE

Further information about the programs described in this administrative report is available from:

*Office of Public Information
National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894
(301) 496-6308*

e-Mail: publicinfo@nlm.nih.gov

NATIONAL
INSTITUTES
OF HEALTH

NATIONAL
LIBRARY OF
MEDICINE

PROGRAMS & SERVICES • FISCAL YEAR 1994

National Library of Medicine Catalog in Publication

Z National Library of Medicine (U S)
675.M4 National Library of Medicine programs and services --
U56an 1977- -- Bethesda, Md The Library, [1978-
v ill , ports
Report covers fiscal year
Continues National Library of Medicine (U S) Programs and services Vols for
1977-78 issued as DHEW publication , no (NIH)
78-256, etc , for 1979-80 as NIH publication , no 80-256, etc
Vols for 1981-available from the National Technical Information Service,
Springfield, Va
ISSN 0163-4569 = National Library of Medicine programs and services

1 Information Services - United States - periodicals 2 Libraries, Medical -
United States - periodicals I Title II Series DHEW publication , no 80-256, etc

DISCRIMINATION PROHIBITED: Under provisions of applicable public laws enacted by Congress since 1964, no person in the United States shall, on the ground of race, color, national origin, sex, or handicap, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance. In addition, Executive Order 11141 prohibits discrimination on the basis of age by contractors and subcontractors in the performance of Federal contracts. Therefore, the National Library of Medicine must be operated in compliance with these laws and executive order.

Contents

Preface	v
High Performance Computer and Communications Program	1
Visible Human	
Awards for HPCC Medical Applications	
Unified Medical Language System	
Office of Health Information Programs Development	2
NLM Long Range Plan	
Outreach	
International Programs	
Library Operations	7
Planning and Management	
Collection Development	
Bibliographic Control	
Network Services	
Reference Services	
National Network of Libraries of Medicine	
Special Onsite Programs	
Specialized Information Services	18
TOXNET	
Databases Under ELHILL	
AIDS	
Other Programs	
Lister Hill Center	23
Computer Science Branch	
Information Technology Branch	
Communications Engineering Branch	
Educational Technology Branch	
Audiovisual Program Development Branch	
National Center for Biotechnology Information	38
Database Building	
Database Access	
Basic Research	
User Support and Outreach	
Extramural Programs	
Biotechnology Information in the Future	
Extramural Programs	42
Training	
Research	
Resource Grants	
Publication Grants	
Other Support	
Scientific Review	

Office of Computer and Communications Systems	50
Major Developmental Activities	
Programming Support Services	
Administration	54
Streamlining	
Reinvention Initiative	
Financial Resources	
Personnel	
Appointments	
Awards	
Special EEO Initiative	

Appendixes

1. Regional Medical Libraries	57
2. Board of Regents	58
3. Board of Scientific Counselors/LHC	59
4. Board of Scientific Counselors/NCBI	60
5. Biomedical Library Review Committee	61
6. Literature Selection Technical Review Committee	62

Tables

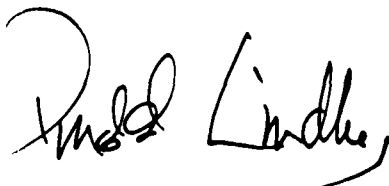
Table 1. International MEDLARS Centers	6
Table 2. Growth of Collections	13
Table 3. Acquisition Statistics	13
Table 4. Cataloging Statistics	14
Table 5. Bibliographic Statistics	14
Table 6. Circulation Statistics	14
Table 7. Online Searches	15
Table 8. Offline Searches	16
Table 9. Reference Services	16
Table 10. History of Medicine Activities	17
Table 11. Extramural Grants and Contracts Program	49
Table 12. Financial Resources and Allocations	54
Table 13. Staff	56

Preface

Fiscal year 1994 saw a number of exciting advances in the programs of the National Library of Medicine:

- In January 1994, NLM announced that its AIDS-related databases would be provided without charge to all users. Usage immediately shot up and, more important for the longer term, many community-based organizations joined the NLM online network. This action was well received both in the library community (the February 15, 1994 *Library Journal* carried a signed editorial supporting the move) and the HIV/AIDS community.
- In April 1994 the Library announced 11 awards for medically related High Performance Computing and Communications projects, which, together with one announced the previous fall, totaled some \$26 million over 3 years.
- Also in April, NLM announced the first full-text retrieval service for health services research information. "HSTAT," as it is called, contains clinical practice guidelines and NIH consensus development statements.
- In June 1994, the Library announced the award of 19 purchase orders of up to \$25,000 each to enable local community-based groups and public libraries to design their own programs for improving AIDS information access to targeted groups in their community.
- In September 1994, NLM, along with the Agency for Health Care Policy and Research, announced five new awards totaling \$1.5 million for developing computerized medical records.
- In September, HHS Secretary Donna E. Shalala named to the NLM Board of Regents three leaders in the medical community: Michael E. DeBakey, M.D., of the Baylor College of Medicine; Marion J. Ball, Ed.D., of the University of Maryland; and George H. Nolan, M.D., of the Henry Ford Hospital in Detroit. [A list of all Regents is in Appendix 2.]

I would like to thank the Library's staff for their dedication and diligence in making these advances possible. Also, much thanks go to our friends and collaborators in the National Network of Libraries of Medicine, without whose help our outreach efforts would be futile. Finally, I would like to express the appreciation of all of us at NLM for the guidance and assistance we receive from our advisors on the Board of Regents, Boards of Scientific Counselors, and other bodies.



Donald A. B. Lindberg, M.D.
Director

High Performance Computer and Communications Program

The High Performance Computing and Communications initiative continues to be the rubric under which NLM carries out several important technology-related programs. The NLM provides facilities for the HPCC National Coordination Office, of which NLM Director, Donald A.B. Lindberg, M.D., is the head. The HPCC initiative is a major interagency program that involves a number of Federal agencies, including all the major science agencies. The goal of the program is to develop computers with scalable performance up to a trillion operations per second and a digital network capable of transmitting a billion bits per second. The present Internet is the forerunner of the more extensive system of the future.

Visible Human

For the last several years, the NLM has been supporting a project to create two 3-dimensional, computer-generated human beings—male and female. The Visible Human is an example of a dataset that will require high performance computing and communications technology for effective use. The Visible Man will be introduced to the health community in November 1994; the Visible Woman will be completed in 1995. The project, which will be described in next year's report, is managed and funded by the NLM; the work is being carried out by scientists at the University of Colorado Health Sciences Center in Denver.

Awards for HPCC Medical Applications

In April 1994, the NLM funded 11 projects (under a mechanism known as the "Broad Agency Announcement") designed to help physicians practice better medicine by utilizing advanced computing and networking capabilities, including the Internet. These, and one funded in FY 1993, will amount to an expenditure of \$26 million over three years. The

projects are funding such HPCC health care applications as test bed networks to share information resources, computerized patient records and medical images, telemedicine projects to provide consultation and medical care to patients in rural areas, and advanced computer simulations of human anatomy for training via "virtual surgery." The projects will also grapple with such thorny issues as how to store, access, and transmit patient medical records while protecting their accuracy and privacy.

The development of a system of computerized medical records is seen as crucial if we are to be successful in applying the latest in computer and communications technology to medical care. In September 1994, NLM made five awards, totaling \$1.5 million, to conduct research in such areas as developing clinical medical terminology and vocabularies, automated tools to support health services research, and Internet access to multimedia electronic medical records systems. More awards are expected to be made in FY 1995. The Library is funding several of these projects jointly with the Agency for Health Care Policy and Research.

Unified Medical Language System

The UMLS is an ambitious program sponsored by the NLM in collaboration with other research groups around the country the goal of which is to help health professionals and researchers to retrieve and integrate biomedical information emanating from a variety of electronic sources. To do this, it must overcome obstacles of language (various information sources use various vocabularies), disparate telecommunications and search procedures, and wide distribution of computer-based information across the electronic landscape. The UMLS seeks to build "knowledge sources" that may be used to develop interfaces by those creating medical information systems. These knowledge sources map concepts and terms from many different biomedical vocabularies and classifications and also provide machine-processable descriptions of the contents of health-related databases. UMLS applications are now being developed and tested in a variety of environments, including linking automated patient records to related information, such as clinical practice guidelines.

Office of Health Information Programs Development

Elliot R. Siegel, Ph.D.
Associate Director

NLM Long Range Plan

The 1987 NLM Long Range Plan, along with its updates on Outreach (1989), Electronic Imaging (1990), and Toxicology and Environmental Health (1992), recognized the importance of developing a cadre of highly trained health sciences librarians, and other information professionals, to adapt modern information technologies to the needs of the biomedical community. This requirement is also a major underpinning of a key component of the Presidential initiative in High Performance Computing and Communications (HPCC), which is concerned with the training needs of individuals capable of creating and utilizing emerging computing and networking technologies in the national interest. *Platform for Change*, a recent planning document prepared by the Medical Library Association, lists areas in which health sciences librarians believe they will need proficiency in the future.

In response to these perceived needs, the NLM Board of Regents convened a Long Range Planning Panel on the Education and Training of Health Sciences Librarians. The Panel, chaired by Dr. Thomas Detre, Senior Vice Chancellor for Health Sciences, University of Pittsburgh, met three times at NLM: September 28-29 and December 8-9, 1993, and March 15-16, 1994.

The purpose of this panel, the fourth to be charged by the Board of Regents since the completion of the original Long Range Plan in 1986, was to analyze what NLM and others might do over the next 10 years to assure that our society benefits from the skills of health sciences librarians, and to help assure that persons who choose medical librarianship will be properly educated and trained, and that they have an opportunity to engage in the most important work concerning information and health care.

The final report of the Panel was reviewed by the Board for at its September 1994 meeting, and will be officially incorporated into the Long Range Plan of the Library and widely distributed.

The report states that in order for students and practicing health sciences librarians to acquire the knowledge and skills necessary to prepare them for leadership roles in the application of currently emerging information technologies to health care, graduate schools of library and information science, professional associations, health sciences librarians, and the National Library of Medicine must work together. It makes recommendations for each of these groups in the general areas of evolving roles for the health sciences librarian, professional educational, lifelong learning, and broaden-

ing recruitment into health sciences librarianship.

The recommendations include a program of NLM "challenge awards" to support planning for the implementation of specific report recommendations deemed to be of exceptionally high priority and that require further study. The essential objective of these awards is to identify specific groups and institutions that are prepared to take responsibility for a portion of the problem, to form alliances and to work collaboratively toward achievement of our common goals.

Outreach

Since FY 1990, Congress has appropriated outreach funds to NLM to support its efforts in bringing the benefits of its information systems to all American health professionals. Based on the recommendations of the Planning Panel on Outreach that was convened to formulate a plan to guide the Library's efforts, NLM developed a long-term, wide-reaching effort to make health professionals across the country more aware of the medical information resources that NLM provides, to facilitate their access to these resources, and, whenever possible, to link them to local library services. In these endeavors, NLM relies greatly upon the involvement and participation of medical librarians and information specialists at the 4,000 institutions that make up the 8 regions of the National Network of Libraries of Medicine. Much of outreach to individual health professionals is undertaken by these libraries that serve as NLM's field force. A special focus is placed on outreach to health professionals who serve minority populations.

Outreach Management and Evaluation

Because outreach is a high priority at NLM, the NLM maintains oversight through the Associate Director for Health Information Programs Development in the Office of Health Information Programs Development (OHIPD). Within OHIPD, the Office of Outreach Development has responsibility for planning, developing, coordinating and evaluating NLM's outreach programs. NLM is currently conducting an evaluation of all 200 outreach efforts involving over 500 institutions since 1989 to provide a better understanding of those strategies and approaches that have been especially effective and that can be utilized in further outreach efforts. This information will be shared with current and future collaborators who will benefit from this knowledge in the course of undertaking new outreach initiatives. An interdivisional Outreach Coordinating Committee has also been formed to foster communications among the various components of NLM that are sponsoring and supporting outreach.

Much has been accomplished with the funds that Congress has appropriated for outreach. The Library, in conjunction with the 8 Regional Medical Libraries and the 4,000 NN/LM members, has undertaken a variety of approaches to reach tens of thousands of health professionals across the country. Since our outreach efforts began, NLM

has seen a 5-fold increase in user codes issued (to almost 100,000) and a doubling of searches conducted each year (to 7 million). However, it is also clear that NLM's work is not done. There are still far too many health professionals throughout this country who are not aware that the National Library of Medicine and the NN/LM exist and work together to provide them with access to the most up-to-date medical information—without regard to where they are located or to the time of day—and to make the information readily and easily obtainable. Outreach will not be completed until every health professional in this country knows about NLM and the information resources it produces and makes available. Moreover, as NLM looks ahead to the future, it is anticipated that members of the public will demand even greater access to health-related information; NLM has already begun a modest shift towards satisfying this need in the areas of toxicology and environmental health and, most recently, in HIV/AIDS.

HIV/AIDS

Everyone agrees that to deal with the AIDS pandemic, it is necessary that the latest information be provided quickly and efficiently to those who need it. NLM and the NIH Office of AIDS Research (OAR) co-sponsored an invitational conference on June 28-30, 1993, organized by NLM/OHIPD, to examine the role that the NIH is playing in providing that information. The meeting involved broad-ranging discussions among members of the various constituent groups who need HIV/AIDS information. The conference report contains recommendations that are intended to reflect the views of the HIV/AIDS community of users.

One resultant action taken by NLM was to eliminate all online charges for searching AIDSLINE, AIDSDRUGS, AIDSTRIALS, and DIRLINE, effective January 25, 1994. This is in response to concerns voiced at the HIV/AIDS information services conference that even the existing modest fees were a financial burden that was inhibiting access to AIDS-related information.

During FY 1994, NLM began a program of small contracts to support access to HIV/AIDS electronic information resources by local community-based organizations and public libraries. Nineteen awards have been made during this first year of funding. Support was provided for a consortium-building effort in Detroit through funding of a large-scale demonstration project to link libraries (hospital, academic and public) and community groups for HIV/AIDS information sharing. A model demonstration project in Philadelphia has been funded to develop exportable training protocols and tools to instruct persons in the affected community, plus public and school librarians, to access and use electronic information resources. Projects have been developed by the Regional Medical Libraries to address special needs in their areas.

New Phase in Outreach: Federal-State Partnership

In FY 1993, meetings were held in six southern

states in an attempt to identify and establish linkages with key state government officials, health professionals, and local community organizations. Maternal and child health was the focal point. NLM sought to identify programs in the states in which access to information might become an integral part of the local programs. Resulting from this effort is a project at the Texas Department of Health which is establishing 10 outreach sites at public health facilities in south Texas. This effort focuses on making access to biomedical information easier and more readily available to a diverse group of health professionals engaged in addressing public health concerns in the Texas-Mexico border area.

Users at these sites will be connected to the Internet and trained to use Grateful Med. The integration of access to NLM's databases and resources dovetails with the installation of the new telecommunications technology to bring medical information to a part of Texas that is geographically remote and in which information resources are scarce. The Texas Department of Health Library is being aided in its effort by the medical libraries at the University of Texas Health Science Center at San Antonio, the Texas Medical Association, and the Texas Tech University Health Sciences Center.

International Programs

In FY 1994, NLM marked another year of active international programs with individual countries, international government organizations such as the World Health Organization (WHO) and the Pan American Health Organization (PAHO), and international nongovernmental organizations such as the International Council for Scientific and Technical Information (ICSTI). The Special Foreign Currency Program was active in the support of critical reviews and history of medicine projects. Other NLM international activities included training for colleagues from abroad, the NLM publication exchange program (with 169 institutions in 51 countries, including the U. S.), as well as receiving numerous professional visitors from abroad.

To assist international health professionals in accessing NLM MEDLARS databases, the Library has signed MEDLARS Center agreements with partners in 17 foreign countries and with two international organizations (Table 1).

Country-to-Country Activities and Bilateral Agreements

Former Soviet Union

With the assistance of Department of State funding, progress has been made in helping scientists in the Former Soviet Union (FSU) to obtain access to American scientific information. Since the dissolution of the USSR, each independent country of the FSU has been undergoing political and economic reforms that have had drastic consequences for local scientists. Scientists are eager to read the latest scientific papers and to develop contacts with their colleagues in the West. The depletion of foreign currency resulted in cancella-

tion of subscriptions to western scientific and technical journals in libraries; and the poor communication infrastructure has resulted in a sense of intellectual isolation among scientists. A minimum goal of this project is to improve access to U. S. biomedical information from Belarus, the Ukraine, central Asia, and the Baltic countries through their national medical libraries.

To achieve this objective, it is necessary to establish electronic mail for communication by biomedical librarians and scientists. E-mail can be used to search MEDLARS databases, exchange correspondence with American scientists, and to obtain information from other libraries. This practical concept was implemented by providing a PC system, a printer, a modem and software to the national medical libraries of Belarus, Ukraine, Kazakhstan, Krgygstan, Uzbekistan, Estonia, Latvia and Lithuania. In addition, workshops were conducted for librarians from these libraries to master the use of computing and telecommunications and to search the MEDLARS databases.

To remedy the shortage of U. S. scientific journals, two arrangements were made on behalf of these libraries. One was to pair an FSU national medical library with a medical library in the U. S. for interlibrary loans and another was to provide libraries with subscriptions of some important journals. The medical libraries of the University of Massachusetts Medical Center, University of Kansas Medical Center, University of Maryland at Baltimore, and the University of Pennsylvania agreed to provide interlibrary loans. The American Association for the Advancement of Science (AAAS) has an ongoing journal project for FSU libraries, and an agreement was reached for the AAAS journal project to include additional medical libraries supported by this NLM project.

Israel

In April 1993, the Israel MEDLARS Center was formally established at the Berman National Medical Library at the Hebrew University-Hadassah Medical Center in Jerusalem. This is the first of a new generation of foreign centers that will access NLM's databases exclusively over the Internet using the Grateful Med software. During the past year, the Center staff successfully implemented a variety of means for establishing Internet connectivity using different hardware configurations and local area network arrangements found in health science libraries located throughout Israel. This experience will be quite valuable to NLM as a prototype for other candidate centers wishing to establish comparable service arrangements. The Israel Center and its librarian staff have assumed an important leadership role in advancing the use of high speed computers and telecommunications networks as a means for obtaining access to health information resources in Israel.

Egypt

The USAID-supported feasibility study in Egypt

with the Academy of Scientific Research and Technology was terminated without an immediate commitment for the next phase. The goal of this project was to modernize medical information services in Egypt.

Other Countries

Others experimenting with the use of Grateful Med through the Internet include Jordan and Hong Kong. Jordan sent three professional staff to NLM for a week-long search training course and has obtained a block of NLM passwords for their scientists. The Chinese University in Hong Kong has demonstrated successfully from many hospitals, doctor's offices, and libraries how to use Grateful Med and the Internet to search NLM MEDLARS databases.

Activities with International Organizations

Pan American Health Organization (PAHO)

NLM has a Memorandum of Understanding with PAHO, an intergovernmental health organization. In 1989, PAHO amended its leasing agreement with NLM to provide online access to MEDLARS databases from Argentina, Chile, Jamaica, and Costa Rica. In 1990, NLM began a collaborative project with PAHO and the University of Chile to improve a gateway system named "BITNIS." The objective of the BITNIS project is to provide the MEDLINE to health professionals in all Latin American countries where the high cost of international communication services inhibits access. This new system demonstrated the capability for health professionals to conduct MEDLINE searches from Argentina, Chile, Costa Rica, Mexico, and Venezuela.

In 1992, the BITNIS gateway software was ported into a Sun Spartan Workstation which includes many new features. A Beta test was conducted from June to October 1992 by 12 participating institutions. In 1993, the software for searching the PDQ database on BITNIS was completed. As of July 1994, there are 173 codes issued to institutions in Argentina, Brazil, Chile, Colombia, Costa Rica, Ecuador, Mexico, and other countries with a daily average of 25 searches being logged onto the gateway. To execute BITNIS, a MEDLINE search command is initiated by using Grateful Med on a personal computer. Using e-mail the search commands created by Grateful Med are transmitted to NLM through over the Internet. The search results obtained from the NLM computer are transmitted back to the originator through Internet e-mail; and Grateful Med is used again on the PC to edit and present search results.

World Health Organization (WHO)

The National Library of Medicine and WHO cooperate in the publication of the *Quarterly Bibliography of Major Tropical Diseases* and the *Bibliography of Acute Diarrhoeal Diseases*. NLM prepares camera-ready copy

from the MEDLINE system, and WHO prints and distributes these to thousands of institutions in the developing countries.

NLM and WHO continued to expand their collaboration in other ways. The newly established Internet host at WHO provides a variety of information services, including NLM information resources. The medical library at the WHO headquarters in Geneva has online access to the NLM MEDLARS databases by using the Internet. It is used mostly by medical librarians for the demonstration of Grateful Med to WHO staff, although search services are also provided to health professionals on demand.

NLM and WHO also continued their collaborative interlibrary loan arrangements in which photocopies of journal articles are provided to WHO-referred requestors at a reduced rate. Library resources in developing countries are usually insufficient and the need for biomedical and health information can be met only by drawing on the collections of the developed world. Even though NLM and WHO continue to provide some photocopies of journal articles to developing countries, this arrangement can only partially meet the demand. Unless other resources in developed countries can be found, the need for interlibrary loans to developing countries will continue to grow.

Special Foreign Currency Program

Authorized under Public Law 83-480, as amended, the Library's Special Foreign Currency Program utilizes U.S.-owned local foreign currencies to prepare and publish biomedical scientific publications for the health-science community. Active since 1962, this program will terminate with the end of this fiscal year.

During FY 1994, Bernhard Naunyn's *Memories, Thoughts and Convictions*, edited by Dr. David L. Cowen, was published. It was originally published in Germany in 1925, and is considered a major medical autobiography that centers on the development of scientific medicine. In book production in India is the last publication supported by the program: Jacques Tenon's *Memoirs on Paris Hospitals*, edited by Dr. Dora B. Weiner. The book, originally published in 1788 in France, is a classic text of the French Enlightenment providing the model for the public hospital.

International Scholarships

NLM accepted the second international associate into its 1993-94 Associate Program. A medical librarian with

computer experience from the Lithuanian National Library of Medicine in Vilnius, Lithuania, was selected from a group of many qualified applicants. Also, two Fulbright Fellows and one WHO Fellow were accepted for short term, specialized training in librarianship.

International Meetings and Visitors

The Library is a member of the International Council for Scientific and Technical Information (ICSTI). This organization serves as a forum for commercial and governmental information and abstracting agencies from a number of countries. Common interests include the economics of primary and secondary publications, transborder flow of information, electronic publication, and standardization of the information needs of developing countries. At the 1994 General Assembly and Council meeting held on July 8-11, 1994 in Cambridge, UK, the NLM Deputy Director continued his term as ICSTI President, and the Associate Director for Health Information Programs Development represented membership from NLM and delivered a paper entitled "Improving Access to ICSTI Information Resources in Developing Countries: A Cooperative Project with UNESCO/ITU."

The Library continues to attract many foreign visitors, including medical librarians, health professionals, and government officials on study tours in the U. S. Many of these visitors are responsible for medical, scientific or technical information in their own countries. Their interest in NLM is more than cursory, and they are officially received and briefed on relevant aspects of NLM operations and research. Among the visitors in 1993 were: the Hubert H. Humphrey Fellows (USAID) in Public Health, local and state health officers from China, and pharmacy students from Korea. Other visitors came from the following countries:

Argentina, Australia, Belarus, Chile, China, Colombia, Croatia, Czechoslovakia, Egypt, Estonia, Germany, Greece, Grenada, Hungary, India, Israel, Japan, Kuwait, Latvia, Lithuania, Marshall Islands, Macedonia, Mexico, Moldova, Philippines, Poland, Qatar, Romania, Russia, Saudi Arabia, Singapore, Slovak Republic, South Africa, Spain, Syria, Thailand, Ukraine, United Kingdom, Venezuela.

Table 1**International MEDLARS Centers**

Tapes	Tapes/Software	Online NLM
France*	China	Australia
Germany	Sweden	PAHO*
Japan		Canada
India*		Egypt
PAHO (BIREME)*		France*
Switzerland*		India*
		Israel
		Italy
		Korea
		Kuwait
		Mexico
		South Africa
		Switzerland*
		Taiwan
		United Kingdom

* Combined online/tapes

LIBRARY OPERATIONS

Lois Ann Colaianni
Associate Director

The Library Operations Division (LO) carries out NLM's fundamental library services: building and preserving a comprehensive collection of biomedical literature; organizing this literature through indexing and cataloging; disseminating NLM's authoritative bibliographic data online, in machine-readable products, and in publications; and providing document delivery, reference, and research assistance as a back-up to services available from the other U.S. health sciences libraries. LO also coordinates the National Network of Libraries of Medicine (NN/LM), maintains an active research program in the history of medicine, conducts research and evaluation related to LO programs and services, and directs or participates actively in a number of NLM-wide initiatives. The services provided by LO form an essential foundation for NLM's Outreach Program and for other special programs including those in biotechnology, AIDS, and health services research information.

LO is the largest of NLM's Divisions, employing more than 250 librarians, library technicians, technical information specialists, subject matter experts, historians, and administrative support personnel. These staff members are organized into four main components: Bibliographic Services, Public Services, Technical Services, and History of Medicine; three smaller units: the Medical Subject Headings (MeSH) Section, the NN/LM Office, and the National Information Center on Health Services Research and Health Care Technology (NICHSR); and a small administrative staff in the Office of the Associate Director.

Planning and Management

During FY 1994, LO contributed to NLM's successful application to become a Federal "Reinvention Laboratory" and helped to develop and to oversee the NLM-wide System Reinvention project. LO staff members also participated in a number of special programs designed to increase diversity in all levels of the workforce. In keeping with the current mandate for the Federal government to do more with less, LO continued to enhance automated systems to reduce the staff time required for some tasks and to cut back on little-used products to free resources for new or enhanced customer services. Specific program adjustments are reported elsewhere in this chapter.

Collection Development

To build and maintain a comprehensive collection of biomedical literature, NLM establishes and updates literature selection policy; acquires and processes relevant biomedical literature in all formats and languages, and maintains

and preserves the materials acquired. As of September 30, 1994, the Library owned 2,103,223 printed books, journal volumes, theses, and pamphlets and 2,925,353 non-print items, including audiovisuals, computer software, microforms, prints, photographs, and manuscripts (Table 2).

Selection

LO staff members select materials for the Library's collection according to guidelines published in the Collection Development Manual of the National Library of Medicine. NLM conducts periodic collection assessment studies to determine how successfully it is applying its collection guidelines and to identify subject areas in which the guidelines may need revision or clarification. In FY 1994, NLM Library Associates conducted collection assessments of state health insurance literature and materials published in the states of the former Soviet Union. Both studies revealed some collection deficiencies that will be addressed in the coming year.

Acquisitions

In FY 1994, NLM received and processed 176,645 modern books, serial issues, audiovisuals, and software packages (Table 3). The net increase in the Library's collection was 13,613 volumes and 32,344 other items (e.g., audiovisuals, microforms, software, pictures, manuscripts). A total of 20,807 volumes was removed from the collection; most were duplicate monographs or older general reference works. Important additions to NLM's rare book collection included: three works by Paracelsus: *Archidoxorum* (Basel, 1570), *Buchlein von der Tinctura Physica* (Basel, 1570), and *Medicina Diastatica* (London, 1653); the *Promptuarium Medicinae* (Magdeburg, 1483), believed to be the first edition of the first German herbal as well as the first vernacular herbal; and *Destructio Fundamentorum Medicinæ Vopisci Fortunati Plempii...* (Rotterdam, 1657) by James Primrose, a leading opponent of William Harvey's theories. Gifts and purchases for the Library's historical image collections included 201 post-cards depicting U.S. hospitals, four Lithuanian public health posters, four Christmas Seal posters from the late 1920s which show African-American patients and physicians, and two 1956 films dealing with schizophrenia.

To conform to changes in National Institutes of Health procurement policies, the LO staff changed the method of procurement of serial subscriptions from blanket purchase arrangements to competitive contracts. This was a major effort, involving cancellation and reissue of thousands of serials subscriptions. The NLM automated systems that support acquisitions and receipts of literature continue to be modified to reduce the time required to process incoming materials. Recent improvements include automated keyword sort capability in the Master Serials System and transfer of data from the serials processing system to the cataloging system.

Collection Preservation and Maintenance

To preserve and maintain the collection, NLM binds incoming journal issues, microfilms brittle volumes, provides conservation treatment for rare and unique items, and maintains appropriate storage facilities and conditions for all types of library materials. The Library promotes the use of permanent paper in new biomedical publications and explores the use of new technology to preserve library materials.

In FY 1994, the Library bound 28,789 volumes, microfilmed 2.7 million brittle pages, and gave conservation treatment to 444 items in the special historical collections. Following the installation of additional shelving, the entire book collection was weeded for duplicates, shifted, and compacted as the Library implemented last year's decision to shelve new books added to its closed stacks by accession number rather than subject classification. The construction of an offsite vault for storage of historical audiovisuals was completed. During FY 1994, staff from LO and the NLM Office of Administration directed disaster recovery efforts which averted serious damage to the NLM collection when burst pipes caused severe flooding in LO office areas and the stacks.

Bibliographic Control

To allow users to identify relevant information contained in the world's biomedical literature, NLM maintains the Medical Subject Headings (MeSH) thesaurus for use in describing the subject content of indexed or cataloged items, develops the NLM Classification for subject arrangement of books on library shelves, and produces authoritative cataloging and indexing records for newly published or acquired items.

Thesaurus

MeSH now contains 18,022 subject headings. Its supplementary chemical file includes about 79,000 additional records for substances. In FY 1994, the LO staff added 403 new MeSH heading and 836 new entry terms, updated the terminology for 90 existing headings, and added more than 1,000 scope notes. Eighty percent of MeSH headings now have definitions. Terminology was expanded and enhanced in the fields of history of medicine, space life science, immunology, molecular biology, health services research, public health, and epidemiology. The hierarchical arrangements for neuropharmacology and dermatologic and anti-inflammatory agents were substantially reorganized and improved.

MeSH is a major component of the UMLS® Metathesaurus®, and the MeSH Section plays a key role in editing additions to the Metathesaurus. The 1994 edition of the Metathesaurus incorporated all preferred names of procedures from the International Classification of Diseases, 9th edition. Clinical Modifications; the procedure section of the

Systematized Nomenclature of Human and Veterinary Medicine. SNOMED International; the complete Universal Medical Device Nomenclature System; and a number of other significant additions. In late FY 1994, the MeSH Section issued a contract for editing support for the Metathesaurus and MeSH.

Cataloging

NLM catalogs biomedical literature both to describe what is available in the Library's own collection and to provide authoritative cataloging and name authority records for use by other health sciences libraries. In FY 1994, the Library cataloged 21,006 modern books, serials, nonprint items, and cataloging-in-publication (CIP) galleys, using a combination of inhouse staff, contracts, an interagency agreement with the Library of Congress, and assistance from the International MEDLARS Center in China. There was a net reduction in the working inventory of uncataloged books of 418 items. To make newly acquired materials available more quickly, the Library began limited cataloging of some serial publications, developed a priority scheme for nonprint materials, and improved the processing flow for CIP books.

NLM's Lister Hill Center collaborated with the History of Medicine Division to make "Images from the History of Medicine", an integrated catalog and image database covering NLM's historical picture collection, available on an Internet World Wide Web server via the Mosaic multimedia client. This is the first time that NLM's picture catalog has been generally accessible to remote users. A beta-test of a videodisc version provided useful suggestions on ways to improve the catalog records and to expand the collection. NLM will not pursue public distribution of the image catalog on videodisc because Internet provides a more flexible and powerful access mechanism.

The Cataloging Section completed work on the fifth edition of the *National Library of Medicine Classification* which is used by many health sciences libraries for the shelf arrangement of books and other types of literature. A number of librarians in the National Network of Libraries of Medicine and elsewhere assisted by providing comments on drafts of the revised Classification, which will be printed and distributed in FY 1995.

Indexing

To assist users in identifying articles on specific biomedical topics, NLM indexes nearly 3,800 journals. If indexed articles are retracted, corrected, or challenged in subsequently published commentaries, the Library also updates and annotates the indexed citations to these articles accordingly. LO staff members also index gene sequences and edit author sequence submissions for incorporation in the National Center for Biotechnology's databases (see NCBI chapter).

An NIH-chartered committee, the Literature Selection Technical Review Committee (LSTRC—see Appendix 6 for list of members), advises NLM on the journals that should be indexed in MEDLINE, Index Medicus, and other NLM databases. In FY 1994, the LSTRC reviewed 353 journal titles, rated 85 sufficiently highly for NLM to begin indexing them, and recommended that NLM cease indexing four titles. After considering recommendations prepared by professional societies in the fields of cardiology, medical records, and respiratory therapy, the LSTRC advised NLM to index nine additional titles and to stop indexing eleven titles in these subject areas.

MEDLARS indexing is done by a combination of LO staff members, commercial contractors, international MEDLARS centers, and cooperating organizations such as the American Hospital Association, the American Journal of Nursing Co., and the American Dental Association. In FY 1994, NLM added 368,000 indexed citations to MEDLINE. A larger than normal or desirable backlog of unindexed articles developed, primarily because the indexing contractors had difficulty in hiring people with suitable subject and language expertise. Seventy-six percent of the citations added to MEDLINE contained English-language abstracts. NLM updated previously indexed citations to reflect 26 retractions, 3,360 published errata, and 30,417 substantive commentaries. Indexers at several international MEDLARS centers have begun to use NLM's online indexing system via the Internet. This will simplify indexing for the international centers and should also improve the currency of citations in MEDLINE.

As part of its collaboration with NIH's Office of the Medical Applications of Research (OMAR) to improve access to information on both completed and ongoing clinical trials, NLM began tagging MEDLINE citations for retrospective articles reporting clinical trials supplied by the Baltimore Cochrane Center. The Library also studied the degree to which current clinical trials are accurately tagged by MEDLARS indexers and examined a set of clinical trial articles for which the MEDLINE citations contained no data on the sex of the human subjects involved. Both projects revealed that in general NLM's indexing accurately reflected the content of the articles.

In FY 1994, the History of Medicine Division began to use the MeSH vocabulary to index citations for the HISTLINE database. PC-based client server software is now employed to create and maintain HISTLINE records and to make changes and corrections to the MEDLINE files. The new software simplifies maintenance and also provides enhanced data validation.

As a pilot project, a publisher is supplying NLM machine-readable data in SGML (standard generalized markup language) format for ten indexed journals so that the Library may experiment with its use as an alternative to keyboarding. In an effort to identify an appropriate strategy for converting its pre-1966 indexing data to machine-readable form, NLM has arranged for keyboarding the author

portion of the 1965 *Cumulated Index Medicus* and has also enlisted the assistance of the Department of Energy to determine if available OCR technology is suitable for scanning earlier Index Medicus volumes.

Network Services

To carry out its mission to improve access to biomedical information, NLM distributes its indexing and cataloging data online, in machine-readable formats, and in publications; provides reference and research assistance to onsite and remote users; delivers documents to onsite and remote requesters who need items from the NLM collection to supplement the resources of other libraries; and directs the National Network of Libraries of Medicine (NN/LM). The primary goal of NLM's outreach initiative is to link more health professionals to these services.

Online Services

Nearly 100,000 individuals and institutions now have codes for searching the 48 online databases mounted on NLM's systems. As a result of continuing outreach efforts by NLM and its partners in the NN/LM, the Library distributed 23,000 online codes in FY 1994, more than in the first 17 years that NLM offered online searching. Many of the new codes were issued under flat-rate per code arrangements with professional associations or fixed-fee arrangements with hospitals and medical schools, which offer alternatives to the regular billing method. NLM's online users conducted about 6.9 million searches during 340,000 connect hours (tables 7 and 8). These figures do not include searches performed on the computer systems or CD-ROM products of other organizations that lease data from NLM.

Most of NLM's online users search via the PC or Macintosh versions of Grateful Med, a user-friendly micro-computer-based software package. Since Grateful Med first appeared in 1986, the National Technical Information Service has distributed 75,585 copies (62,782 PC; 12,823 Macintosh). Purchasers receive new versions of the software at no extra charge. As the number of online users increases so does the number of telephone and Internet requests for assistance from online searchers. Analysis of the questions received indicates that 25 percent relate to technical issues such as software installation, telecommunications, etc.

With its expanded appropriation for AIDS-related services, NLM began to offer free access to its AIDS databases in January 1994. Feedback from community-based AIDS organizations which participated in the NIH HIV/AIDS Information Services Conference (June 1993) indicated that NLM's modest online charges were a serious barrier for many AIDS patients and the groups who serve them. The Library offers free "AIDS only" online codes, as well as free access to AIDS files for regular code holders. AIDSLINE was enhanced by the addition of more meeting abstracts. In

response to a recommendation from the Conference, NLM has also identified a set of AIDS newsletters that will be indexed for AIDSLINE.

NLM released several new health services research databases in FY 1994: HSTAR (Health Services and Technology Assessment Research), a bibliographic database which includes relevant records from MEDLINE, HEALTH, and CATLINE as well as citations to specially indexed journals, government documents, and technical reports; HSTAT (Health Services and Technology Assessment Text), a full-text database of practice guidelines supported by the Agency for Health Care Policy and Research, the Guide to Clinical Preventative Services, and NIH Consensus Development statements and technology assessments; and HSRProj (Health Services Research Projects), a research-in-progress database including projects funded by government agencies and by private foundations. NLM and the National Aeronautics and Space Administration made substantial progress on the development of a new joint database on space medicine and life sciences which will become publicly available in FY 1995.

NLM conducted a satellite broadcast in January 1994 to update health sciences librarians on new online system features and databases. The more than 1,700 people who watched the broadcast heard about new mechanisms for obtaining offline prints via the Internet as an alternative to mailed printouts and the increasing amount of NLM information available via ftp, Gopher, and World Wide Web as well 1994 additions to MeSH and the new health services research databases. Response to the broadcast was very favorable. LO completed a major review of the online search training program and developed a plan for a transition to a more computer-based, student-centered approach to training search intermediaries. In FY 1994, NLM and the three RML online training centers conducted 89 classes for a total of 1,152 search intermediaries and other students.

Machine-readable databases

To ensure the widest possible access to its authoritative data, NLM leases its databases in machine-readable form to commercial database vendors, international MEDLARS centers, universities, and other organizations which then make the data available online or in CD-ROM products. In FY 1994, NLM distributed data from one or more of its databases to more than 100 different licensees. At the end of the year, 12 licensees were producing CD-ROM products containing data from 7 different NLM databases. When NLM implemented free online access to its AIDS databases, these files also became free to licensees. Fourteen organizations lease one or more of NLM's AIDS databases.

In April, NLM issued a revised License Agreement which altered provisions related to re-distribution of data. In addition, the Library notified licensees of changes in the foreign surcharge for use of NLM data effective October 1, 1994. These included an option for licensees with fewer than 26 concurrent users to pay a flat-rate surcharge instead of a

surcharge based on actual online usage.

The MARC (Machine Readable Cataloging) version of NLM's cataloging and name authority records is now produced directly from NLM's internal online cataloging system, rather than from the CATLINE, AVLINE, and Name Authority files. Since the internal system is more compatible with the MARC format, NLM's MARC data are now more accurately tagged. The Library is exploring more cost-effective ways to distribute its machine-readable data. In FY 1994, licensees were surveyed about their preferences for various distribution media, including 8mm and 4mm tapes and ftp. As a result of the survey and successful experiments to ftp some test MeSH data in MARC format, NLM expects to initiate some new distribution mechanisms in 1995.

Publications

As the electronic options for access to NLM's authoritative bibliographic data increase, the Library continues to review its publication program and to modify or eliminate specific publications that have outlived their usefulness. The *National Library of Medicine Current Catalog* and the *National Library of Medicine Audiovisuals Catalog* ceased publication with the 1993 annual cumulations. The widespread availability of NLM's cataloging data in bibliographic utilities and through NLM Locator sharply decreased the need for what were once essential tools for many health sciences libraries. Also in FY 1994, agencies and organizations that cooperate with NLM to produce recurring bibliographies in different subject areas were asked to re-examine the need for these publications. As a result, nine recurring bibliographies have ceased publication, and some others will be produced less frequently.

LO continues to collaborate with various NIH institutes to produce issues of Current Bibliographies in Medicine to support NIH Consensus Development Conferences or other special meetings. FY 1994 bibliographies addressed such topics as Hispanic-American health, ovarian cancer, Persian Gulf experience and health, and psychosocial aspects of AIDS. These publications are also available on NLM's Internet servers. The number of NLM publications accessible via the Internet continues to increase. The 1993 edition of the *Collection Development Manual of the National Library of Medicine*, the *List of Journals Indexed in Index Medicus* and the *List of Serials Indexed for Online Users* are now available for electronic transfer.

Reference Services

NLM provides reference service and assistance to onsite users and to remote requesters as a backup to the service available from other U.S. health sciences libraries. In FY 1994, NLM's Reference Section responded to 65,188 requests for reference assistance, from onsite and remote requesters (table 9). The number of inquiries received via the Internet is increasing.

The Library's Reference Section provided major support for the Administration's initiative to identify and review Federally supported research projects that involved radiation of human subjects. NLM has representatives on both the NIH and the PHS Human Radiation Task Forces and also assists the White House Advisory Committee on Human Radiation Experiments. Since much of the work in question was conducted from 1944 to 1965, before the advent of MEDLINE and other online databases, searching for literature resulting from Federally funded projects is difficult and time-consuming. To identify the most cost-effective approaches, NLM staff members conducted sample searches by various methods and compared the results.

NLM Locator, the online public access catalog interface for onsite and remote users, was enhanced by the addition of access to AIDS databases and the DIRLINE (Directory of Information Resources onLINE) as well as support for onsite circulation (see Document Delivery section below). The Reference Section reviewed onsite services to determine how accessible they are to users and staff members with disabilities. As a result of the review, some special equipment and software have been ordered.

Document Delivery

NLM provides a national back-up document delivery service to supplement the service available from other libraries in the NN/LM. The Library also delivers documents from its closed stacks to onsite users of its Reading Rooms (table 6). NLM's Collection Access Section received a total of 539,988 interlibrary loan and onsite requests for post-1913 documents in FY 1994. The History of Medicine Division received 11,077 remote and onsite requests for items from the historical collections. NLM received 324,670 interlibrary loan requests for post-1913 materials and filled 71 percent of them. If requests for which the requester was unwilling to pay are excluded, the fill rate was 76 percent. Seventy-nine percent of filled requests were processed within one day of receipt. During FY 1994, NLM began document delivery via Internet to libraries that have the Ariel software. Demand for this service is growing rapidly.

NLM received 87 percent of its interlibrary loan requests via DOCLINE, its automated document request and routing system. DOCLINE was enhanced to permit automated referral of requests that NLM cannot fill, which reduces the staff time required to handle such requests. A total of 2,671 libraries used DOCLINE in FY 1994, its tenth year of operation. Sixteen percent of these libraries accessed the system via the Internet. DOCLINE participants entered 2.8 million document requests and filled 87 percent of them. NLM filled an additional 7 percent for an overall fill rate of 94 percent. Individual Grateful Med users may employ the Loansome Doc feature to route requests for documents identified in MEDLINE searches to a network library that has agreed to serve them. These requests are referred on by

DOCLINE if that library is unable to fill them. In FY 1994, Grateful Med users initiated 196,795 Loansome Doc requests, a 93 percent increase from the previous year. Some institutions with fixed-fee arrangements for use of NLM's online system encourage their employees to use the Loansome Doc feature to route document requests to the institution's library.

An analysis of all DOCLINE requests filled by NLM and other NN/LM libraries in 1991 and 1992 revealed that large numbers of different journal titles and articles are requested each year and relatively few are requested many times. Of two million article requests filled by DOCLINE participants in FY 1992, three quarters were for articles that were requested only once in the entire year. One by-product of this study was the addition to NLM's automated serials system of summary statistics on DOCLINE requests for journal titles for the last several years.

Onsite users requested 215,318 documents from NLM's closed stacks in FY 1994. The Library was able to supply 86 percent of them. During FY 1994, NLM implemented an extension to NLM Locator which automates onsite document requests, provides better tracking of items provided to onsite patrons, and produces statistics on request traffic. For the first time, the Library is collecting onsite use data that are comparable to its data on the interlibrary loan workload.

National Network of Libraries of Medicine

The purpose of the NN/LM is to make current biomedical information readily available to U.S. health professionals irrespective of their geographic location. There are 4,130 Network members, including health sciences libraries of every size and type located in all parts of the country. NLM's NN/LM Office oversees and coordinates Network programs which are administered by the eight Regional Medical Libraries (RMLs). The NN/LM Office and the RMLs communicate regularly via e-mail and audio-teleconference to ensure that new and enhanced programs and services are introduced smoothly throughout the country. The RML contracts will be recompeted in FY 1995. In FY 1994, NLM staff spent considerable time and effort in examining the current program and considering how it should be refocused for the coming next five-year period.

The NN/LM program is a critical component of NLM's outreach initiative. The RMLs, as well as many individual network members, carry out specific outreach projects to underserved rural and inner-city health professionals. In FY 1994, a number of network members began special AIDS-related outreach efforts in conjunction with community-based organizations and public libraries. In the health services research area, NLM directed a special consultant project to assess the extent to which academic health services researchers currently have access to and make use of NLM and NN/LM services. The results of this project will be used to design an outreach strategy to reach this group. The

Library also engaged a consultant to help design a health services research training program for librarians in the NN/LM.

The RMLs now perform most of the exhibiting and demonstration of NLM's products and services at health professional meetings around the country. NLM staff the exhibits at the meetings in the Washington, D.C. area. In FY 1994, NLM and NN/LM services were highlighted at about 170 exhibits at national, regional, and state association meetings throughout the country. A number of these were chosen so that the new health services research databases could be demonstrated to groups for whom they are specially relevant.

Another focus of the NN/LM program is technology transfer. All of the RMLs collect and disseminate information about regional Internet service providers and sources of funding for start-up costs for Internet connections. RMLs also arrange special sessions on new information technologies, either as free-standing programs or in conjunction with other professional meetings. Some of these conferences have been co-sponsored by the Friends of the National Library of Medicine as well as other organizations. In FY 1994, some of these sessions focused on High Performance Computing and Communications, the National Information Infrastructure, and health care applications such as telemedicine.

Special Onsite Programs

In addition to reference and document delivery services, NLM offers a variety of special programs and services to those who visit the Library in Bethesda, including guided tours, briefings on NLM's operations and services, and historical exhibits and symposia. NLM also has a one-year post-master's training program for librarians with potential for substantial contributions to health sciences information services.

Public Tours and Briefings

NLM is a popular attraction for domestic and international visitors with an interest in biomedical communication, health sciences librarianship, and information technology. In FY 1994, LO staff members conducted 165 regular daily tours for a total of 559 visitors. The Office of Public Information (Office of the Director) arranged 98 special group tours and orientation programs for 1,115 visitors. NLM staff members also arranged special briefings on library programs and services for many individual visitors.

Historical Programs

In FY 1994, NLM presented a special lecture by Dr. Allen DeBus on the occasion of the 500th anniversary of the death of Paracelsus. This was part of a joint Paracelsus observance with Hahnemann Medical College and Washington University (St. Louis), which involved exhibits at all three institutions and a joint exhibit brochure. The Library held a symposium, *Medicine and Health Since World War II: Four Federal Achievements*, in honor of the 50th anniversary of

NLM's History of Medicine Division. The program featured presentations on important contributions by the Veterans Administration Tuberculosis Cooperative Study, the Food and Drug Administration, the U.S. Army Institute of Surgical Research, and the Centers for Disease Control and Prevention. In celebration of the 900th anniversary of the oldest item in its collection, an Arabic manuscript, NLM engaged Dr. Emilie Savage-Smith to assist in preparing an exhibit and an afternoon program on *Islamic Culture and the Medical Arts*. All three of these programs were supported by the Friends of the National Library of Medicine.

The Friends of the NLM also arranged for an exhibit entitled *Perez on Medicine*, featuring the artist's depiction of different medical specialties. Other exhibits included "*If You Knew the Conditions...: Health Care to Native Americans*," which traced the history of Federal involvement, and the *Art of Medicine at the 21st Century*, paintings by May Lesser. In observance of African-American History Month, NLM sponsored a lecture by Clifford Muse, Howard University Archivist, on "Howard University: 'Capstone' of Black Education" and prepared a small exhibit entitled *Howard University College of Medicine: 125 Years of Educational Excellence*.

Stephen Strickland, Ph.D., visiting historian, completed a draft of his work on the history of the Regional Medical Programs (RMPs). Members of NLM's History of Medicine Division continued their research using NLM's collections. Staff research results were published and presented at professional meetings and invited lectures throughout the year.

NLM Associate Program

The NLM Associate Program is a one-year competitive program that allows library school graduates to become familiar with NLM's operations, to gain an understanding of key issues facing health sciences libraries, to use new information technologies, and to develop their skills by conducting special projects. In addition to the collection assessments previously mentioned, FY 1994 Associates expanded and publicized NLM's minority applicant locator file, examined NTIS technical report records to determine their compatibility with NLM's bibliographic files, obtained feedback from users on NCBI information services, and compared the utility of the UMLS Metathesaurus to that of MeSH alone in automatic assignment of MeSH headings based on words in titles and abstracts. Associates also have an opportunity to visit other national libraries and various types of health sciences libraries and information centers and to attend professional meetings.

Five associates completed the 1993/94 program. The four U.S. associates accepted positions at the Region 1 Regional Medical Library, the Kevric Company in Rockville, MD, the Miami Valley Hospital in Dayton, Ohio, and Academy for Educational Development in Washington, D.C. The international associate returned to Vilnius State Library in Lithuania. Four new Associates began the program in September.

Table 2
Growth of Collections

<i>Collection</i>	<i>Previous Total (9/30/93)</i>	<i>FY 1994</i>	<i>New Total (9/30/94)</i>
<i>Book Materials</i>			
<i>Monographs:</i>			
Before 1500	576	0	576
1501-1600	5,773	7	5780
1601-1700	10,095	5	10,100
1701-1800	24,414	8	24,422
1801-1870	39,983	19	40,002
Americana	2,341	0	2,341
1870-Present*	581,989	14,650	596,639
Theses (historical)	281,794	0	281,794
Pamphlets	172,021	0	172,021
Bound serial volumes	998,464	27,040	1,025,504
Volumes withdrawn	(35,149)	(20,807)	(55,956)
Total volumes	2,082,301	20,922	2,103,223
<i>Nonbook Materials</i>			
<i>Microforms:</i>			
Total microforms	325,482	30,840	356,322
Reels of microfilm	58,484	10,198	68,682
Number of microfiche	266,998	20,642	287,640
Audiovisuals	55,736	1,490	57,226
Computer software	649	13	662
Pictures	56,600	1	56,601
Manuscripts	2,454,542	0	2,454,542
Total nonbook	2,893,009	32,344	2,925,353
Total book and nonbook	4,975,310	45,957	5,028,576

*FY 1993 figure adjusted to reflect limited cataloging materials underreported last year.

Table 3
Acquisition Statistics

<i>Acquisitions</i>	<i>FY 1992</i>	<i>FY 1993</i>	<i>FY 1994</i>
Serial titles received	21,863	22,397	23,250
Publications processed:			
Serial pieces	157,882	154,069	154,076
Other	25,753	23,682	22,569
Total	183,635	177,751	176,645
Obligations for:			
Publications	\$4,358,439	\$4,129,478	\$4,456,480
Included for rare books	(\$193,193)	(\$149,829)	(\$207,575)

Table 4
Cataloging Statistics

	<i>FY 1992</i>	<i>FY 1993</i>	<i>FY 1994</i>
Completed Cataloging	21,221	21,835	21,006

Table 5
Bibliographic Services

<i>Services</i>	<i>FY 1992</i>	<i>FY 1993</i>	<i>FY 1994</i>
Citations published in MEDLINE	401,562	376,312	367,877
For <i>Index Medicus</i>	380,485	358,993	351,958
Recurring bibliographies	23	22	12
Journals indexed for <i>Index Medicus</i>	3,048	3,058	3,127
Abstracts entered	295,803	280,599	274,514

Table 6
Circulation Statistics

<i>Activity</i>	<i>FY 1992</i>	<i>FY 1993</i>	<i>FY 1994</i>
Requests Received	528,288	522,472	539,988
Interlibrary Loan	302,271	307,481	324,670
Onsite	226,017	214,991	215,318
Requests Filled:	401,565	401,162	410,453
Interlibrary Loan	219,831	220,464	229,949
Photocopy	207,685	207,442	217,627
Original	10,726	11,493	10,864
Audiovisual	1,420	1,529	1,458
Onsite	181,734	180,698	180,504

Table 7
Online Searches

DATABASES	FY 1992	FY 1993	FY 1994
AIDSDRUGS	389	582	1,288
AIDSLINE	38,165	38,485	54,596
AIDSTRIALS	925	1,377	2,319
ALERT		1,923	2,027
AVLINE	15,435	22,298	27,269
BIOETHICS	13,536	15,450	15,075
BIOTECHSEEK	738	781	725
CANCERLIT	79,562	83,805	88,077
CATLINE	211,834	279,474	363,805
CCRIS	4,038	4,763	4,048
CHEMID	9,961	10,782	10,608
CHEMLINE	20,928	18,784	16,588
DART	4,168	3,338	3,061
DBIR	1,302	115
DENTALPROJ	205	120	146
DIRLINE	10,072	11,036	17,064
DOCUSER	10,982	13,082	14,353
EMIC		1,082	3,577
EMICBACK	3,794	4,473	2,995
ETICBACK	1,264	1,076	896
GENETOX	1,778	1,496	1,808
HEALTH	172,124	192,083	186,701
HISTLINE	5,697	4,658	4,374
HSDB	36,934	33,239	35,767
HSRPROJ	93
HSTAR		3,301	27,789
INFORM	1,202
IRIS	20,710	23,244	21,453
LOAN STATUS	3,821	7,475	14,495
MEDLINE	4,215,477	4,421,825	4,989,911
MESH VOCABULARY FILE	35,936	38,355	37,221
NAME AUTHORITY FILE	3,644	2,585	1,760
PDQ	25,748	24,342	23,794
POPLINE	18,696	17,328	24,610
REFLINE	47,347	43,301	42,518
RTECS	16,291	15,122	14,685
SDILINE	52,666	51,733	44,259
SERLINE	81,232	178,945	418,162
STORED SEARCH	127	248	336
TOXLINE	75,722	69,271	69,944
TOXLIT	17,914	12,924	11,766
TRI	30,207	25,519	21,320
TRIFACTS	392	663	577
USERS		4,581	3,912
YEAR86	4	3	2
Total	5,290,967*	5,685,067	6,625,774

* Revised figure

Table 8**Offline Searches**

<i>DATABASES</i>	<i>FY 1992</i>	<i>FY 1993</i>	<i>FY 1994</i>
AIDSLINE	1,998	1,971	1,710
AVLINE	157	82	75
BIOETHICS	23	3	5
CANCERLIT	3,408	3,967	3,177
CATLINE	565	526	504
CHEMLINE	2	0	0
DENTALPROJ		1	0
HEALTH	10,983	10,509	9,943
HISTLINE	7	4	6
MEDLINE	13,990	6,376	3,733
MESH VOCABULARY FILE	0	1	0
POPLINE	4,627	4,684	4,692
SDILINE	227,103	243,713	210,612
SERLINE	4	1	3
TOXLINE	5,612	4,334	3,557
TOXLIT	4,194	2,887	2,234
Total	272,673	279,059	240,251

Table 9**Reference Services**

<i>Activity</i>	<i>FY 1992</i>	<i>FY 1993*</i>	<i>FY 1994</i>
Reference Section:			
Offsite requests	23,037	24,015	22,706
Onsite requests	49,511	47,901	42,482
Total	72,548	71,636	65,188

*Revised figures for FY 1993

Table 10
History of Medicine Activities

<i>Activity</i>	<i>FY 1992</i>	<i>FY 1993</i>	<i>FY 1994</i>
Acquisitions:			
Books	152	115	126
Modern manuscripts	538,125	56,475*	193,725
Prints and photographs	0	0	313
Historical audiovisuals		38	61
Processing:			
Books cataloged	355	297	340
Modern manuscripts processed	44,175	31,940	0
Pictures cataloged	0	0	1
Citations indexed	5,777	4,801	3,697
Public Services:			
Reference questions answered	9,528	12,352	13,516
Onsite requests filled	6,447	5,751	7,866

* Corrected figure

Specialized Information Services

George Cosmides, Ph.D.
Acting Associate Director

Worldwide concern about hazardous chemicals and their potential health effects remained high in 1994. Despite sporadic efforts at environmental cleanup and restoration, and toxic chemical release reduction, the ubiquity of hazardous chemicals remains, for all practical purposes, undiminished, and comprehensive environmental strategies which will attack pollution and poisoning have yet to evolve. Increasing emphases on pollution prevention, risk assessment, and sustainable development will surely play roles in keeping people healthy within, and in harmony with, their environment. Another important part of the solution lies in information; generating, organizing, and disseminating it, widely and efficiently. The much touted information superhighway is becoming less hypothetical daily. "Knowledge is power" it has often been said, but knowledge comes from information, and the Specialized Information Services (SIS) Division of NLM is continuing its almost 30-year tradition of providing a variety of toxicological and environmental health information to the public.

SIS's primary program, the Toxicology and Environmental Health Information Program (TEHIP), has undertaken or completed several important projects in 1994. IRIS (Integrated Risk Information System), a U.S. Environmental Protection Agency (EPA) file and one of the TOXNET system's most heavily used files, became available for menu interface searching as part of NLM's Grateful Med software package, thus making EPA's much-sought-after risk assessment values more widely available than ever before.

TRI92 (Toxic Chemical Release Inventory - 1992) is the newest of the TRI series to join TOXNET, which now contains 6 years of data (TRI87-92) tracking environmental releases and transfers to waste sites by industrial facilities throughout the United States. TRI92 is the second year to contain source reduction and recycling data. TOXNET's CROSSFILE search capabilities make the TRI series all the more valuable by allowing the analysis of trends from year to year.

Federal and state drinking water standards and guidelines have been added to TOXNET's flagship file, the HSDB (Hazardous Substances Data Bank, thus expanding the already broad scope of this file and its 4,488 chemical records. HSDB may be the world's only peer-reviewed online encyclopedia of toxicological and environmental data.

1994 also saw the implementation of a prototype Toxicology and Environmental Health Information Server using the Mosaic Internet browser on the World Wide Web. This is an important first step in linking Internet users to sources of toxicology information not only within, but out-

side, NLM. A further capability of the server is providing more intuitive and graphical user interfaces for searching TOXNET and its files. A TEHIP Gopher is also being developed, as a component of the NLM Gopher, to provide users with various resources and documentation about NLM products and services.

TEHIP is keeping up with the spread of, and demand for, toxicological information. By taking advantage of new computer and communication technologies, it is able to provide this information in a timely fashion to a wide audience. By developing new search capabilities, TEHIP is also lessening the need of users for prerequisite search knowledge, thus putting them more directly and efficiently in touch with the data they are seeking. Finally, TEHIP is moving beyond the bounds of NLM and exploring ways to link users to relevant sources of toxicological information wherever these sources are located.

This approach to the deployment of specialized information services is also applied to other domains of information. In accordance with the recommendations of the NIH HIV/AIDS Information Services Conference, services in this area have been expanded and an outreach effort to assist community-based organizations in improving access has been initiated. Databases containing such information, e.g., DIRLINE, AIDSLINE, AIDSDRUGS, and AIDSTRIALS, are now offered to the public free-of-charge. The Division also participates in providing the PHS AIDS Clinical Trials Information Service.

TOXNET

Enhancements were made to three primary modules of the Toxicology Data Network (TOXNET): the Library's networked microprocessor system (including the Remote Data Entry File Building module), the Scientific Review Panel (SRP) module for peer reviewing records, and the Search/Retrieval module. All access to the TOXNET system was upgraded by means of high-speed modems across all terminal servers. The preliminary use of 586 microprocessors has shown that the speed of the system will double, and new high-density storage devices will increase the capacity of TOXNET to handle not only full-text files but graphics as well. Work has continued on the further enhancement and implementation of the Windows workstation for building and updating the Hazardous Substances Data Bank records, and will be expanded to other files maintained on TOXNET.

During FY 1994, one new file was added to the TOXNET system, bringing the total to 16. The additional file was the 1992 database for Toxic Chemical Release Inventory (TRI92), including pollution prevention data. The most important developments during FY 1994 have been the implementation of an experimental Mosaic Boolean search interface and a "Mosaic Graphical Map" concept search interface to TOXNET. A presentation of these prototypes was made to the NLM Board of Regents and was received with great enthusiasm and encouragement. Additional work has been

conducted to implement a prototype of a fully associative graphical user interface, using the Forms Software imbedded in Mosaic/World-Wide Web, to facilitate literature searches by TOXNET users who also need to search ELHILL databases, such as MEDLINE and TOXLINE. More work is required to complete this new approach of searching all files within the NLM.

Databases under TOXNET

The Hazardous Substances Data Bank (HSDB) continues to be the most highly used data bank on the TOXNET system, averaging over 10,000 searches each month. During FY 1994, 116 new records were created for this file, bringing the total to 4488 records. A total of 264 records were enhanced/updated and processed through peer-review. Through staff efforts, 1598 records received special enhancements, including updates for Threshold Limit Values, and were processed and released to the search system. A collaborative effort with EPA's Office of Drinking Water was undertaken to add standards and guidelines for Federal and state drinking water to HSDB records was undertaken. These values were publicly released in May 1994.

The Toxic Chemical Release Inventory (TRI) series of files now includes six online files, TRI87 through TRI92. These files remain an important resource for environmental release data and continue to attract high usage. Mandated by the Emergency Planning and Community Right-to-Know Act (Title III of the Superfund Amendments and Reauthorization Act of 1986), these EPA-sponsored databases contain environmental release data to air, water, and soil for some 330 EPA-specified chemicals. Starting with the TRI91 file and continuing with the most current TRI92 file, the reporting facilities were required to report source reduction and recycling activities, in addition to the environmental releases. These additional reporting requirements were mandated by the Pollution Prevention Act of 1990 and have considerably increased the size and complexity of the database.

In spite of the additional data, the TRI92 file was released in April 1994, the earliest date yet, and TRI93 may, if all goes well, be completed by March 1995. The early release dates have been made possible by increased TOXNET system efficiencies and improvements in data handling at the EPA. EPA is planning yet other increases to the TRI files, with expansion by another 300-plus chemicals in 1995 and the inclusion of submissions from Federal facilities and military installations. TRIFACTS, a companion file to the TRI series, supplies users with information related to the health and ecological effects, as well as on the safety and handling of the TRI chemicals. These records are supplied by EPA.

The Chemical Carcinogenesis Research Information System (CCRIS) continues to be built and maintained and made publicly accessible via TOXNET. This data bank

is supported by the National Cancer Institute (NCI) and has grown to almost 6,000 records. The chemical-specific data covers the areas of carcinogenesis, mutagenesis, tumor production and tumor inhibition.

The Integrated Risk Information System (IRIS), EPA's official health assessment file, continues to be very popular with the TOXNET user community. In recognition of the significance of this information, NLM developed Grateful Med form screens for the 6.5 version (IBM compatible) to facilitate searching by inexperienced users. Creating a search mechanism with the TOXNET Concept Menu is also under consideration.

The GENE-TOX file continues to be built and updated directly on TOXNET by EPA scientific staff. GENE-TOX contains peer-reviewed genetic toxicology (mutagenicity) studies for about 3,000 chemicals. The file is of particular interest to users in other countries. Most of the inquiries received this year came from Germany, France, Russia, and India.

The Registry of Toxic Effects of Chemical Substances (RTECS) is a data bank based upon a National Institute for Occupational Safety and Health (NIOSH) file by the same name which NLM has restructured and made available for online searching. SIS continues to add new data to this file that NIOSH makes available. Changes are expected this year in the mutation tests reported in RTECS. This year, NIOSH announced that the full RTECS will no longer be available as a printed or microfiche product, so online access via NLM will become even more important. In addition, SIS continues to assist in the location of CAS Registry Numbers for records in RTECS so that these important identifiers can be present on records to help users locate information across all databases.

The Developmental and Reproductive Toxicology (DART) database now contains over 21,000 citations from literature published since 1989 on agents that may cause birth defects. DART is funded by the Environmental Protection Agency, the National Institute for Environmental Health Sciences and the FDA's National Center for Toxicological Research. Records in DART contain bibliographic citations, abstracts (when available), Medical Subject Headings, and the names and Chemical Abstracts Services Registry Numbers (CAS RN) for all chemicals mentioned in the publications. Over half of the records are derived from MEDLINE and supplemented with additional chemical index terms. Records not found in MEDLINE, such as citations to meeting abstracts, articles from journals not indexed for MEDLINE, books, and technical reports make up the remainder of the database. This year, a significant effort was made to add relevant technical reports to DART. Records from DART are also included in the TOXLINE database and are added on a quarterly basis. DART is a continuation of the Environmental

Teratology Information Center Backfile (ETICBACK) database. ETICBACK contains almost 50,000 citations to literature published from 1950 to 1989. ETICBACK citations are also found in TOXLINE.

SIS is investigating the possible addition of the Columbia Hospital for Women's ReproTox database to TOXNET, either as a separate database or as part of HSDB. ReproTox is a factual database covering the potential reproductive and teratological effects of chemical substances and would complement the DART database. The Division is also considering an expansion of the scope of DART to include citations to literature on agents that may affect fertility, lactation, and the childhood development.

The Environmental Mutagen Information Center (EMIC) database contains citations to literature on agents that have been tested for genotoxic activity. It is produced by the Department of Energy's Oak Ridge National Laboratory and managed by NLM. EMIC has been publicly accessible since June 1993 and now contains over 8,000 citations literature published since 1991. Like DART, many of the citations for EMIC are derived from MEDLINE. All of the records contain bibliographic citations, abstracts (when available), EMIC special keywords, chemical names, and CAS RNs. ORNL is also responsible for locating and adding citations to relevant publications not found in MEDLINE.

A backfile for EMIC (EMICBACK) has been publicly available since June 1989. EMICBACK contains over 70,000 citations to the literature published from 1950 to 1991. Records are similar in format to the EMIC records, however none of the records contain abstracts. Records from EMICBACK are included in TOXLINE. Plans are under way to add the records from the new EMIC database to TOXLINE, as well. EMIC is funded by the Environmental Protection Agency and the National Institute of Environmental Health Sciences and managed by NLM.

Databases under ELHILL

ChemID (Chemical Identification File) is an NLM online chemical dictionary which contains over 267,000 records, primarily describing chemicals of biomedical and regulatory importance. ChemID allows users to search by a variety of chemical and biological identifiers. It also allows users to formulate strategies for locating and searching other MEDLARS files or external data sources, e.g., the EPA's Toxic Substances Control Act Inventory database, which may contain more information about a chemical.

ChemID contains an important set of regulatory data, collectively known as SUPERLIST. Over 9,000 records are augmented with the name and an indication of source for chemicals mentioned in one or more of 19 lists of regulatory or biomedical importance. Coverage includes such lists as the Department of Transportation Hazardous Materials List and the Priority List of the Agency for Toxic Substances and Disease Registry. These data allow users to determine if a

certain chemical is mentioned on a given list and under what name, as well as to search for chemical classes on these lists. In FY 1994, the systems used to create ChemID were reprogrammed to allow more facile file regeneration and to allow easier monitoring of data quality.

CHEMLINE (Chemical Dictionary Online) is an online chemical dictionary and directory file that allows users to identify chemical substances via nomenclature and other identifiers and to formulate optimum search strategies for other NLM files. Each chemical record has pointers to other NLM files that contain information about that chemical substance. CHEMLINE is updated every two months and regenerated annually. The basic foundation of CHEMLINE's data is supplied by the Chemical Abstracts Service from its Registry System, and this is augmented heavily by NLM with nonproprietary data from a variety of sources. CHEMLINE now contains over 1,200,000 records on chemical substances. During FY 1994, new File Locators and name data were added to CHEMLINE, including data from the new edition of the Cosmetic Toiletries and Fragrance Association Dictionary, and a pointer to HSTAR, NLM's new Health Services/Technology Assessment Research file.

TOXLINE (Toxicology Information Online) is an NLM online bibliographic retrieval service, produced by merging "toxicology" subsets from some 17 secondary sources. TOXLINE and its backfile, TOXLINE65, contain data from sources that do not require royalty charges based on usage. The Chemical Abstracts Service requires usage royalties; therefore, information from this source is used to create two additional online bibliographic files, TOXLIT and TOXLIT65. The four databases in the TOXLINE family of services now contain nearly four million records. Approximately 20,000 records are added with each monthly update.

During FY 1994, the TOXLINE files were rebuilt to add current MeSH indexing vocabulary to the subset derived from MEDLINE, the Biological Abstracts subfile, and the DART (Developmental and Reproductive Toxicology) subfile. Some other subfiles were also replaced with updated information at this time. Work was begun to change the conversion programs at NLM to accept data and format changes made by the producers of the International Pharmaceutical Abstracts subfile. Changes are also expected in the indexing vocabulary used by Biological Abstracts, producers of the BIOSIS subfile. We are exploring ways that the Unified Medical Language System (UMLS) can be utilized to accommodate vocabulary changes in TOXLINE subfiles.

DIRLINE (Directory of Information Resources Online) is an online directory of information resources including organizations, databases, electronic bulletin boards, as well as programs and projects with special biomedical subject expertise. These resources provide information that is not available from NLM's bibliographic databases. NLM continues to maintain the general biomedical subfile of

DIRLINE. Over 2,000 records have been revised with almost 100 new records added. Approximately 100 records were deleted as out of scope or no longer in existence.

DIRLINE became available free of charge in January 1994 as part of NLM's efforts to make HIV/AIDS-related information more readily available. There has been a concomitant increase in the use of the database. Further, in August 1994, DIRLINE became available through Locator, NLM's online service on the Internet. This allows those connected to the Internet to access DIRLINE without obtaining an NLM user code.

The **Directory of Biotechnology Information Resources (DBIR)** has been available online since January 1989. Originally DBIR was available as a separate database and as a subfile of DIRLINE. Currently DBIR is only available as a subfile of DIRLINE. The DBIR subfile contains over 2,000 records describing databases and other information services, organizations, collections, publications, and sanctioned nomenclature committees, all related to biotechnology and molecular biology.

AIDS

In response to the recommendations of the NIH HIV/AIDS Information Services Conference held in 1993, NLM expanded its HIV/AIDS efforts substantially in 1994. NLM continues its successful participation in the PHS AIDS Clinical Trials Information Service (ACTIS) along with the National Institute of Allergy and Infectious Diseases, the Food and Drug Administration, and the Centers for Disease Control and Prevention. The use of the two databases produced as part of this effort, AIDSTRIALS and AIDS DRUGS, increased substantially after NLM's announcement in January 1994 eliminating all charges for their use.

The *Guide to NIH HIV/AIDS Information Services* was printed and widely distributed to libraries and community-based AIDS organizations. This was done in direct response to one of the recommendations of the conference. The conference report itself, *Information Services for HIV/AIDS: Recommendations to the NIH*, was also widely distributed.

NLM initiated a solicitation for AIDS outreach purchase orders to assist community-based organizations, patient advocacy groups, and public libraries in improving access to HIV/AIDS-related information. In addition to these outreach efforts, additional funding has been provided to expand the existing Historically Black Colleges and Universities outreach program to include an AIDS-specific task. Building upon the existing model for training in accessing information related to toxicology and environmental health, health professionals will be training in accessing AIDS-related information resources. SIS participates in the NLM AIDS Planning Group which is coordinating the Library's efforts in AIDS.

Other Programs

Outreach

SIS continued its support of the Toxicology Information Outreach Project. The objective of this initiative is to strengthen the capacity of Historically Black Colleges and Universities (HBCUs) to train medical and other health professionals in the use of NLM's toxicological, environmental, occupational health and hazardous wastes information resources. In addition to providing workstations, training, and free online access to nine HBCUs participating in a pilot training development project, NLM has collaborated with the Agency for Toxic Substances and Disease Registry to train representatives from 52 additional schools in the use of NLM's valuable online resources.

During the past Fiscal Year, one of the training classes was hosted by Howard University and was coordinated by Howard's participant in a "Train-The Trainers" class designed and conducted by the Oak Ridge Institute for Science and Education (ORISE) to prepare trainees for conducting classes at their respective institutions. The three instructors for this introductory-level training class were from Howard, ORISE, and NLM. Institutional reports given at the fourth meeting of the Toxicology Information Outreach Panel indicated successful implementation of this project at each institution participating in the pilot. Encouraging reports indicate that Medical Informatics will become part of the medical school curriculum of Howard University.

User Support Activities

Development of a multimedia, microcomputer-based tutorial for the ELHILL software was completed in fall 1993. Like the tutorials for TOXLINE and MEDLINE it was also written for delivery in a Windows environment. It takes full advantage of Windows' graphical user interface incorporating color graphics, photography, animation, and audio. It is anticipated that this tutorial will be released as part of a package containing the multimedia versions of TOXLEARN, MEDTUTOR, and ELHILL LEARN.

As part of a collaborative project with NCI's International Center for Cancer Information, SIS directed the development of a microcomputer-based tutorial for the CANCERLIT database and is directing development of a similar product for the PDQ database. Both are written for delivery in a DOS environment. The CANCERLIT tutorial was completed in winter 1993 and has undergone extensive beta testing. It is to be released to the public in October 1994. PDQ is expected to be completed in spring 1995. Additionally, SIS has repackaged its DOS versions of TOXLEARN, MEDTUTOR, and ELHILL LEARN using the CANCERLIT LEARN as a model. They are to be released both on the Internet and for sale through the National Technical Information Service.

International Activities

SIS is represented on the Committee on Data for Science and Technology (CODATA) of the International Council of Scientific Unions (ICSU), as well as on its Commission on Standardized Terminology for Access to Biological Data. SIS is also represented on the Committee on Receptor Nomenclature and Drug Classification of the International Union of Pharmacology, one of the Scientific Unions of ICSU which develop guidelines and promote and provide interdisciplinary and international access to the terminology and nomenclature of biology. Support and funding to promote such international activities, which are relevant to both NLM's Medical Subject Headings (MeSH) and its Unified Medical Language System (UMLS), have been recommended by the Planning Panel on Toxicology and Environmental Health. SIS was represented at the XIIth International Congress of Pharmacology in Montreal in July 1994; participating in the General Assembly, several symposia on the nomenclature of receptors and ion channels, and a meeting of editors of pharmacology, toxicology, and medicinal chemistry journals.

Alternatives to Animal Testing

SIS continued to compile and deploy references

from the MEDLARS files that were identified as relevant to methods for reducing, refining, or replacing animals in biomedical research and toxicological testing. Requests for these quarterly bibliographies have increased, as has the number of articles deemed relevant to the field. SIS is making the bibliographies issued during the past three years available on the Internet through an NLM Gopher.

Because of the Division's commitment since 1984 to providing literature support for animal alternatives in research and testing, NLM is now represented on the Interagency Coordinating Committee for the Use of Animals in Research, which is mandated under the NIH Revitalization Act of 1993. The Committee is responsible for an annual inventory of activities relevant to the use of animals in biomedical research and toxicological testing.

Information Services to Other Agencies

SIS provided support for special or mandated information services to the following agencies: NIOSH, NCI, NIEHS, NIAID, DOE, CDC, ATSDR, FDA, and EPA. SIS also provided training to staff of these organizations in the use of information resources accessible through NLM.

LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS

Harold M. Schoolman, M.D.
Acting Director

The Lister Hill National Center for Biomedical Communications (LHNCBC) was established by a joint resolution of Congress in 1968. An intramural research and development division of the NLM, LHNCBC research programs apply state-of-the-art computer and communications technologies to the management of biomedical knowledge. Such knowledge can take the form of images, electronic signals, sounds and standard information. LHNCBC programs create innovative methods for acquiring, storing, retrieving, analyzing, communicating, and presenting the knowledge of the life sciences.

A Board of Scientific Counselors meets to review the quality and contents of the intramural research programs within the Lister Hill Center. The Board is composed of scientific and technical experts (see Appendix 3 for a list of members) who are prominent leaders in the fields of medicine, computer science, engineering, and health professions education.

The Center is organized in five branches:

- Computer Science Branch
- Information Technology Branch
- Communications Engineering Branch
- Educational Technology Branch
- Audiovisual Program Development Branch

The research and development programs of the LHNCBC fall into three categories:

- Computer and information science as applied to the problems of the Library, of biomedical research, and health care delivery;
- Biomedical image engineering, including image acquisition, processing, storage, retrieval, and communications; and
- Use of computer and image technologies for health professions education.

In FY 1994 the Lister Hill Center continued with its second group of students in the Undergraduate Research Study Program. This program provides 2-year scholarships and research experience for sophomore students majoring in electrical engineering, computer science, computer engineering, or physics at participating historically black colleges and universities. The program takes advantage of the Lister Hill

Center's staff and facilities to provide experience in biomedical image processing and computer visualization.

Computer Science Branch

Research projects of the Computer Science Branch (CSB) concentrate on the application of artificial intelligence techniques to problems in the representation, retrieval, and manipulation of biomedical knowledge. CSB projects involve both basic and applied research in such areas as expert systems, intelligent database systems, multimedia hypertext information delivery, machine learning, and machine-assisted indexing for information classification and retrieval. The research addresses issues in knowledge representation, knowledge base structure, knowledge acquisition, the validation of automated consultant systems, and the human-machine interface for complex systems. Important components of the research include multimedia knowledge-based systems with interactive video capability, and embedded intelligence systems which combine local reasoning with access to large-scale mainframe databanks.

Branch staff members participate in individual and team research projects within the branch. Several are principals in the Access Model project of NLM's System Reinvention Program and in the development of the Metathesaurus, the Semantic Network, and the Information Sources Map of NLM's Unified Medical Language System initiative. They participate actively in the medical informatics and information science research communities and professional specialty societies. And recognizing the importance of addressing the future of medical informatics by helping to train new researchers, Branch Chief Dr. Lawrence Kingsland directs the 8-week NIH "Medical Informatics" elective for third-year and fourth-year medical students each spring.

Expert Systems Program

Expert systems are computer programs combining knowledge of a particular subject area with inferencing mechanisms that enable them to use this knowledge in problem-solving situations. An artificial intelligence research program concentrating in expert systems was established at LHNCBC in 1984. The objective of the Expert Systems Program is to facilitate computer-assisted access to knowledge. The great variety and creativity of human endeavor is such that this knowledge may reside in different forms, in different places, on different media, with different structures and naming conventions.

The primary research projects of the Expert Systems Program for FY 1994 have been the NLM Access Model, the Coach Expert Search Refinement System and Coach Metathesaurus Browser, the AIRHEUM consultant system in rheumatology, the CTX "criteria engine" shell and its family of automated testing and validation tools, the medical expert systems evaluation project, and the Rheumatology Image Library videodisk and its associated browser program.

The Coach Expert Search Refinement System was developed as an in-house research project to apply the UMLS Metathesaurus and other knowledge sources to help users of NLM's Grateful Med program improve retrieval when searching MEDLINE. The system works interactively with the user, with Grateful Med, and with the ELHILL system on NLM's mainframe. Coach emulates a number of the actions of an expert human searcher in diagnosing user search problems and determining which of a series of functions to invoke for their solution. It has access to multiple knowledge sources built to help augment or replace the user's query terms or to map to new terms in helping users get more retrieval or better focused retrieval. The Coach Metathesaurus Browser provides users with multiple views of the UMLS Metathesaurus, an extremely rich and increasingly comprehensive resource. Concept definitions, semantic types, synonyms, lexical variants, related terms, child and sibling terms, co-occurring terms and other information can be displayed, helping users to visualize search terms in MeSH tree context and to choose appropriate additional terms to incorporate into their queries. The Coach Metathesaurus Browser was updated with Meta-1.4 and distributed on CD-ROM to more than 450 users as part of the 1994 release of the UMLS Knowledge Sources. It has approximately 182 Mbytes of files.

The Coach Expert Search Refinement System analyzes user searches, applying MeSH and its own special knowledge sources in performing such functions as detecting occupational specialty headings and synonyms of topical subheadings and mapping them to better terms or to conceptual clusters of subheadings. Coach knows which MeSH terms can be "exploded" to include their child terms and which are available pre-exploded by ELHILL. It spots appropriate opportunities to offer MeSH "consider also" terms and "forward see related" cross reference terms to improve a search. Coach can guide the user in applying subheading qualifiers for focusing retrieval, displaying dynamically created subheading "pick lists" specific to the MeSH heading in question. The usage of each subheading qualifier displayed in the pick list is explained. The functions created for the systems of the Coach research project and the lessons learned in their alpha- and beta-testing have been applied to a new system for intelligent Internet-based delivery of information which evolved from this work: the client software of the NLM Access Model.

The Access Model project is designed to help users find what they need from NLM's database services. The Access Model includes all the software tools and systems between an NLM user (which may be a person at a machine, or a program or intelligent agent that person has launched) and an NLM database service. An applications gateway at NLM called the Request Manager accepts incoming data streams, scans them, translates them into database-specific command sets when necessary, and routes them to appropriate servers running back-end retrieval systems with the information the user wanted. The Access Model group is building client

software that interacts with the Request Manager and an Information Sources Map to help direct users to the right database or set of databases. The Request Manager has available a Library of Intelligent Search Aids and a series of Knowledge Sources (chief among them, the UMLS Metathesaurus) to help users create good searches. A proof-of-concept Access Model prototype is now running.

The knowledge base of the AIRHEUM consultant system is being updated and its data entry process further streamlined. The system has been field tested with collaborating clinicians in family practice settings in five sites in Missouri and in internal medicine practices in five sites in Utah. Focus group discussions held last year with participating physicians provided important feedback on the way a user interaction must flow if the system is to be helpful in real-world medicine. In consultation with expert rheumatologists in Utah, NLM staff made system changes directly responsive to the clinicians' comments. The new version of AIRHEUM was retested in the Utah clinics with very positive reaction from the clinical users. This final phase of the study concluded the data acquisition portion of the expert systems evaluation project. A manuscript that pulls together the history of this multi-year project is nearly complete.

AIRHEUM is the best known of a series of knowledge-based medical consultant systems using the criteria table form of knowledge representation pioneered by NLM researchers. CTX, a multimedia expert system shell written at NLM for the development of criteria-based reasoning systems, has been tested with projects in several new subject domains. Discussions with the Agency for Health Care Policy and Research (AHCPR) explored the use of CTX to implement developmental versions of the AHCPR's Clinical Practice Guidelines. The CTX shell, with its explicit and unusual multimedia links to knowledge sources in different forms in different places, even on different machines, is designed to serve as a useful resource to developers in multiple subject areas. Several software tools written as adjuncts to CTX provide utilities assisting the developer in manipulating multi-thousand-frame videodisk image banks and in automating the performance evaluation of CTX-based consultant systems against benchmark sets of test cases. The refinement of a new version of CTX running under Microsoft Windows is now under way.

Dr. Kingsland of the Expert Systems Program served again in FY 1994 as coordinator for the 8-week NIH "Medical Informatics" elective for third-year and fourth-year medical students. The course included a seminar series of 90-minute lectures, independent research projects under the direction of NIH preceptors, and oral and written presentations of research results. Several of these extremely bright, highly motivated students have made important contributions to Expert Systems Program projects. This year, one of the students created a laboratory data reporting system that fed data to the CTX shell running as a "callable inference engine" on another network-connected machine.

Information Sources Map Project

The Information Sources Map (ISM) is a component of NLM's Unified Medical Language (UMLS) Project. Its goal is to employ the natural language processing tools put in place by the other components of the UMLS (principally the Metathesaurus and the Semantic Network) to facilitate automated identification of (and connection to) computer-based information resources appropriate to user-generated queries. The ISM was originally conceived at a time when most telecommunications involved modem-based communications between the user's terminal or PC and a large, centrally managed computer. Increasingly, the transport of computer-based information is carried out over high-speed networks. The ultimate computer network is the Internet, a hierarchical network of networks that spans the globe. The Internet, which has been growing exponentially, is currently associated with approximately 3 million computers, more than 40,000 subnetworks, and more than 20 million users.

One reason for the staggering success of the Internet is its reliance on "open standards"--mutually agreed-upon technical and social rules of behavior defined by broad consensus and made freely available, allowing individuals and institutions to develop software tools that will intercommunicate successfully. Such technical rules are sometimes referred to as protocols. One fundamental underpinning of the Internet at present is a communications protocol known as TCP/IP. TCP/IP traffic on the National Science Foundation communications backbone is currently measured in terabytes (millions of billions of bytes) per month. Another successful paradigm introduced by widespread computer networking is the "client-server" model, in which the computer user employs a program known as a client to communicate with a program on a remote computer, known as a server. The client program is often little more than a display and communications tool, with most of the information and computing being done by the remote server application providing information to the user.

Even where modems remain in use they are themselves increasingly used in the context of networks. By employing programs such as SLIP (Serial Line Internet Protocol) and PPP (Point-to-Point Protocol), modems can be used to communicate using network protocols such as TCP/IP. The central focus of the ISM has thus shifted over the past year to embrace fully the use of the Internet, client-server computing, and appropriate open standards. Within the past two years, several important new standards for sharing network-accessible information resources have appeared.

Wide-Area Information Servers (WAIS systems) are focused on the indexing of text-based documents, although they have limited multimedia capabilities. WAIS relies on a protocol known as Z39.50. Initially developed for the sharing of bibliographic data, Z39.50 has broadened in scope and can be used for a wide variety of information. It is a stateful protocol, which means that a client can have multiple sequential interactions with a server dealing with the

same issue. The server "maintains state" by remembering the user's sequence of transactions, allowing a user to accomplish such tasks as refining a bibliographic search by repetitive interactions with a remote search engine. Another type of system for online access to information from multiple providers is called Gopher. Gopher, developed at the University of Minnesota, uses hierarchical text-based menus that allow a user to browse information.

The World Wide Web (WWW) is a network-distributed multimedia hypertext system that has access to WAIS, Gopher, and more traditional Internet-based retrieval systems (telnet, ftp) through software gateways. The most successful WWW client to have appeared is NCSA Mosaic, by the National Center for Supercomputing Applications. Mosaic is an important development for at least two reasons: it is available for a high fraction of the computers in current use (IBM PC-compatibles running Microsoft Windows, Macintosh computers, and UNIX workstations), and it provides a scripting language that allows a simple customized graphical user interface (GUI) to be created, thus bypassing the considerable time and expense formerly associated with creating a cross-platform GUI. In the past year, the use of WWW has rapidly overtaken that of WAIS and Gopher.

The protocols used by gopher and WWW are "stateless." Stateless protocols are designed to accept a single request from a client (for example, a request that a specified file be sent), service that request, and then terminate communication with the client. WWW, however, does allow an information provider to develop original services that utilize the WWW communications protocol. Through these means it is possible to develop sophisticated stateful search-and-retrieval systems that can be accessed by clients such as NCSA Mosaic.

The ISM project team has developed a highly successful World Wide Web server for the NLM called HyperDOC. HyperDOC demonstrates all of the features of WWW technology including the delivery of text, images, audio, and full-motion video. HyperDOC offers access to a variety of NLM information online, as well as being the sole source for multimedia hypertext versions of present and past exhibitions from the NLM's History of Medicine Division. Use of this experimental service is growing by about 20 percent per month. At the present time, it receives connections from approximately 500 distinct sites per day, transferring 3500 files in the process. HyperDOC has been used as an internal educational aid (and thereby has contributed to the initiation of Web-based research and development efforts elsewhere within the NLM). It has been used as a demonstration platform for the outside (introducing many other components of the NIH, including NCI, to this technology). Finally, HyperDOC has served as a development platform for original technology.

To create the sort of stateful search engine required by the ISM, a prototype Web-based cataloged image retrieval system, OnLine Images (OLI) was developed. OLI/HMD, offering Internet access to nearly 60,000 images from the

History of Medicine Division, is serving 100 distinct sites per day and transferring 2000 files per day. Twenty outside sites have expressed interest in using OLI to provide network-based access to their own image archives, and the ISM team is now assembling an informal consortium of users to continue development and dissemination of OLI.

The stateful searching technology of OLI is now being applied to the more ambitious goals of the ISM. In collaboration with the NSF-funded Clearinghouse for Networked Information Discovery and Retrieval (CNIDR), a simple (stateless) proof-of-concept Web-based ISM application known as Sourcerer has been developed. A more sophisticated stateful second-generation version is currently under development. Sourcerer accepts user queries, identifies information sources likely to have relevant information, and connects to each of these sources, conducting a query and returning the findings to the user. A second related software tool, Apprentice, is in the early stages of development. It will allow information providers to register information resources with the centralized database employed by the ISM.

Dr. R.P.C. Rodgers, leader of the ISM Project, has been active in tracking relevant developments in the Internet and network-based information retrieval communities, through participation in the Internet Engineering Task Force, one of the standards-making bodies for the Internet Society, as well as the Z39.50 Implementor's Group. He organized the third meeting of the Special Interest Group for Networked Information Discovery and Retrieval, hosted by the NLM in November 1993. He was appointed as the first Chairperson of the newly formed NCSA Mosaic Federal Consortium, a coalition of federal agencies assembled by the National Science Foundation to help fund and guide further development of NCSA Mosaic. He is a founding member of the International World Wide Web Conference Committee, and is a member of the editorial and organizing committees for the Second International World Wide Web Conference to be held in Chicago in October 1994.

Machine Learning Project

The Lister Hill Center has been exploring the application of machine learning technology to biomedical problems since 1989. Machine learning encompasses a variety of mechanisms for creating computer programs that improve their performance with use. The objective of this project is to develop and apply methods by which programs can automatically acquire knowledge and put it to work.

The underlying motivation for this work arises from the explosion of available biomedical information and the less well acknowledged explosion of the analytical tools and techniques applied to that information. The National Library of Medicine has long recognized the need for automated assistance to help researchers and clinicians gain access this extremely valuable corpus of knowledge, and has supplied the community with a wide variety of databases. However, to take full advantage of the anticipated exponential growth of

biomedical data and of the increasingly evident interrelationships among previously disparate information sources, dramatic improvements in automated knowledge manipulation, analysis, and inference will be necessary.

Programs like expert systems have already moved from the manipulation of information to the manipulation of knowledge. The Machine Learning Project creates computer programs that not only manipulate knowledge, but also can acquire it themselves. Ideally, a researcher or clinician with a question should be able to have a machine learning program identify where to find relevant information; retrieve that information (possibly from multiple data sources), and analyze and assemble the information into a complete, accurate, and comprehensible representation of the desired knowledge.

Machine learning research may also help transcend the traditional computer interaction: a user issuing commands and a program responding. In a world of rapidly advancing knowledge, programs will have to do more than retrieve information when asked; they will have to manage information retrieval and inference over time. Once a user has specified a question of interest, a machine learning program should be able continuously and intelligently to track evolving knowledge sources for potentially relevant information. When the program finds relevant information it should automatically assemble it, analyze it, and send a report to the questioner.

These visions are the driving force behind the Machine Learning Project. Currently, machine learning technologies focus primarily on inducing concept definitions from externally specified datasets. To pursue the vision, the Project endeavors to advance the state of the art in machine learning, creating a computationally tractable theory of how to use diverse sources of knowledge, and deploy diverse (and complex) analytical tools in pursuit of explicitly stated goals. This approach, called knowledge acquisition planning, is in an early stage of development. Although achieving the vision described above is clearly a long term goal that will require fundamental advances in basic computer science, the process of developing the theory and implementing prototypes has already produced some useful results.

The primary testbed for research in knowledge acquisition planning at LHNBC is a program that selects and manages the use of computerized analytical tools and database searchers to achieve specific goals. This experimental program, INVESTIGATOR, operates by selecting other programs, such as statistical analysis packages and database search engines, which can be applied to achieve its human-provided knowledge acquisition goals. Each program INVESTIGATOR deploys must have an internal representation describing the preconditions for executing the program, the data formats it requires, its expected outputs, how long the program takes to run, and so forth. From this information, INVESTIGATOR's planning mechanism can select the appropriate tools and databases to accomplish a variety of tasks.

INVESTIGATOR has been programmed to use several computerized analytical tools, and to plan to acquire

knowledge in several different domains. The analytical tools include inductive category formation, heuristic Bayesian classification, various kinds of neural networks including back-propagation and conjugate gradient descent training, temporal difference learning, analysis of variance, regression and other techniques. The databases INVESTIGATOR has accessed include MEDLINE, GenInfo (the sequence database of the NCBI), the Protein Information Resource, Brookhaven National Laboratory's Protein Databank of crystallographic structure information, and others. Results of INVESTIGATOR-managed knowledge acquisition plans have addressed questions in diverse domains including exploring the causes of the disease Osteogenesis Imperfecta, identifying patterns in protein substructure, and providing an objective method for identifying the protein superfamilies and functional domains on the basis of sequence analysis. These machine learning results have led to several significant publications.

The machine learning research application to understanding the molecular mechanisms underlying the genetic bone disorder Osteogenesis Imperfecta have been particularly significant. The relationship between mutation and disease severity has been under intense scrutiny for several years. Published expert hypotheses about the relationship were contradicted by additional data, and traditional statistical and computational methods were unable to draw conclusions about the complex molecule given such a small amount of data. However, a method developed in the Machine Learning Project called the focus-induce-extract discovery strategy was used successfully to find patterns in mutation lethality in the data. These patterns are now being used to direct more detailed molecular modeling studies and may shed light on questions about the structure of collagen in addition to the disease mechanisms.

In addition to embedding existing machine learning tools into a broader framework, Machine Learning Project staff have developed new tools to address unmet needs for inferential abilities. All previous unsupervised clustering methods were limited to handling a few thousand examples. Since the biosequence databases now contain more than 60,000 sequences and are doubling rapidly, project staff developed a novel way of clustering very large numbers of protein sequences efficiently. Other active areas of research include methods for finding limited aspects of large databases that are relevant to particular problems (the "focus of attention" problem), and models of parallel, cooperating learning programs.

The general problem of selecting and coordinating diverse and complex sources of knowledge touches on many open questions in cognitive science. The only available model for designing a system that might accomplish these tasks is human beings. A significant component of the project's research is therefore the analysis of human subjects as they acquire knowledge. Machine Learning Project personnel work with computer-sophisticated biomedical researchers to gather data on how people manage knowledge acquisition

tasks. Protocols of researchers using computer tools and devising retrieval and analysis strategies have been gathered and analyzed to provide insight into this complex cognitive process. Results from these experiments have led to the identification of connections between attentional phenomena in cognitive and social psychology and computational complexity considerations in the design of machine learning systems. Potential implications of this research for the understanding of human cognitive phenomena are also being pursued. The machine learning research program places a strong emphasis on the use of cognitive models in the design of artificial intelligence systems.

MedIndEx Project

The MedIndEx Project develops and tests interactive knowledge-based systems for computer-assisted indexing of medical literature currently indexed in the MEDLINE database using terms from the Medical Subject Headings (MeSH) thesaurus. The main objective of MedIndEx is to facilitate expert indexing that goes into the MEDLINE product. Another focus of this research has been developing intelligent retrieval systems utilizing the same representations and environment of the indexing system. MedIndEx is a continuing project that has been described in previous annual reports.

Recent developments include the completion of the MedIndEx evaluation project proposal, "Evaluation of a Knowledge-Based Expert System for Subject Indexing," produced in collaboration with the Index Section of NLM's Division of Library Operations. This proposal was submitted late in 1993 for support from the NIH 1 Percent Evaluation Set-Aside Fund and approved early in 1994 as Evaluation Project NIH/NLM 94-305. The primary objectives of the project are to determine whether MedIndEx is sufficiently promising as an approach to MEDLINE indexing to warrant further development and to identify ways in which MedIndEx can be improved. The methodology is based on an earlier evaluation design project (NIH/NLM 92-308) which also received support from the NIH Set-Aside Fund.

Other recent work includes further enhancement of the indexer interface and knowledge base to more fully encompass the indexing task, improve efficiency, and improve displays and help messages; enhancing the interfaces for assignment of documents to indexers and generating indexing score reports; and updating the MedIndEx knowledge base to reflect new editions of MeSH.

Information Technology Branch

The Information Technology Branch pursues applied R&D in computer and information science with an emphasis on electronic information generation, storage, and retrieval. Major program areas at present are targeted toward the development of object oriented full text and fielded data retrieval systems for both online and CD-ROM-based appli-

cations. Areas of activity within these current programs include: development of generalized windowing interfaces across multiple platforms; object-oriented retrieval systems encompassing fielded data, full text, and graphics objects; editing workstations for manuscript preparation; computer-based publication; and CD-ROM technology. Within these activities, many areas of applied computer science must be addressed, including portability, object-oriented programming, multi-processing, client/server distributed processing models, and advanced memory management.

Full Text Retrieval

The Full Text Retrieval Program is targeted specifically to address the needs of searching, retrieving from, and updating online medical reference works. A medical reference work may contain voluminous amounts of text, structure (chapters, sections, sub-sections, etc.), and a variety of objects in addition to standard text, such as table of contents, figures, tables, and footnotes. The initial full text retrieval system developed by the Branch, IRx (Information Retrieval Experiment 1), allowed full text retrieval and maintenance of linear, nonstructured, text. IRx has been extensively employed to deliver online full text medical reference material to an international biomedical community by Johns Hopkins University Medical School, the NCBI, and others.

Present R&D efforts are targeted toward the evolution and further development of an object-oriented full text retrieval system, FTRS, designed to address the needs of more general medical reference works. Version 1 of FTRS became operational in 1993. It provided a client/server architecture and many advanced features including Table of Contents browsing and Natural Language Query searching. FTRS 1 employed an object-oriented encapsulation of the IRx search engine. FTRS 2.0 became operational in 1994; it is no longer dependent on any IRx code and provides a complete object-oriented programming structure for modular development. Major user-oriented enhancements in both the FTRS server and clients were achieved including phrase searching, an improved ranking algorithm, online comments and annotations, and Internet as well as dial-up access. In addition, two new prototype FTRS clients were completed: a prototype full windows-based client, FTWIN, and a FTRS/Mosaic client. The former, developed under MS-DOS Windows will be ported to other personal computers and workstations such as Macintosh and Sun. The latter will allow Mosaic users to have access to full text databases with the more advanced full text search and structured search features of FTRS.

The development of tools and methodologies for preparing text for online delivery has, of necessity, proceeded in concert with the development of FTRS. The area of text preparation is much broader than just preparation for use by FTRS, alone. The more general problem is one of preparing and maintaining text in a form from which both traditional hardcopy and online publications can be derived without highly redundant manual labor and/or processing. The prob-

lem is compounded by the fact that different online access methods most often require different formats and/or tagging for the same full text.

In May 1994, NLM announced public online access to the Health Services/Technology Assessment Text databases (HSTAT). HSTAT initially included full text from three sources: (1) AHCPR Guidelines, (2) NIH Consensus Development Conference and Technology Assessment Reports, and (3) the PHS monograph, "Guide to Clinical Preventive Medicine." Addition of full text from a fourth source, Aids Treatment Information System, is in progress. Online access to HSTAT is available via standard modem dial-up and Internet. Internet access is available via FTRS, WWW clients (in particular, Mosaic), Gopher, and FTP. The Branch is supporting HSTAT and gaining invaluable feedback from the biomedical community regarding needs and shortcomings of online full text delivery from the user perspective.

CD-ROM Program

NLM has a growing need to disseminate large databases that include not only full-text and fielded data but data such as digitized images or audio in a number of program areas including UMLS, NCBI, and SIS across multiple platforms including MS-DOS, Macintosh, and Sun-Unix. CD-ROM is a uniquely appropriate storage and dissemination medium for such information.

The Information Technology Branch's CD-ROM pre-mastering and simulation laboratory, established in 1991, supports pre-mastering and full simulation and testing facilities for DOS and Macintosh operating systems as well as pre-mastering and emulation capabilities for Sun-UNIX and Windows. The purpose of the laboratory is to: (1) provide in-house expertise in CD-ROM technology; (2) address the constraints imposed by CD-ROM technology; (3) advise NLM program areas on the best approach for individual problems; (4) assist in optimizing chosen implementations; and (5) address research issues and problems in the effective use of this technology. Primary research issues include identifying alternative storage organizations on CD-ROM and caching for purposes of optimization.

The research activities of the CD-ROM laboratory continue to focus on the design and implementation of FDRS (Fielded Data Retrieval System), a prototype, object-oriented fielded data search and retrieval system. This system will not only provide CD-ROM access to a variety of databases but will also serve as an experimental workbench for testing the suitability of different indexing and search options for CD-ROM distribution. Planned FDRS system development includes incremental updates on magnetic disk, and an integration with FTRS.

Communications Engineering Branch

Projects in the Communications Engineering Branch focus on R&D in image engineering: the capture, storage,

processing, online retrieval, transmission, and display of both biomedical documents (mainly journals) and medical imagery. Data types of interest include bitmapped bitonal document images, digitized color documents, digitized x-rays, and motion video. Areas of active investigation center on image compression, image enhancement, image understanding, pseudo-grayscale rendition, image transmission and networks, omnifont text recognition, and man-machine interface design. The underlying motivation for this applied R&D is the set of NLM's mission-critical tasks such as document delivery, archiving, and preservation, as well as future imaging applications in support of medical educational packages employing digitized radiographic, dermatological, and other imagery.

DocView

The goal of the DocView program is to investigate the role of advanced prototypes in providing access to, and delivery of, bitmapped images of biomedical documents over the Internet. The prototype DocView system is being developed to allow remote end users access to document images over Internet. Not only does DocView overcome some of the inherent limitations of fax in terms of image resolution and transmission speed, it represents a state-of-the-art approach to empowering an end user to seek and use electronic documents stored remotely. The components of the prototype DocView system are a Unix-based server holding bitmapped document images, and a PC-based Windows client for the end user. The first version of the user workstation employed commercial hardware, a Kofax image processing board, and some commercial software modules, but currently an effort is under way to develop a version independent of commercial products entirely, so that all of DocView's image processing functions are implemented by in-house-developed software.

The client software running under Windows 3.1 on a 486 platform allows a user to select a server located anywhere on the Internet, the current options being a Sun 690MP in the lab, a 3-processor C3830 Convex supercomputer on the NIH campus and a Sparcstation across the country at the University of Arizona. After connecting to any one of these servers, the user receives information on the number of documents available and document citations. The user may key on a citation to retrieve a document, and choose to preview just the first page of the desired document on the screen before deciding to download the entire document. If after previewing, the document is not of interest, the user may delete it by selecting a *trashcan* icon. The user may navigate through a document selecting pages of interest by invoking an *electronic bookmark* feature, rotate pages to view those in landscape mode more conveniently, zoom in for better legibility, copy sections of interest into a *notebook*, and add text from a wordprocessing program, and either print or store electronically items of interest.

While a DocView user may access and retrieve document images from a store of document images at a Unix

server, either in-house-developed as well as public domain (such as World Wide Web, FTP and Gopher servers), document images can also be received from Ariel. Ariel is a commercial system that, in a manner analogous to conventional fax transmission over telephone lines, transmits document images over Internet, and is increasingly being used by medical and other libraries. In the situation where a WWW server sends document images through the Mosaic user software, DocView is designed to serve as a viewer.

Following development, DocView is to be alpha- and beta-tested to investigate performance and design issues. This will evaluate ease of use and utility of image manipulation functions. Also an evaluation of both client and server performance will be done in terms of document transfer rates from jukebox to server, error rates, and delivery throughput over the Internet. The DocView system is described in the literature: Walker FL, Thoma GR. Access to document images over the Internet. *Proc. 9th IOLS meeting*. Medford NJ: Learned Information, 1994; 185-97.

System for Automated Interlibrary Loan (SAIL) Program

This program seeks to investigate the technical feasibility and role of automated document delivery to meet the requirements of the NLM's interlibrary loan (ILL) service. This program is motivated by the increasing burden faced by the Library in servicing the interlibrary loan requests in the traditional manual way. The research staff designed and built a prototype system consisting of a networked complex of PC-based workstations. This system automatically retrieved ILL requests from the DOCLINE request routing package in the NLM mainframe computer, parsed these, and used fielded data contained in them to retrieve document images from optical disks to be automatically faxed or printed for mailing. Operators used document capture systems developed in-house to scan and store about 200,000 pages of biomedical journals selected according to criteria that predicted high use. Design considerations appear in: Thoma GR, Walker FL: *Essential Functions in an Electronic Document Delivery System*. Chapter in: *High-Performance Medical Libraries, Advances in Information Management for the Virtual Era* (N. Broering, ed.), Meckler Publishing, Westport CT, 1993, pp. 77-88.

The system is operated in a pilot test mode to investigate performance and cost issues. Cost on a unit basis (per article delivered) turned out to be comparable to the cost of delivering documents the conventional (manual) way, even considering the disparity in volume (SAIL handled 5 percent of the total ILL load). The prime component in the cost figure is the labor necessary to convert paper documents to bitmapped electronic images. In terms of performance, delivery is in minutes and hours rather than days or weeks, but there is variation due to ambiguities in the requests, or not having the disk containing the requested article currently mounted in a drive. Other problems were identified in operations, hardware, software, and indexing inconsistencies. Examples of

solutions developed to address these problems: an artificial neural network solution to the *NLM Do Not Fill* problem; page orientation detection algorithm (forwarded to NIH to seek patent protection); skew angle detection algorithm based on preprocessing the bitmapped data followed by the Hough Transform.

The pilot operation revealed that the 64 titles preselected for this pilot project delivered 5 percent of the total ILL requests to the NLM, proportionately a high figure considering the size of the journal collection at the Library. However, of the articles stored, only one-third were accessed to serve the ILL service. This has motivated a second look at the way articles are entered into the system, prompting an investigation of a point-of-request or delivery-on-demand system. The design of such a system is based on a 486 platform running under Microsoft Windows 3.1. The idea is that the only human operation involved should be one that cannot readily be automated, viz., scanning the requested document. All other operations, e.g., faxing, printing, transmitting over Internet, retrieving ILL requests from DOCLINE, extracting information from the ILL requests, and updating DOCLINE as to status, are to occur automatically in the multitasking environment of Windows. The design of this integrated system is under way.

A number of engineering studies are being pursued in support of SAIL development and the library's ILL activity. Among these are the following:

Artificial Neural Network (ANN) to Reduce Ambiguities. Work begun in previous years on the application of ANN to classification problems encountered in SAIL operation continued. One such problem, centered on the ambiguities in ILL requests resulting from remarks made by users in the unstructured comments field in the requests. Such ambiguities caused SAIL to automatically fill requests that the requesters did not want filled. As described in last year's report, a subsystem consisting of a parser and an ANN of the back error propagation type was developed. The conclusion was that 87% of the ILL requests which have unstructured comments can be handled automatically, and that the remaining 13% may be referred to a human operator for a decision.

Artificial Neural Network for Journal Identification. Image analysis is the subject of another ANN project intended to aid scanning operators by automatically identifying a journal, and thereby reduce operator error in the selection of journal titles while scanning. This is accomplished by processing the image characteristics of a journal's cover page. In one approach the histogram of row and column black pixel counts is used as the input signature to an ANN. In the second approach, the black pixel distribution is initially processed by a Fast Fourier Transform whose first 35 coefficients serve as the input. The first approach was found to be successful in correctly classifying 70 of 75 different journals, but was slow to train. The second approach correctly classified 66 of 75 journals but is faster to train. Most of the errors

arose because of journals whose cover pages change style with time. This research is reported in: Hauser SH, Cookson TJ, Thoma GR: Using Back Error Propagation Networks for Automatic Document Image Classification. *Proc. SPIE Conference on Intelligent Information Systems*, 1993, Vol. 1965, pp. 142-50.

Simulation Studies. To predict a migration path for a scaled-up SAIL system, a discrete event simulation language, GPSS/H, is being used to model the image retrieval subsystem. This model allows a representation of varying numbers of fax servers, optical disk drives, magnetic disk drives and jukeboxes. Independent variables are the rate of ILL requests, the fraction of requests that are for fax service, and the distribution of requests over the optical platter set. The model will enable the testing of strategies on how the articles should be distributed over magnetic and optical media; for example, older articles could be on optical disks and more recent ones on magnetic disks. The results of the simulation will establish theoretical bounds on the number of system components and the overall system architecture for different levels of service. Studies on jukebox performance have been done and appear in the literature: Hauser SH, Roy G, Thoma GR. Optical disk jukebox performance in multi-user applications. *Proc. 1994 Optical Data Storage Topical Meeting*; Vol. 10: pp. 53-5.

Automated Portrait/Landscape Mode Detection. As part of research into automated document imaging, an algorithm was developed to detect the orientation (portrait vs. landscape) of a binary page image. This work, the subject of a patent filing, was described in last year's report.

Automated Document Skew Detection. Rescanning of documents is a time consuming and costly step, but often necessary in document conversion. One reason for rescanning is skew in the scanned image. A multistage technique was designed to automatically detect page skew. The principal elements of this algorithm are component labeling, a procedure to reduce the amount of data to be processed, a technique to minimize the effect of nontextual data (graphics, forms, line art, large fonts and dithered images), and the Hough transform. The algorithm is characterized by the following: (1) it uses the bottom part ("feet") of the objects (characters); (2) the data to be processed is reduced by a factor of 15 for a typical page of text, and more than 80 for a compound page; (3) the detection process can be running while a page is scanned; and (4) it is independent of text dominance. The algorithm was tested with several hundred images of medical journal pages, and found to detect skew with an accuracy of about 0.5 degrees. This work appears in the literature: (1) Le DX, Thoma GR: Document Skew Angle Detection Algorithm. *Proc. SPIE Visual Information Processing II*, 1993, Vol. 1961, pp. 251-62. (2) Le DX, Thoma GR, Wechsler H. Document classification using connectionist models. *Proc. 1994 IEEE International Conference on Neural Networks*.

Orlando FL: June 1994, vol. 5: pp. 3009-14.

Digital X-ray Prototype Network (DXPNET) Program

This program whose acronym also stands for Digital X-ray Prototype workstations linked via InterNET, aims to investigate the technical feasibility and design issues in developing, maintaining, and operating an archive of digitized radiographs. It is a collaborative program in which the Communications Engineering Branch, on behalf of NLM, serves as Technical Manager. The other participants are the National Center for Health Statistics (NCHS) and the National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS). The impetus for the program is to support the National Health and Nutrition Examination Surveys (NHANES) which NCHS periodically conducts to produce statistics on the health status of the U.S. population. One element of the collected data consists of radiographs, 17,000 from a survey already completed and an expected additional 10,000 from a current survey. The radiographs are of cervical and lumbar spine, and hands and knees.

The problems that DXPNET addresses are: preservation of the radiographic collection, extracting information from this collection, and providing remotely located users with access.

Specific program objectives include: (1) develop an Electronic X-ray Archive (EXA) for digitized NHANES x-rays implemented via an optical disk jukebox controlled by a UNIX workstation interfaced to the Internet; (2) develop image retrieval and display workstations, also called Standardized Readings Workstations, suitable for radiologists to remotely access the images from the archive, retrieve, display and manipulate them for reading; and (3) evaluate technical aspects of the design including compression technique, response time, and use patterns.

At the inception of this program, as described in last year's report, the Communications Engineering Branch developed and tested an affordable prototype PC-based workstation that enables technicians from NCHS to perform quality control on the digital images produced by scanning the x-rays. In addition, the Branch provided direct hardware and software system integration support to create a Sun-based QC workstation. This workstation runs Imview, an in-house-developed software package.

The next steps under way are: to develop the Electronic X-ray Archive based on a recently acquired 144-disk robotic optical disk jukebox, to develop a pair of Unix-based workstations that will allow NCHS radiologists to create standardized readings by viewing the x-ray images on a high resolution (2048 x 2560 pixel) monitor, to develop the capability of linking the workstations over Internet, and to deploy one of the two workstations at NCHS to allow image retrieval from the archive located at NLM.

Current investigations focus on a new method for faster transmission over the Internet using a multisoocket transmission concept. This technique relies on the multitasking

feature of Unix that allows segments of a large image to be sent concurrently over multiple socket pairs rather than using one socket pair for the entire image. We have developed prototype multisoocket software, and began performance tests to determine the optimum number of sockets.

Recent papers on the DXPNET activity include: (1) Thoma GR, Long LR, Berman LE. Design issues for a digital x-ray archive over Internet. In: Niblack W, Jain RC, eds. Proc. SPIE: Storage and Retrieval for Image and Video Databases II. San Jose CA: 1994; vol. 2185, 129-38. (2) Long LR, Berman LE, Thoma GR. Design considerations for wide area distribution of digital x-ray images. In: Jost RG, ed. Proc. SPIE Medical Imaging '93: PACS Design and Evaluation. Newport Beach CA: 1993; vol. 1989, 383-94.

Image Compression Studies. Since image compression is needed to maximize storage capacity and minimize transmission time, key to making such projects as DXPNET practical, compression studies continue in-house and with a collaboratory of external investigators. As described in last year's annual report, the participants in this collaboratory are researchers at Stanford University, Monash University (Australia), the Canada-France-Hawaii Telescope Consortium, NASA Ames Research Center, IBM Almaden Research Center, and NIH.

JPEG Evaluation Tool (JET). The degree of compression attained in lossy JPEG compression technique largely depends on the specific quantization table used. JET was developed by Branch researchers to study the effects of this selection on the radiographs in the DXPNET collection. The description and use of this tool were presented at the 1993 International Symposium of the Society for Information Display, and appears in the proceedings: Berman LE, Nouri B, Roy G, Neve L: Interactive Selection of JPEG Quantization Tables for Digital Xray Image Compression.

Machine-Readable Archives in Biomedicine (MRAB) Program

While bitmapped images are essential for the implementation of the ongoing document delivery and other projects, it is recognized that machine-readable text coupled with graphics also has a role in several NLM missions such as indexing and cataloging, as well as in efficient document delivery. Converting bitmapped document images to machine-readable ASCII characters involves optical character recognition (OCR). However, the reliability of OCR is compromised by noise in the image and the presence of unwanted page edge effects. As described in last year's annual report, the near term goal of this program is to conduct research into techniques to enhance bitmapped images to a quality level commensurate with OCR requirements by reducing noise in the bitmapped images and to automatically remove unwanted borders (page edge effects). The long term goal is to develop a prototype system that will implement the conversion from

paper scanned by a bitmapping engine to a database of segmented text/image files.

Biomedical Digital Image Processing (BDIP) Program

Biomedical images include monochrome and color still images as well as motion video. The information content of these images is a key to understanding the disease conditions in many medical disciplines. A prerequisite for "understanding" this content, or its extraction or analysis, is a set of image processing activities, including: image capture, segmentation, compression, image manipulation, image file format conversion, and related areas. This requires the digital capture of images at full 24 bits/pixel color and the subsequent processing to pare the portions of the image down to 8 bits/pixel and 1 bit/pixel in those areas where there is gray material and two-tone text respectively. It also requires the digitization of video and providing access to selected sequences through textual databases. The fundamental goal of this program is to further the existing in-house capability to acquire, compress, retrieve, manipulate, segment, analyze, display and transmit digitized biomedical images, both still and moving image sequences.

Recent work includes research in automated extraction and classification of features in digital x-ray images. This work is motivated by the possibility of using features (e.g., texture, shapes, edges) as descriptive elements in images for image retrieval. The principal tool used is mathematical morphology. The approach is to determine if texture or shape can be used to discriminate between, say, vertebrae and other areas of the image. Morphological operations (e.g., erosion, dilation, etc.) and kernels as structuring elements to identify the vertebrae were investigated.

Other work in this program focuses on digital moving video. This data type, usually in the form of videotape, is not uncommon in the archives of medical libraries and hospitals. A prototype system was developed to demonstrate the storage of digital video and retrieval via Mosaic. Video clips from NLM videotapes were selected for digitization and the resulting MPEG-1 data was stored in the optical disk jukebox. A perl script was developed to link citations retrieved from AVLINE with the video sequences. The objective is to investigate issues related to video-on-demand, especially the data compression issues (compare MPEG-1, MPEG-2 and motion JPEG), and implementation of easy linkages between citations and video sequences.

Engineering Laboratories

The R&D conducted in the Branch relies on laboratories which are designed, equipped and maintained by the Communications Engineering Branch. These are described below.

Signal Processing Laboratory. Housed in this laboratory are advanced systems to electro-optically capture the

digital images of documents, both in bound volume form and as loose sheets. Subsystems are available to perform image enhancement, segmentation, compression and storage on digital optical disk media. The laboratory also includes workstations for performing quality control and tagging for the captured documents. These bitmapped images may be retrieved in conjunction with a search of NLM's bibliographic databases or the DOCLINE document request system that serves the interlibrary loan function at NLM.

Specific equipment developed in-house includes high performance loose-leaf and bound volume scanners using charge coupled devices; optical disk drives; laser printers; high resolution (200 dpi) softcopy display devices. These are configured into systems that serve as laboratory testbeds to support research into automated document delivery, document preservation and document archiving, and techniques for image enhancement, manipulation, segmentation, compression for high density storage and high speed transmission, omnifont text recognition, and related areas.

Networks include both Ethernet and Token Ring at backbone rates of 10 Mbps connected to other local area networks throughout the building and to the Internet.

Image Processing Laboratory. This laboratory supports the investigation of image processing techniques for both grayscale and color biomedical imagery at high resolution. It consists of computer resources and image processing equipment to capture, process, and display such high-resolution digital images. The equipment includes Sun/Unix workstations, Sun-compatible Solbourne workstations, and IBM PC-compatible computers, all linked via an Ethernet local area network. All Unix machines run under Solaris 2.3.

The Sun/Unix machines include a Sun 690MP SPARCserver with a large magnetic disk storage capacity (6.4 Gigabytes). This machine operates as a file server for Unix machines both within and outside the Image Processing Laboratory. Additional storage for the 690MP is provided by a 9-track tape drive and an 8mm tape drive.

Large-volume storage is provided by a jukebox containing 144 5-1/4 inch rewritable optical platters, each formatted to contain files in the format required by the Unix file system. Each platter has a storage capacity of 586 Megabytes, for a total jukebox storage capacity of 81.5 Gigabytes. The host computer for the jukebox is a Sun 670MP computer with 2.4 Gigabytes of magnetic disk storage, a CD-ROM player, and an 8 mm tape drive. In addition to the jukebox, mass storage is also provided by a Sun SPARCstorage unit, a RAID system, consisting of 18 1.2 GB hard drives, six independent fast buffered SCSI-2 buses, and connected to an Sbus card hosted by a SPARC 20 model 612 via a 25 MB/s fiber channel connector.

Sun SPARC 10 workstations are used in the lab for DXPNET project development. These machines have 400 Megabytes of magnetic disk storage. One of the other Unix workstations is a Sun 4/260. This system supports development of the Standardized Readings Workstation for the

DXPNET project. A very high resolution Megascan monitor is attached, capable of displaying 2048x2560 8-bit grayscale pixels. This monitor is intended primarily for displaying x-ray images for DXPNET. A conventional Sun color monitor is also attached to provide the user interface for DXPNET.

The lab also contains Sun-compatible Solbourne workstations and PCs which are used for DXPNET project development. In addition there are three PCs in the lab attached to the Ethernet. They are used for display and grayscale image processing operations, an Internet file server in order to provide access to medical images for researchers in remote locations, and a platform for the capture and display of grayscale or color images. Other capabilities include the capture of 8 bit grayscale, and 8 or 24 bit color, from both paper and 35mm slides, using Ricoh and Howtek scanners.

Educational Technology Branch

The Educational Technology Branch conducts research and development in educational and information technologies, disseminates information about these technologies to NLM's various constituencies, and supports their application in health professions education. Systems developed both internally and elsewhere are demonstrated and information about the technology is disseminated through publications, seminars and workshops.

Branch staff identify and demonstrate technologies and applications which meet needs in health science instruction and information transfer. They regularly evaluate the effectiveness of educational and information technologies, and they assist health professions educators and others in improving the use of technology in health sciences education through training, demonstration, and consultation activities.

Branch programs involve managing a Learning Center for computer-based technologies, and carrying out research in educational and information technologies.

The Learning Center for Interactive Technology

The Learning Center for Interactive Technology is the setting for the use, investigation, and demonstration of new and effective applications to faculties and staff of health sciences institutions. The establishment of the Center reflects a commitment by the NLM to develop and support innovative, computer-based approaches to training health care professionals. Computer-based technology offers these professionals a means of improving their own knowledge of the discipline as well as an opportunity to incorporate this technology in the curriculum for the benefit of their students.

The Center has three components: (1) a demonstration area; (2) an R&D area; and (3) an interactive training facility. The demonstration area has 15 carrels that allow Center staff to demonstrate a wide range of interactive multi-media programs. Large group demonstrations are presented from a carrel configured for video projection, and are also conducted in the training facility.

The Center is visited by large numbers of individuals. Since its inception in 1984, the Center has been visited by some 8,000 individuals, including physicians, nurses, scientists, librarians, and other health care professionals.

Research in Educational Technologies

During FY 1994 the Educational Technology Branch launched its own World Wide Web Server. The server provides information relevant to health professions education and information technologies. The server, whose URL is <http://www.wetb.nlm.nih.gov>, features information about Branch research projects, resources, and monographs and provides links to closely related Web and Gopher servers.

Access to several databases maintained by the Branch is provided on the server. The Learning Center courseware database provides brief descriptions, system requirements, and contact information for the Center's courseware. The courseware collection includes state-of-the-art microcomputer and optical disc hardware and educational software, high quality working prototypes, and commercially available products for health science education.

During FY 1994, AuthorBase, a database of authoring system software created in 1991 by Branch staff, was updated by the U.S. Army Research Institute (ARI) according to an agreement that was entered into with the ARI in FY 1993 to jointly maintain the database. AuthorBase includes information on over 100 authoring tools for developing courseware and other materials for classroom use. Users can browse the database for information about currently available software and can identify those that work with different platforms, operating systems, and multimedia. Authorbase is accessible and searchable through the Web server.

A database providing sources of generic videodisks and CD-ROMs for health professions educators who wish to develop interactive teaching materials is also provided on the server. Using authoring software, generic videodisk/CD-ROMs can be tailored to meet individual teaching needs. This process is often referred to as repurposing, i.e., using a resource such as a generic videodisk or CD-ROM to manipulate or re-sequence individual pictures or motion video to meet the needs of students, faculty or practitioners in the health professions. These multipurpose, generic videodisks and CD-ROMs provide an extremely flexible and cost-effective means for users to define the equipment, environment, and instructional design that best meets their individual needs.

A subset of NLM's AVLINE database, relating to videodisk programs and computer software, much of which has been designed for instructional purposes, is also provided on this Web server. The Web server is updated regularly and new features continue to be provided by project staff.

During FY 1994 the Digital Microscopy System (DMS) project was initiated. The long-term goal of the project is to produce an indexed digital image system that will provide a retrieval mechanism for medical images that is equal to that currently provided for print materials. During FY

1995 the project seeks to provide an index to the existing NLM microscopic anatomy and pathology photomicrographs teaching collection. Project staff will investigate the appropriateness of using UMLS Metathesaurus terminology to index the collection. The index should provide the medical teacher and, eventually, the clinician, with a tool that will search and display images from the database, given certain specified search criteria. Project staff are exploring the possibility for digitization and compression of the existing image collection. Preliminary observations suggest that it will be possible to produce quality digital images from the existing analog collection.

The Computer-based Curriculum Delivery Systems (CCDS) was completed during FY 1993. The primary goal of the CCDS project was to design, produce, and test experimental technology-deliverable curricula for the health professions. The project produced 10 videodisks and accompanying computer software. The programs are now in use in 47 U.S., eight Canadian, two Caribbean, one European (U.K.), and two Philippine health professions schools. Some schools have used the videodisks to replace lecture and laboratory sessions, but the majority have used them as enrichment or remediation tools. During FY 1994 distribution of the videodisks and computer programs was begun through NTIS.

The Digital Anatomy Browser client software developed by the University of Washington was installed in the Learning Center in September 1992. The system is regularly shown to visitors from around the world. The Macintosh client software remotely accesses a structural information server running in Seattle. The server continues to be developed at the University of Washington and supports images, graphics, and a knowledge base of anatomic names and their relationships. System performance analysis of this distributed client-server based information system is done for the purpose of measuring performance at both the client and server computers and across the national Internet network. Branch staff made presentations at several major educational technology conferences and organized and delivered a number of workshops during the year. The workshops involved interactive multimedia, repurposing of existing videodisk systems, authoring tools, and network technologies. In November 1993, Branch staff actively participated once again in the annual Radiologic Society of North America meeting by offering demos, theater presentations, and workshops.

Research in Information Technologies

The Natural Language Systems (NLS) Program is concerned with investigating the contributions that Natural Language Processing (NLP) techniques can make to the complex task of mediating between the language of users and the language of the databases they attempt to access. The focus of the NLS program is the development of SPECIALIST, an experimental NLP system for the biomedical domain. The system includes a parsing system supported by a large lexicon, modules that access the extensive UMLS knowledge

sources, and a module for experimentation in information access and retrieval. The SPECIALIST system has been under development for several years, and in FY 1994 several of its major components were released to interested researchers as part of the UMLS program. It is expected that feedback from these researchers will result in continued improvements to the system.

The SPECIALIST parser includes several modules based on the components of natural language. The morphological component is concerned with the structure of words and the rules of word formation. The syntactic component treats the constituent structure of phrases and sentences, and the semantic component is concerned with the meaning of words, sentences, and discourses. All three rely heavily on the lexical component, which encodes the information specific to the words in the language.

The lexicon currently contains over 60,000 lexical records, with more than 118,000 lexical forms. Lexical entries are created using a lexicon building tool that incorporates a grammar of all the allowable slots and their values. The lexical grammar serves to constrain the possible choices that must be considered when entering an item, and it also serves as an automatic check of the correct form of completed lexical records.

The morphological component includes rules of inflectional and derivational morphology. Inflectional morphology deals with the different forms of a given lexical item. In English, this is used to mark nouns for number, verbs for tense, and adjectives and adverbs for their comparative and superlative forms. Derivational morphology relates alternates of lexical items that are grammatically related by affixation, but that generally do not share the same word class.

The parsing system exploits the syntactic information available in the lexicon, and generates a single, underspecified, syntactic representation for each linguistic structure encountered. This representation then depends on flexible semantic interpretation and robust domain knowledge processing to produce a final interpretation, or conceptual structure. At the heart of this approach is a mapping to concepts in the UMLS Metathesaurus and Semantic Network.

The problem of providing users with the information they seek can be viewed as the problem of mapping the language of the user to the language of the database. Users formulate queries in ways that reflect their knowledge and understanding of the topic and expect it will be sufficient to retrieve relevant information from the database. However, because of the richness and diversity of natural language, mapping between requests and information is rarely straightforward. An important goal of the NLS research program is to establish a more precise understanding of the relationship between user queries and the information that may be relevant to those queries. The type of semantic processing currently being explored in the context of the SPECIALIST system has shown some early promise for enhancing existing information retrieval methods.

NLS Program staff participates actively in NLM's UMLS project. During FY 1994 staff members worked with other UMLS participants in the preparation of the 1994 release of the UMLS knowledge sources. The SPECIALIST lexicon was released as a fourth UMLS knowledge source and will continue to be released with the other knowledge sources annually.

The SPECIALIST morphology module was released with the 1994 version of UMLS knowledge sources, as well. Morphological analysis is most successful if it is interleaved with look-up in a lexicon. This ensures that part-of-speech and other crucial information is available for the morphologic processing. The morphology module used together with the SPECIALIST lexicon should be a powerful tool for those who are conducting research in biomedical language processing and information retrieval.

NLS staff has developed a Unix-based retrieval system for browsing, navigating, and extracting information from the Metathesaurus and Semantic Network. The system can be used for batch processing, and it can also be used interactively to answer individual user questions. It can be used to query any of the information in the Metathesaurus for a particular concept. In addition, global searches can be executed that report all concepts that have a particular feature or set of features. An API for systems programmers is included. The system is being tested by several UMLS research groups in the United States and abroad.

The Electronic Imaging Document System (EIDOS) project involves research and development in electronic document imaging, indexing and retrieval methods. The project seeks to develop methods to facilitate computer-assisted access to information in textual and image form. The project investigates document management and retrieval methods based on the client-server, wide area network paradigm.

EIDOS project staff works with the National Coordination Office staff of the High Performance Computing and Communications (HPCC) initiative in the development and maintenance of an in-house HPCC library. During FY 1994 Branch staff created Web and Gopher servers for the NCO office. In the fall of 1993, the HPCC "Blue Book" (*High Performance Computing and Communications: Technology for the National Information Infrastructure*) was released on the server in advance of its printed publication. During the first three months of its use in late 1993 the servers were accessed by some 4,000 users with over 50,000 interactions, from over 40 countries. The server provides current information on a wide-range of activities related to the HPCC initiative. Branch staff monitor these activities and add information to the server regularly.

During FY 1994 the Radiation Articles Database project was initiated. In January 1994, President Clinton established an Advisory Committee on Human Radiation Experiments. The Advisory Committee is tasked with reviewing experiments that were conducted between 1944 and 1974. The experiments of interest are those that involved intentional

exposure of human beings to ionizing radiation and those that involved intentional environmental releases of radiation, in particular, those that were designed to test human health effects of ionizing radiation or that were designed to test the extent of human exposure to ionizing radiation. The NIH has established a Human Radiation Studies Task Force to collect and analyze data relating to these experiments. Branch staff will build a database of literature which resulted from the experiments performed.

Branch staff participate in the national and international health professions education, computer and information sciences, and medical informatics communities. Program staff have published several research papers during FY 1994. Acting Branch Chief, Dr. Alexa McCray, has been co-editor of the International Medical Informatics Association (IMIA) *Yearbook of Medical Informatics* since its first volume was published in 1992.

Audiovisual Program Development Branch

The Audiovisual Program Development Branch conducts media development activities with three specific objectives. As its most significant effort, the Branch supports the LHCBC's research, development, and demonstration projects with high-quality video, audio, and graphic materials. From initial project conceptual images, through actual project implementation with image preservation, transfer and display, to project evaluation and reporting, all forms and formats of imaging are produced. In FY 1994, the video production facility was enhanced by the installation of replacement CCD, solid state, color cameras. An obsolete electronic still store capability was replaced, and the video graphics system was expanded with the addition of new networked workstations and a graphics animation disc recorder.

Creative consultation and materials development are also provided by the Branch for the NLM's educational and information programs. With the added mission requirement of the Library to increase its outreach activities, the support that the Branch provides to these type of programs has increased significantly. From optical media technologies to teleconference support, the graphic, video, and audio materials requirement has increased in quantity and diversified in format.

The third area that the Branch concentrates on is technical development issues such as image resolution, color fidelity, media transportability, media storage and image communication. In addition to technique development by the staff, the facilities and hardware systems must reflect state-of-the-art standards in a very rapidly changing field. High Definition Television (HDTV) is a development area being explored that represents the future for improved electronic image quality. Multimedia techniques are being pursued for the educational and cost advantages that they offer. Three dimensional computer graphics, animation techniques, and photorealistic rendering methods have changed the tools and

products of the graphic artists in the branch. Digital video and image compression techniques are central to projects being pursued to improve image storage and communication. With all of the technologies which are being brought together in the development and support projects of the branch, the central core expertise remains the creative, artistic, communication skills of the staff.

Cervical Cancer: Success in Sight

During FY 1992 and FY 1993, in collaboration with the Early Detection Branch of the National Cancer Institute, LHCBC developed an award-winning educational program on the early detection of cervical cancer. Initially, an interactive laser videodisk was designed and produced for the IBM M-Motion system. Subsequently, the same videodisk was programmed for operation on the IBM Infowindow system. The Branch then redesigned the subject material and developed a Compact Disc Interactive (CD-i) version, in order to make use of the audio, video and programmatic capabilities of that all-digital platform. During FY 1994, the Branch developed a menu option on the CD-i program, which allows the viewer to use a CD-i modem to dial up the NCI's cancer information database, Physician's Data Query (PDQ), and to download the "Screening" and "Treatment" files for cervical cancer onto the CD-i unit, for display on an attached TV monitor.

Cardiac Embryology

APDB continues to play a leading role in a demonstration project in which 3-dimensional computerized representations of the embryonic development of the human heart are being created and tested in an educational setting. Beginning in FY 1993, a beta test version of the program was mounted on a Macintosh Quadra computer system in the Instructional Resources Center in the PreClinical Teaching Building of the Johns Hopkins University School of Medicine. Supported by the School's Medical Informatics Department, the program has been used and tested by medical students and faculty members, whose comments and suggestions are helping to further refine the project. During FY 1994, two faculty pediatric cardiologists borrowed the beta test version and demonstrated it, most successfully, to attendees of a national Pediatric Pathology Convention. Utilizing the latest computer graphics systems and high-speed communications networking, the final product of this project will be a widely accessible, desktop computer-based, interactive hypermedia presentation, available in videodisk or CD-ROM format.

The Making of a Medical Illustrator

Beginning with the 1993-1994 academic year, APDB began videotaping selected learning experiences of beginning medical illustration students in the Johns Hopkins

University School of Medicine's 2-year master's degree program. Portions of these tapes were edited into a 22-minute highlights program which was presented to attendees at the 1994 World Congress on Biomedical Communications. These students will be featured, along with their instructors and other medical illustration experts, in a documentary program covering the entire 2-year medical illustrator program.

Educational and Information Program Support

- "Online Clinical Practice Guidelines," a 5 1/2 minute videotape program, was produced for outreach use when the guidelines information was mounted on the Library's computerized databases and available to users via the Internet.
- Another short videotape was produced, called "NLM and the Internet," explaining how NLM information databases can be accessed over the so-called "Information Superhighway." This was produced both as a "stand-alone" informational videotape and as a part of the NLM-sponsored teleconference described below.
- The Branch produced 12 pre-videotaped presentations, plus an animated NLM Net opening sequence and the 4 1/2 minute "NLM and the Internet" program, mentioned above, which were integrated into a live nationwide satellite teleconference, called "NLM Update '94." APDB also provided technical advice to the Library on the selection of production facilities, talent, and the establishment of teleconference transmission and reception sites around the country.
- Selections from five existing films were transferred to videotape, edited, and converted to DRAW (Direct Read After Write) videodisk format on the subject of "Health Care to Native Americans." The videodisk, mounted in a special kiosk, was a featured part of a History of Medicine Division exhibit in the Library's Main Lobby.
- A 20-minute program was designed, produced, and edited by the Branch on "Islamic Calligraphy". The finished videotape was shown during an all-day "Islamic Culture and the Medical Arts" symposium, held in the Lister Hill Center Auditorium. The same day, a DRAW disc version was kiosk-mounted as a key part of another HMD-coordinated NLM Lobby exhibit.
- An updated program highlighting the educational materials held in the Educational Technology Branch's Learning Center was produced, edited, and manufactured as a Level III computer-controlled laser videodisk. Utilizing all four available audio channels, to permit highest quality sound on state-of-the-art playback equipment, this program is called "The Learning Center Videodisk-1994."

- A presentation by the Educational Technology Branch Chief was videotaped, in advance of a national HESCA meeting to be held in California, and was forwarded to the conference producers. APDB followed up by directing a live appearance of the ETB Chief in Washington, D.C. studios, where her appearance was integrated into the national satellite teleconference.
- In addition to providing normal audiovisual support of NLM-approved meetings held in the Lister Hill Center Auditorium, the 1994 Joseph Leiter Lecture

and several NLM-sponsored conferences and symposia were videotaped for documentation and outreach use by the Library.

Audiovisual Support Activities

The branch continues to upgrade the equipment used to support meetings held in the Lister Hill Center Auditorium and the NLM Board of Regents Room. APDB also provides preventive maintenance for audiovisual recording, playback, and projection equipment used in other meeting rooms throughout the Library.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

David Lipman, M.D.
Director

The National Center for Biotechnology Information (NCBI) was established by Public Law 100-607 in November, 1988, as a division of the National Library of Medicine. The establishment of the NCBI reflects the importance of information science and computer technology in the understanding of the molecular processes that control health and disease. The Center has been given the responsibility to

- Create automated systems for storing and analyzing knowledge about molecular biology, biochemistry, and genetics,
- Perform research into advanced methods of computer-based information processing for analyzing the structure and function of biologically important molecules and compounds,
- Facilitate the use of databases and software by biotechnology researchers and medical care personnel; and,
- Coordinate efforts to gather biotechnology information worldwide

There are presently 48 senior scientists, postdoctoral fellows, and support staff working at the NCBI. These scientists have backgrounds in medicine, molecular biology, biochemistry, genetics, biophysics, structural biology, computer and information science, and mathematics.

NCBI programs are divided into three areas: (1) creation and distribution of sequence databases, primarily GenBank; (2) basic research in computational molecular biology, and, (3) dissemination and support of molecular biology databases, software, and services. Within each of these areas, NCBI has established a network of national and international collaborations and also closely coordinates its activities with other NLM divisions. NCBI integrates data from NLM databases such as MEDLINE into specialized data resources for the molecular biology community.

Database Building

GenBank—The NIH Sequence Database

NCBI is responsible for all phases of GenBank production, support, and distribution. GenBank is the NIH genetic sequence database, an international database that collects all known DNA sequences, and a critical research tool in the analysis and discovery of gene function. The

database contains more than 215,000 sequences, 58,000 of which are human. Some 70,000 sequences were added during the year—a 33% increase in the size of the database—and growth is projected to continue at a rapid rate.

NCBI is directly involved in the creation and revision of sequence records and distribution of GenBank via the Internet, e-mail servers, and CD-ROM. Indexers with specialty training in molecular biology create records from direct submissions by scientists worldwide. NCBI staff and on-site contract personnel are involved in the record building and revision process, and NCBI scientists review the records for accuracy of biological information. Every effort to improve the biological accuracy of submitted data and to correct existing entries is being made by the GenBank team. New releases of GenBank are made every two months, daily updates are made available via the Internet.

Comprehensive coverage of all sequence data, protein as well as DNA, is provided by GenBank along with the corresponding MEDLINE bibliographic information, including abstracts. At the NLM, more than 4,000 journals are scanned for sequence data, sequences in journals that have not been submitted by authors are added to the database by a team of indexers in Library Operations. NLM has expanded its journal coverage to include all journals that regularly contain sequence data even if they are in nonmedical domains, e.g., applied biotechnology. An integral component of the database is the inclusion of abstracts and indexing terms from the MEDLINE records of sequence-containing articles.

GenBank is a key component in an integrated sequence database system that NCBI has developed in order to serve as a single, comprehensive source of all known DNA and protein sequence information. The integrated database offers researchers the capability to perform seamless searching across all available data including the MEDLINE abstracts linked to the sequence data.

An international collaboration with the EMBL Data Library in Germany and the DNA Database of Japan ensures comprehensive collections of sequence data worldwide. Cooperative arrangements are used to augment the in-house data capture operation. The National Agricultural Library, as part of the Plant Genome Project, furnishes coverage of the plant sequence literature. Similarly, an agreement with the U.S. Patent and Trademark Office has been established to capture sequence information from issued patents. Further enhancement of GenBank with specialized sequence data, such as vector sequences, repetitive DNA, and protein sequences from the older journal literature, has been provided by R&D contracts.

Other Specialized Databases

NCBI staff also are active in creating special-purpose databases, including an integrated view of *E. coli* genetic map and sequence data. The Expressed Sequence Tags database (dbEST), initiated in 1992, collects the growing number of gene fragments obtained through cDNA sequencing and

now contains over 50,000 entries, about 50% of which are human. Participants submit data electronically to NCBI and are provided with access to the BLAST network service, plus software tools to assist in their cDNA analyses. NCBI began accepting submissions for the Sequence Tagged Sites database (dbSTS) in April 1994, and both dbEST and dbSTS are available for public queries via the Internet.

NCBI, with assistance from taxonomy experts, has undertaken a comprehensive review of the GenBank taxonomy in order to correct errors, identify inconsistencies, and incorporate new scientific knowledge. A merged view of existing taxonomies from PIR-International, GenBank, EMBL, Swiss-Prot, and DDBJ databases was prepared using TaxMan, a taxonomy database management tool developed by NCBI. Specialized taxonomies such as the ICTV international standard taxonomy for viruses, the U.S. Department of Agriculture taxonomy for plants, and the FlyBase taxonomy for Drosophilidae were added. International collaborators representing the major sequence databases and taxonomy experts specializing in different branches of the taxonomic tree will continue to work together to maintain a current and accurate taxonomy resource.

Database Access

Software Toolkit

Equally important as building databases for molecular sequence information is the ability to access and retrieve the information using automated systems. The NCBI software toolkit concept addresses this need by creating software modules that provide a set of high-level functions to assist developers in building application software. Among these tools are a Portable Core Library of functions in the C language that facilitate writing software for different hardware platforms and operating systems, and AsnLib, a collection of routines for handling ASN.1 data and developing ASN.1 applications. The ASN.1 (Abstract Syntax Notation) tool is an International Standards Organization data description language that provides a mechanism for defining and structuring data as well as a set of program definitions that can interact with databases structured in ASN.1.

NCBI's adoption of ASN.1 for database output has several advantages for users as well as developers. The data definitions in ASN.1 for biological objects enable the representation and structuring of complex biological data in data files without the need for a specific database management system. Manipulation of the complex objects is performed through the ASN.1 software tools that are freely distributed to the biology community. Thus, complicated analysis programs can be readily constructed from pre-existing sets of modular tools, saving considerable time and programming effort. In order to disseminate information about NCBI software and database projects, and provide technical support, NCBI has organized a Professional Software Developers Forum. This forum is intended for professional software

developers in the biotechnology and pharmaceutical industries.

Entrez Retrieval Software

A major application based upon the toolkit approach is a retrieval tool, called Entrez, that searches nucleotide and protein sequence databases and related MEDLINE citations. With Entrez and the integrated database on a CD-ROM or a local network, a user can search several hundred megabytes of sequence and literature data using techniques that are fast and easy. A key feature of the system is the concept of "neighboring," which permits a user to locate related references or sequences by asking for all papers or sequences that resemble a given paper or sequence. The ability to traverse the literature and molecular sequences via neighbors and links provides a very powerful yet intuitive way of accessing the data. New releases of the database are produced every two months.

The Entrez software is also available over the Internet. Versions for multiple computer platforms are available that allow a user (client) anywhere on the Internet to formulate queries on a local workstation and send off requests to a server at NCBI. Network Entrez has grown from 250 sites in October 1993 to more than 800 in September 1994. More than 3,600 queries from Network Entrez are processed every day.

Other Network Services

The BLAST sequence searching server is one of NCBI's most heavily used services. BLAST compares a user's unknown sequence against the database of all known sequences to determine likely matches. Sequence similarities found by BLAST have been critical in several gene discoveries. Hundreds of major sequencing centers and research institutions around the country use this software to transmit a query sequence from their local computer over the Internet to a BLAST server running on a computer at the NCBI. In a few seconds, the BLAST server compares the user's sequence to the database and returns the results to the client program for viewing by the user. More than 3,100 search requests are processed by network Blast each day, with a response time of under 15 seconds for many queries. The BLAST network server has been recognized by research groups as an essential laboratory tool not only to analyze data but to aid in setting directions for genome research.

NCBI's electronic mail servers, RETRIEVE and BLAST, are widely used by the scientific community. The first, RETRIEVE, is used to retrieve records from several sequence databases, including GenBank, EMBL, Swiss-Prot, and PIR, by sending a mail message containing the query to the server. The second, BLAST, is used to search the nucleic acid and protein sequence databases for matches to a user's sequence using the BLAST algorithm to identify sequence similarity. The service has a public key encryption option to guarantee the confidentiality of user data as it traverses the

public networks. Any user in the world with e-mail access can submit a query to the servers and have an answer returned within minutes. More than 4,800 queries are handled daily by the two electronic mail servers.

In December 1993, NCBI expanded Internet access to NCBI information and search tools by making available a public World Wide Web (WWW) server. The server provides capabilities for Entrez and BLAST searches. Many other WWW servers have links to the NCBI server to conduct searches and obtain the latest GenBank records. Information about all NCBI services, data submission and update, and individual scientists research projects is also available.

GenBank is also distributed over the networks through the standard File Transfer Protocol (FTP) program. More than 1,500 requests are processed daily for downloading files from NCBI's public data repository, including 150 of the daily GenBank update files. The entire database of over one gigabyte, as well as daily updates, are available to network users. Thirty additional databases of interest to the molecular biology community are distributed through the NCBI Data Repository.

NCBI uses a cluster of Unix-based multi-processor servers to provide network services such as BLAST and Entrez over the Internet. The multi-processor architecture has enabled NCBI to keep pace with the growth of the database and with a two-fold increase in the number of searches by increasing the number of CPUs per server and dedicating new servers to specific search functions. The systems are in constant use 24 hours per day, seven days per week performing searches and processing data for inclusion in the database. Additional servers are used for the internal computing needs for GenBank database production and for NCBI's basic computational biology research. The servers are supplemented by scientific workstations that are used for database applications, software development and molecular modeling programs.

Basic Research

Basic research is at the core of NCBI's mission. The Basic Research and Information Engineering Branches at the NCBI are made up of a multidisciplinary group of scientists who carry out research on fundamental biomedical questions at the molecular level by developing and utilizing mathematical, statistical, and other computational methods. The approach is both theoretical and applied. These two lines of research are mutually reinforcing and complementary. The basic research has led to new practical methods and the application of these methods has opened new areas of research.

Several analytical algorithms and methods have been developed to complement the investigation of biomedical problems. Analytical methods and algorithms focus on sequence similarity, structural modeling and genome analysis. Pattern matching and sampling methods for biopolymer sequence data, which include the statistics of sequence com-

parisons, have been applied to investigate such protein families as the ras-line GTPases, steroid receptors, cold shock domain proteins, HMG-1 box-containing proteins, and NTPases, as well as other analyses of different sequence motifs. Biomolecular structural investigations include such projects as the analysis of packing contacts in protein crystals, 2-D lattice models of proteins, 3-D modeling of ribonucleic acids, protein threading and the energy distribution of compact states of a peptide. Genome analyses have been a significant part in both the methods development and the scientific analysis of several different genomes including that of *E. coli*, RNA viruses, and humans.

The intramural group is engaged in 40 projects, many of which involve collaborations with NIH and other research laboratories. The work is reviewed by a Board of Scientific Counselors of distinguished extramural scientists (see Appendix 4 for list of members). The high caliber of the work has been evidenced by the number of peer-reviewed publications, some 50 in FY 1994, and the requests for outside collaborations and invitations for talks at international meetings.

User Support and Outreach

As part of its mandate to support the development of new information technologies of relevance to biology and genetics, the NCBI has exercised a leadership role in sponsoring forums for the exchange of information among leading scientists from the fields of computer science and biology. NCBI has also extended its outreach to the library science community by invited presentations and workshops on biotechnology information topics.

The Visitors Program continues to be successful in bringing members of the scientific community to the NCBI to engage in collaborative research in the bioinformatics area as well as joint activities in database design and implementation. This program, administered in conjunction with Oak Ridge Associated Universities, facilitated over 30 visits from senior researchers this past year.

As the number of database services and number of users have increased, user support services have expanded. Contract staff have been added to supplement existing NCBI staff in the Information Resources Branch who provide responses to telephone, e-mail, letter, and Fax queries for information and assistance. The three main areas of user support include: information about GenBank services and data submission; technical assistance for submission of new GenBank data and revision of existing records; and technical assistance with Entrez and other data retrieval services. Most responses are immediate and nearly all answers or information are provided within 24 hours of receipt of a message. Likewise, authors who submit their sequences to GenBank are furnished with accession numbers for publication within 24 hours.

To increase awareness of NCBI and its programs, NCBI staff participate in exhibits, seminars, workshops, and

courses, both nationally and internationally. In FY 1994 NCBI staffed exhibits at scientific society meetings, including the American Society for Biochemistry and Molecular Biology; American Society for Cell Biology; Genome Sequencing and Analysis Conference V; Human Genome 1994: The International Conference on the Human Genome Project; and the First Single Chromosome Workshop on Chromosome 1. In addition, senior NCBI members participated as faculty at courses sponsored by the American Association for Cancer Research, the World Health Organization, and Johns Hopkins School of Medicine/The Jackson Laboratory. Three issues of a newsletter were distributed to a mailing list of 20,000 biologists and institutions, and new Fact Sheets on programs and services were distributed at all public forums where NCBI was represented. On the NIH campus, more than 350 scientists are supported through online access to 20 databases under the IRX system.

The NCBI also participates in an advisory and collaborative role with other government agencies such as the Patent and Trademark Office and the Department of Agriculture on programs involving biotechnology information. Within the NIH, the NCBI coordinates with other institutes and particularly with the National Center for Human Genome Research on databases and informatics programs that impact information exchange on a national level.

Extramural Programs

The NLM's Extramural Program (a separate division of NLM) offers a program of grants for computer analysis of molecular biology data. A wide variety of work in computational biology has been supported through the program, including methods and algorithms for sequence analysis, structure and function prediction, new machine architectures and specialized databases. Postdoctoral training in the cross-disciplinary areas of biology, medicine, and computer science is also funded through the NLM's informatics fellowship program.

Biotechnology Information in the Future

The explosive growth in the fields of genetics and molecular biology reinforces the need to build and maintain a strong infrastructure of information support. NCBI continues to be engaged in developing and employing new methods for disseminating knowledge to the biomedical community. Based on a core of advanced intramural research in several areas of computational biology, NCBI can rapidly address the evolving informatics needs for genome research with state-of-the-art software and databases. Genomic information resources such as NCBI have repeatedly demonstrated their value as indispensable discovery tools for basic research. The value of these resources will only continue to grow as they support the breakthroughs in basic research and provide us with better understanding and treatment of human disease.

EXTRAMURAL PROGRAMS

Milton Corn, M.D.
Acting Associate Director

Introduction

The Extramural Programs Division provides support to the health science community in the biomedical areas for which the National Library of Medicine takes particular responsibility. NLM support for extramural programs stems from two sources: from the Medical Library Assistance Act (MLAA) of 1965 and its extensions, and from Section 301 of the Public Health Service Act as amended. The dual basis of the funding sources as well as the historic mission of the Library explain the eclectic variety of the funded projects for which the Division takes responsibility.

Overview

The Extramural Programs Division is deeply involved in such major NLM themes as Outreach, the High Performance Computer and Communications (HPCC) initiative, and Biotechnology. However, because much of EP's existing program structure is already related to Outreach, HPCC, and Biotechnology-related activities, new funds provided to support such areas are for the most part included among the "traditional" programs described in specific sections below rather than in an artificial separate listing. An exception is the Connections program, a new program supported with HPCC funds and designed to promote connections of health science sites to Internet; it is described below in the Resource Grants section. Another exception, also described below, concerns a new program directed towards the problems of creating a computerized patient record; during FY 1994 EP, in association with the Agency for Health Care Policy and Research, published a Request for Applications in the field of Electronic Medical Record Systems. The grants will be administered as Cooperative Agreements.

As developed by the Board of Regents' Planning Panel, the Outreach program involves a number of EP initiatives, including professional training, IAIMS, and resource grants to improve access to biomedical information by health libraries and physicians.

Biotechnology is supported by resource grants and research grants that are reviewed by a subcommittee of the BLRC.

An important goal in FY 1994 concerned an ongoing effort to broaden support for informatics by encouraging other NIH Institutes and other agencies to fund informatics-related projects, and by cooperation between NLM's Extramural Programs Division and other organizations.

Although budget restrictions, including a mandated

cutback in travel funds, were experienced by all divisions, the relative lack of travel funds caused particular hardship for EP's vital site visit program. Critical visits to evaluate applications were made, albeit with minimal staff attendance. However, judicious administrative site visits are also of great importance for evaluation of ongoing grants, of prospective grantees, and for promulgating the activities of NLM's grant programs; such trips were significantly curtailed.

The Research Grants Section of this report summarizes some recent activities in the area of basic and applied information science. The application of computers to biomedical information storage and retrieval has revolutionized the operations of biomedical libraries and has engendered the useful term, medical informatics, to describe the theory and practice of providing information and decision support accurately and usefully to health workers. Such research is vital now when the volume of biomedical information is growing at a rate that threatens our ability to keep track of what we know, and to use what we do know most efficiently.

Training efforts also merit specific description. Training of competent professionals in medical informatics must remain an important goal of the Division. This new field needs scientists who can exploit the enormous potential for improvement in health delivery which medical informatics is capable of providing. Applying information science to modern health care and research poses complex problems whose solution will depend on well-trained specialists. NLM supports both institutional training programs and fellowship programs.

The ten five-year institutional training grant awards made in 1992 entered their third year in FY 1994. Of special interest in FY 1994 was the relatively heavy response to the newly created fellowship in applied informatics, which is open to almost all health care professionals, and is designed to promote training of experts in applying informatics to the clinical, educational, research, and administrative problems of health centers. The relatively large interest in the program by librarians is gratifying.

The NLM's Integrated Advanced Information Management Systems (IAIMS) program addresses the insufficiently appreciated but vital issue of integrating usefully the myriad information systems which have sprung up at most of our medical centers. Such databases are useful, to be sure, but all too often are unrelated, isolated, and very far from taking advantage of the synergism that can be realized by linkage of the various information systems present in the library, research sites, administration offices, medical education operations, hospitals, and outpatient areas.

To respond appropriately to NLM experience and changing biomedical culture/technology during the past ten years, IAIMS was extensively revised during FY 1992 with important changes in scope, objectives, and funding levels. The three FY 1992 IAIMS/HPCC demonstration grants entered their final year in FY 1994.

Medical Library Resource Grants have been an essential element of the Division's activities for years. It is

clearly an NLM mission to make biomedical information easily available to all health professionals. This emphasis was heightened when the NLM adopted Outreach as a major new initiative. Improvement of access by physicians to medical information was specifically addressed by a recent revision of the Resource Grant Program designed to expand the ability of hospital libraries, particularly in rural, inner city, or other underserved areas, to establish facile contact with the national biomedical library system.

A revised program announcement describing Resource Grants was issued in 1994 to emphasize the current NLM policy of promoting access to national networks as the cornerstone of biomedical information management in the future.

Grants in support of publications have little to do with medical informatics but are a time-honored, important commitment by the Division to the scholarly activities that lie at the heart of libraries everywhere.

The support provided for the bioethics center was converted from a grant to a sole-source contract in FY 1994, with the award being made to Georgetown University, which has maintained the bioethics center for a number of years. The sections on the Division's committee activities, and on the conferences supported are self-explanatory.

EP made an initial foray into the new and complex area of database support in FY 1994 with a Request for Applications for maintenance of databases in certain biotechnology-related areas. The problem of who will maintain the growing number of on-line databases which are of importance nationally to biomedicine will need to be addressed. These databases represent an electronic evolution of the reference material traditionally maintained by libraries in print form, but the cost of supporting such digital reference sources is high and getting higher, and no national decision has yet been made about Federal policies in this arena.

Regional Medical Library support, as authorized by the Medical Library Assistance Act, is described in the Annual Report's chapter on Library Operations. The Special Foreign Currency Program, administered by the Extramural Programs' International Programs Branch, was discontinued in FY 1994 as described elsewhere in the annual report under International Programs.

Budget information is summarized in Table 11.

Training (MLAA)

The NLM continues its support of research training in the fields of medical informatics and biotechnology information. It is clear that these disciplines go beyond the use of the computer as a computational tool and extend into the process of knowledge representation, storage, retrieval, and manipulation to support inferential reasoning, and to rationalize decision making in the health sciences. There remains a need for qualified, talented investigators, well equipped to address fundamental issues in the use of computers and

automated information systems in health care, health professions education, and biomedical research. These investigators will contribute to the growth of science by their studies of knowledge management and by advancing the frontiers of the computer sciences for organizing, retrieving, and utilizing health knowledge. NLM also intends to foster medical informatics and biotechnology information as growing disciplines with an appropriate place in academic medicine. It is expected that the trainees will become able, cross-disciplinary translators, taking the computer sciences to all of biomedicine.

Approximately 100 trainees were supported on the institutional training grants. The currently supported sites are listed below. In addition to the institutional training grants, NLM supported ten individual fellows. These fellows were receiving training in the conduct of informatics research in order to become professional informaticians or they were receiving training which would allow them to apply the techniques and technology of informatics to their respective disciplines.

Institutional training grantees:

1. Harvard Medical School
Massachusetts Institute of Technology
Brigham and Women's Hospital
Massachusetts General Hospital
New England Medical Center
Robert A. Greenes, M.D., Ph.D., Director

Major research emphases are computer-based decision support systems, modeling of physician decision making, representation and structure of medical knowledge, application of information technology to medical education, database and data analysis systems, computer graphics, and the development and evaluation of digital imaging systems.

2. University of Minnesota
Lael Gatewood, Ph.D., Director

The focus of this interdisciplinary program is to provide training in cognitive, information, and computer sciences. Current research includes physician decision making, diagnostic classification, nurse decision making, electronic communications for health professionals, physician training, and health information systems.

3. University of Pittsburgh
Randolph A. Miller, M.D., Director

Operating under the Intelligent Systems Program at the University of Pittsburgh, this program uses the faculty and services of the following: the School of Medicine, the Graduate School of Business, the Department of Computer Science, and the Interdisciplinary Department of Information Science.

4. Stanford University
Edward H. Shortliffe, M.D., Ph.D., Director

This formal program in medical informatics offers Masters and Ph.D. degrees to individuals with a career commitment to this field. The specialized curriculum focuses on the development of a new generation of researchers with a commitment to developing practical, computer-based solutions to problems in the optimal management of biomedical knowledge.

5. Yale University
Perry L. Miller, M.D., Ph.D., Director

This training program will prepare individuals for careers in medical informatics research. The program will include both postdoctoral and predoctoral training. In addition to multidisciplinary research opportunities, the program will also offer didactic experiences.

6. Rice University
Baylor College of Medicine
University of Houston
G. Anthony Gorry, Ph.D., Director

The W.M. Keck Center for Computational Biology brings together scientists in one of the nation's premier organizations for the exploration of molecular and genetic structure in conjunction with biomolecular design and engineering. The center addresses fundamental problems at the frontiers of biological research through the extensive use of advanced parallel computing and imaging.

7. Oregon Health Sciences University
Kent Spackman, M.D., Ph.D., Director

The Biomedical Information Communication Center of OHSU trains individuals who are committed to a career in medical informatics. The program has its focus on end-user informatics with areas of concentration that include design and delivery of information resources and knowledge bases, organization and representation of health information, information retrieval, design and construction of health professional workstations, health outcomes research, image analysis, administrative informatics, and informatics training and education.

8. University of North Carolina
Duke University
Charles P. Friedman, Ph.D., Director

The Duke-UNC program stresses both the basic sciences underlying medical informatics and an equally important body of practical knowledge necessary to design and implement computer applications that function in support of health care. The program also stresses the methods used to study rigorously the effectiveness of these information technology applications. To these ends, degree programs include courses

specifically addressing topics in medical informatics; courses in biomedical engineering, computer science, information science, biostatistics, and other basic disciplines; electives tailored to each student's special interests; and internships that provide experience with computer applications installed in clinical settings.

9. University of Missouri
Joyce Mitchell, Ph.D., Director

The University of Missouri offers a flexible program to train postdoctoral and predoctoral candidates for an academic career in medical informatics. The training program and curriculum prepares individuals to develop, use, and evaluate applications of innovative information methods and computers in the health care environment. The program emphasizes the synthesis, retrieval, organization, and effective management and communication of knowledge.

10. Columbia University
Paul D. Clayton, Ph.D., Director

Areas of special interest include clinical information system development, clinical decision support, computer-based learning, database design, natural language processing, medical vocabularies, medical knowledge representation, information retrieval strategies, physician data entry, genome mapping, molecular modeling, and high performance computing and communication. Opportunities also exist in the areas of computer applications in medical librarianship and computer aided learning. A particular strength is the ability to provide research opportunities in the context of working production systems.

Research (PHS 301)

Through its research grants, the National Library of Medicine seeks new understandings of medical knowledge and new ways of using knowledge and information for health care, research, and education. For many years, the Library has emphasized research in informatics. Informatics is the field, or science, of managing information in a computer and telecommunications context. Representing knowledge or information, particularly medical knowledge, offers numerous challenges to computer science, but because information has personal and cultural dimensions, informatics includes large elements of social and behavioral science. In NLM's research program there are further distinctions among medical informatics, biotechnology informatics, and health library information science. Within these areas are also distinctions between research on fundamental questions and on more immediately applicable work.

Medical Informatics

This year NLM directed its medical informatics

research and development program towards electronic medical record systems. Together with the Agency for Health Care Policy and Research (AHCPR), NLM published a request for cooperative agreement applications. Cooperative agreements differ from grants in that the sponsoring program has a more active partnership role with the principal investigators.

The importance of electronic or computerized medical records has been recognized for some time. NLM and AHCPR seek new, practical approaches to vocabulary issues and to articulation of the records with patient care guidelines. NLM is also concerned with transportability and generalizability of systems. Patients' records must be accessible at many care locations in ways that facilitate health care decisions but do not compromise privacy or confidentiality. Record systems must also facilitate the analysis of accumulating data for purposes of health care research. NLM believes that the field is mature and large enough to influence the health provider community significantly in these areas of program interest.

The response to the request for applications, 94 altogether, was gratifyingly large. NLM conducted a merit review, using a special review committee and a triage process. Reviewers agreed that many of the applications were outstanding or excellent. Together with AHCPR, NLM hopes to fund at least eight of them over a two-year period. In FY 1994, NLM awarded cooperative agreements to James Cimino, Columbia University; Michael Kahn, Washington University; and Isaac Kohane, Children's Hospital, Boston. NLM and AHCPR jointly funded Christopher Chute, Mayo Foundation. AHCPR also funded Clement McDonald, Indiana University.

In another initiative, NLM continued to participate in the Human Brain Project, a multi-institute program to sponsor informatics and related neuroscience research for mapping the brain. The various institutes collaborate through a Federal Interagency Coordinating Committee. This year, NLM joined the National Institute on Deafness and Other Communication Disorders in a research award to Dr. James Brinkley of the University of Washington. He is investigating an information system for organizing, visualizing, and managing intra-operative cortical language mapping data. His integration of neuroscience and informatics represents the kind of excellent research that the Human Brain Project seeks to promote.

In other Medical Informatics activities, NLM was able to award eight new grants, a modest increase from the six new awards in FY 1993. One of these was a Research Career Development Award to Dr. Frank A. Sonnenberg of the University of Medicine and Dentistry of New Jersey/Robert Wood Johnson Medical School. These awards convey support for time and effort for five years. In return, the applicant must describe a research career plan and show that adequate resources are available for the investigations planned. Dr. Sonnenberg's major research interest is knowledge management for clinical decision analysis.

One of the research project awards was a competitive renewal of support for Professor Eta Berner of the University of Alabama at Birmingham. She will study the effect of computer decision support systems on physicians' differential diagnoses. Her report of her earlier work on assessment of decision support systems was published in the *New England Journal of Medicine*, with an accompanying editorial. (Berner ES, Webster GD, Shugerman AA, Jackson JR, et al. Performance of four computer-based diagnostic systems. *N Engl J Med*. 1994 June 23; 330(25): 1792-6.)

Biotechnology Informatics

Biotechnology information research is supported to investigate effective methodologies for organizing and analyzing data about molecular control of life processes. Relevant problems include design and management of large databases, more powerful methods to retrieve information from multiple factual databases, and general pattern matching algorithms for biological sequences.

The need for conveniently available information continues in the molecular biology community. In FY 1994 NLM issued a Request for Applications to support the maintenance of protein databases in the field of molecular biology. The input data could derive from the published literature or other databases and may contain sequence data, NMR structural data, or other types of data useful by molecular and protein biologists. The database will be highly responsive to a national user community.

In FY 1994 NLM supported 13 grants in the field of biotechnology informatics expending \$3.7 million. These grants were of a wide range of types including investigator initiated research grants, FIRST awards for investigators just getting started in the field, research resource grants where specialized information resources are made available to the user community, and two conferences were supported.

Health Library Information Science

This program area concerns information issues and problems directly related to medical bibliography and medical library services. No new grants were awarded in this area in FY 1994, but two earlier projects received continued support.

Resource Grants (MLAA)

NLM offers four categories of grant to encourage the utilization of computer and communications technologies to access medical information, including bibliographic information. These grants range from introducing these technologies via the Information Access Grant to more sophisticated uses as demonstrated by Information Systems grants. Most complex and generally restricted to large institutions with many subsidiary systems are the IAIMS Grants.

Access and Systems Grants

In FY 1994 two Information Access Grants were awarded. The first, to the Rural Health Projects, Inc. (formerly the Northwest Oklahoma AHEC), will provide funding for organizing a consortium of health-related institutions to assess their technological needs for accessing Grateful Med and Loansome Doc. Approximately 10 institutions are potential members. The second Information Access Grant was to the Laurel Highlands Health Science Library Consortium in Johnstown, Pa., to give the members of this established consortium the technological means to access online catalogs as well as Grateful Med and Loansome Doc.

Three Information Systems Grants were awarded. The one to the University of Mississippi Medical Center Library will further develop the statewide "Mississippi Health Sciences Information Network" (MisHIN). The goal is to establish a statewide electronic network involving a coalition of hospital and academic libraries, academic departments, various professional organizations, and some state health agencies. The project consists of Grateful Med, DOCLINE, Loansome Doc, Internet, union catalog, and long-distance education. Considerable training will be provided in using the resources and services of MisHIN. The second Information Systems Grant was awarded to Harvard's Countway Library of Medicine to enter its 19,000 serial titles into electronic databases in order to be accessible nationally and to a wider audience. The third Information Systems Grant was awarded to the University of New Mexico Medical Center Library to add image-based information access and presentation to its current suite of system functions and to add novel state-wide information sources and access. Implementation includes an infrastructure upgrade, an upgrade of the catalog system, an access system to the New Mexico Tumor Registry (NMTR) and to the Office of the Medical Investigator (OMI), an online source of Native American history, and a Geographical Information System (GIS).

IAIMS Grants

Integrated Advanced Information Management Systems (IAIMS) are institution-wide computer networks that link and relate library systems with a variety of individual and institutional databases and information files, within and external to the institution, for patient care, research, education, and administration. Resource grants are made to assist medical centers and health science institutions and organizations in planning and development projects that will lead to the implementation of IAIMS. The goal of the program is to create organizational mechanisms within health institutions to manage more effectively the knowledge of medicine, and to provide for a system of comprehensive, direct information access for health professionals.

NLM provides grant support for (1) an institution-wide planning phase where support may be for up to \$150,000 per year for one or two years, and (2) an operation phase in

which IAIMS plans are implemented. Operation phase grants may be for up to \$500,000 per year for up to five years, or for \$550,000 per year if support of an IAIMS apprenticeship option is approved. Total IAIMS program funds awarded in FY 1994 were \$3,114,658.

In FY 1994, 28 IAIMS applications were reviewed: 25 requested support for planning, and 3 were for operations. Twenty-one applications were scored by the reviewers and recommended for consideration of funding. Eleven applications were from hospitals; all three operations applications were from academic medical centers. New IAIMS grant awards were made in the year for Phase II operations at the University of Washington, and for Phase I planning at Louisiana State University, Medical College of Wisconsin, University of Chicago, University of Missouri, University of Rochester, and the University of Virginia. Continuation grant awards were made to Baylor College of Medicine and Duke University. In FY 1994 NLM continuation awards were made for two High Performance Computing and Communication grants resulting from a one-time special initiative begun in FY 1992 to certain IAIMS institutions: Dr. Edward L. Chaney at the University of North Carolina, and Dr. Gordon K. Springer at the University of Missouri at Columbia.

Internet Connections Grants

For the third successive year NLM participated with the National Science Foundation in offering Internet Connections Grants to health science institutions. These grants are available in two forms: (1) \$30,000 to single institutions for an initial Internet hook-up and (2) \$50,000 to institutions willing to extend their existing connection to other sites, for example, teaching hospitals. In FY 1994, 17 Internet Connections Grants were awarded.

Special Minority Support Grant

In January 1994, NIH's National Institute of General Medical Sciences (NIGMS) issued a Request for Applications (RFA) for Predoctoral Fellowship Awards for Minority Students. NLM participated for its fields of interest and made one award to Alexander R. Carter, who is pursuing studies in cognitive neuroscience. He is enrolled in the M.D./Ph.D. program at Harvard Medical School.

Publication Grants (MLAA)

The Publication Grant Program provides selective short-term financial support for not-for-profit, biomedical scientific publications. Studies prepared or published under this NLM program include critical reviews or monographs on special areas of medical research and practice; research monographs in the history of medicine; writings on medical informatics, health information science and biotechnology; and, in certain cases, secondary literature tools and scientifically significant symposia. Because funds for publication

support are limited, available resources in recent years have been used principally for history of medicine projects.

During FY 1994 NLM 8 Publication Grants were awarded, totaling \$239,300. Of these, five were new awards. This small grant program has a current self-imposed annual ceiling on direct costs per grant of \$25,000. The average grant awarded, including both direct and indirect costs, was under \$30,000.

Among the new awards made in FY 1994 was a two-year study of the impact of AIDS on the experience of American physicians during the first decade of the epidemic by Dr. Ronald Bayer of Columbia University. The research will be based on oral histories with some 50 physicians, and will explore the way AIDS influenced the attitudes, practices, reputation, linkages, and sense of mission of physicians. The result will be a book-length monograph and an oral history archive that will preserve for future researchers an invaluable resource for understanding how AIDS affected the practice of medicine and the impact pioneering physicians had on the public understanding of the demands imposed by the HIV epidemic.

Another new award, for one year, was made to Dr. Christopher C. Sellers of the New Jersey Institute of Technology to complete research and a book on the history of occupational disease research in the United States between 1900 and 1930. The book will examine the technological, economic, and social influences that led medical and other scientists to take up the study of occupational disease during the period, the intellectual and technical resources upon which they drew to assemble a new discipline, and the role that their new science came to play in the American workplace. It will offer important insights into the beginnings of applied biomedical disciplines, and will be a unique contribution to the history of occupational health.

Among the books published in FY 1994 funded through the Publication Grant Program in previous years was Dr. Paul F. Basch's *Vaccines & World Health: Science, Policy, and Practice*. The book deals with the application of biotechnologic innovations to improve health throughout the world, and focuses on the role of vaccines, existing and proposed, in preventing and controlling diseases. It clarifies the interrelations of science, technology, economics, epidemiology, and other disciplines needed to take a vaccine from concept to reality. Another book published was *Trials of an Ordinary Doctor: Joannes Groenevelt in Seventeenth-Century London* by Dr. Harold J. Cook. This microhistory shows how a medical malpractice case against an otherwise obscure Dutch physician in London became the center of one of the era's great medical controversies. Cook uses the specifics of a single life to illuminate how society and politics, as well as the scientific and professional uncertainties and jealousies of the early Enlightenment, combined to change the authority of the physicians of London.

Other Support

Bioethics (MLAA)

NLM continued its support for a National Reference Center for Bioethics Literature at Georgetown University. The resources of a highly specialized collection are offered to inquirers nationally by means of an outstanding reference service. NLM completed the fiscal year with a conversion from grant to contract support. The goal is to integrate services more closely with other NLM activities, such as a recently renewed contract for indexing the material the Center collects. The indexing is made available as Bioethicsline, one of NLM's group of computerized medical databases.

Small Business Grants

These grants offer small businesses an opportunity to undertake research and development leading to marketable products within the scope of the funding program. Funds for these grants are mandated by law and cannot be used for any other purpose. This year NLM awarded two Phase I SBIR grants. Both are in the area of molecular biology: one is to develop a client-server approach for accessing these databases by Internet; the other is to develop computer tools for exploring and parsing DNA sequences. NLM's funding for the STTR program (a related "technology transfer" offering designed to promote collaboration between small businesses and academic institutions) was too small to support a single project. NLM therefore contributed its allocation to another institute. NLM will doubtless support one of these grants at some future time as more funds are allocated by Congressional formula.

Conference Support (MLAA, PHS 301)

The Extramural Programs sponsors small workshops as a means of focusing peer comment on research needs and opportunities within NLM's program scope. Such workshops also assist the field by serving as non-authoritative consensus groups. In March 1992, NLM, the National Cancer Institute, and AHCPR co-sponsored a workshop on medical imaging databases. In November 1993, NLM and NCI followed with a similar conference on the need for common standards so that exchange of information among database sites becomes practicable and easy. In radiation therapy planning, for example, the problem is especially acute. The 17 participants reached agreement on general principles. The conference chair, Dr. Carl Jaffe of Yale University, has since reported the conclusions at national society meetings.

Scientific Review

The Biomedical Library Review Committee (BLRC)

is NLM's initial review group for grant applications. The Chairperson of the full committee during FY 1994 was Dr. Sherrilyne Fuller. This committee reviews about 90 percent of all grant applications which have a primary assignment to the Library. Dr. Michael Vannier was appointed Chairperson of the BLRC for FY 1995.

Over the years, this committee has been concerned with the evaluation components of grant applications. In an attempt to explicate the various types of evaluation schemes that are applicable at various stage of project development, some committee members recently published a paper on the topic. (Stead WW, Haynes RB, Fuller S, et al. Designing medical informatics research and library resource projects to increase what is learned, *J Am Med Informatics Assoc*; 1994; 1:28-33.) The paper attempts to provide some guidance to potential applicants, and to assist the review committee in their deliberations of specific applications. Dr. William Stead, the primary author on this paper, was formerly a member and chairman of the committee.

In a similar vein, Virginia Bowden and other members of the Committee published a paper (Bowden, VM. National Library of Medicine resource grants: application and review, *Bull Med Libr Assoc* 80 (2) April 1992, 157-168.) seeking to be of assistance to potential applicants to the resource grant program. The paper details and explains the criteria used by the committee when reviewing resource applications.

The current committee activity, which is ongoing, is to produce a document that will attempt to explicate the review process and what feedback an applicant can expect. This document probably will take the form, at least in part, of questions and answers.

Highlights of Committee Activities

The Biomedical Library Review Committee met three times in FY 1994 and reviewed 132 applications; 100 were recommended for further consideration. The Committee operates as a "flexible" review group, i.e., it is composed of three standing subcommittees: Medical Library Resource Subcommittee, Medical Informatics Subcommittee, and Biotechnology Information Subcommittee.

A final peer review of applications is performed by the Board of Regents, which meets three times a year, approximately three months after the Biomedical Library Review Committee. One of the Board's subcommittees, the Extramural Programs Subcommittee, meets the day before the full Board for the review of "special" grant applications. Examples of "specials" include applications for which the recommended amount of financial support is larger than some predetermined amount; at least two members of the scientific merit review group dissented from the majority; a policy issue is identified; or an application is from a foreign institution. The Extramural Programs Subcommittee makes recommen-

dations to the full Board which votes on the applications.

Special Activities

Mrs. Francis Howard, as guest or speaker, represented EP and NLM at a number of federal and private sector organizations for purposes of obtaining information of use to NLM and for providing information about NLM to others. A partial listing of organizations involved includes the World Health Organization, Pan American Health Organization, National Health Council, Woodrow Wilson Center for Scholars, Research America, National Endowment for the Humanities, Cystic Fibrosis Foundation, U.S. Congressional Wives, Center for Disease Control, and the Special Olympics Committee.

Plans for FY 1995

- The retirement of the Division's Administrative Officer/Grants Management Officer left a vacancy that could not be filled because of hiring restrictions. Accordingly, the position was split, and recruitment is now under way to hire someone from within PHS for the newly defined Administrative Officer slot. The incumbent will be expected to serve as computer systems officer, and office manager in addition to the customary administrative duties. Grants Management, which now has two staff specialists to do the work formerly done by four, will need to have additional help during FY 1995. Ways and means are now under consideration.
- The biotechnology database RFA, initiated in FY 1994, will be completed with one to two more awards in FY 1995.
- If additional funds are provided, four more grants will be funded for the Cooperative Agreement on Electronic Medical Record Systems. Several meetings of the working committee of all P.I.s and two federal officers will be held during the year to ensure intercompatibility of the various projects.
- The feasibility of a grant program as followup to the recent NLM Panel on the Education and Training of Health Science Librarians is being evaluated by EP.
- Also in progress is the transfer to EP of an educational specialist with particular interest in nursing. It is expected that new programs and/or modifications of existing programs will come about.
- The Internet Connections Program, previously administered by NSF for the National Library, will be administered by EP beginning in FY 1995.

Table 11**Extramural Grant and Contract Program
(dollars in thousands)**

<i>Category</i>	<i>FY 1992</i>		<i>FY 1993</i>		<i>FY 1994</i>	
	<i>No.</i>	<i>\$</i>	<i>No.</i>	<i>\$</i>	<i>No.</i>	<i>\$</i>
Resource projects	21	5,295	25	4,878	38	5,150
IAIMS	(3)	(3,938)	(5)	(3,150)	(10)	(3,115)
Access	(7)	(493)	(9)	(589)	(2)	(80)
Systems					(9)	(1,283)
Connection					(17)	(672)
Research	51	11,863	51	11,674	53	12,436
Informatics*	(29)	(6,894)	(31)	(6,899)	(19)	(5,681)
Biotechnology**	(13)	3,759	(13)	(3,915)	(13)	(3,776)
Cooperative agreements	(3)	(3,938)	(5)	(3,150)	(3)	(1,008)
Career awards					(16)	(1,725)
Library science					(2)	(246)
Training	19	3,818	16	4,138	19	4,372
Institutional	(10)	(3,520)	(10)	(3,878)	(10)	(3,980)
Fellowship	(9)	(298)	(6)	(260)	(9)	(392)
Publications	10	283	9	269	8	239
Bioethics					1	400
SBIR/STTR					2	145
Regional Medical Library	8	5,482	8	5,500	8	5,678
Totals:	109	\$26,741	109	\$26,459	129	\$28,420

* Includes chairman's grants

** Includes two interagency agreements

Office of Computer and Communication Systems

Fernando Burbano
Director

Major Developmental Activities

Developmental activities during the past year include: a major new release for the Technical Services System (TESS), a new release of Grateful Med and expansion of Grateful Med services, extensions to the Local Area Network (LAN) services, Information Systems Laboratory (ISL) projects, a new NLM Locator/Collection Management System, and starting the development of a Z39.50 server for NLM, and enhancements to various NLM program support systems.

TESS

The Technical Services System (TESS) is being developed to integrate various functions of the Technical Services Division. TESS is being developed as a distributed processing system that is comprised of mainframe computer, personal computer, database, and local area network (LAN) technologies.

TESS development started in 1989. The initial implementation under TESS provided for the creation and maintenance of original cataloging. The next major release of TESS provided the framework for the integration of acquisition and cataloging activities (1990). In 1991, authority control for the cataloging function was integrated into TESS. This release included the capability for creation and maintenance of the Name Authority File.

Late in 1993 a major release of TESS included the conversion of the entire CATLINE/AVLINE files to be supported within the TESS environment. TESS was enhanced to support all file maintenance activities of CATLINE/AVLINE, and to support all cataloging distribution activities, principally the distribution of catalog records in MARC format to MARC subscribers. Work on this release began in 1992 and continued through 1993. This release of TESS was implemented in December 1993. Additional efforts to support MARC distribution of data continued in 1994.

Grateful Med

During 1994, over four million ELHILL searches were performed using Grateful Med. The number of direct or command-language searchers decreased from 1994 with almost 80% of all ELHILL users accessing NLM with Grateful Med. Also, the number of Grateful Med sessions using TCP/IP increased to 36% by the end of 1994. Grateful Med sales increased during 1994. There were 10,828 PC versions and

2,894 Macintosh versions of Grateful Med sold during this time. An update for each version was created in 1994. These updates were distributed to 35,000 PC users and 9,200 Macintosh users.

The development of a windowing version of Grateful Med continued during 1994. The Graphical User Interface (GUI) was designed by an NLM working group with input from the Human-Computer Interaction Laboratory and Center for Automation Research at the University of Maryland. The implementation of this interface, as well as the implementation of the "search," "communications," and "preferences" modules, were started in 1994. It is anticipated that a limited (only MEDLINE access) windowing version of Grateful Med will be available by the end of 1995.

The two contracts providing Grateful Med software development support were extended to October 31, 1995. A Request for Contract Action (RFCAs) for a recompetition for Grateful Med support was submitted. The new contract(s) will be for 4 years and should be awarded by November 1, 1995.

Local Area Networks

Local Area Networks (LANs) form the internal communications systems of NLM. 10BaseT (10 Mbps) Ethernet segments, linked together with 100Mbps FDDI (Fiber Distributed Data Interface) backbones, form the network topologies throughout NLM. Novell IPX, TCP/IP, and Appletalk protocols transport data over the network topologies among the various NLM computer systems. Together, these LANs support approximately 700 NLM staff, contractors, and visiting patrons. They provide access to all NLM data processing resources as well as access to external computer networks and data systems. Access is provided not only to MEDLARS, but also to file servers, minicomputers, and other systems used for research, development, and office automation. As computing systems continue to become more distributed and powerful, increasing demands are placed on the communications facilities in terms of speed and throughput.

During FY 1993, NLM commenced a program to upgrade its LAN systems from an older coaxial broadband facility to deploy newer technology. This project was completed in FY 1994. The purpose was to accommodate the requirements for growth, higher speeds, better reliability and management, and more efficient sharing of resources. Buildings 38 and 38A were rewired with level 5 unshielded twisted pair (UTP) copper for 10BaseT Ethernet. Dark (spare) fiber optic cable was installed to the desktop in most hard-to-wire areas of Building 38. A fiber optic backbone links hubs to the main computer room, where an FDDI backbone provides connectivity to other local and remote communications links such as the Internet and the NIHNet. This system will be enhanced and expanded over the years to continue to meet communications requirements for NLM.

Information Systems Laboratory

The Information Systems Laboratory (ISL) was created within the Development Branch in 1991. It is intended as a core facility to help OCCS modernize and enter the emerging technology domains of distributed processing, open systems, high speed networks, and worldwide connectivity and service provision. The ISL released its first software product, TC_COMM V1.0, in September 1992.

During FY 1993, the ISL supported the development and implementation of Locator, the NLM public access catalog; NLM PUBLS, the anonymous FTP service providing online copies of NLM technical publications; and Implement, a meta-DBMS toolkit designed to address the special problems of bibliographic data storage and retrieval. The ISL also supported developments for remote cataloging and indexing activities.

During FY 1994 the ISL is continuing to introduce open systems computers and workstations to support operational requirements. Efforts continue to redesign existing systems and develop new systems to use multiplatform open system servers, TCP/IP communications and Internet connectivity. The Internet is becoming a major domestic and international access pathway.

NLM Locator

In 1991 OCCS developed a prototype Online Public Access Catalog (OPAC) which became a full-scale development effort in FY 1992. The operational system, named NLM Locator, was introduced to the Reading Room on February 22, 1993. In May 1993 the system was made available over the Internet and usage continues to grow. NLM Locator provides direct access to the NLM collections through the ELHILL databases CATLINE (monographs), AVLINE (audiovisuals), SERLINE (serials), and DIRLINE (information resources) which was added in 1994.

This project was OCCS's initial Unix-based client/server development effort, and the development team was challenged to acquire skills in new technologies. The complete project included mastering new skills in new computer hardware, system software, networking, programming languages, and a number of utility tools.

One of the important concepts of the client/server architecture is the ability to adapt to change without re-engineering the application. NLM Locator utilizes a workstation client communicating with function servers which in turn communicate with a data server. The client workstations in the Reading Room are DOS PC's, while Internet users log in as VT-100 terminal sessions to a Unix client process executing in the function servers. The function servers are Sun and IBM computer systems running the Unix operating system. Having multiple function servers permits reliability as well as additional capacity, should usage require it. If necessary, another function server could be added without software modification. The data server is the NLM mainframe comput-

er utilizing the ELHILL retrieval system. No changes to the legacy systems were required to implement NLM Locator.

The system continues to function well and comments received via the electronic mail feature have been most complimentary. Reading Room and NLM staff usage has stabilized. Internet usage continues to grow with an average of 20 first time NLM Locator users per day.

The second phase of the project was to provide circulation control and collection management features for a combined system: the LOCATOR/Collection Management System (LOCATOR/CMS). This effort was completed in 1994 and it provides online patron registration, availability data, status information, requests for library materials, and a great number of management reports to the Public Services Division.

NLM Z39.50 Project

The project to make MEDLINE available to Internet clients via the Z39.50 protocol continues on schedule. Coordinated by OCCS since September 1994, the development effort has been reaching its milestones in all three areas: the ELHILL retrieval agent (the "back-end" that interfaces with the Z39.50 engine); the Z39.50 protocol engine (the "front-end" that delivers MEDLINE information over the network in Z39.50 format); and dynamic ELHILL record conversion.

The ELHILL retrieval agent supports simple searching and full Boolean searching. It has been ported to run under Solaris. A primitive line-mode client program has been adapted from the freely available Stanford client and used in testing with different negotiated parameters, such as record formats, buffer sizes, and record composition. The server protocol engine comprises some 9000 lines of C program code.

A simple World Wide Web-based front-end has also been devised for this test client. The table-driven ELHILL record conversion procedures have been modified for the delivery of "brief" records. A MARC-encoding for MEDLINE has been devised.

In May the server went into beta-test on a separate machine and has been shown to interoperate with at least five independently developed Z39.50 client programs.

Programming Support Services

The Office of Computer and Communications Systems supports the various NLM programs and serves as the nucleus of all automated programming support services. In FY 1994:

(1) Enhancements to the DOCLINE Interlibrary Loan system, NLM's online facility for requesting library materials, consisted of a full panel version of DOCLINE with help features that was installed as a more efficient alternative to the line-by-line terminal version. Enhanced system administrator

functions were a by-product of this effort. Also, software was modified to allow the World Health Organization to enter requests from Geneva and be billed at the domestic rate.

(2) A set of software enhancements was installed to streamline **Collection Access Section (CAS)** operations:

- Barcodes are now printed on NLM receipts to reduce keyboarding time. The receipts are sorted for maximum efficiency and requests are handled electronically whenever possible by interpreting comments. Requests are referred to History of Medicine Division for material before 1914.
- Referral of network requests not filled by NLM is now handled electronically by software that finds a potential lending library from holdings files.
- UPS mailing labels are now produced at noon so material can be mailed out the same day that the request is received.
- For better quality control, programs were written to produce daily reports about keyboarder activity and requests sent to NLM without call numbers.
- A capability for CAS to send standard messages through DOCLINE without retyping was installed.

(3) Major enhancements to the **publications systems** consisted of:

- Geographic print headings were added to the AIDSLINE bibliography.
- ASCII, Postscript, and WordPerfect versions of the *List of Journals Indexed* and the *List of Serials Indexed for Online Users* were made available on the NLM PUBLIS server.
- MeSH publications were produced with larger print.
- Several major catalog publications were discontinued.

(4) Major enhancements to subsystems of the **Automated Indexing Management System (AIMS)** became operational. Some of the subsystems affected were: journal control, indexing, CHECKIN, bibliographic processing, binding, and gapping. AIMS is an IBM mainframe computer application that runs under the Customer Information Control System (CICS). It provides access to the Inquire Data Base Management System for storage/retrieval of new records or old records to be maintained. The data entered, verified, and validated is NLM bibliographic data that becomes part of the MEDLARS databases and associated publications. Items like a capability to interactively view master serials, binding or gapping data while being attached to other data entry sub-

systems such as the CHECKIN subsystem was implemented. The online, interactive, full-screen processing mode of the various AIMS input subsystems greatly facilitates timely capturing and accuracy of NLM bibliographic data.

(5) More than 10% of the NLM bibliographic data of over 7 million records) were class maintained this year. **Class maintenance** is the adding of new terms, deleting old terms and replacing terms with more preferred ones in the MEDLARS data base records. Moreover, new data fields are introduced to the records as required.

(6) NLM is collaborating with Springer-Verlag (publishers) and writing conversion software to convert their journal citation data encoded in the **Standard Generalized Markup Language (SGML)** to MEDLARS II format as an alternative to keyboarding the data. SGML is part of the International Standards Organization (ISO) initiative.

(7) Enhanced versions of client-file server **Citation Maintenance Systems (CMS)** became available this year. These systems use templates for creating and editing individual citations on the client side. The editor software is written in "C." Additionally, the client software is DOS based. Individual citations to be maintained are extracted from the IBM mainframe files, formatted into a MARC-like format and transmitted to the PC for maintenance. The following CMS were re-engineered with additional functionality or became new systems:

- **Bioethics Citation Maintenance System (BCMS)**—provides its users with the ability to import citations from MEDLINE and the HEALTH database, adds additional controlled vocabulary, and stores them on its own network holding file. It will import *non*-NLM citations to the Bioethics database. This will reduce typing and improve quality of records. The new system also operates on common platforms as other Citation Maintenance Systems.
- **History of Medicine Citation Maintenance System (HCMS)**—provides its users with the ability to prepare and transmit new or modified bibliographic information to the HISTLINE database. The HCMS communicates directly with the MEDLARS on NLM's IBM mainframe computer. Communication between the PC-based HCMS application and the MEDLARS is effected through the use of TCP/IP. In addition to its own input, the HCMS imports citations from MEDLINE, CATLINE, AVLINE, and HEALTH, adds additional control vocabulary, and stores them on its own network holding file.
- **Cumulated Index Medicus Citation Maintenance System (INDX65)**—provides its users with the ability to input the bibliographic data from the 1965 *Cumulated Index Medicus* into a DBMS file. A publicly available database will be built

with data derived from INDX65 CMS. The INDX65 can be considered an independent system, since it does not communicate directly with MEDLARS and operates in a stand-alone mode on a local network. It communicates with the holding file stored on the network file server.

- **Health/Services Research Citation Maintenance System**

(RCMS)—provides its users with the ability to prepare and transmit new or modified bibliographic information to the HSTAR database. The system creates and maintains information on research projects from a variety of smaller organizations and agencies. It also provides the flexibility needed to maintain the overall database.

ADMINISTRATION

Kenneth G. Carney
Executive Officer

Streamlining

The National Performance Review and the President's initiative to reduce the size of the government challenge agencies to develop a smaller yet responsive workforce that works better and costs less. These do not represent altogether new challenges for the NLM. The Library continually seeks improved methods to accomplish the increasing demands for its programs and services with reduced staff, and to make its services even more responsive to the nation's health information needs. Nevertheless, over the next several years the Library must reduce its staff, increase its supervisory ratio, reduce the number of positions in certain designated categories, and modernize its information systems. To approach to this streamlining task, a small working group has been established, co-chaired by the Deputy Director and the Associate Director for Library Operations. The working group will gather information, hear recommendations from staff at all levels, and make its report in FY 1995.

Reinvention Initiative

The National Performance Review also identified a set of recommendations to improve services to the public. Federal agencies were asked to establish "reinvention labs" to function as test sites for agency streamlining and deregulation. The NLM was identified as such a test site and is now involved in the NLM System Reinvention Initiative. This initiative will redesign existing NLM computer systems to take advantage of new directions in software, hardware, and communications that now make it possible to provide new functions that will help users in ways not available under earlier circumstances. A Systems Reinvention Coordinating Committee, an Executive Bureau, and working groups with representatives from all divisions of NLM will focus on making the transition to new information systems as smooth as possible for both users and staff.

Financial Resources

In FY 1994, the Library had a total appropriation of \$118,019,000. Table 12 displays the FY 1994 budget authority plus reimbursements from other agencies, and the allocation of these resources by program activity.

TABLE 12

Financial Resources and Allocations, FY 1994
(dollars in thousands)

Budget Authority:	
Appropriation, NLM	\$118,019
Plus: Reimbursements	11,849
Total	\$129,868
Budget Allocation:	
Extramural Programs	28,419
Intramural Programs	93,107
Library Operations	(56,049)
Lister Hill National Center for Biomedical Communications	(20,426)
National Center for Biotechnology Information	(9,196)
Toxicology Information	(7,436)
Research Management and Support	8,342
Total	\$129,868

Personnel

The NLM closed FY 1994 with 597 full-time equivalents (FTE). For most of FY 1994 the Library operated under an employment freeze imposed by the Public Health Service. A hiring and promotion freeze into the senior grade levels (GS-14 and above) was also imposed by PHS during this period.

In May 1994, the NIH received authority to offer buyouts under the Voluntary Separation Incentives Program to employees at the GS-13 and above grade levels. Four NLM employees took advantage of the offer.

In August 1994 the reorganization of the Public Services Division of Library Operations was effective. The reorganization resulted in the movement of the collection maintenance function and various employees into the Preservation Section, which was renamed the Preservation and Collection Management Section.

On September 1, 1994, Dr. Daniel Masys, Director of the Lister Hill National Center for Biomedical Communications, retired from the PHS Commissioned Corps and his position at the NLM. Dr. Masys served as Director of the LHCNBC since July 1986. Dr. Harold M. Schoolman, Deputy Director for Research and Education, was appointed to serve temporarily as Acting Director in addition to his regular duties.

Appointments

Fernando Burbano joined the NLM as a member of the Senior Executive Service staff and serves as Director, Office of Information Systems. Mr. Burbano also serves as the Director, Office of Computer and Communication Systems.

Colombe Chappey, Ph.D., was appointed a Visiting Fellow with the National Center for Biotechnology Information. Dr. Chappey is a graduate of the University of Paris where she carried out her thesis work in developing computer methods for assessing the variability of related amino acid sequences. While with the NCBI, Dr. Chappey is conducting research in the area of virus and molecular evolution.

Shu-Chin Wu, Ph.D. was also appointed as Visiting Fellow with the NCBI. Dr. Wu received his Ph.D. in chemical engineering from the Veterinary Pathobiology Department at Texas A & M University. Dr. Wu is an expert in the laboratory methods relevant to immunodeficiency viruses and the computer-assisted mathematical modeling of viral infection.

Jean-Francois Gibrat, Ph.D. joined the NLM as a Visiting Associate with the NCBI. Dr. Gibrat received his Ph.D. in physical chemistry from the University of Paris of Marie Curie where he conducted his thesis work on the computer modeling of 3-dimensional structure of proteins. At NCBI, Dr. Gibrat is developing programs for 3-dimensional structure of proteins.

The NCBI has recruited four additional Postdoctoral Fellows and one Predoctoral Fellow through the Intramural Research Training Award (IRTA) Program. They are

- Thomas Madej, Ph.D., appointed a Postdoctoral IRTA on December 1, 1993. He received his Ph.D. in computer science from the University of Illinois at Urbana-Champaign. While with the NCBI, Dr. Madej is working on the linking of sequence and 3-D structure databases.
- Andreas Baxevanis, Ph.D., appointed a Postdoctoral IRTA on June 1, 1994. Dr. Baxevanis received his Ph.D. in biochemistry from Johns Hopkins University. While at NCBI, he is conducting research on high-mobility group, histone, and other families of nucleoproteins.
- Douglas Bassett, appointed a Predoctoral IRTA on June 1, 1994. Mr. Bassett is a second-year student in the Predoctoral Training Program in Human Genetics at Johns Hopkins University. At NCBI, Mr. Bassett conducts research in the designing a relational database for comparative mapping data for yeast, murine, and human genes.
- Heidi Sofia, Ph.D., appointed a Postdoctoral IRTA on July 1, 1994. She received her Ph.D. in biochemistry from

the University of Wisconsin at Madison. While at NCBI, Dr. Sofia is performing both basic and applied research in computational molecular biology.

- Myung S. Chung, Ph.D. appointed a Postdoctoral IRTA on August 1, 1994. She received her Ph.D. in Mathematics from the University of Illinois at Urbana-Champaign. While at the NCBI, Dr. Chung performs research in the protein folding problem using graph theory and other mathematical and statistical methods.

Dr. Stephen Altschul was converted to a permanent appointment as a mathematician with NCBI. Dr. Altschul had been serving as a Senior Staff Fellow with the Center since May 1989. Dr. Altschul is responsible for initiating and executing research on the analysis and comparison of biological sequences.

Awards

The NLM Board of Regents Award for Scholarship or Technical Achievement was awarded to Lawrence E. Hunter, Lister Hill Center, for outstanding achievements in the field of artificial intelligence and for conception, inspiration, and contribution to the book *Artificial Intelligence and Molecular Biology*.

The PHS Commendation Medal was presented to Richard P. Rodgers, M.D. for the creation of advanced demonstration projects for network-based information discovery and retrieval, with exemplary use of graphical interfaces to multimedia resources.

The NLM Director's Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to two employees this year: Alvin J. Barnes, Division of Library Operations, and Dr. Milton Corn, Division of Extramural Programs. Mr. Barnes was recognized for furthering NLM's mission by applying remarkable library science skills on behalf of the public and library staff. Dr. Corn was recognized for his contributions to the Nation's health through vigorous and exemplary leadership of NLM's Extramural Programs.

The NIH Merit Award was presented to six employees this year: Angela B. Ruffin, Ph.D., Catherine Soehner, and Christa F. B. Hoffmann with the Division of Library Operations, Richard K. C. Hsieh, D.P.H., Office of the Director, Theodore E. Youwer, Office of Administration, and David Nash, Office of Equal Opportunity. Dr. Ruffin was recognized for her exceptional contributions to the outreach programs of the NLM. Ms. Soehner was recognized for her leadership and management skills promoting utilization of biomedical information by health professionals using Grateful Med. Ms. Hoffmann was recognized for significant leadership in developing policies and practices that have substantially increased NLM's cataloging output and the currency and availability of national level bibliographic records. Dr.

Hsieh was recognized for his accomplishments to promote worldwide access to the NLM's information resources, particularly in developing countries. Mr. Youwer was recognized for demonstration of superior managerial skills which greatly enhanced the quality of support services provided the NLM. Mr. Nash was recognized for significant ingenuity and leadership in advancing equal opportunity for all employees.

Special EEO Initiative

On February 7, 1994, the NLM and the Calvin Coolidge High School signed a "Declaration of Partnership." The Library is now one of more than 300 organizations providing partnership to an inner-city school in the District of Columbia Public School System under the "adopt-a-school" program—Partners in Education. NLM EEO Officer David Nash and various program staff, including Cynthia Gaines of

the Specialized Information Services (and former chair of the NLM EEO Committee) helped to create the program, under which NLM will assist the high school in establishing an Internet node and to use Internet-accessible resources (including NLM's databases). NLM staff participated in individual tutoring, guest lecturing, mentoring, and curriculum development. In addition, ten students and three faculty members were provided summer appointments at NLM.

NLM will also help the school to improve its basic library collection in science and technology and will collaborate with the Coolidge faculty and staff in a variety of programs designed to encourage students to pursue careers in science and medicine. A Media Resource Center will be opened at the school in FY 1995. In recognition of the success of the program the NLM was awarded the "Outstanding Volunteer Service Award" from the District of Columbia Public Schools.

Table 13

Staff, FY 1994 Full-Time Equivalents

Program	Full-Time Permanent	Other
Office of the Director	17	1
Office of Public Information	6	1
Office of Administration	53	11
Office of Computer and Communications Systems	60	7
Extramural Programs	12	5
Lister Hill National Center for Biomedical Communications	72	7
National Center for Biotechnology Information	20	15
Specialized Information Services	34	3
Library Operations	242	48
Total	516	98
TOTAL FTEs	614	

Appendix 1: Regional Medical Libraries in the National Network of Libraries of Medicine

- | | | | |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. | <p>MIDDLE ATLANTIC REGION
 The New York Academy of Medicine
 2 East 103rd Street
 New York, NY 10029
 (212) 876-8763 FAX (212) 534-7042
 States served: DE, NJ, NY, PA</p> | 5. | <p>SOUTH CENTRAL REGION
 Houston Academy of Medicine-Texas
 Medical Center Library
 1133 M.D. Anderson Boulevard
 Houston, TX 77030
 (713) 790-7053 FAX (713) 790-7030
 States served: AR, LA, NM, OK, TX</p> |
| 2. | <p>SOUTHEASTERN/ATLANTIC REGION
 University of Maryland at Baltimore
 Health Sciences Library
 111 South Greene Street
 Baltimore, MD 21201-1583
 (410) 706-2855 FAX (410) 706-0099
 States served: AL, FL, GA, MD, MS,
 NC, SC, TN, VA, WV, DC, Virgin
 Islands, Puerto Rico</p> | 6. | <p>PACIFIC NORTHWEST REGION
 University of Washington
 Health Sciences Center Library, SB-55
 Seattle, WA 98195
 (206) 543-8262 FAX (206) 543-2469
 States served: AK, ID, MT, OR, WA</p> |
| 3. | <p>GREATER MIDWEST REGION
 University of Illinois at Chicago
 Library of the Health Sciences
 P.O. Box 7509
 Chicago, IL 60680
 (312) 996-2464 FAX (312) 996-2226
 States served: IA, IL, IN, KY, MI,
 MN, ND, OH, SD, WI</p> | 7. | <p>PACIFIC SOUTHWEST REGION
 University of California, Los Angeles
 Louise Darling Biomedical Library
 10833 Le Conte Avenue
 Los Angeles, CA 90024-1798
 (310) 825-1200 FAX (310) 825-5389
 States served: AZ, CA, HI, NV and U.S.
 Pacific Territories</p> |
| 4. | <p>MIDCONTINENTAL REGION
 University of Nebraska Medical Center
 Leon S. McGoogan Library of Medicine
 600 South 42nd Street
 Omaha, NE 68198-6706
 (402) 559-4326 FAX (402) 559-5482
 States served: CO, KS, MO, NE, UT, WY</p> | 8. | <p>NEW ENGLAND REGION
 University of Connecticut Health Center
 Lyman Maynard Stowe Library
 263 Farmington Avenue
 Farmington, CT 06030-4003
 (203) 679-4500 FAX (203) 679-1305
 States served: CT, MA, ME, NH, RI, VT</p> |

Appendix 2: Board of Regents

The NLM Board of Regents meets three times a year to consider Library issues and make recommendations to the Secretary of Health and Human Services on matters affecting the Library.

Appointed Members:

WALKER, H. Kenneth, M.D. (Chair)
Professor of Medicine
Emory University School of Medicine
Atlanta, GA

ALLEN, Beverly E.
Director, Multi-Media Center
Morehouse School of Medicine
Atlanta, GA

BOOKER, Naomi C.
Chair and President
Marketing and Management Innovations
Baltimore, MD

CORTEZ, Edwin M., Ph.D.
Asso. Professor, School of Library and Information Science
University of Wisconsin
Madison, WI

JOYNT, Robert J., M.D., Ph.D.
Vice President and Vice Provost for Health Affairs
University of Rochester
Rochester, NY

NEWTON, Carol M., M.D., Ph.D.
Professor of Biomathematics
University of California
Los Angeles, CA

PHILLIPS, Steven J., M.D.
Senior Heart Surgeon
Mercy Hospital Medical Center
Des Moines, IA

Ex Officio Members:

Librarian of Congress

Surgeon General
Public Health Service

Surgeon General
Department of the Air Force

Surgeon General
Department of the Navy

Surgeon General
Department of the Army

Under Secretary for Health
Department of Veterans Affairs

Assistant Director for Biological Sciences
National Science Foundation

Director
National Agricultural Library

Dean
Uniformed Services University of the Health Sciences

Appendix 3: Board of Scientific Counselors/ Lister Hill Center

The Board of Scientific Counselors meets periodically to review and make recommendations on the Library's intramural research and development programs.

Members:

MUN, Seong Ki, Ph.D. (Chair)
Director, Division of Imaging Physics
Georgetown University
Washington, D.C.

BRINKLEY, James F., M.D., Ph.D.
Research Assistant Professor
Department of Biological Structure
University of Washington
Seattle, WA

CIMINO, James J., M.D.
Assistant Professor of Medicine
College of Physicians and Surgeons
Columbia University
New York, NY

HUNTLEY, Joan S., Ph.D.
Research and Development Project Leader
Weeg Computing Center
Iowa City, IA

KAHN, Michael G., M.D., Ph.D.
Assistant Professor of Medicine
Division of Medical Informatics
Washington University
St. Louis, MO

LEHNERT, Wendy G., Ph.D.
Professor of Computer and Information Science
Department of Computer and Information Science
University of Massachusetts
Amherst, MA

MITCHELL, Joyce A., Ph.D.
Director, Information Science Group
University of Missouri-Columbia
Columbia, MO

PETERSON, George D., Ph.D.
Asst. Vice President for Academic Affairs
Morgan State University
Baltimore, MD

Appendix 4. Board of Scientific Counselors/ National Center for Biotechnology Information

The National Center for Biotechnology Information Board of Scientific Counselors meets periodically to review and make recommendations on the Library's biotechnology-related programs.

Members:

SAUER, Robert T., Ph.D. (Chairman)
Professor, Department of Biology
Massachusetts Institute of Technology
Cambridge, MA

CANTOR, Charles R., Ph.D.
Director, Center for Advanced Research in Biotechnology
Boston University
Boston, MA

FITZGERALD, Paula, M.D., Ph.D.
Senior Research Fellow
Department of Biophysical Chemistry
Merck Sharp & Dohme
Rahway, NJ

HUNKAPILLER, Michael W., Ph.D.
Executive Vice President
Applied Biosystems Division
Perkin-Elmer Corporation
Foster City, CA

WILLIAMS, Myra N., Ph.D.
Vice President, Information Technology
Glaxo Research Institute
Triangle Park, NC 27709

Appendix 5. Biomedical Library Review Committee

The Biomedical Library Review Committee meets three times a year to review applications for grants under the Medical Library Assistance Act.

Members:

VANNIER, Michael W., M.D.
(Chair)
Professor of Radiology
Washington University
St. Louis, MO

BROERING, Naomi C.
Director, Biomedical Info. Resource
Center
Dahlgren Memorial Library
Georgetown University Medical
Center
Washington, D.C.

BUNTING, Alison
Associate University Librarian for
Science
Louise Darling Biomedical Library
University of California
Los Angeles, CA

CLEVELAND, Ana D., Ph.D.
Professor of Information Science
School of Library and Information
Sciences
University of North Texas
Denton, TX

EZQUERRA, Norberto F., Ph.D.
Associate Professor
College of Computing
Georgia Institute of Technology
Atlanta, GA

FIELDS, Christopher A., Ph.D.
Laboratory Director
Comparative Genomic Department
The Institute for Genomic Research
Gaithersburg, MD

IYENGAR, S. Sitharama, Ph.D.
Professor and Chairman of Computer
Science

Louisiana State University
Baton Rouge, LA

JACKSON, Sara Jean
Director, Research Medical Library
M.D. Anderson Cancer Center
Houston, TX

KULIKOWSKI, Casimir A., Ph.D.
Professor of Computer Science
Rutgers University
New Brunswick, NJ

LOVE, Erika
Director, Medical Center Library
University of New Mexico
Albuquerque, NM

**MAVROVOUNIOTIS, Michael L.,
Ph.D.**
Associate Professor
Chemical Engineering Department
Northwestern University
Evanston, IL

MUSEN, Mark A., M.D., Ph.D.
Assistant Professor
Depts. of Medicine and Computer
Science
Stanford University
Stanford, CA

PEAY, Wayne J.
Director, Eccles Health Science
Library
University of Utah
Salt Lake City, UT

RANKIN, Jocelyn A.
Professor of Library Science and
Director, Medical Library and
Learning Resources Center
School of Medicine
Mercer University
Macon, GA

**REYNOLDS, Robert E., M.D.,
Dr.P.H.**
Senior Associate Vice President for
Health Sciences
University of Virginia
Charlottesville, VA

ROBERTS, Richard J., Ph.D.
Research Director
New England Biolabs
Beverly, MA

ROE, Bruce A., Ph.D.
Professor of Chemistry and Biochem
istry
University of Oklahoma
Norman, OK

ROSSE, Cornelius M. Ph.D., D. Sc.
Professor and Chairman
Department of Biological Structure
School of Medicine
University of Washington
Seattle, WA

SEARLS, David B., Ph.D.
Research Associate Professor
Department of Genetics
University of Pennsylvania
School of Medicine
Philadelphia, PA

VRIES, John K. M.D.
Assoc. Vice Chancellor for Medical
Informatics
University of Pittsburgh Medical
Center
Pittsburgh, PA

WRIGHT, Barbara A.
Director, Library and Information
Services
Fayetteville Area Health Education
Center
Fayetteville, NC

Appendix 6. Literature Selection Technical Review Committee

The Literature Selection Technical Review Committee meets three times a year to select journals for indexing in *Index Medicus* and MEDLINE.

Members:

WEAVER, William Lynn, M.D., F.A.C.S. (Chair)
Chief, Surgical Service
VA Medical Center
Buffalo, NY

ALTAMORE, Rita Ann, M.D., M.P.H.
Senior Lecturer
Department of Health Services
University of Washington
Seattle, WA

COLON, Angel Rafael Jr., M.D.
Professor of Pediatrics
Georgetown Univ. School of Medicine
Washington D.C.

FURANO, Anthony V., M.D.
Chief, Section on Genomic Function and Structure
Laboratory of Biochemical Pharmacology
Nat. Inst. of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Bethesda, MD

GOLDBERG, Herbert S., Ph.D.
Associate Dean for Research and Academic Affairs
University of Missouri
Columbia, MO

HOAK, John C., M.D.
Director, Div. of Blood Diseases & Resources
National Heart, Lung, and Blood Institute
Bethesda, MD

PINCUS, Harold A., M.D.
Director, Office of Research
American Psychiatric Association
Washington, DC

PITTMAN, Constance Shen, M.D.
Professor of Medicine
University of Alabama
Birmingham, AL

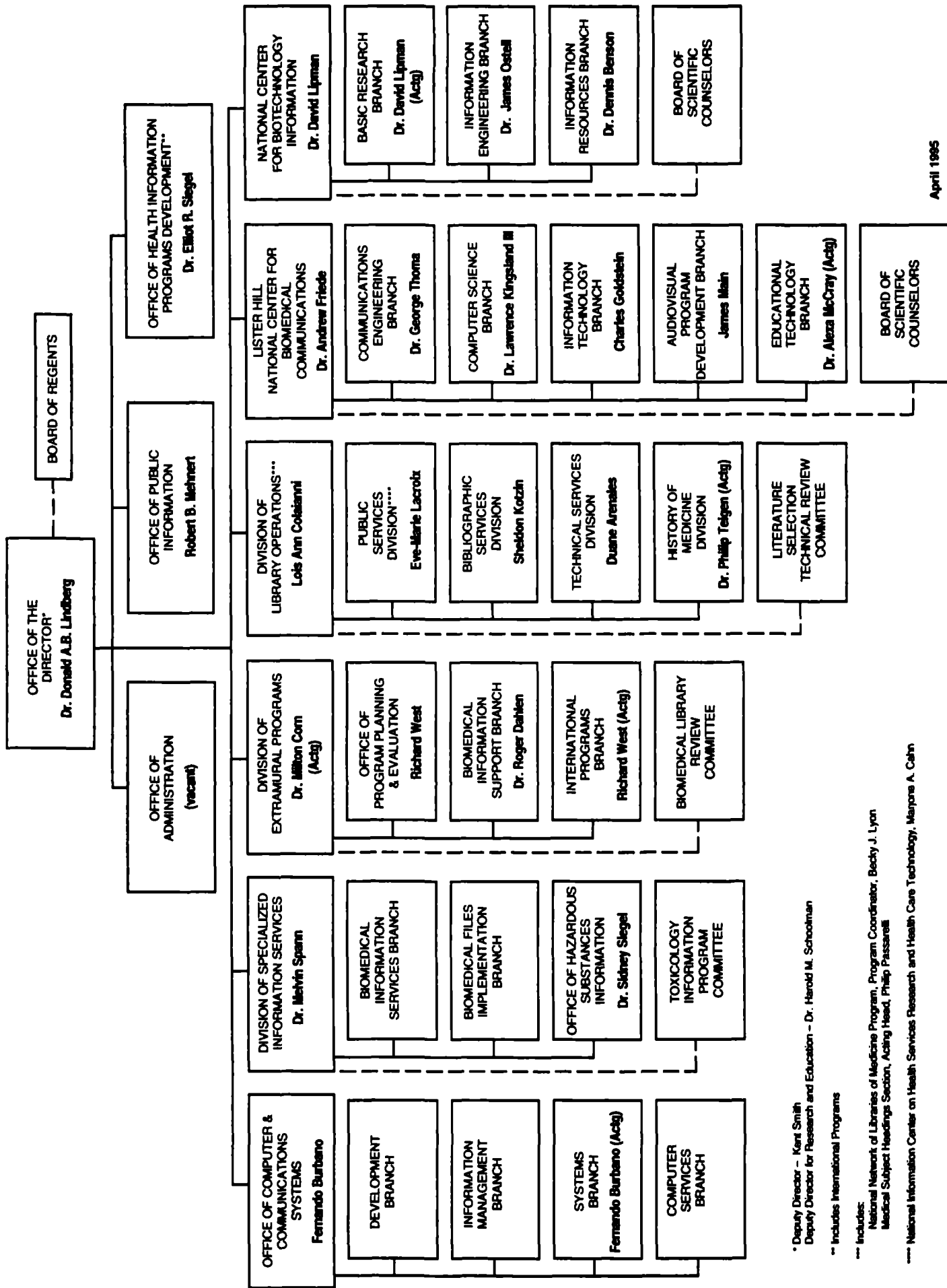
THOMAS, John A., Ph.D.
Vice President for Academic Affairs
University of Texas Health Science Center
San Antonio, TX

WALTER, Pat L.
Acting Biomedical Librarian
Louise Darling Biomedical Library, UCLA
Los Angeles, CA

WHITE, Charles A., M.D.
Professor and Head Emeritus
Department of Obstetrics and Gynecology
School of Medicine
Louisiana State University
New Orleans, LA

WILSON, Frank C., M.D.
Professor and Chief of Orthopaedics
University of North Carolina
School of Medicine
Chapel Hill, NC

National Library of Medicine



* Deputy Director - Kent Smith
Deputy Director for Research and Education - Dr. Harold M. Schootman

** Includes International Programs

*** Includes:
National Network of Libraries of Medicine Program, Program Coordinator, Becky J. Lyon
Medical Subject Headings Section, Acting Head, Philip Passarelli

**** National Information Center on Health Services Research and Health Care Technology, Maryona A. Cahn

