NATIONAL INSTITUTES OF HEALTH

NATIONAL LIBRARY OF MEDICINE

Programs and Services
Fiscal Year 1997

U.S. Department of Health and Human Services
Public Health Service
Bethesda, Maryland

National Library of Medicine Catalog in Publication

Contents

Appendixes

Tables

Preface

The most important event of the 1997 fiscal year involving the National Library of Medicine was the unveiling, on June 26, of "free MEDLINE" via the World Wide Web. The announcement was made in the U.S Capitol at a public forum sponsored by Senators Arlen Specter (R-PA) and Tom Harkin (D-IA). Vice President Al Gore did the ceremonial "first free MEDLINE search."

A new feature introduced was the ability to link a MEDLINE user who wishes to get a full article directly to the home page of a number of medical publishers, where articles may be viewed or ordered. So far, about 100 journals are so linked to MEDLINE and the number continues to increase. The search system that makes this possible is PubMed, which, together with the Internet Grateful Med is now providing most of NLM's MEDLINE access. By the end of Fiscal Year 1997, it was estimated that the rate of MEDLINE searching had increased tenfold over previous years.

Other significant events this year were the introduction of the Human Gene Map, the logging of the one billionth base to the GenBank DNA sequence database, the introduction of the National Cancer Institute's Cancer Genome Anatomy Project (in which NLM scientists played an important role), the funding of 19 telemedicine projects through the Lister Hill Center's Office of High Performance Computing and Communications, the introduction of a new toll-free number (1-888-FINDNLM) that puts callers quickly in touch with a variety of NLM service areas, and the welcoming of the State Central Scientific Medical Library in Moscow as the 21st International MEDLARS Center. These advances, and more, are described in this year's report.

I would like to express my appreciation to the staff of the Library for their dedication and hard work, to the community of health science librarians who provide the "front line" of medical information services to the health professions and to the public, and to the members of the Board of Regents and other NLM advisory groups who provide us with their expert counsel.

_____

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

Office of Health Information Programs Development

Elliot R. Siegel, Ph.D.
Associate Director

NLM Long Range Plan

In 1987 NLM's Board of Regents published a Long Range Plan to guide the Library in using its human, physical, and financial resources to fulfill its mission. NLM continues to use this plan in the day to day management of the Library and to update it in specific areas as necessary; planning reports prepared as supplements to the 1987 Plan contain recommendations on outreach to health professionals (1989), electronic imaging (1990), information services for toxicology and environmental health (1992), and the education and training of health science librarians (1995). The most recent planning effort examined NLM's international programs.

Planning Panel on International Programs

The Long Range Planning Panel on NLM International Programs, chartered by the Board of Regents and chaired by Donald S. Fredrickson, M.D., former NIH Director, met three times between September 1996 and June 1997. The purpose of the panel was to advise NLM on the relative priority of its international activities and responsibilities, and to assist in the development of appropriate strategies consistent with the Library's statutory mission and availability of resources. The panel had 24 distinguished members and a number of expert consultants in areas such as health sciences librarianship, electronic publishing, and telecommunications. The Panel framed its recommendations to the Board of Regents in terms of three major goals for NLM's future international activities:

- Strengthen and expand global access to the world's health-related literature;
- Chart new routes to biomedical knowledge generation and its utilization; and
- Enable NLM to fulfill its international mission

The report of the panel will be formally published in 1998.

International Programs

In FY 1997, the NLM conducted work with individual countries, international government organizations, and international nongovernmental organizations. International collaboration played a prominent role in NLM's leadership of the Multilateral Initiative on Malaria (MIM) Communication Working Group.

Other international activities included training for colleagues from abroad, exchanging publications program with libraries in other countries, and receiving numerous professional visitors from abroad.

Communications Research in Support of the Multilateral Initiative On Malaria

Three million deaths per year (mostly children), one death every 20 seconds, and a punishing impact on the health and economy of Africa—this is the cost of malaria as a re-emergent infectious disease in Africa. Traditional means of malaria prevention and treatment are failing due to drug resistance, insecticide resistance, and new and dramatically different patterns of disease transmission. International research agencies, donor organizations and scientists from Africa and other locations worldwide came together in Dakar, Senegal in January 1997 to plan a concerted effort to combat malaria in Africa—a new Multilateral Initiative on Malaria. One pressing need is to enhance the capacity of African scientists to communicate electronically with colleagues in Africa and around the world, and to access needed scientific information from local libraries, remote databases, and the Internet.

Communications in support of the scientific mission can take many forms. Investigators in basic research, for example, need access to genetic sequence databases such as GenBank and EMBL, maintained in the U.S., the U.K., and Japan. Developers of vaccines, drugs, and diagnostics need access to discussion groups for product development work and coordination of clinical trials. Fundamental to all research is access to the various published literatures identifiable through computer databases such as MEDLINE, BIOSIS and PASCAL. The NLM's new PubMed system supports Internet access to an increasing number of electronic versions of printed journal articles, most valuable where local library collections are poor. Epidemiologists engaged in vector control and surveillance programs use geographic information systems (GIS) and satellite-based remote sensing technologies. It is highly desirable that collaborating units maintain contact through e-mail and other visual means of communication that are becoming increasingly possible over the Internet.

At a follow-up funders meeting in July 1997 at The Hague, Netherlands, the NLM accepted the assignment to organize and lead a MIM working group on communications and to develop a strategic plan to

guide its efforts. In addition to infrastructure requirements relating to communications links, equipment and repositories, other MIM groups are addressing research priorities, manpower and training needs (WHO/TDR, Geneva, October 1997), and new drug development (Wellcome Trust, London, September 1997).

The strategy to enhance communications is to identify and work in African countries in which there is high scientific priority for MIM research funding. The communications objectives of MIM are to effect the "last mile" connection between the actual research center and the closest Internet node; establish means to support staff training in communications and information management at the research site; and to assure sustained funding for communications as part of the scientific mission.

In April 1997, NLM undertook a site visit to Mali and the Malaria Research and Training Center just outside the capital city of Bamako. It resulted in the selection of this NIAID/NIH-supported research facility as the initial pilot location at which to test this communications plan. The country's first Internet gateway node was recently installed there with funding under USAID's Leland Initiative. The procurement and installation of hardware linking the Center's laboratories and library to the gateway is presently under way.

In January 1998, a meeting of the Communications Working Group will be convened in Bethesda, Maryland at the NLM. The purpose will be to review and assess the pilot implementation of the communications plan in Mali, and to develop implementation and site visit plans for the additional countries that have been targeted as high priority in this first round. Scientific and technical persons from each of these locations will participate in plenary discussions and focused work groups. They will address issues of "connectivity," "content" and "training" from the perspective of basic science and/or applied research needs at each site.

An international consulting team is being engaged comprising individuals skilled in information management resources and training from Africa and the U.S. funding agencies already active at pilot project sites include NIH, Institut Pasteur, CDC, Wellcome Trust, US Walter Reed Army Institute of Research, JICA, USAID, WHO/TDR, and the World Bank.

International MEDLARS Centers

To assist international health professionals in accessing NLM MEDLARS databases, NLM has signed Memoranda of Understanding with partners in 19 foreign countries and with two international organizations (Table 1).

On February 5, 1997 a Memorandum of Understanding was signed establishing an International MEDLARS Center in Russia. The event culminated several years of discussion and planning by representatives of the National Library of Medicine and the State Central Scientific Medical Library (SCSML) in Moscow.

In a ceremony held at the NIH in Bethesda, the Memorandum of Understanding was signed by NLM Director Dr. Donald A. B. Lindberg and by Dr. Boris Loginov, SCSML Director. Looking on were Donna E. Shalala, Secretary, Department of Health and Human Services, and Dr. Tatyana Dmitrieva, the Russian Minister of Health. Dr. Siegel and Lois Ann Colaianni, Associate Director for Library Operations, were instrumental in moving events forward to this successful outcome.

The ceremony took place in conjunction with the fifth meeting of the Health Committee of the Gore-Chernomyrdin Commission, an initiative by U.S. Vice President Al Gore and Russian Prime Minister Viktor Chernomyrdin that seeks to identify long-term goals for federal level cooperation between the U.S. and Russia.

World Health Organization (WHO)

The National Library of Medicine and WHO continue their cooperative effort 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.

International Council For Scientific And Technical Information (ICSTI)

Working in collaboration with the International Council for Scientific and Technical Information, NLM led the development of a new pilot demonstration project intended to reduce existingtechnological and tariff barriers to the flow of scientific and health-related information over the Internet to developing regions of the world. Focused in the eastern Caribbean, the NLM/ICSTI effort was carried out in conjunction with a number of other international organizations, including UNESCO, PAHO, the International Telecommunications Union, and the United Nations Development Programmes.

Global Internet Connectivity

At their January meeting, the G7 Global Healthcare Applications Project national coordinators approved a U.S. proposal for Global Internet

Connectivity advanced by NLM. The Internet connectivity proposal was premised on the fact that many G7 healthcare applications projects depend in whole or in part on the availability of the Internet. As the preferred technical means for the exchange of biomedical information with and between the G7 nations and beyond, the Internet is a critical component of the emerging global health information infrastructure.

NLM has begun to lay the groundwork for collaborative Internet testing that should improve our understanding of Internet connectivity from the user's perspective, and ultimately help better inform users when selecting Internet services and solving Internet problems. NLM has identified a need for objective and widely accepted methods of characterizing and diagnosing the quality and performance of Internet connections under different conditions (e.g., locations, distances, time periods, data volumes). Also, we have established collaborative relationships with an initial group of domestic and international partners to both validate testing methods and conduct tests of Internet connectivity between selected locations.

International Visitors

The Library continues to be a focal point for visitors of the international community from a variety of disciplines. Many of these visitors are responsible for medical, scientific or technical information in their own countries. Visitors are officially received and briefed on relevant aspects of NLM operations and research. Among the visitors in 1997 were representatives from:

Argentina, Australia, Belarus, Bulgaria, Canada, China, Croatia, Czech Republic, Georgia, Hungary, Israel, Jamaica, Japan, Korea, Kyrgystan, Latvia, Lithuania, Malaysia, Nigeria, Poland, Romania, Russia, Slovakia, Slovenia, South Africa, Spain, Sweden, Taiwan, Ukraine

Outreach

The NLM has a longstanding commitment to the effective dissemination and use of biomedical information within the health community. To help achieve this goal, NLM has, since 1989, collaborated with its National Network of Libraries of Medicine (NN/LM) to conduct outreach to health professionals and especially those in rural, minority, or other underserved communities. The objectives of NLM-sponsored outreach are to: 1) make health professionals aware of the information products and services NLM provides; 2) facilitate health professionals' access to and use of biomedical information; 3) provide training in the searching of electronic databases; 4) assist health professionals in developing new information-seeking behaviors; and 5) improve heath care practices through the use of authoritative, up-to-date information.

Between 1990 and 1995, NLM supported close to 300 outreach projects that involved more than 500 institutions across the country. In 1996, NLM published a 5-year review of its outreach program and activities.[1] The review concluded that NLM's outreach program has made significant progress overall. However, the review also recommended that the methodology for evaluating outreach be more fully developed.

The ability to conduct this 5-year review was, to a great extent, facilitated by data collecting and reporting by the RMLs and NN/LM. This is an important first step in assessing the strategies that have been undertaken so far. However, to evaluate outreach approaches better and demonstrate the effectiveness of outreach projects in the future, several additional steps should be taken. NLM should develop benchmarks against which the RMLs and the NN/LM can measure the effectiveness of their outreach efforts. In addition, NLM and the RMLs should work together to develop further expertise in evaluation methodology. Evaluation components should be an integral part of all NLM-sponsored outreach.

The 5-year review envisioned that strengthened outreach evaluation would help NLM and the NN/LM more clearly discern lessons learned from past experience, better plan for future outreach activities, and design future outreach with a built-in evaluation component to the extent feasible.

NLM determined that a logical next step in outreach would be to undertake a special project to develop a framework or model of outreach planning and evaluation. This approach is similar conceptually to that used by NLM with regard to telemedicine research, where the Institute of Medicine was commissioned to conduct a study and prepare a guide on evaluating the use of telecommunications in health care. NLM has required that its telemedicine contractors use the guide as part of a built-in evaluation component.

The working hypothesis is that the medical library community would benefit from knowledge of evaluation studies of outreach-like activities that have been carried out in related disciplines. NLM is especially interested in exploring related fields which have evaluated efforts directed toward minority populations because outreach to minority and other underserved populations is one of NLM's highest

priorities, and, at the same time, an area in which success has been most difficult to achieve.

For this project with its focus on outreach evaluation, NLM selected the NN/LM's Pacific Northwest Regional Medical Library, located at the University of Washington Health Sciences Library. NLM decided that a subfocus on Native Americans in the Pacific Northwest would be useful to better understand outreach to underserved populations and test the outreach planning and evaluation model to be developed by the project. Initial project plans were presented at an outreach panel as part of the 1997 Medical Library Association meeting in Seattle. The project has established an advisory panel consisting of representatives from academia, libraries, health professionals, and Native American groups. Finally, a separate but related tribal connectivity project also to be administered by the Pacific Northwest RML will provide support for new or enhanced Internet connections for selected American Indian tribes and Alaska Native villages in the Pacific Northwest. Some of these sites will be used for pilot tests of the outreach planning and evaluation guide.

[1] See K.T. Wallingford, A. B. Ruffin, K.A. Ginter, M. L. Spann, F.E. Johnson, G. A. Dutcher, et al., "Outreach Activities of the National Library of Medicine: A Five-Year Review," Bulletin of the Medical Library Association, Vol. 84, No. 2 (Supplement), April 1996, pp. 1-60.

Table 1

International MEDLARS Centers

COUNTRY	CENTER
Australia	National Library of Australia
Canada	Canada Institute for Scientific & Technical Information (CISTI)
China	Institute of Medical Information, Chinese Academy of Medical Sciences (CAMS)
Egypt	Egyptian National Scientific & Technical Information Network, (EINSTINET)
France	Institut National de la Sante et de la Recherche Medicale (INSERM)
Germany	Deutsches Institute for Medical Documentation and Information (DIMDI)
Hong Kong	The Chinese University of Hong Kong (CUHK)
India	National Informatics Center Planning Commission
Israel	Hebrew University, Hadassah Medical School, Berman National Medical Library
Italy	Ministry of Health, Istituto Superiore di Sanita
Japan	Japan Science and Technology Corporation (JST)
Korea	Medical Library, Seoul National University, College of Medicine
Kuwait	Ministry of Public Health, Kuwait Institute for Medical Specialization (KIMS)
Mexico	Centro Nacional de Informacion y Documentacion sobre Salud (CENIDS)
Russia	State Central Scientific Medical Library
South Africa	South African Medical Research Council
Sweden	Medical Information Center, Karolinska Institute
Switzerland	Documentation Service (DOKDI) of the Swiss Academy of Medical Sciences (SAMS)
United Kingdom	The British Library
*Pan American Health Org.	Centro Latino Americano e de Caribe, Bibliotec Regional de Medicina (BIREME)/PAHO
*Taiwan	Science and Technology Information Center

*International/Intergovernmental Organizations

Library Operations

Lois Ann Colaianni
Associate Director

The oldest and largest of NLM's Divisions, Library Operations (LO) performs the basic NLM functions that ensure access to the scholarly record of biomedicine and the health professions. LO acquires, preserves, and organizes the world's historical and contemporary biomedical literature; develops a thesaurus and a classification designed for use in organizing biomedical information; produces authoritative indexing and cataloging data; builds and disseminates bibliographic, directory, and full-text databases; provides national back-up document delivery, reference, and research assistance; helps health professionals, researchers, and health sciences librarians to make effective use of NLM's services; and coordinates the 4,500-member National Network of Libraries of Medicine (NN/LM) which provides primary library services to U.S. health professionals and researchers. LO's basic information services provide an essential foundation for NLM's Outreach Program and for its biotechnology, AIDS, and health services research information programs.

LO's multi-disciplinary staff of librarians, technical information specialists, subject experts, health professionals, historians, and technical and administrative support personnel also directs the National Information Center on Health Services Research and Health Care Technology (NICHSR); plays a key role in the development of NLM's Web site; carries out an active research program in the history of medicine; conducts research and evaluation studies related to the Library's programs and services; directs and participates in research in advanced automated information storage and retrieval; and contributes to the development of Federal health data policy. LO staff members participate actively in NLM-wide efforts to improve the quality of worklife at NLM, including the Diversity Council and development of the NLM Intranet.

Planning and Management

In FY 1997, LO worked with other NLM program areas to establish effective mechanisms for handling the many technical, resource, and policy issues that result from NLM-wide System Reinvention projects. The implementation of free Web-based access to MEDLINE and other NLM databases and the transition away from the ELHILL mainframe-based

retrieval system necessarily require significant changes in the way the Library supports and trains its online users and creates and disseminates its machine-readable data. Likewise the replacement of the legacy systems currently used in acquisitions, serials control, cataloging, and circulation with a commercial integrated library system (ILS) will have a major impact on NLM's internal library functions. These efforts must be accomplished against a background of ongoing production pressures and service demands and continuing investigation and experimentation with new approaches to building and delivering information services.

LO established a temporary position of Special Assistant for ILS Implementation in the Office of the Associate Director to coordinate the Division's participation in what will be a very complex and resource-intensive process. In FY 1997 LO's ILS activities focused on functional evaluation of two commercial ILS products that were possible candidates for use by NLM and on identifying and initiating data clean-up projects that will facilitate the merger of bibliographic data from the current serials control, monograph acquisitions, and cataloging systems.

A self-managed LO team was appointed to implement plans for Customer Service Reinvention. This effort is testing the use of automated systems and streamlined procedures to help LO respond to inquiries from offsite users more quickly, accurately, and consistently, to capture useful data on the nature of user inquiries, and to obtain and make use of feedback from users to improve the Library's services. Specific Reinvention activities are described elsewhere in this chapter and in other chapters of the report.

In many respects, FY 1997 was an atypical year for LO's basic production and service activities. LO eliminated the indexing and data entry backlogs caused by the extended suspension of the data entry contract in FY 1996. As reported elsewhere, staff from other NLM divisions collaborated to stabilize alternative input streams using optical character recognition and direct publisher submission of electronic SGML data to provide insurance against similar disruptions in the future. The unusual indexing backlog inherited from last year is just one of several reasons that LO's FY 1997 statistics are not strictly comparable to FY 1996 figures. Service demand was artificially low in FY 1996 due to the combination of Federal government furloughs and closures due to snow storms. The introduction of free Web-based access to MEDLINE in July 1997 not only increased online searching, but also made previous methods of measuring search volume no longer possible. The introduction of a new routing system for incoming customer service telephone calls also affected the

Collection Development and Management

NLM's comprehensive collection of biomedical and health professional literature is the core of many of the Library's basic services. To ensure that the collection will meet the needs of current and future health professionals and researchers, LO develops and updates selection policy, acquires and processes literature that meets its selection guidelines in all languages and formats, organizes and maintains the collection for efficient current use, and preserves the materials acquired for future generations. At the end of FY 1997, NLM's collection contained 2.22 million volumes and 3 million other items, including audiovisuals, microforms, computer software, pictures, and manuscripts. (Table 2)

Selection

LO and its agents select new and retrospective literature for the NLM collection according to guidelines in the *Collection Development Manual of the National Library of Medicine*, which has typically undergone a major review and revision every 5 to 7 years. In between major revisions, the LO staff develops operational guidelines for materials that are posing selection difficulties and for emerging subject areas. LO also conducts assessments of the extent to which it has been successful in adhering to selection guidelines in particular subject fields.

In FY 1997, NLM completed a 9-month pilot test of draft selection guidelines for the electronic information sources that will be described in NLM bibliographic databases or its online directory services. The guidelines were revised based on the results of the test; NLM is developing plans for their implementation. As a cost-saving measure, LO reviewed all currently received serial bibliographies and serials in the field of environmental sciences and canceled those that did not meet selection guidelines.

Acquisitions

LO received and processed 172,000 contemporary books, serial issues, audiovisuals, and software packages. (Table 3) A total of 41,433 volumes and 32,671 other items (e.g., microforms, audiovisuals, software) were added to NLM's collection during FY 1997. NLM re-competed its four major serials acquisition contracts, developed and tested a new model for evaluating the performance of monograph acquisition vendors, and identified a new acquisition source for materials from China. Staff members who

acquire and process books and journals were also heavily involved in evaluating commercial integrated library systems against NLM's functional requirements.

The Library's collection was enriched by the addition of important historical books and manuscripts that reflect the great breadth and depth of NLM's holdings in the history of medicine and public health. Historical book acquisitions included: Rolando Capulleti's *De Curatione Pestiforum Apostematum* (Rome, ca. 1475), one of the first printed works on the plague; Heinrich Vogtherr's *Auszlegung unnd beschreibung der Anathomi* (Strasbourg, 1539), the rare first edition of a text on anatomy intended for the ordinary person; Nicolas Monardes' *Primera y Segvnda y Tercer a Partes de la Historia Medicinal...* (Seville, 1580), the second edition of his classic work on medicinal plants from the New World; Thomas Spackman's *A Declaration of Such Greivous Accidents as Commonly Follow the Biting of Mad Dogges, Together with a Cure* (London, 1613), considered the first English book on rabies; Tamba Gentoku's *Kokei Saiku Ho* (Kyoto, 1790), an extremely rare complete copy of a three-volume work on emergency treatments; and Pedro Albuquerque de Olivera's *Pathogenesia Homoepathica Brasileira, Contendo a Descripcao dos Medicamentos Indigenas Conhecidos ee Analysados...* (Rio de Janeiro, 1856), a work by a practicing homeopathic physician who describes the pharmaceutical effects of over 100 Brazilian plants. Interesting manuscript acquisitions included several French works: *Manuel des Operations de Chirurgie* (1757) by Jean Cochon-Dupuy, a manuscript textbook of surgical procedures and instruments; *L'hypnotisme en therapeutique: Guerison d'une contracture hysterique*, an autographed work by the 19th century French pathologist and neurologist, Jean Martin Charcot (1825-93), on the physiological manifestations of hysteria and hypnosis; an autographed rough draft of trial testimony on a sensational malpractice suit in obstetrics by one of the founders of scientific gynecology, Joseph C. A. Recamier (1774-1852); and a lavishly calligraphed and illustrated manuscript by Y. J. E[douard] Le Ber, *Phrenologie des gens du monde* ...(Paris, 1833), written during the heyday of phrenology in France.

Additions to NLM's 20th century manuscript collections included the papers of the noted virologist, Telford Hindle Work, which document over 50 years (1942-94) of international research and teaching in the field of arbovirology; the papers of two distinguished NIH scientists, John B. Calhoun, Laboratory of Brain Evolution and Behavior, National Institute of Mental Health, and Abraham Goldin, Biochemical Pharmacology Section, Laboratory of Chemical Pharmacology, National Cancer Institute; and the

medical papers of William Carlos Williams (1883-1963), Pulitzer prize-winning poet and practicing physician. The Library also received the records of the National Institutes of Health's Human Radiation Studies Task Force; the transcripts of the Primary Care Oral History Project (ca. 1994-97) which includes interviews conducted by Fitzhugh Mullen, M.D. with primary care leaders and practitioners in the U.S. and the United Kingdom; the initial set of materials for the Nursing Informatics History Collection; the first installment of records from the American Academy of Physician Assistants; and additions to the manuscript collections of Nobel Laureate Marshall Nirenberg, the Editorial Files of the Encyclopedia of Bioethics, the Western Surgical Association Records, the FDA Oral History Series, and the Society for Research in Child Development Records.

Preservation and Collection Management

LO undertakes many activities to extend the life of NLM's archival collection of biomedical literature and to keep materials in good order so they are accessible for current use. Preservation and collection management functions include binding, microfilming, conserving rare and unique items, maintaining appropriate storage conditions and facilities for all segments of the collection, and preventing and responding to emergencies that threaten library materials. LO also distributes data about what the Library has preserved in MARC records and also provides other preservation information useful to health sciences librarians on NLM's Web site. NLM continues to explore the use of new technology to preserve library materials and to promote the use of more permanent media in new biomedical publications.

In FY 1997, LO bound 26,330 volumes, microfilmed 6,472 volumes, conserved 391 items from the historical collections, and repaired 3,064 volumes in the newly completed onsite Book Repair and Conservation Laboratory. The Library has completed microfilming of all brittle volumes of *Index Medicus* titles and is now focusing on other titles listed in [Garrison Morton] and in the *Index-Catalogue.* The microfilming contract was modified to include special procedures for microfilming rare books, obviating the need for separate procurements to handle these materials.

Several projects improved the condition of NLM's closed stacks. The collections of indexing and abstracting tools and microforms were moved to new locations; more than 10,000 "problem" items, e.g., unlabeled, mislabeled, and damaged items, were handled; and unneeded duplicate serial issues were pulled from the shelves. Housekeeping contract procedures were revised to reflect preservation concerns. Pre-cut plastic sheeting for protecting the collection from water damage was distributed to all stack levels as part of LO's ongoing effort to review and improve the Library's disaster plan and procedures. After a long period of renovation and testing, an onsite audiovisual vault is now being used to store historical motion pictures under more appropriate environmental conditions.

Bibliographic Control

LO creates authoritative indexing and cataloging records for literature in the NLM collection so that health professionals, researchers, and other interested users can easily find the journal articles, books, films, pictures, and electronic media that are relevant to particular inquiries. LO also maintains the Medical Subject Headings (MeSH), the thesaurus used by NLM and many other institutions to describe the content of biomedical literature, and links MeSH to other biomedical vocabularies in the Unified Medical Language System (UMLS) Metathesaurus. The LO staff also maintains the *National Library of Medicine Classification*, a scheme for arranging physical library materials according to their subject matter that is used by health sciences libraries around the world.

Thesaurus Development

The 1998 MeSH contains 18,937 main subject headings, 787 subheadings or qualifiers, 83 publication types, and more than 97,000 supplementary records for chemicals and other substances. For the 1998 edition, LO staff added 365 new descriptors, replaced 432 main headings with more current terminology, and created 1311 new cross-references. Special efforts were undertaken to modify psychiatric terminology to reflect the fourth edition of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental and Behavioral Disorders* and to update enzyme nomenclature. New MeSH hierarchies were created for Archaea and Food and Beverages. The practice of designating some hierarchical terms as "Non-MeSH" was discontinued. All terms that previously belonged to this category were reviewed and, if necessary, revised or deleted. Many useful cross-references previously found only in text notes were converted to entry terms that can be automatically mapped to main headings to assist online searchers.

Most of the editing for the 1998 edition of the Metathesaurus was completed during FY 1997 under MeSH Section supervision. New additions to the 1998 Metathesaurus will include: the remainder of the *Read Clinical Classification*, which is developed by the

National Health Service in the United Kingdom; the *International Classification of Diseases, 10th edition,* as issued by the World Health Organization; the *International Classification of Primary Care*; the Health Care Financing Administration's Common Procedure Coding System (HCPCS), which includes the American Dental Association's *Current Dental Terminology*; the *Nursing Outcomes Classification*; the Food and Drug Administration's medical device vocabulary; and detailed anatomical terms for the thorax developed by the University of Washington's Digital Anatomist project. The 1998 Metathesaurus will contain all but one of the administrative data coding systems that are being recommended as standards for 2000 as part of the implementation of administrative simplification provisions of the Health Insurance Portability and Accountability Act of 1996.

During FY 1997, NLM and the Agency for Health Care Policy and Research sponsored a large-scale test to determine the extent to which the set of existing controlled vocabularies in the 1996 Metathesaurus (plus some others that have since been added for the 1997 or 1998 editions) contain the detailed clinical concepts and terms needed in computer-based patient record systems. LO collaborated with NLM's Lister Hill Center (LHC) to analyze the test data which were collected via a special LHC-developed Web-interface to the UMLS Knowledge Source Server. More than 60 participants nationwide contributed more than 40,000 test terms. The results indicate that the bulk of terms needed to describe patient conditions are present in a combination of existing vocabularies. More work is needed to reduce unnecessary duplication of effort by vocabulary developers, to establish a stable mechanism to support the development of key clinical vocabularies, and to build a tighter feedback loop between vocabulary developers and software developers who are attempting to implement controlled vocabulary in clinical information systems.

Cataloging

LO catalogs the biomedical literature acquired by the Library to document what is available in the NLM collection and to provide cataloging records that can be used by other health sciences libraries to minimize the level of effort required to organize their own collections. Many health sciences libraries also use MeSH and the NLM Classification to catalog materials of local interest for which no NLM catalog record is produced.

In FY 1997, the Technical Services Division cataloged 20,296 contemporary books, serials, nonprint items, and cataloging-in-publication galleys, using a combination of in-house staff, contractors, and an interagency agreement with the Library of Congress. To speed cataloging of newly received items, NLM now makes heavier use of cataloging copy generated by other libraries. The working inventory of uncataloged contemporary books has been reduced to 3-4 months workload and is largely composed of items in languages other than English.

The History of Medicine Division cataloged 558 early books and began a long term project to catalog NLM's large collection of historical pamphlets. HMD staff members worked with the Lister Hill Center to mount a pilot Digital Manuscript project, scanning, organizing, and providing Web-based access to the papers of Donald S. Fredrickson, M.D., former Director of the National Institutes of Health. The success of this effort led to the decision to develop a broader Digital Manuscripts program that will focus on the papers of NIH Nobel Prize winners.

During FY 1997, LO developed and tested new cataloging policies for electronic publications. NLM's catalogers obtained workstation access to the Library of Congress's "Catalogers' Desktop," which provides electronic access to LC's cataloging rule interpretations and USMARC format documentation. Major changes and enhancements to NLM's online cataloging system have been suspended in anticipation of the move to a commercial integrated library system. Cataloging staff members assisted in evaluating the functional capabilities of candidate systems. The picture catalog and scanned Images from the History of Medicine collection were used to test the robustness of commercial ILS catalog features.

Indexing

LO indexes articles from nearly 4,000 biomedical journals to build MEDLINE and other NLM databases which allow health professionals, scientists, and others to locate articles on specific topics. A combination of in-house staff, contractors, International MEDLARS Centers, and cooperating organizations perform the indexing under the supervision of the Index Section.

In addition to indexing newly published articles, LO also annotates previously created MEDLINE records when the articles to which they refer have been retracted, corrected, or challenged in subsequently published notices or commentaries. Until mid-FY 1997, LO assisted the National Center for Biotechnology Information (NCBI) by indexing gene sequences and editing author sequence submissions. The success of the author submission process allowed NLM to phase out the separate gene sequence indexing unit that had previously been essential to maintaining

the quality and currency of GenBank.

The Literature Selection Technical Review Committee (LSTRC - see Appendix 6 for members), an NIH-chartered committee of outside experts, advises NLM about which journals should be indexed for MEDLINE, *Index Medicus*, and other NLM databases. In FY 1997, the LSTRC reviewed 398 journal titles and rated 63 sufficiently highly for immediate inclusion in MEDLINE. The Committee also reviewed recommendations on journal coverage from professional societies in the fields veterinary medicine and gastroenterology. As a result of these reviews, the LSTRC recommended that NLM index 8 additional titles and discontinue indexing 11 titles. LO initiated a review of consumer health magazines and newsletters by NLM staff and outside experts as the first step in determining whether the Library should index some of these publications for MEDLINE. Ten were selected for addition to MEDLINE in 1998.

By March 1997, NLM had cleared the data entry and indexing backlogs stemming from the several-month suspension of the keyboarding contract in FY 1996. As a result, the Library added 519,000 indexed citations to MEDLINE in FY 1997—an all-time record that is unlikely to be duplicated or surpassed for some time to come. Significant progress was made in the LO/LHC/OCCS effort to establish scanning and optical character recognition (OCR) as a second high-volume data input stream, and data for a total of about 65,000 MEDLINE citations and abstracts were captured by this method in FY 1997. The number of citations and abstracts input via this method is expected to at least double in FY 1998, as the Library continues to customize the OCR software to reflect the specific characteristics of the literature indexed for MEDLINE. NCBI and LO are also working with publishers to increase the volume of citations and abstracts received in appropriately tagged electronic format. NLM received electronic data from publishers for about 20,000 articles in FY 1997 and expects this number to increase significantly over the next several years.

LO is an active participant in an NLM-wide effort directed by the Lister Hill Center to determine whether more intensive use of automation can produce comparable retrieval performance while reducing the level of human effort needed to index articles. Called the Indexing Initiative, this project is experimenting with a range of automated techniques and making intensive use of the UMLS Knowledge Sources.

Information Products

NLM's comprehensive collection of biomedical literature and its authoritative indexing, cataloging, and thesaurus data support the development of online databases, electronic products, and print publications. LO collaborates with other NLM Divisions to produce and disseminate some of the world's most heavily used information tools.

Online Databases

Online use of NLM's retrieval systems increased dramatically after the June 1997 announcement of free Web-based access to MEDLINE and other NLM databases via PubMed and Internet Grateful Med. Online users are no longer required to register for codes to search NLM's databases, and the majority of MEDLINE searching is now done against PubMed's underlying retrieval system rather than on ELHILL, the Library's legacy mainframe retrieval system. As a result, NLM had to alter its methods of counting online users and the number of searches they perform on the system. In September 1997, users with an estimated 400,000 different Internet addresses conducted about 4 million searches on NLM's systems. This represents roughly a seven-fold increase from the amount of searching conducted in September 1996. The figures do not include searches performed on the computer systems and CD-ROM products of the many organizations that lease data from NLM.

The new PubMed retrieval system provides integrated access to MEDLINE and genetic information sources and links to the full-text of some journal articles when available from cooperating publishers, in some cases with subscription fees. A new version of Internet Grateful Med introduced in FY 1997 allows users to search 10 NLM databases (MEDLINE, AIDSLINE, HealthSTAR, AIDSDRUGS, AIDSTRIALS, DIRLINE, HISTLINE, HSRProj, OLDMEDLINE and SDILINE). The Library also released Grateful Med for Windows in early FY 1997, which provides access to MEDLINE only.

NLM's database offerings were expanded by the addition of OLDMEDLINE, a database of pre-1966 *Index Medicus* data which currently covers 1964-5 but will gradually cover earlier years as they are converted to machine-readable form. The conversion of HISTLINE data to the current file format was completed with the addition of more than 11,000 records for book chapters and meeting papers. HSTAT (Health Services/Technology Assessment Text) now provides integrated searching of full-text guidelines on servers at NLM and at the Centers for Disease Control and Prevention.

With the implementation of free Web-based access to its databases, LO reviewed and modified a number of related pricing policies and training activities. The fixed-rate and flat-rate programs were phased out or modified to cover searching of databases

not yet available on the Web. The foreign surcharge was eliminated effective October 1, 1997. Student codes and the $20 practice credit for new users were eliminated. In FY 1997, NLM and the National Online Training Center at the New York Academy of Medicine taught a total of 579 students, most of whom were search intermediaries. Intensive work began on revamping the NLM online training programs for search intermediaries to focus on Web-based searching via PubMed and Internet Grateful Med.

Machine-Readable Data

To promote the broadest possible access to its electronic information, NLM leases its databases in machine-readable form to commercial database vendors, international MEDLARS centers, universities, and other interested organizations. Licensees then make the data available online or in CD-ROM products. In FY 1997, NLM distributed data from one or more of its bibliographic databases to 67 different organizations. Hundreds of individuals and organizations obtained MeSH data.

LO continues to increase electronic distribution of data via the Internet. In FY 1997, ftp distribution of MARC cataloging records was extended to monthly subscribers. NLM also began weekly ftp transfer of serial cataloging records to OCLC. Those interested in licensing MeSH data may now register online on NLM's Web site and then ftp the data in a variety of formats. LO assisted OCCS in setting up a test of alternative means for distribution of MEDLINE data. The results of the test will lead to new MEDLINE distribution mechanisms in FY 1998.

Print and Electronic Publications

NLM publishes some of its authoritative data in printed publications, of which *Index Medicus* is the oldest, largest, and most widely distributed. As NLM's primary clientele becomes able to use more flexible electronic access mechanisms, the Library reviews and modifies or eliminates specific publications that have outlived their usefulness in that format. In FY 1997, LO ceased publication of the *MeSH Supplementary Chemicals* and the microform *Health Sciences Serials* and decided to stop publishing the monthly *Abridged Index Medicus*, effective with the December 1997 issue.

The NLM World Wide Web site continues to grow as a primary vehicle for providing information to a wide range of NLM users. After the site was redesigned in the spring of 1997, June NLM Web site users were asked to fill out a voluntary customer satisfaction survey, which ended when the target number of respondents (500) was reached. Thirty

percent of those responding self-identified themselves as health practitioners, 19% as librarians, 16% as scientists/researchers, 12% as students, 6% as educators, and 17% as belonging to some other category. Seventy percent of the respondents found the site very or extremely easy to use to find the information they were seeking. Free Web-based access to MEDLINE, implemented just after the survey was completed, should address many of the concerns expressed by those who were less satisfied with NLM's Web site. In FY 1997, users obtained 600,000 copies of NLM publications from the Web site. In order to increase customer awareness of new additions to the site, NLM introduced a new files alerting service in August. More than 900 people now receive weekly messages announcing the documents added and updated on the site during the previous week.

NLM's series of *Current Bibliographies in Medicine* are available on the Web site and, in some cases, also in print. Each bibliography addresses a topic of current interest that may be difficult to search in NLM's databases or be spread across the literature of multiple disciplines. The bibliographies cover subjects important to specific programs of NLM, NIH, and other Health and Human Service agencies and are often produced in conjunction with NIH Consensus Development Conferences or other special meetings. LO staff members collaborate with outside experts to produce each bibliography. Among the eight FY 1997 additions to the series were: *Malaria: Prevention and Therapy (with Emphasis on Africa)*, *Acupuncture*, *Domestic Violence Assessment by Health Care Practitioners, American Indian and Alaskan Native Health, Genetic Testing for Cystic Fibrosis,* and *Unified Medical Language System,* produced to mark the tenth anniversary of the UMLS project.

User Services

LO provides document delivery, reference, and customer service as a national back-up to services available from other health sciences libraries and information suppliers; coordinates the National Network of Libraries of Medicine (NN/LM) which works to equalize and enhance access to information services and technology for all U.S. health professionals; arranges exhibitions and programs on the history of medicine and related fields; and directs training programs for health sciences librarians.

Document Delivery

LO provides copies of documents to other libraries across the country and around the world to fill requests from health professionals and researchers for

items that are available in NLM's comprehensive collection of biomedical literature and cannot be readily obtained from other NN/LM libraries or document suppliers. LO also retrieves documents from NLM's closed stacks for use by onsite patrons. In FY 1997, the Collection Access Section received 630,190 requests for documents (Table 6). Total document request traffic has increased by about 10 percent since FY 1995, when NLM also experienced a "normal" production year. The volume of requests from other libraries has increased just 2 percent over the past 2 years, but onsite request traffic has grown by 21 percent. NLM filled 75 percent of the interlibrary loan requests it received and processed 91 percent of the filled requests within 24 hours of receipt.

During FY 1997, NLM installed and tested Relais, a new document delivery processing system originally developed for the Canada Institute for Scientific and Technical Information (CISTI) and modified to meet NLM's internal requirements by its creator, Network Support, Inc. A part of NLM's broader System Reinvention initiative, Relais captures incoming DOCLINE requests directed to NLM and automated document requests from NLM's onsite users, sorts and prints the requests for processing, and allows the Library to track their whereabouts in its processing stream. All documents are scanned rather than photocopied, and the temporarily stored images are sent either to the high-speed printer or directly via Internet or fax, depending on the delivery mechanism selected by the requesting institution. (In FY 1997, 22 percent of NLM's filled ILL requests were delivered via Internet or fax.) When implemented in FY 1998, Relais should reduce NLM's ILL processing time significantly.

There are now 2,929 libraries using DOCLINE, NLM's automated system for routing ILL requests among U.S. and some international health sciences libraries. DOCLINE users entered 2.87 million requests in FY 1997, which represents a stable request level for the past three years. Ninety-four percent of the requests were filled. During FY 1997, DOCLINE was enhanced to allow users to request a specific delivery method (mail, fax, Ariel (Internet), pick-up, e-mail, or printer) for each request and to permit users to quickly generate requests by entering the unique identifier for citations in additional NLM databases (AIDSLINE, CANCERLIT, SPACELINE, HealthSTAR, PREMEDLINE, OLDMEDLINE). The RMLs may now make changes to the DOCUSER file via a Web page, which supports more rapid updating of addresses, phone numbers, e-mail addresses, and contact names. NLM and the RMLs continue to encourage DOCLINE participants to access the system via the Internet to reduce the Library's telecommunications costs, and, in FY 1997, 34 percent of them did so. Financial and

quality of service issues continue to limit the ability of some users to switch to Internet access.

DOCLINE routes requests based on the 1.4 million holding statements for 3,202 libraries, primarily in the U.S. and Canada, in the SERHOLD database. In FY 1997, OCCS and LO implemented streamlined procedures for processing SERHOLD tape input and also set up mechanisms for SERHOLD updates to be sent electronically via ftp. These steps improve the currency of holdings data in SERHOLD and therefore the accuracy of the automated routing of requests by DOCLINE.

Reference and Customer Service

LO provides biomedical and health-related reference and research assistance to onsite and remote users as a backup to services available from other health sciences libraries. The LO staff also helps the Library's users to make effective use of NLM information products and services. In FY 1997, staff in the Reference Section, the National Information Center on Health Services Research and Health Care Technology (NICHSR), and the History of Medicine Division responded to 75,842 requests for information from onsite and offsite users. The MEDLARS Management Section handled 40,360 inquiries from users of NLM's online databases and retrieval systems. As expected, the introduction of free MEDLINE service via PubMed and Internet Grateful Med has increased the number of inquiries NLM receives from the general public. During FY 1997, LO assumed responsibility for answering many user inquiries related to PubMed and Internet Grateful Med which were previously handled by staff in the National Center for Biotechnology Information and the Lister Hill Center.

The Customer Service Reinvention project is working toward a more streamlined and integrated approach to handling all inquiries from offsite users, including requests related to the use of NLM's online services and other products as well as reference requests for information on particular topics. During FY 1997, the Customer Service Team implemented a new, streamlined phone service, 1-888-FINDNLM; installed the Teloquent call distributing system to handle incoming calls for both the MEDLARS Management and the Reference Sections; completed a customer survey to obtain the baseline user satisfaction level with service received in both Sections; and installed and tested the CustomerQ software for logging and tracking customer service inquiries via e-mail and surface mail. LO is now working with OCCS and Quintus, the developer of CustomerQ, to modify the system to better suit NLM's requirements handling telephone inquiries.

During FY 1997, NLM's services to onsite

users were expanded to include access to full MEDLINE via Internet Grateful Med and PubMed. The main reading room was reorganized, and the Reference Desk and patron workstations were moved from the Rotunda to the reading room itself. The Reference Section held an open house to highlight and explain NLM's special equipment and facilities for people with disabilities.

In addition to onsite users of NLM's collection and information products, NLM attracts many domestic and international visitors interested in biomedical communication, information technology, health sciences librarianship, and the history of medicine. In FY 1997, LO staff members assisted in conducting 139 regular daily tours of NLM's facilities for a total of 536 visitors. The Office of Public Information (Office of the Director) made arrangements for 125 special group tours and briefings for 1,919 visitors and for press conferences in conjunction with the implementation of free MEDLINE service via PubMed and Internet Grateful Med.

National Network of Libraries of Medicine

The goal of the NN/LM is to provide U.S. health professionals, researchers, educators, and administrators with timely, convenient access to biomedical and health information resources. The NN/LM strives to ensure equal access to up-to-date information irrespective of the user's location or institutional affiliation. The network includes more than 4,500 health sciences libraries of all sizes and types located throughout the country. LO's NN/LM Office oversees the network programs which are coordinated and administered by eight Regional Medical Libraries. (See Appendix 1 for a list of the RMLs.)

Because Internet connectivity facilitates access to many high-quality information sources, the NN/LM program is focusing on ways to facilitate connectivity for member libraries and health professionals. In FY 1997, LO's NN/LM Office, assisted by the RMLs, conducted a telephone-response survey of all network member libraries to determine the extent of their Internet connectivity. The survey resulted in 3,491 usable responses, which provide a good picture of the variations in Internet connectivity in the network. The results will be used to target RML efforts to assist hospital libraries which lack Internet access or have connections that are too slow to permit efficient use of Web-based resources.

The NN/LM program is a core component of NLM's outreach initiative. The RMLs and other NN/LM members develop and conduct many projects to reach underserved health professionals and patients in both rural and inner city areas. Some of the NN/LM

outreach projects funded in FY 1997 will facilitate patient access to HIV/AIDS information for patients in middle Tennessee; provide Internet and online searching instruction to health care providers in 18 rural Alaskan hospitals; provide computers and training to nursing faculty, home health care staff, and nursing students in the St. Louis inner-city; provide Internet training to community health professionals in Vermont and New Hampshire; and improve information services and establish collaborative relationships with grassroots organizations, health personnel, and libraries in a federally designated Enterprise Community in South Carolina.

The NN/LM Office, NLM's National Information Center for Health Services Research and Health Care, the RMLs, the Centers for Disease Control and Prevention, the National Association of County and City Health Officials, and the Association of State and Territorial Health Officials launched a joint "Partners in Information Access for Public Health Professionals" program in FY 1997. The objectives of this program are to increase awareness and use of NLM, NN/LM and CDC information services by public health professionals; to increase awareness of public health information needs and resources among NN/LM members; to assist public health professionals in obtaining the information technology needed to access current information sources; and to train public health professionals to use information technology and relevant information services. Initial steps have been taken to establish links between the Web sites and information resources of the participants, to schedule joint exhibits and presentations, to begin development of joint training programs, and to publicize to the public health community the availability of information services, informatics training programs, and funding for Internet connections.

The RMLs and other NN/LM members conduct the majority of exhibits and demonstrations of NLM's products and services at health professional meetings around the country. NLM staffs the exhibits at the meetings in the Washington D.C. area, at the annual meetings of the Medical Library Association and some other organizations, and at some distant meetings focused on health services research and related fields. In FY 1997, NLM and NN/LM services were displayed at about 70 exhibits at national, regional, and state association meetings across the U.S.

Historical Exhibitions and Programs

FY 1997 was a transition year for NLM's historical exhibition program. In the past, the History of Medicine Division (HMD) typically mounted two or three historical exhibits each year in the front lobby of

the library building. In the future, the Library plans to install one major exhibition each year, using both the lobby and at least part of the adjacent rotunda. These major exhibitions may be arranged in collaboration with other organizations, possibly in conjunction with scientific meetings. NLM will continue to mount "mini-exhibits" in two exhibit cases at the entrance to the History of Medicine reading room.

This year, the Library mounted one major exhibition according to the new model, one lobby exhibit similar to those produced in previous years, and one mini-exhibit. The major exhibition, "Emotions and Disease," presented the history of scientific and medical theories about the relationship of emotional states to health and disease. A joint project of NLM and the National Institute of Mental Health, the exhibition was co-sponsored by the Charles A. Dana Foundation, the John D. and Catherine T. MacArthur Foundation, and the Fetzer Foundation. It opened in November 1996 with a reception hosted by Dr. and Mrs. Lindberg and sponsored by the Friends of the National Library of Medicine and prompted substantial media discussion of the relationship between mental and physical health. A smaller exhibit, "So What's New in the Past? The Multiple Meanings of Medical History," was on display in the NLM lobby from July to September 1997. It explored some of the different meanings people have found in the history of medicine in the United States— history as weapon, as inspiration, as edifice, as profession, and as today's news. HMD also collaborated with staff members in the Bibliographic Services Division and the Lister Hill Center to design and mount a mini-exhibit on "William Blake at the National Library of Medicine," which highlighted Blake's works from NLM's collection as well as a Blake life mask and engraving tools owned by an NLM staff member.

HMD, the NLM EEO Office, the Washington Society for the History of Medicine, and the office of the Historian, Public Health Service co-sponsored a lecture by Professor Spencie Love on "One Blood: Charles R. Drew, M.D. and a Mythic Prescription for Our Times" in February in observance of African-American History Month. HMD and the NLM EEO Office also sponsored a presentation by Professor Judith W. Leavitt on "Personal Liberty or the Public Health? The Story of Typhoid Mary" during Women's History Month. HMD also arranged a series of nine public historical lectures for the NLM and NIH communities.

In FY 1997, the National Information Center on Health Services Research and Health Care Technology began work on the production of a short video history of the field of health services research to be used in its continuing education programs for health sciences librarians. The film will draw on footage from

a larger project to interview and videotape leaders in the establishment of the field of health services research.

HMD staff members presented historical papers and lectures during the year and also published the results of their scholarship in books, chapters, articles and reviews.

Training Programs for Health Sciences Librarians

In addition to the online services training program for search intermediaries previously described, LO also directs a post-masters training program for librarians and a continuing education program in health services research.

1997 was the fortieth anniversary of the NLM Associate Program. It provides individuals with a master's degree in library and information science with an understanding of NLM's programs and operations, key issues facing health sciences libraries, and new information technologies, and offers them an opportunity to develop and apply their skills by conducting special projects. NLM Associates also have an opportunity to visit other national libraries and a range of health sciences libraries and to attend professional meetings. As a follow-up to the report of the NLM planning panel on education of medical librarians and results of a series of NLM challenge grants to develop proposals for new librarian training initiatives, NLM has decided to expand the Associate program effective with the participants who enter the program in September of 1998. The number of Associates will be increased from 4 to 8 and an optional second year will be added to the program, to be spent in an academic medical center, hospital, or other health-related institution. The purpose of the second year is to give librarians experience in working with multi-disciplinary teams to integrate library and information services into the patient care or professional education programs of the parent institution.

NICHSR continues to locate, develop, field-test, and teach continuing education courses for health sciences librarians in the field of health services research and to make related training materials available via NLM's Web site. In FY 1997, NLM's *Introduction to Health Services Research* training manual and *Health Statistics Sources* by Michael Newman, Stanford University, were added to the NICHSR Web page. The conversion of *Introduction to Technology Assessment* by Clifford Goodman, Ph.D., Lewin Group, into html format for World Wide Web distribution is nearly complete. Classes on the Introduction to Health Services Research and the Use of National Center for Health Statistics Datasets were taught this year, and a Health Services Research

module, including introductory classes on health services research and epidemiology, will be presented at the annual meeting of the Medical Library Association in 1998.

The Future of LO Services

The rate of change in the way LO delivers services is accelerating, driven by advances in information technology and in the use of this technology in health care, publishing, and every day life. As described throughout this report, NLM is in the midst of reinventing its automated systems to provide a better technical platform for more advanced information services—resulting in changes that are having a substantial effect on our users. At the same time, major changes in the environment in which NLM and other medical libraries operate—for example, increased health professional and consumer access to Web-capable technology, the growth of electronic publishing, the digitization of library collections, the development of computer-based patient record systems—are offering important new opportunities for information services that can improve health care decision-making. These changes in the information environment are accompanied by yet-to-be-resolved problems, such as uneven access to advanced communications among health professionals, medical librarians, and consumers; debate about electronic intellectual property rights; lack of practical mechanisms for large-scale preservation of electronic data; and privacy concerns. The LO staff is working with other NLM divisions to take advantage of opportunities to expand and improve services—and to minimize the confusion and disruption that inevitably accompany significant change.

Table 2
Growth of Collections

Collection	Previous Total (9/30/96)	Added FY 1997	New Total (9/30/97)
Book Materials			
Monographs:			
Before 1500	576	2	578
1501-1600	5,793	11	5,804
1601-1700	10,124	3	10,127
1701-1800	24,446	7	24,453
1801-1870	41,121	10	41,131
Americana	2,341	0	2,341
1870-Present	629,099	15,414	644,513
Theses (historical)	281,794	0	281,794
Pamphlets	172,021	0	172,021
Bound serial volumes	1,072,325	28,939	1,101,264
Volumes withdrawn	(57,857)	(2,953)	(60,810)
Total volumes	2,181,783	41,433	2,223,216
Nonbook Materials			
Microforms:			
Total microforms	431,306	31,001	462,307
Reels of microfilm	83,539	6,459	89,998
Number of microfiche	347,767	24,542	372,309
Audiovisuals	60,723	1,360	62,083
Computer software	857	310	1,167
Pictures	56,601	0	56,601
Manuscripts	2,454,542	0	2,454,542
Total nonbook	3,004,029	32,671	3,036,700
Total book and nonbook	**5,185,812**	**74,104**	**5,259,916**

Table 3
Acquisition Statistics

Acquisitions	FY 1995	FY 1996	FY 1997
Serial titles received	22,600	22,522	22,378
Publications processed:			
Serial pieces	154,342	154,146	150,713
Other	22,743	20,846	21,422
Total	**177,085**	**174,992**	**172,135**
Obligations for:			
Publications	$4,788,181	$5,012,058	$5,282,348
Included for rare books	($189,527)	($209,178)	($267,221)

Table 4

Cataloging Statistics

	FY 1995	FY 1996	FY 1997
Completed Cataloging	22,232	20,365	20,296

Table 5

Bibliographic Services

Services	FY 1995	FY 1996	FY 1997
Citations published in MEDLINE*	392,354	322,825	519,012
For *Index Medicus**	374,907	309,038	499,794
Recurring bibliographies	6	6	----
Journals indexed for *Index Medicus*	3,093	3,205	3,211
Abstracts entered	293,724	242,544	398,576

* Figures for FY 1997 reflect reduction of a backlog carried over from FY 1996.

Table 6

Circulation Statistics

Activity	FY 1995	FY 1996	FY 1997
Requests Received	573,541	584,738	630,190
Interlibrary Loan	345,428	347,992	353,408
Onsite	228,113	236,746	276,782
Requests Filled:	424,169	410,732	463,011
Interlibrary Loan	245,078	227,810	252,830
Photocopy	231,993	215,461	239,575
Original	11,384	10,814	11,670
Audiovisual	1,701	1,535	1,585
Onsite	179,091	182,922	210,181

Table 7*

Online and Offline Searches

	FY 1995	FY 1996	FY 1997
AIDSDRUGS	1,453	1,240	781
AIDSLINE	66,200	59,507	51,159
AIDSTRIALS	3,148	2,203	1,588
ALERT	3,062	3,244	953
AVLINE	27,543	22,712	23,517
BIOETHICS	15,248	14,352	13,722
CANCERLIT	91,324	80,206	64,289
CATLINE	390,067	382,321	418,741
CCRIS	3,168	2,697	2,139
CHEMID	9,970	9,151	9,390
CHEMLINE	13,075	12,150	11,032
DART	3,861	2,556	2,962
DENTALPROJ	142	72	58
DIRLINE	33,427	27,602	18,090
DOCUSER	17,699	14,664	12,610
EMIC	3,787	3,743	4,906
EMICBACK	2,039	1,566	1,317
ETICBACK	1,046	666	622
GENETOX	1,913	2,935	1,414
HEALTH	178,295	93,183	
HealthSTAR		76,857	165,421
HISTLINE	6,806	6,577	8,750
HSDB	30,296	24,947	19,767
HSRProj	1,266	1,756	2,527
HSTAR	41,670	27,263	
IRIS	17,593	15,870	10,031
LOAN STATUS	19,939	21,304	23,773
MEDLINE	5,262,329	5,414,712	6,669,929
MeSH	38,058	39,138	55,810
PDQ	20,303	17,404	10,565
POPLINE	22,864	19,731	12,944
PREMEDLINE		3,001	49,857
REFLINE	39,984	33,541	14,011
RTECS	13,808	12,159	9,535
SDILINE	218,654	188,418	29,977
SERLINE	633,615	664,265	780,136
SPACELINE	1,507	3,119	5,037
TOXLINE	71,631	66,761	55,232
TOXLIT	11,074	9,954	6,566
TRI	12,793	8,959	5,898
TRIFACTS	511	441	420
Total	7,331,168	7,392,947	8,575,476

*Minor corrections made to FY 1995 and 1996 figures. FY 1997 is the last year that usage statistics will be kept on individual databases. The introduction of Web-based searching of MEDLINE and several other NLM databases on June 26, 1997, renders such record-keeping unnecessary.

Table 8

Online Searches—Interim Period

Web-Based Online Searches	FY 1997*
MEDLINE	12,536,000

*PubMed searches for the period June 26 through September 30, 1997. See footnote to previous table.

Table 9

Reference Services

Activity	FY 1995	FY 1996	FY 1997
Reference Section:			
Offsite requests	25,135	24,799	21,297
Onsite requests	43,132	34,796	40,851
Total	68,267	59,595	62,148

Table 10

History of Medicine Activities

Activity	FY 1995	FY 1996	FY 1997
Acquisitions:			
Books	154	188	97
Modern manuscripts	184,000	163,500	750,000
Prints and photographs	2,046	183	284
Historical audiovisuals	147	108	36
Processing:			
Books cataloged	2,451	351	704
Modern manuscripts processed	0	0	210,000
Pictures cataloged	0	0	0
Citations indexed	2,510	1,363	2,430
Public Services:			
Reference questions answered	13,434	12,885	13,701
Onsite requests filled	7,273	5,289	6,313

Specialized Information Services

Melvin Spann, Ph.D.
Associate Director

Thirty years ago, the National Library of Medicine established the Division of Specialized Information Services (SIS) and the Toxicology Information Program (TIP) now known as the Toxicology and Environmental Health Information Program (TEHIP). TEHIP's evolution has kept pace with the spread of, and demand for, toxicological and environmental health information, by taking advantage of new computer and communication technologies. Such mechanisms have enabled us to provide more rapid access to a wider audience. Our development of novel search capabilities means that users require less knowledge of search techniques and thus allows data to be relayed to them more efficiently. Finally, we are moving beyond the bounds of the physical National Library of Medicine, and exploring ways to point and link users to relevant sources of toxicological and environmental health information wherever they may reside. This will be accomplished primarily through the TEHIP and AIDS Web sites recently made public by SIS. Currently, refinements and additions are being made to allow easy access to a range of information collected by this Division of NLM.

In FY 1997 SIS reexamined the scope and coverage of current SIS programs, proposed new opportunities to enhance SIS information services, and investigated emerging areas, including the application of new multimedia technology in the delivery of information services. This examination was initiated using the mechanism of an Institute of Medicine evaluation of the TEHIP Program, and a report from this study, *Toxicology and Environmental Health Information resources: the Role of the National Library of Medicine*, was released in the spring of 1997.

TOXNET

The **Toxicology Data Network** (TOXNET), the Library's networked microprocessor system, has upgraded more microprocessors to Pentium machines to better serve the public. Work continued on the further enhancement and implementation of the Windows workstation for building and updating Hazardous Substances Data Bank (HSDB) records and other TOXNET files.

The TOXNET System Contractor completed a study of new platform alternatives and has recommended the most advantageous hardware environment (a Unix-based platform) to NLM for implementation in FY 1998. A major enhancement to the search/retrieval module is in progress. A World Wide Web search interface to TOXNET is being developed to facilitate searches by TOXNET users coming through the Internet. Initially, the Web interface will allow searching of the HSDB, CCRIS, IRIS, RTECS, and GENE-TOX databanks. Additional files will be added as soon as possible.

Databases under TOXNET

The **Hazardous Substances Data Bank** (HSDB) continues to be the most highly used databank on TOXNET, averaging over 6,000 searches each month. Increased emphasis was placed on providing more data on human toxicology and clinical medicine within HSDB, in keeping with earlier recommendations of the Board of Regents Subcommittee on TEHIP. Changes to the composition of the Scientific Review Panel are being made to accommodate the shift in content emphasis. Newer sources of relevant data are being examined for incorporation into new and existing data fields within the current 4,520 HSDB records. More records are being processed through special enhancements, including source updates from various peer-reviewed files. These enhancements are being made possible by a customized Windows-based PC workstation with enhanced file-building features.

The **Toxic Chemical Release Inventory** (TRI) series of files now includes nine online files, TRI87 through TRI95. These Environmental Protection Agency-sponsored files remain an important resource for environmental release data and continue to attract new users. They contain data on environmental release data to air, water, and soil for over 600 EPA-specified chemicals. Starting with the TRI91 file, reporting facilities were required to report source reduction and recycling activities, in addition to environmental releases, which has considerably increased the size and complexity of the databases.

The TRI95 file was released in May 1997. EPA included over 300 additional chemicals in TRI95, as well as required submissions from federal facilities and military installations. Another major change in TRI95 was the change to two versions of the reporting form, a long and a short form, requiring unit record changes in TOXNET. TRIFACTS, a companion file to the TRI series, supplies users with information related to health and ecological effects and the safety and handling of the TRI chemicals. These records are

supplied by EPA, which also funds the management and maintenance of the TRI files.

The **Chemical Carcinogenesis Research Information System** (CCRIS) continues to be built, maintained, and made publicly accessible by NLM. This databank is supported by the National Cancer Institute and has grown to 7,555 records. The chemical-specific data covers the areas of carcinogenesis, mutagenesis, tumor promotion, and tumor inhibition.

The **Integrated Risk Information System** (IRIS), EPA's official health risk assessment file, continues to experience high usage on TOXNET and to be very popular with the user community. During FY 1996, EPA introduced a version of IRIS on the Agency's Web page. IRIS now contains 666 chemicals.

The **GENE-TOX** file continues to be built and updated directly on TOXNET by EPA scientific staff. This file contains peer-reviewed genetic toxicology (mutagenicity) studies for about 300 chemicals. GENE-TOX is much used by scientists in other countries.

The **Registry of Toxic Effects of Chemical Substances** (RTECS) is a databank based upon a National Institute of 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 as NIOSH makes them available. This year, information on a standard European reference, the Beilstein Number and Reference, was added to the file. Now that NIOSH no longer prints the full RTECS, online access via NLM to this widely used resource including over 139,000 chemical records has become even more important.

The **Developmental and Reproductive Toxicology** (DART) database now contains over 35,000 citations from literature published since 1989 on agents that may cause birth defects. Records from DART are also added quarterly to TOXLINE. DART is a continuation of the Environmental Teratology Information Center backfile (ETICBACK) database, which contains almost 50,000 citations to literature published from 1950 to 1989. ETICBACK citations are also found in TOXLINE. DART is funded by NLM, the EPA, the National Institute of Environmental Health Sciences (NIEHS), and the FDA's National Center for Toxicological Research, and is managed by NLM.

The **Environmental Mutagen Information Center** (EMIC) database contains over 17,000 citations to literature on agents that have been tested for genotoxic activity. A backfile for EMIC (EMICBACK) contains over 75,000 citations to the literature published from 1950 to 1991. 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 EPA and the NIEHS and managed by NLM.

Databases under ELHILL

ChemID (Chemical Identification File) is an NLM online chemical dictionary that contains over 340,000 records, primarily describing chemicals of biomedical and regulatory importance. It also contains an important set of regulatory data, collectively known as SUPERLIST. Over 13,000 records are augmented with the name and an indication of source for chemicals mentioned in one or more of 31 lists, e.g., the Priority List of the Agency for Toxic Substances and Disease Registry (ATSDR). These data allow users to determine if a chemical is mentioned on a given list and under what name, as well as to search for chemical classes on these lists. In FY 1997, new data from the EPA Master Testing List status was added to ChemID. During FY 1998, an extensive quality control effort will be continued on data in this file and new nomenclature data will be added to enrich the file content because of the impending retirement of CHEMLINE.

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. 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 extensively by NLM with nonproprietary data from a variety of sources. CHEMLINE now contains over 1,600,000 records on chemical substances. On January 1, 1988, CHEMLINE will no longer be carried by NLM since it is a royalty-based file requiring payment of usage fees through NTIS.

TOXLINE (Toxicology Information Online) is an NLM online bibliographic retrieval service produced by merging "toxicology" subsets from some 18 secondary sources. TOXLINE and its backfile, TOXLINE65, contain data from sources that do not require royalty charges based on usage. Because the Chemical Abstracts Service requires usage royalties, NLM has separated information from this source into two online bibliographic files, **TOXLIT** and **TOXLIT65**. At the start of FY 1998 all these toxicology bibliographic files together contained over 4 million citations.

In FY 1998, as NLM moves more toward providing even more free Internet access, the two files with CAS licensed data, as in the case with CHEMLINE, will also be discontinued. We also will continue to consider ways that the Unified Medical Language System can be utilized to accommodate vocabulary changes in TOXLINE subfiles and help users who will access our data through Internet Grateful Med. IGM access is expected during FY 1998.

The decision to no longer carry the TOXLIT files has prompted NLM to establish a working group to explore alternative ways to acquire toxicology bibliographic information. During FY 1997, SIS collaborated with NIOSH in evaluating the use of their NIOSHTIC database, which is also included in TOXLINE. We are exploring possible options to replace this information or share efforts with NIOSH to create it should NIOSH decide to discontinue building it as they have in the past.

DIRLINE (Directory of Information Resources Online) is NLM's online directory of resources including organizations, databases, bulletin boards, as well as projects and programs with special biomedical subject focus. These resources provide information to users that may not be available from one of the other NLM bibliographic or factual databases.

The availability of DIRLINE via the Internet through NLM's Locator continues to result in a high level of use of the database and a higher level of recognition of its utility by biomedical librarians, health professionals, and the public. Increased funding for database maintenance has made it possible to improve the quality and timeliness of the content of the database and improve collaboration with other subfile producers. An online version of *Health Hotlines*, NLM's popular publication listing toll-free telephone numbers to health-related sources, was made available this year through the SIS Web site. In FY 1997, DIRLINE was also added to the list of databases available free via the Internet Grateful Med.

AIDS Services

The recommendations from the 1993 NIH HIV/AIDS Information Services Conference remain the guide for NLM's HIV/AIDS information programs and services. In 1996 NLM initiated a World Wide Web Homepage for AIDS (http://sis.nlm.nih.gov/aidswww.htm). This distribution mechanism has enhanced NLM's ability to get important information to the people who need it. In addition to links with NLM's online databases, fact sheets, and publications, links to other NIH components are available as part of a new online version of the Guide to NIH HIV/AIDS Information Services. New databases of abstracts presented at important AIDS conferences have also been made available through this mechanism prior to their addition to AIDSLINE.

Two collaborative projects with other PHS agencies, the AIDS Clinical Trials Information Service (ACTIS) and the HIV/AIDS Treatment Information Service (ATIS) continue to be very successful. While direct online use of the AIDSTRIALS and AIDSDRUGS databases has stabilized, many other information providers are downloading the data and adding it to their services. A number of bulletin boards and World Wide Web servers are providing the data to their users in a variety of formats and the data have also been distributed internationally through this mechanism. In addition, in FY 1997 AIDSTRIALS and AIDSDRUGS became available for searching via Internet Grateful Med.

NLM initiated a fourth round of AIDS Outreach Purchase Orders in FY 1997 and made awards to 20 community-based organizations and libraries. Six previously funded projects that had shown evidence of success received funding to continue or expand their activities.

NLM has continued to work with the Health Information Center at the Wheaton Regional Library (Montgomery Country, Maryland) on the collaborative project in AIDS and toxicology information. Outreach programs have taken place at the library, which have brought in new users for the electronic services. Wheaton Library staff have developed and conducted training for the public in the use of the World Wide Web for information retrieval. Satellite centers are being established in the three other regional libraries that will bring health resources to consumers throughout the county.

Training and outreach to the Historically Black Colleges and Universities has resulted in the development of training materials with culturally specific content. A number of training sessions were held during the year with very positive feedback from the trainees.

Other Programs

Internet

The NLM and SIS gopher servers were taken out of service, since most Internet usage now comes to the World Wide Web servers. All content was converted to the World Wide Web servers.

The SIS WWW servers offered a variety of new services in 1997. New searchable versions of *Health Hotlines*, as well as meeting abstracts of the 4th Conference on Retroviruses and Opportunistic

Infections were made searchable under WAIS-SF. A new Chemical Structure Search Page was made available on a separate server, and allowed searching HSDB structures using structural input and output. Our online training capabilities were expanded with the advent of Toxicology Tutor on the Web, which used the latest Javascript capability to present fundamental concepts in toxicology, and new online file demonstration packages were made available using Javascript. A new HTML version of the Calendar of Events in Toxicology and Environmental Health was also made public. The publication, *Bibliography on Alternatives to the Use of Live Vertebrates in Biomedical Research and Testing* was moved from the gopher to the WWW, and hypertext versions of the latest editions were made available. A new Bibliography on Environmental Justice was also developed and published on the WWW server in HTML format.

Outreach

SIS continues 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 (ATSDR) to train representatives from 61 additional schools in the use of NLM's valuable online resources. During the past fiscal year, one of the training classes was hosted by Texas Southern University and it included HBCUs and community-based organizations from the Lower Mississippi Delta. This class was jointly sponsored by ATSDR, NLM, and the EPA's Environmental Justice Office in support of the Mississippi Delta Project.

User Support Computer-Based Activities

A new version of *Toxicology Tutor,* developed in FY 1996, was released for the Internet this year, and is available on the SIS Web server. Also, SIS developed a computer-based demo for the Internet illustrating the TEHIP databases: CHEMLINE, TOXLINE, RTECS, HSDB, CCRIS, IRIS, TRI, TRIFACTS, GENETOX, EMIC, and DART. It was released in FY 1997, together with an Internet version of the TEHIP slide overview.

Alternatives to Animal Testing

SIS continued to compile and publish references from the MEDLARS files that were identified as relevant to methods or procedures which could be used to reduce, refine, or replace 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. Bibliographies issued during the past four years are available on the Internet through the SIS WWW Server.

LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS

Alexa T. McCray, Ph.D.
Director

Introduction

The Lister Hill National Center for Biomedical Communications was established by a joint resolution of Congress in 1968 and serves as an NLM intramural research and development division. The Center conducts research, provides training opportunities, develops and deploys systems used in NLM production activities, and sponsors and monitors external research through telemedicine and other research contracts.

Center research is carried out through several major research programs, all sharing the same broad purpose of improving health-care information dissemination and use. This involves a wide range of research activities and disciplines. Research staff conduct their research by drawing on a diverse set of scientific fields and methods. Current staff have backgrounds in medicine, computer science, library and information science, linguistics, cognitive science, education, and engineering. Research projects include digital library research, automated indexing techniques, vocabulary and thesaurus research, natural language processing, medical knowledge representation, image processing, database design, machine learning, expert systems, and computer-based learning. The Unified Medical Language System knowledge sources, the Visible Human data set, and MEDLINE data are important sources of knowledge for many research experiments. The most current information about Center activities can be found at the Center's Web site (lhncbc.nlm.nih.gov*)*.

A Board of Scientific Counselors meets regularly to review the quality and contents of the intramural research programs at the Center. The Board is composed of scientific and technical experts (see Appendix 4 for a list of members) who are prominent leaders in the fields of medicine, computer science, engineering, and health professions education.

The Center is currently organized into five components:
- Audiovisual Program Development Branch
- Cognitive Science Branch
- Communications Engineering Branch
- Computer Science Branch
- Office of High Performance Computing and Communications

The **Audiovisual Program Development Branch** conducts media development activities with three specific objectives. As its most significant effort, the branch supports the LHNCBC's research, development, and demonstration projects with high quality video, audio, and graphic materials. From initial project concept, through actual project implementation with image preservation, transfer and display, to project evaluation and reporting, all forms and formats of imaging are produced. Consultation and materials development are also provided by the branch for the NLM's educational and information programs. With the mission requirement of the Library to include effective outreach activities, the support that the branch provides to these programs continues to increase. From optical media technologies to teleconference support, the graphics, video, and audio materials requirement has increased in quantity and diversified in format. The third area of concentration is the development of technical improvements including issues such as image resolution, color fidelity, media transportability, media storage and image communication. Facilities and hardware systems must reflect state of the art standards in a very rapidly changing field. High resolution video is a development area being explored that represents the future for improved electronic image quality. Multimedia systems and techniques, visualization and networked media 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 graphics artists in the branch. Digital video and image compression techniques are central to projects being pursued to improve image storage and communication.

The **Cognitive Science Branch** conducts research and development in computer and information technologies, disseminates information about these technologies to the Library's various constituencies, and supports their application in health professions education. The Learning Center for Interactive Technology is a place where individuals come to learn about the most current technology and its potential application in the health care setting. Branch staff regularly host a large number of visitors to the Center, providing both individual consultation as well as workshops and tutorials for larger groups. Branch staff mentor high school, undergraduate, graduate, and post-

doc students regularly through a variety of programs, including the Center-wide Medical Informatics Training Program. Important research areas for the branch involve the investigation of a variety of techniques, including linguistic and knowledge based methods, for improving information retrieval from complex systems. Branch members collaborate with other NLM research staff in an indexing initiative project whose goal it is to develop automated and semi-automated techniques for indexing the biomedical literature. Branch staff actively participate in most aspects of the Unified Medical Language System project, including the development and maintenance of the UMLS Knowledge Source Server, the Semantic Network, and the SPECIALIST lexicon. The branch conducts research in digital libraries and is currently collaborating with NLM's History of Medicine Division in a document digitization project. The branch maintains a World Wide Web server (wwwcgsb.nlm.nih.gov) that provides information about research projects and resources, including a variety of tutorials on developing Internet-based educational applications and distance learning programs.

The principal focus of the **Communications Engineering Branch** is applied R&D in image engineering and communications engineering, motivated by NLM's mission-critical tasks such as document delivery, archiving, and automated data entry, as well as wide Internet access to mixed text and x-ray image databases, and imaging applications in support of medical educational packages employing digitized radiographic, anatomic, and other imagery. The branch conducts R&D in the capture, storage, processing, online retrieval, transmission and display of both biomedical documents and medical imagery. Data types of interest include bitmapped bitonal document images, digitized color documents, digitized x-rays, color cryosection images from NLM's Visible Human collection, and motion video. Areas of active investigation center on image compression, image enhancement, image understanding, pseudo-grayscale rendition, image transmission and networks implemented via asynchronous transfer mode (ATM) and satellite technologies, optical character recognition and man-machine interface design applied to automated data entry. Current information on Branch projects appears in the branch Web server (archive.nlm.nih.gov).

The **Computer Science Branch** applies techniques of computer science and information science to problems in the representation, retrieval and manipulation of biomedical knowledge. Branch 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. Research issues include knowledge representation, knowledge base structure, knowledge acquisition, and the human-machine interface for complex systems. Important components of the research include embedded intelligence systems which combine local reasoning with access to large-scale mainframe databanks. Other project work involves multimedia knowledge-based systems with interactive video capability and systems which search multiple databases with a single user query. Branch research staff include the team which developed NLM's Internet Grateful Med program and the team which annually produces the Unified Medical Language System Metathesaurus and Information Sources Map. Staff members participate actively in the medical informatics and information science research communities and professional specialty societies. They participate in the meetings of the Internet Engineering Task Force and the International World Wide Web Conference Committee. Finally, recognizing the importance of addressing the future of medical informatics by helping to train new researchers, Branch staff conduct the eight-week NIH elective in Medical Informatics for third-year and fourth-year medical students each spring.

The **Office of High Performance Computing and Communications** coordinates HPCC and Next Generation Internet planning, research and development activities with federal, industrial, academic, and commercial organizations. The Office acts as NLM's contact point for its telemedicine interests and coordinates NLM activities in this area, including the NLM telemedicine contracts. NLM is the lead agency within DHHS for the government's High Performance Computing and Communications initiative and as such has a direct interest in the use and effects of the National Information Infrastructure (NII) on health care. The growth of the NII and the increasing access to high-speed computers and communications by consumers, health care providers, public health professionals, and basic, clinical, and health services researchers is having a fundamental effect on health and human services throughout the nation. The telemedicine program is designed: (1) to evaluate the impact of the NII on health care, research, and public health; (2) to test methods to preserve the privacy of individual health data while also providing efficient access for legitimate health care, research, and public health purposes; and (3) to assess the utility of emerging health data standards in health applications of advanced communications and computing technologies. Coordination of externally funded Visible Human

research projects is a second major focus of the office. A successful and broadly attended conference was held in the fall of 1996, bringing together early adopters, potential users and developers, and other interested individuals for the purpose of discussing the results of using the Visible Human datasets.

Current Center programs and activities fall into several major categories:

- Computer and information science research, including medical language processing, expert systems, machine learning, and Web based application development
- Image processing research, including image acquisition, storage, retrieval, and dissemination
- Education and training in medical informatics and learning technologies
- Resource Support and Development, including core system support, system security and advanced network planning
- External Research Support, including support for National Information Infrastructure Telemedicine projects and the Digital Library Phase 2 Initiative

Computer and Information Science Research

Digital Library Research

This year the Center has initiated a major project in Digital Library research. The project builds on work already accomplished under a special joint project with the History of Medicine Division begun in the spring of 1997, as well as on earlier work conducted under the EIDOS document management project. The research program involves all aspects of creating and disseminating digital collections including proposed and adopted standards, emerging technologies and formats, effects on previously established processes, and protection of original materials. The project explores new methods for creating and accessing digital library collections, including semantic based retrieval methods using the UMLS knowledge sources, the development of modular and open information environments, investigation of the role of well-structured metadata, and the exploration of different "points of view" on the same underlying data set.

An example of early work conducted in this area is the Regional Medical Programs collection. The heart of the collection is approximately 40,000 pages comprising some 1,500 documents related to the Regional Medical Programs of the 1960's and 1970's. All textual materials in the collection can be browsed or searched. An ongoing collaboration with the NLM's History of Medicine Division is expected to result in a

technology transfer which will result in a digital library developed and maintained by the Division's Modern Manuscripts section. The section will house and disseminate historically significant digitized collections, particularly those of prominent biomedical scientists. A work in progress is a sample of the collection of former NIH Director Dr. Donald S. Fredrickson. This collection includes papers, photographs, and multimedia materials relating to Dr. Fredrickson's scientific career and personal history. The collection has been presented from two different points of view. One view highlights the major phases of his career, and the other focuses on telling the story of his scientific achievements.

Indexing Initiative

For more than 150 years, the NLM has provided access to the biomedical journal literature through the analytical efforts of human indexers. Since 1966, access has been provided in the form of electronically searchable document surrogates consisting of bibliographic citations, descriptors assigned by indexers from the MeSH controlled vocabulary and, since 1974, author abstracts of many, but not all, items. The objective of the Indexing Initiative project work is to investigate methods whereby automated indexing methods partially or completely substitute for current indexing practices. The project will be considered a success if methods can be designed and implemented that result in retrieval performance that is equal to or better than retrieval of citations based on humanly assigned index terms.

The project assumes the continued existence and growth of NLM's MeSH vocabulary and of the UMLS Knowledge Sources and is investigating concept-based indexing methods that go well beyond automatic word-based indexing. One project seeks to discover semantic relationships between phrases in text as a way of more accurately representing content. Another identifies Metathesaurus concepts in biomedical text and then maps these through a weighting algorithm based on linguistic and knowledge-based techniques to appropriate MeSH terms. One set of experiments has demonstrated the value of such concepts for automatic query expansion. Another investigation provides an enhanced representation of semantic content by ranking concepts assigned to MEDLINE abstracts on the basis of frequency of occurrence and specificity as measured by hierarchical depth in MeSH. Methods for providing better access to MeSH terminology through the UMLS are also being investigated and these should lead to improved MeSH validation procedures as part of the indexing process.

Unified Medical Language System (UMLS)

Ongoing UMLS projects include the deployment of the redesigned system for supporting on-site and off-site network-based Metathesaurus enhancement, editing and maintenance; the annual updating and refinement of the Metathesaurus, the Semantic network and the SPECIALIST lexicon; the continued development and maintenance of the UMLS Knowledge Source Server, the further development of the prototype Information Sources Map; and the development of application programs such as the Information Sources Map Sourcerer program, the UMLS Knowledge Source Server, and the Internet Grateful Med system.

UMLS Metathesaurus: The Metathesaurus is the foundation product of the UMLS initiative. It is a machine-readable knowledge source representing multiple biomedical vocabularies organized as concepts in a common format. It thus provides an immensely rich terminology resource in which terms and vocabularies are linked by meaning. The two major priorities of the Metathesaurus development group have been to release annual editions rapidly increasing in content and to design and install new software systems which reflect the group's increased understanding of the science and practice of incorporating and linking vocabularies in a unified concept-oriented system.

There are some 100,000 additional meanings in the 1998 Metathesaurus, an increase of more than 30% over the 1997 release. The largest additions include two thirds of the Read Codes from the United Kingdom (completing its inclusion) and all of ICD10, which will soon be used in the United States. The 1998 Metathesaurus includes full current versions of ICD9-CM and CPT, the two most commonly required coding systems in this country.

The software environment in which the Metathesaurus is edited and maintained has been completely redesigned. The redesigned Metathesaurus software system has been installed and is undergoing testing in parallel with the present system. While the initial goal was to implement equivalent functionality and move on from there, the new design already provides considerable improvements in editor support, data management, and reversibility of editing actions. Testing will be complete and full production use of the new system will begin in December 1997 when editing of the next Metathesaurus begins. Ongoing efforts include extension of the new system to provide more automated production of release versions, and study of the requirements and initial design of an integrated vocabulary development environment. The goal of this environment will be to allow the maintenance of multiple autonomous terminologies while meeting technical standards and maintaining inter-vocabulary links to prevent each from reinventing in isolation. Interesting technical challenges include support of remote client-server workstations on Sun and Windows NT platforms, alternative relational database management system (RDBMS) client and server systems, Web-based platform-independent software clients, and standardized supplementary information servers for editing and management.

UMLS SPECIALIST Lexicon and Lexical Programs: The SPECIALIST lexicon, a large syntactic lexicon of medical terminology containing over 100,000 lexical items, is developed by the Natural Language Systems group and is distributed annually as one of the UMLS Knowledge Sources. Lexical items are collected into unit records containing morphological, syntactic and spelling information about each item. Morphological information includes full inflectional information about each item. A database of derivational relationships is also maintained with the lexicon. Morphological and spelling information is important for matching of lexical items and forms a major part of the capability of the lexical tools. Syntactic information includes verb complement patterns for verbs and sequencing information for adjectives. A recent experiment on the lexicon's coverage of MEDLINE text showed that the lexicon was able to account for 95% of the word-tokens in a random sample of MEDLINE abstracts.

The lexical variant programs use data from the SPECIALIST lexicon as they compute different forms of lexical items. The lexical programs consist of several modules that can be combined in a variety of ways to generate variants. For example, users may be interested in seeing only the singular or plural of an input term. In that case, they would choose the inflection option. Or, they may be interested in running their terms against a stopword list and also ignoring word order so as to maximize their chances of finding related terms in a particular vocabulary or text. In this case, they would choose the stopword removal and word sorting options.

The programs allow for a good deal of flexibility in matching one term to another. The basic principle that is involved in using the programs is that any manipulation of a source input term or terms must involve the same manipulation of the target terms. For example, if users want to see if terms in their particular vocabulary source are found in the Metathesaurus target, and if they want to find those terms regardless of whether they appear there in the singular or plural or whether they are in upper or lower case, then they would transform the source vocabulary using the lowercasing and inflectional options and, importantly,

they would do the same for the Metathesaurus terminology. Since some users will prefer to use a method that does not involve additional processing of the Metathesaurus data, a normalization program together with normalized string and word indexes of all Metathesaurus terminology is included with the lexical programs. The normalization process involves splitting a string into its constituent words, lower-casing each word, converting each word to its base form, ignoring punctuation, and sorting the words in a multi-word term into alphabetic order.

UMLS Information Sources Map: A new release of the Information Sources Map (ISM) was generated for the 1997 UMLS distribution CD-ROMs, reflecting substantially altered record structure and content. The Sourcerer ISM testbed was transferred from a workstation to the ISM's own workgroup server, and the database and all Java language components were revised in conjunction with the CD-ROM work as well as being updated with the latest versions of any outside software they depended upon. Experiments in automated document indexing were conducted with the help of an NLM Library Associate in collaboration with the Natural Language Systems group.

ISM staff completed several generations of Java-and CORBA-based interfaces to the Natural Language Group's MetaMap program, the UMLS Metathesaurus, the 50+ MEDLARS databases, and the ISM database. The latest versions of these interfaces comply with the Common Object Request Broker Architecture (CORBA) 2.0 specifications. All four servers have been integrated into a common CORBA repository on the principal ISM workgroup server. Java's component software architecture technology, Java Beans, was evaluated for potential application to ISM work. The Visibroker commercial CORBA implementation and the CORBA request broker built into the Netscape Navigator Web client were explored as well. Other staff members extended work on Elvis, the Java ISM client. Work on source selection algorithms progressed to the point that Elvis was able to make source selections for the MEDLARS family of databases that correlated well with those made by human experts. Selections were less successful for TOXLINE, where additional work will be directed in the coming year.

UMLS Knowledge Source Server: The Internet-based UMLS Knowledge Source Server provides access to all of the UMLS knowledge sources and can be accessed through three different client interfaces. The World Wide Web interface allows users to browse and explore the data and to see how those data are organized in the UMLS. Users can request

information about a particular concept, for example, retrieving all its synonyms, its definition, its semantic type, and all the other concepts that are saliently related to it in the Metathesaurus. Users can navigate the Semantic Network, exploring its structure, and they can, for example, retrieve all the concepts that refer to medical devices or diseases in the Metathesaurus. The SPECIALIST lexicon can be searched, and syntactic and morphologic information about the lexical item will be displayed, together with a link to a Metathesaurus definition if there is one. The Web interface to the Information Sources Map currently allows users to see a description for each of the databases and also to browse a sample record for each database.

The command-line interface to the server is best suited for batch processing. Researchers can submit a list of terms to the server to see if they can be found in the UMLS; they can search for various attributes of the terms that are found; and they can filter the results, limiting the result set by attribute, for example, to just those terms that have a particular semantic type or a particular lexical tag.

The Application Programming Interface (API) allows developers at remote sites to embed calls in their application programs to the Knowledge Source Server, thereby accessing the UMLS data directly over the Internet. The API has been designed to be simple, consisting of functions for establishing connections to the server, posting queries, and retrieving the query results.

Natural Language Systems

The Natural Language Systems research team is investigating the contributions that natural language processing methods can make to the task of mediating between the language of users and the language of the databases they attempt to access. The team has developed the prototype SPECIALIST system, an experimental natural language processing system for the biomedical domain.

The group is developing a text analysis system that is heavily based on the SPECIALIST lexicon and associated lexical tools. This system is modular in design to allow for flexible use and continuous enhancement. The modules are servers which will be available to a variety of clients for a variety of uses. When complete, the system will consist of 6 or more modules: a tokenizer module to analyze text into tokens and label them; a sentence identification module to analyze text into sentences; a lexical look-up module to find lexical items in the text; a shapes module to identify items in the text that do not occur in the lexicon but have types recognizable from their form; and a parser module to assign phrase structure to the

sentences of the text. A structured text analysis module is also planned. This module will deal with structured text such as text with SGML markup, MEDLINE record structure or other structuring devices. Preliminary versions of each module, with the exception of the shapes module, have been created.

The tokenizer and sentence identifier modules of this early version use regular expressions to break text into tokens and sentences and to identify the type of each token. The pattern used to identify sentence breaks was selected after some preliminary empirical studies of sentence breaks in MEDLINE text. The lexical look-up module uses a Berkeley DBMS database. Multi-token lexical items (multi-word terms) are retrievable by their first token allowing the lexical look-up module to detect multiple token lexical items without combinatorial explosion. Lexical look-up currently uses a strategy of finding the longest left anchored match to cover a span of text.

The design of the parser module is strongly based on the syntactic information encoded in the SPECIALIST lexicon. After consideration of several grammar paradigms, Categorial Grammar was chosen as a grammatical system that expresses this lexically based approach to sentence structure. The lexical record for each lexical item will be translated into a Categorial Grammar category symbol which expresses how that lexical item combines with other lexical items and phrases to form a larger phrase. In the preliminary version of the system the place of the Categorial Grammar Parser is being held by a simple phrase extraction program. This program extracts phrases containing nouns and their pre-modifiers. An early application of the text processing system will be an experiment to test the hypothesis that the SPECIALIST lexicon encodes a large part of the information needed to parse medical English text.

Lexical Systems

A lexicon building tool is used to facilitate the building of the SPECIALIST lexicon and maintain its consistency. The tool enforces a complete and consistent form for lexical entries, and provides users with a menu based approach to entering lexical information. A lexicon grammar specifying the correct form of lexical records assures that lexical records created by the tools are correctly formed. The first large-scale effort in lexicon building was initiated in 1990; it stands today at over 100,000 lexical items. The lexicon consultants are working off-site with the new Web-based tool, LexBuild. New lexical records are downloaded to NLM via the Internet biweekly. Each of the lexicon building projects has emphasized quality control and correction of errors in the existing lexicon

as well as its growth. The current effort has concentrated on terms from the UMLS Metathesaurus that bear lexical tags. In the future lexical tags will be removed from the Metathesaurus and the lexicon will carry that information. NLM's consultants, Dr. Lynn McCreedy and Dr. Susan Hoyle have also been reviewing the derivational variation tables used by the lexical tools. Both consultants are professional linguists with broad knowledge of the issues involved in lexicon coding.

The SPECIALIST lexical tools are designed to help users abstract away from a wide range of lexical variation. The three tools lvg, norm and wordind, which offer this sort of capability, were described in last year's report.

Semantic Knowledge Representation

Access to biomedical information depends on reliable representation of the knowledge contained in text. The Semantic Knowledge Representation project is concerned with developing programs to provide usable semantic representation of biomedical free text by building on existing resources. Such knowledge representation supports improved access to biomedical documents, including automatic indexing for information retrieval, retrieval of images, structured browsing and navigation facilities, document profiling, and the mining of information from textual sources.

The programs being developed by the project assign semantic structures as a representation of the information contained in text and depend crucially on the knowledge sources available in the UMLS and on the SPECIALIST natural language processing tools. The parser provides syntactic analysis for biomedical free text. MetaMap maps syntactically analyzed text to Metathesaurus concepts. The SemRep program then uses the syntactic structure and the UMLS knowledge to map to the Semantic Network, thereby providing a semantic conceptual structure to represent the information contained in the text being processed.

One important aspect of enhancing the MetaMap algorithms is improving the treatment of ambiguous abbreviatory expressions in medical text. Programming is currently under way to discover abbreviations and their expansions in the text being processed. The correct identification of such expansions contributes significantly to the resolution of ambiguous abbreviations. In improving SemRep, the rules which map linguistic structures to relations in the Semantic Network need to be extended. A semi-automatic method for discovering such rules, based on patterns of semantic types occurring in sentences in biomedical text, is being developed, and machine learning techniques are being investigated as a possible means of

discovering the regularities in text upon which such rules depend.

The Parser and MetaMap (with supporting resources) are currently being used in several research projects. They serve as a natural language front end for the Sourcerer prototype which accesses the UMLS Information Sources Map. They also support research in the Indexing Initiative, which is investigating improved methods in indexing and retrieval.

Artificial Intelligence and Expert Systems

Several projects use artificial intelligence methods as the basis for the research questions being asked. The Expert Systems program has developed the AI/RHEUM consultant system in rheumatology and the CTX multimedia shell. AI/RHEUM and the CTX shell are used for the exploration of issues in knowledge representation, knowledge acquisition, user interface design, and the evaluation of medical expert systems. The primary goal of the Machine Learning Project is research into artificial intelligence systems that can use diverse sources of data and diverse inference methods in the automatic creation of large, useful knowledge bases or novel program abilities. The project concentrates on research issues in knowledge acquisition planning and on the development of prototype learning systems in biotechnology domains. The MedIndEx project is in its final phases and will be completed in Fiscal Year 1998. The remaining tasks are to provide a publicly available installation package and minimal support for outside users.

Expert Systems Program

The objective of the Expert Systems Program is to build computer programs that facilitate access to knowledge. Such programs could serve as surrogates for the human expert or as high level tools for the expert in the presence of information overload. The Expert Systems Program builds intelligent multimedia expert systems to investigate issues in knowledge representation and knowledge base structure, knowledge acquisition and knowledge base maintenance, the evaluation of knowledge-based systems, and the creation and delivery of knowledge-based systems over the Internet.

The Expert Systems group has built a multimedia expert system shell called CTX, for Criteria Table Expert. The shell includes a knowledge base compiler, a run-time system which can access information from multiple knowledge sources, a knowledge base editor, a case editor and a suite of automated testing programs for analyzing system performance against sets of benchmark test cases. It has been tested successfully with knowledge bases in several widely varying domains.

A version of CTX for MS-DOS machines, a version called WinCTX for machines running Microsoft Windows, and a CTX inference engine which runs on Unix and DOS machines are complete and have been used by a number of outside collaborators. Software development in this fiscal year has concentrated on Internet-CTX, a version of the CTX program that delivers criteria table based applications over the Internet. The run-time system is complete and has been made available to several outside collaborators.

The Internet versions of the knowledge base editor and compiler are currently being implemented to allow development of expert systems over the Internet. The Internet-CTX software was used earlier this year to implement a prototype dermatology expert system. Potential collaborators in other institutions have indicated interest in continuing this work. Development on Internet-CTX will continue in the coming year to complete the Internet-based knowledge base editor and compiler. Design and development of new applications in pediatric rheumatology, dermatology, emergency medicine and psychiatry will be explored with collaborators from several U.S. medical schools and the National Naval Medical Center. Expert Systems Program staff will offer knowledge engineering support in the initial stages of these projects.

The AI/RHEUM knowledge-based diagnostic system in rheumatology has been a research testbed for the Expert Systems Program for many years. Expert Systems Program staff together with collaborating rheumatologists Dr. H. James Williams of the University of Utah and Dr. Balu Athreya of the Thomas Jefferson Medical College in Pennsylvania have completed updates of the AI/RHEUM knowledge base. This year, the system's More-about-Conclusion component has been completed. This involved writing text describing all 54 diseases in the knowledge base for presentation to the user after the user has entered patient data and the system has arrived at a conclusion. Additional videodisc images were added to the Show-Me-More component of the program. Modifications necessary to deliver the AI/RHEUM system over the Internet were completed, including the digitization of all the original videodisc images and motion sequences. Collaborator Dr. Athreya completed his test of the pediatric version of AI/RHEUM with 100 cases from clinics at the University of Pennsylvania and Thomas Jefferson University. The manuscript describing this evaluation is in the final stages of preparation.

In a related effort, the report summarizing the workshop "Evaluation of Knowledge-Based Systems"

held at the Library in late 1995 is complete and ready for distribution from the NLM Web site and in printed form. A longer manuscript discussing the issues involved in the evaluation of knowledge based systems, also resulting from the Workshop, is being prepared for publication in a peer reviewed journal.

Machine Learning Project

The objective of the Machine Learning project is to design knowledge-based computer programs that can formulate and execute plans to learn desired knowledge. The ultimate goal of the project is to build systems that can accept abstractly stated questions and use programmed knowledge to select, assemble and analyze data to address those questions. A subsidiary goal is to increase the performance, ease of use and reliability of automated inference methods. Project research involves the development of novel machine learning algorithms and the creation of systems capable of managing generalized complex automated inference from data in pursuit of goals for biomedical knowledge. At many steps in the process, the significance and reliability of such automated inference methods are evaluated. Since the work is computationally intensive, the project must be prepared to take full advantage of as much computing power and network capacity as can be made available.

The Machine Learning project has significantly advanced its coevolution (COEV) learning method in the last year. Major landmarks include the filing of a patent application and first publication of the method in the scientific literature. COEV has now been applied to several significant biomedical datasets with results that are in several cases superior to those of all previous applications. The project group is currently developing a Web-based user interface for the COEV system, anticipating a version of the program that can be installed and used at multiple sites in early 1998. This will facilitate the testing and validation of the coevolution learning method. The group is also testing new learning methods and new methods for combining the results of multiple trained learners.

In 1997 Machine Learning project staff and collaborators at George Mason University initiated work in new methods for biologically realistic simulations of mammalian neural circuits. The collaboration has resulted in a new methodology called the qualitative reasoning neuron. Recent implementations suggest that this method may be significantly more computationally efficient than any previous simulator. Current Machine Learning project efforts involve building compartmental level models of cerebellar circuits involved in saccade generation and adaptation.

Machine Learning project staff have been doing exploratory work in applications of machine learning technologies relevant to library operations. This work has involved attempting to characterize large bibliographic databases with unsupervised learning methods and attempting to discover mapping between databases. Work has recently begun on extracting biochemical relationships among proteins from MEDLINE abstracts. This work involves multiple inference methods (statistical and linguistic) and multiple databases (MEDLINE and GenBank).

MedIndEx Project

The objective of the MedIndEx Project has been to develop and test interactive knowledge-based systems for computer-assisted indexing of medical literature currently indexed in the MEDLINE database using terms from the MeSH thesaurus. The system is designed to facilitate expert indexing by, in effect, combining MeSH and indexing rules into a knowledge base to assist indexers in a sophisticated workstation environment.

Difficult scheduling problems and changes in circumstances have resulted in the cancellation of an evaluation to compare MedIndEx with NLM's operational indexing system, AIMS (Automated Indexing Management System). The MedIndEx project will be concluded in FY 1998. At the conclusion of the project, NLM plans to make MedIndEx available to the public. This includes the source code written at NLM for the window based subsystems, the public domain components, and Perl scripts for scoring indexing consistency and quality. Those who wish to use the system will need Lucid Common Lisp and a Sun SPARCstation running the Solaris operating environment. NLM staff will be available to assist in the initial installation under the environment described and to help people understand what MedIndEx is supposed to do and how it works. NLM will not support continued development of the system.

Internet Grateful Med

Internet Grateful Med (IGM) is a development of the User Access Services Project, a component of NLM's System Reinvention initiative. IGM is an intelligent gateway system designed to provide assisted searching to NLM's users with a consistent look and feel across multiple database systems as NLM's primary retrieval systems evolve. The gateway architecture has proven a successful means of transparently connecting users to several different types of retrieval systems while insulating them from the specifics of differing command languages. For its approach to the problem of

providing attractive Internet-accessible interfaces to legacy systems, Internet Grateful Med won one of several Federal Showcase Site awards at the Federal Webmaster Workshop in August 1996.

Tens of thousands of physicians have been introduced to Internet Grateful Med through outreach programs of the National Network of Libraries of Medicine (NN/LM) and professional specialty conferences such as the annual conferences of the Radiological Society of North American, the American College of Physicians, and the American Medical Informatics Association. The initial version of Internet Grateful Med searched only in MEDLINE. Searching in AIDSLINE, HealthSTAR, and the new PREMEDLINE database was added to Internet Grateful Med early in September 1996. A special version was made available in NLM's Reading Room in March 1997.

The most important development for Internet Grateful Med in FY 1997 was NLM's announcement of free Web-based searching in MEDLINE and other databases. This announcement, made in June 1997, quickly resulted in the tripling and then quadrupling of searches made by users through Internet Grateful Med. The ELHILL mainframe retrieval system for MEDLARS broke records on an almost daily basis through the late summer and early fall as the IGM user load increased.

To handle this load, the Internet Grateful Med team developed and put in place a server farm system architecture for IGM. This system runs on eight small Sun Ultra servers connected through a high-speed ATM (Asynchronous Transfer Mode) network. Individual servers can be added as user load increases, or taken offline for servicing. In addition to the server farm architecture, other major IGM enhancements included searching in seven more databases (AIDSDRUGS, AIDSTRIALS, DIRLINE, HISTLINE, HSRProj, OLDMEDLINE and SDILINE). This made a total of 11 of NLM's databases available to the IGM user. A personal journal subset function patterned after that in Grateful Med for Windows was added, as was the capability to search by journal with assistance in finding accurate journal names from the list of more than 3,800 journals represented in MEDLINE.

An incoming connection to IGM from the Health Services Technology Assessment Texts (HSTAT) system was created in August 1997. This linkage enables the HSTAT user who has found a journal citation of interest to go directly to Internet Grateful Med to view the complete MEDLINE record of that article. Finally, the ability to invoke NIH Clinical Alerts from the front screen of Internet Grateful Med was added in September 1997.

By the end of FY 1997, usage of the Internet Grateful Med system was running at more than 1.6 million World Wide Web connection requests per week. Users from dozens of countries made more than 635,000 searches in MEDLINE and other NLM databases during September 1997.

Object Oriented Information Retrieval Project

The Object Oriented Information Retrieval project is moving toward an object-oriented distributed database system while continuing the development and support of NLM's publicly available HSTAT service. The interface to HSTAT and the Web-centered capabilities of the system have been improved and expanded throughout the year, laying the ground work for the next stage of the system's evolution.

The HSTAT collection continues to grow in size. During FY 1997 documents were added in several categories: Morbidity and Mortality Weekly Reports (MMWR); AHCPR Guidelines, Technology Assessments and Technology Reviews; Treatment Improvement Protocols; Consensus Development Conference Reports; and Continuing Medical Education exams for the Consensus Program.

The Agency for Health Care Policy and Research is no longer sponsoring the creation of Guidelines but is developing a series of Evidence Reports to be used as the basis for guideline creation. Release of the first complete Evidence Report is expected in October 1997. The project team has prepared a Document Type Definition to define the structure of Evidence Reports. The Evidence Reports will be included in HSTAT as they become available. As an introduction to this evolving venture, the Executive Summary of the first report on Colorectal Cancer Screening is now available on HSTAT.

Milestones in the development of HSTAT this year have included the replacement of the character-based HSTAT client with Lynx, the character-based Web browser from the University of Kansas. Users who formerly used the HSTAT character-based client are provided with a copy of Lynx which runs on the HSTAT server. This has reduced data maintenance efforts considerably. HSTAT access is still being provided to a steady 200-400 users per month through the Lynx interface. The Standard Generalized Markup Language (SGML) and associated Document Type Definitions (DTDs) for all HSTAT data have been revised and improved. HSTAT users can now initialize an Internet Grateful Med search from a MEDLINE-indexed citation found in any of the AHCPR guidelines.

As a first step in moving toward a distributed database system, the project group has worked with the Centers for Disease Control and Prevention to allow

HSTAT users to include the CDC Prevention Guideline Database in an HSTAT search. The CDC database is indexed by the Verity system at CDC and returns a ranked list of document titles in response to a query. This information is provided to the HSTAT user as if it were another HSTAT collection. The user can go directly to the CDC database to view documents or stay within HSTAT.

Late in FY 1997, the project team released for beta testing a new Java-based interface to HSTAT. This interface provides a number of features not normally supported by the standard Web interface such as maintaining a search history, creating a new search based on previous searches, menu options, improved navigation, and simpler download capabilities. The Java HSTAT interface will be further developed in FY 1998.

Image Processing Research

Visible Human Project

Several Center projects involve research on image processing methods, in many cases working with the Visible Human data set. The initial aim of the Visible Human Project was the acquisition of transverse CT, MRI and cryosection images of a representative male and female cadaver at an average of one millimeter intervals. The larger, long-term goal of the Visible Human Project is to produce a system of knowledge structures that will transparently link visual knowledge forms to symbolic knowledge formats. Basic research is still needed in the description and representation of structures and the connection of structural-anatomical to functional-physiological knowledge.

The Visible Human Project data sets are designed to serve as a common reference point for the study of human anatomy, as a set of common public domain data for testing medical imaging algorithms, and as a test bed and model for the construction of image libraries that can be accessed through networks. The Visible Human data sets are being made available through a free license agreement with the NLM. They are being distributed to licensees over the Internet at no cost, and on DAT tape. The data sets are being applied to a wide range of educational, diagnostic, treatment planning, virtual reality, artistic, mathematical and industrial uses by over 850 licensees in 27 countries. The Visible Human Project has been featured in well over 800 newspaper articles, news and science magazines, and radio and TV programs worldwide.

In the fall of 1996, NLM convened a public conference of Visible Human data sets license holders. The conference served as a forum for researchers and developers to exchange information on their applications of the data sets and to share with NLM their successes and problems in using the data sets. Fifty papers were presented to over 250 attendees.

The data sets are having their greatest effect on health care and health education and thus benefit the general public. The data sets are used as a normal reference and as an aid in the diagnostic process. Programs under development will be used to educate patients about the need for and purpose of surgery and other medical procedures as well as to permit physicians to plan surgery and radiation therapy. The images from the Visible Human data sets are used in several prototype virtual reality surgical simulators. Educational materials that make use of the Visible Human data sets are beginning to be used by students from the primary grades to practicing health care professionals. The data sets are being used to form the basis of interactive games to entertain as well as to educate. Some automobile manufacturers now include passenger injury models based on Visible Human data in their vehicle crash simulation models. Engineers and physicists are creating models to quantify human exposures to various forms of electromagnetic radiation. The data provided by the Visible Human data sets are being used by mathematicians as an application for what were previously only theoretical mapping theories. Several artists are using the data set as the basis for new multi-media art forms.

3D Systems Project

Current NLM efforts within the Visible Human Project are directed at turning the data sets into an interactive digital image library. As a pilot project, each object in each cross-section of the thorax is being identified and segmented. A lexical hierarchy of anatomical objects within the thorax is also being developed. This hierarchy includes the relationship of each object to other nearby and related objects. An object-oriented database is being constructed which will link the segmented image objects to the hierarchical lexical objects. This will allow the user to retrieve a labeled image of any object within the thorax. The project's immediate task includes the development of an Internet Web-based image retrieval and delivery system. Initially the anatomical database will contain a collection of images segmented from the Visible Human 70mm male thorax.

Four image file types are defined for the prototype system being developed by the group. First, the database will store rendered image files. These image types are representations of 3D projections, and can be generated in a wide range of image types, including volume and surface rendered, animation, audio-visual, virtual reality, etc. The second image type

is the volume of interest, which is a collection of segmented images within a defined volume of the human body. These images can be imported into visualization systems to produce rendered images. The third image types stored are single bitmap types, such as anatomical cross sections. The fourth and most basic image types are the segmented binary bitmasks, which represent the form of the individual segment images.

Anatomical image data is stored in structures using UMLS identifiers to create linked relationships. Database queries are formatted within a Web client, and transmitted to the Web server that communicates with the database to retrieve records from persistent storage.

Image Storage and Transmission Optimization (ISTO)

ISTO's objective is to investigate compression and transmission techniques to improve access to, and delivery of, data-intensive biomedical images, with specific focus on the Visible Human color image set. The CCD captured male and female data amounts to 55GB, and when the 70mm photographs are scanned at much higher resolution, that will total about 235GB. These datasets will strain both storage and transmission resources, and ISTO was proposed as an investigation of both compression and advanced communication techniques to alleviate these problems. In Fiscal Year 1997, work was initiated toward the development of prototype lossy and lossless compression techniques. The eventual goal is to design a system combining both techniques so that storage is achieved losslessly, and data to be delivered to a user is compressed lossily, at a quality level (and therefore compression ratio) required by the user. This user-defined quality level will define the wavelet/quantization parameters suitable for lossily compressing the data prior to transmission. Specific current goals are to: (a) analyze alternative lossy and lossless compression methods for the Visible Human image data, design and develop two alternative modules, and conduct a comparative evaluation; and (b) evaluate alternative transmission techniques for the Visible Human image data using NLM multisocket technique and RFC 1323 over the current Internet and ATM.

Preliminary investigations of compression and communications techniques were completed. A sample of the Visible Human image set was compressed both lossily and losslessly using public domain software for wavelet transform, DCT (lossy JPEG) and lossless JPEG. A recent publication documenting this early study: Thoma GR, Long LR. Compressing and transmitting Visible Human images. *IEEE Multimedia*; New York: IEEE Computer Society. Vol. 4, No. 2; April-June 1997; 36-45.

In addition to compression resulting from

eliminating the background pixels in the images (over 3.5), the project team hypothesized that for lossless compression pixels in adjacent slices would be highly redundant and therefore promising for compression. Studies of interframe redundancy using the RGB color images but also with the images transformed to the YIQ and HSV planes were conducted. The studies showed that interframe redundancy is low irrespective of color plane. The results of the RGB study have been documented in: Meadows S, Thoma GR, Long LR, Mitra S. Entropy encoding of difference images from adjacent Visible Human digital color photographic slices for lossless compression. In Medical Imaging 1997: Image Display, Yongmin Kim, editor, *Proc. SPIE Vol. 3031*, 1997; 749-55. Further work in comparing the interframe redundancy using the YIQ and HSV color planes has been documented in a paper recently accepted by SPIE for publication.

For lossy compression, studies have been initiated toward the selection of wavelet transforms suitable for VH images. The statistical properties of the coefficients in wavelet transformed subimages of selected VH slices were investigated using filter taps from 4 to 20 to establish the optimality of specific wavelets for this class of images. The tentative conclusion is that both Daubechies and Symlet wavelets with 12 filter taps introduced the least variance from the generalized Gaussian distribution function for optimal coding at all levels of decomposition. The project will continue the development of a lossy compression technique using the wavelet transform, quantization and Huffman coding; conduct data analysis to quantify compression ratio and speed performance to select specific characteristics of signal decomposition and quantization alternatives.

ISTO will also continue to evaluate file transfer performance of NLM multisocket technique vs. RFC 1323 over the Internet and ATM networks, and evaluate client/server database application performance of NLM multisocket technique vs. RFC 1323 over the Internet and ATM networks.

Audiovisual Research

The Audiovisual Research team is developing methods for using the data set from the scanned visible human film images to produce 3D renderings of anatomical structures. These rendered objects will utilize a new Onyx2 workstation upgraded with sufficient memory to allow the massive files from the Visible Human project to be rendered at full resolution.

The videotape, "The Visible Humans—A Step Toward Tomorrow" completed in March 1996, was awarded a 1997 Cine Golden Eagle Award by the Council on International Nontheatrical Events. Initial

videotaping for the new Visible Humans videotape began at the Visible Human Project conference held at the NIH in October 1997. During the course of the conference, 15 interviews were recorded with Project licensees from around the world. These new interviews will be included in the applications segment of the expanded Visible Humans program, and it was decided to develop the updated version as both a linear videotape and as an interactive format appropriate for a variety of platforms.

The National Geographic Society Imaging Lab made several test duplicates of the 70mm film of the Visible Human Male. The quality of these duplicates was verified by scanning them using the KODAK 5057 scanner. After verification of the quality of the process, a complete copy of the 70mm Visible Human Male was made and the dupe was delivered to NLM in September 1997. This copy will be reviewed by sampling images, scanning, and viewing for quality.

In collaboration with Dr. Gary Hack, Professor at the Baltimore College of Surgery, University of Maryland, a method to visualize the Sphenomandibularis muscle, and the Dura-posterior capitis minor muscle connection using the Visible Human data set was developed. An article detailing the collaboration with Dr. Gary Hack, and the use of the Visible Human dataset to verify the anatomic discoveries, has been completed. The article includes 21 color plates showing various dissections, MR scans and renderings from the Visible Human Project dataset. Slides, posters, and a Web site describing the research will now be developed. Future collaboration will include a 3D simulation of these structures to show their possible functions.

A new Visible Human Web page was demonstrated at this year's Radiological Society of North America on December 1-6, 1996, in Chicago. It consists of sample images from the Visible Human Male and Female as well as applications that were developed based on the data set.

A demonstration of the Visible Human Project was given at the Inaugural Technology Pavilion. It was organized by the Inaugural committee to promote cutting edge technologies in various areas such a medicine, aviation and Internet applications and was open to the public on the mall in Washington, D.C.

Automating Data Entry into MEDLINE

The data entry of citations and abstracts of medical journal articles for the MEDLINE database has been traditionally done by manual keyboarding. In response to a crisis in early 1996 in this data entry, the project team developed a system, temporarily code named "MARS" for Medical Article Record System.

MARS combines the keyboarding of citation data which may appear in different areas of a journal issue (journal name, date, author, title, affiliation, page numbers, etc.) with scanning and automatic text conversion by OCR of abstracts which, if keyboarded, proves very labor-intensive.

MARS consists of multiple workstations of three types: scanner, reconciling, and keyboarded citation entry. In addition, the system requires three servers: a network file server, an OCR server, and one to match double keyboarded citations. In parallel with software development, the project team specified and acquired the hardware, designed a floor plan to ensure smooth work flow, installed additional power and laid the Ethernet cabling for LAN connection. All workstations, networked via a LAN, are installed and operating at the work site. The current volume of production is over 300 completed records a day compared to 65 records/day when the system was installed in October 1996.

Briefly, the MARS system works as follows. The scanning operator scans the pages on which the abstracts appear, and zones the titles and abstracts, and the bitmapped TIFF files are sent to the network server. The OCR server performs text conversion on these TIFF files, and produces text files of the abstract and title. A citation entry operator keys in all the citation fields in a template for the journal issue and each article, and a second operator repeats this process for the same journal issue and articles. A CITATION MATCH module in the matching server compares the two citation entries, and produces a "citation difference" file highlighting inconsistencies. A MATCH module in the server then matches this difference file and the OCR'ed abstract, correlating the two article title texts (one from the scanned page and the other from the keyboarded citation). At this point, both the citation and the abstract are available for reconciling (validation and proofing). Following this step, the completed record files are FTP'ed to a module in the NLM mainframe computer from which the files are later accessed by indexers who add the appropriate descriptive information such as Medical Subject Headings.

Recent work has focused on increasing the level of production. To this end, incremental improvements have been made, for example, by: (a) modifying the scanning workstations to allow the operators to enter journal issue identification numbers by barcode readers instead of keyboard entry; (b) providing on-screen tables containing Greek and special symbols to allow the reconciling operators to click from a list in the table without having to type in the words; (c) providing a highlight-and-click capability to capitalize or de-capitalize words or entire text lines; (d) changing icons, error messages, and buttons for clarity;

(e) developing an interface with lexicons and ad hoc rule sets to unhighlight correct words (from the OCR) while retaining highlights for the incorrectly recognized words, and thereby reducing the burden on the reconcile operators; and (f) including the capability of entering diacritical marks directly by keyboard operators.

The system was recently described in: Thoma GR, Le DX. Medical database input using integrated OCR and document analysis and labeling technology. *Proc. 1997 Symposium on Document Image Understanding Technology.* College Park MD: University of Maryland, Institute for Advanced Computer Studies; 1997; 280-1.

Document Image Analysis

Dr. Daniel Le and NLM received a U.S. patent in January 1997 for Dr. Le's development of an algorithm to detect the orientation (portrait vs. landscape) of a binary page image. Orientation detection is an essential preprocessing stage for optical character recognition, skew detection or skew correction. In addition, page orientation is crucial for automated document data entry in which the contents of a printed document is first segmented into such regions as headlines, author or title fields, text columns, graphics or footnotes.

The algorithm is based on an analysis of projection profiles, vertical and horizontal variances on a page, and a technique to reduce the impact of nontextual data (blanks, graphics, forms, line art, large fonts and dithered images). Using a sample of several thousand images of medical journal pages, the algorithm was found capable of detecting page orientation at an accuracy rate of 99.92%. The algorithm is described in Automated Portrait/Landscape Mode Detection on a Binary Image, U.S. Patent Number 5,592,572; January 7, 1997.

Other document image analysis work includes the development of an efficient algorithm for automated skew detection. Skew in a scanned image often requires rescanning the document, a time consuming and costly step, but often necessary in document conversion. The research team has designed a multistage technique whose principal elements 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; (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.

DocView

The goal of the DocView program is to design, develop and investigate the role of advanced prototype client and server systems in enabling access to, and delivery of, biomedical documents over the Internet. The three key elements of this project are that it addresses the NLM mission of providing document delivery to end users and libraries, uses the Internet, and incorporates digital imaging techniques. The documents may be in electronic form in servers or scanned and sent over the Internet by Ariel systems, widely used in the library world. While libraries and document suppliers use Ariel routinely to send documents via Internet to similar organizations, there are few options for end users to directly receive them. DocView fills that niche.

The DocView client software, which runs under any version of Microsoft Windows, enables an end user to receive documents over the Internet at the desktop, retain them in electronic form, view the images, organize the received documents into "folders" and "file cabinets", electronically bookmark selected pages, manipulate the images (zoom, pan, scroll), copy and paste images, and print them if desired. DocView also serves as a TIFF viewer for compressed images received through the Internet by other means, such as World Wide Web client software. DocView is undergoing extensive beta testing in early FY 1998.

A major new function recently incorporated is to enable DocView users to *send* documents. This may be done by both MIME e-mail as well as the Ariel protocol. One advantage of MIME e-mail over Ariel protocol is that document format is not restricted to Group 4 compressed black and white images. With MIME email, any document format may be used. A second advantage of MIME e-mail is that e-mail is more universal than Ariel.

Four papers describing DocView were presented at the 1994, 1995, 1996 and 1997 Integrated Online Library Systems (IOLS) meetings, and published in their proceedings. The latest publication is Walker FL, Thoma GR. Internet document delivery: An end user survey. *Proc. IOLS'97.* Medford NJ: Learned Information, 1997; 145-53. In recognition of his contributions, Mr. Frank Walker, the principal designer of DocView, was awarded the NLM Board of Regents' Award for Scholarship or Technical Development in September 1997.

Workstation for Interlibrary Loan (WILL)

In 1997 the design and development of a prototype system called WILL was completed. The system delivers bitmapped document images on demand. This was in response to a key finding of an earlier project, System for Interlibrary Loan (SAIL), in which journals were preselected, scanned and stored on optical disks, and automatically delivered by fax or printed out and mailed. The SAIL project found that preselection of a wide range of documents was less practical than scanning and sending documents *on demand*, hence the need for an integrated multi functional system to enable operators to scan and transmit documents on request. The pilot SAIL system, after 4 years of successful operation, was discontinued in September 1995.

WILL consists of a Pentium-class computer with internal fax, Ethernet and image processing boards. The software is based on the Windows operating system. The only activity WILL requires of an operator is to scan requested documents. It eliminates other operator decisions on recipient and delivery mode. WILL automatically retrieves ILL requests from DOCLINE (the request routing system used by medical libraries nationwide), parses the request into data fields, provides a GUI for the operator to scan requested documents, and automatically delivers document images by all three conventional delivery modes (print for mail, fax, and Internet), and updates DOCLINE with status. In addition, a database in WILL retains operational information for administrative purposes. The only process that is not automated is the physical handling and scanning of the documents.

To investigate the real-world performance of WILL in document delivery, a beta test was conducted at the NIH Library on the NIH campus, as described in last year's report. The beta test was concluded in 1997 and yielded data on a cost estimate for document input, cost of document delivery, input and delivery throughput, improvements possible in delivery speed, image quality, system reliability, GUI layout and design, database functions, and operator procedures. Following the beta test, a Cooperative Research and Development Agreement (CRADA) was developed between NLM and a private company. Under the CRADA, the private company will enhance the existing system and take the steps required to make WILL a viable, practical system for widespread use in biomedical document delivery. In parallel, the project group will provide the company with inhouse-developed source code and documentation and complement WILL's transmission capability by developing end user software for receiving documents sent by WILL.

Digital X-ray Imaging Projects

These projects aim to address fundamental questions that arise in the handling, organization, storage, access and transmission of very large electronic files in general and digitized x-rays in particular. The approach adopted is to address these questions by the design, development and evaluation of prototype systems which serve as testbeds to investigate: (a) image compression techniques, especially high yield lossy methods, and tools to interactively select compression parameters; (b) techniques to organize images and associated textual data for ready retrieval and use; (c) procedures and algorithms to implement transparent hierarchical storage using heterogeneous storage systems and media to match usage patterns; and (d) multisocket transmission methods to segment large images and to send the pieces concurrently over multiple socket pairs to overcome the inefficiencies of conventional transmission protocols.

As described in earlier *NLM Programs and Services*, a three-agency joint project named DXPNET for *Digital X-ray Prototype Network* was begun to serve as a vehicle to address these engineering goals while creating an archive of digitized radiographic images accessible over the Internet. In this project, a collaboration with the National Center for Health Statistics (NCHS) and the National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS), the Communications Engineering Branch on behalf of NLM takes on the role of technical manager and developer. The x-rays, consisting of about 17,000 cervical and lumbar spine films, were collected during the second National Health and Nutrition Examination Survey (NHANES), one of a series of nationwide surveys conducted by NCHS designed to provide a snapshot of the nation's health. As films they are relatively inaccessible, a major motivation for digitizing them. All 17,000 films were digitized with a Lumisys laser scanner resulting in images of size 5 MB (cervical) and 10 MB (lumbar). These were quality checked and archived in a 144-platter optical jukebox.

A goal of the project is to determine the essential design characteristics required in systems that provide remote access to such a medical image collection, and to design and develop systems that satisfy these requirements. The systems developed include: (a) a Standardized Readings Workstation (SRW), the tool for radiologists to retrieve the images over the Internet and to enter their readings, (b) an Electronic X-ray Archive (EXA) consisting of an optical disk jukebox controlled by a Sun 670MP server and interfaced to the Internet, and (c) Quality Control Workstations (QCW) to enable NCHS technicians and radiologists to perform quality checks at different levels

of the scanned x-rays.

An initial prototype Standardized Reading Workstation was developed on a Sun 4/260 platform, with image display on both the standard Sun monitor (1K x 1K) and a high resolution Megascan monitor (2K x 2.5K). An evaluation by NIAMS staff radiologists and rheumatologists yielded design information used to refine the software. Based on these data a new SRW was developed that would be suitable for remote deployment over the Internet. This system has a Sparcstation 20 as the platform, S-bus connected high resolution monitor (2K x 2.5K), and a finalized design of the user interface under Motif, X-Windows, TAE Plus GUI development tool, client software incorporating custom-designed Berkeley sockets, and image enhancement functions. This implementation includes the functionality required to create stand groups for which modules were designed for the duplication of images for QC, the inclusion of gold standard images and the randomization of images for the readers. For the reading groups, three software modules were implemented: one each for the technician, medical consultant and medical adjudicator, the three types of reviewers envisioned doing the standardized readings. Region-of-interest (ROI) functions for zooming and histogram equalization have also been completed, but these routines are general enough to accommodate other image processing functions. As an element of the SRW, the project group designed and implemented a radiographic atlas subsystem using images selected by NIAMS collaborators, and implemented it as a reference tool for the SRW user. NIAMS staff selected images and provided them to an advisory radiology subcommittee for review and final selection.

As an extension of the DXPNET work, NCHS collaborators have proposed a future project to archive 10,000 hand and knee x-ray images from the third NHANES survey. A key issue here is the scan density (spatial resolution) required to read this class of x-rays when digitized. A study to determine the required scan density was conducted by two radiologists each using a Sparc 20-based workstation developed in-house. The workstations run on Java software, the dialog framework built entirely in Java's Abstract Windowing Toolkit, and a C++ image display module. Several hand and wrist images from a population of Pima Indians was used to determine the best scan density to detect bone erosions as a way to establish resolution requirements for the large NHANES collection. The Pima x-rays were digitized at three resolution levels: 50, 100 and 150 microns. The study concluded that the 100 micron level would provide the resolution needed, and this has provided the information required for NCHS to proceed with the digitization of hand x-rays from NHANES.

In contrast to the SRW and related systems developed specifically for the DXPNET project, another project was motivated by the potentially widespread interest in the heretofore inaccessible x-rays from NHANES. This project, described in last year's report, focused on the development of a prototype system named MIRS (*Medical Information Retrieval System*).

Recent publications describing work in this area are: (1) Long LR, Goh G-H, Neve L, Thoma GR. Architecture for biomedical multimedia information delivery on the World Wide Web. *Proc. SPIE Multimedia Storage and Archiving Systems II,* SPIE Vol. 3229, Dallas TX, November 1997. (2) Long LR, Ostchega Y, Goh G-H, Thoma GR. Distributed data collection for a set of radiological x-ray interpretations. *Proc. SPIE Storage and Retrieval for Image and Video Databases V*, SPIE Vol. 3022, San Jose CA; February 1997; 228-37.

Advanced Medical Imaging Tools

In line with the increasing trend for medical information databanks to incorporate images, tools must be available to enable users to easily search and retrieve such data over the Internet, and to evaluate the returned image data against "gold standard" or reference images. The objective of this project is to address fundamental questions in the design of such tools that are required in the handling, organization, storage, access to, and transmission of very large electronic files in general and digitized x-rays in particular. This project arose out of design activities in the DXPNET project described above.

The current focus of the project is the building of two tools, one as a medical image reference aid, and the other a multimedia database access tool. The first is a platform-independent *digital radiological atlas* of the cervical and lumbar spine building on prior work involving digitized spine x-rays from the NHANES II survey. The approach is to use Java software technology to create the atlas whose images, initially obtained from the NHANES image dataset, will be displayable on conventional monitors as well as high resolution Megascan monitors. This tool would find application in PACS systems and in medical teaching and research environments. The atlas work has been described in: Long LR et al. A digital atlas for spinal x-rays. *Proc. Medical Imaging 1997: PACS Design and Evaluation, Engineering and Critical Issues.* SPIE Vol. 3035, Newport Beach CA, February 1997.

The second tool is an advanced Web-enabled medical image database tool for searching and retrieving contents of biomedical databases containing both text and images, building on the Internet-

accessible, client/server MIRS system developed in the DXPNET project. Techniques incorporated in this development include CGI scripts, C++ classes, the Informix DBMS and Virage datablades. Java applets are used for client enhancements. This tool called *WebMIRS,* is basically a Java-based, WWW prototype biomedical informatics tool for database access. Using this platform-independent tool, access to the text and images in the NHANES databases will be demonstrated and evaluated. These databases, based on the Informix DBMS, will contain: (a) NHANES information relating to the national prevalence of osteoarthritis including both textual information and cervical and lumbar spine x-ray images; (b) NHANES information related to osteoarthritis, osteoporosis, functional health of the elderly, and other NHANES focus areas of interest as defined by NIH researchers. An early prototype of WebMIRS was demonstrated over an ATM OC-3 link in July 1997 at the CDC Joint Meeting of the Public Health Conference on Records and Statistics and the Data Users Conference in Washington, DC.

Education and Training

Medical Informatics Training

The Medical Informatics Training project supports medical informatics training opportunities at the Center. Trainees are on site at the NLM and participate directly in Center research activities. The training program supports undergraduate, graduate, and post-doc students in a variety of areas of medical informatics. Participants interact with outstanding professionals, have access to on-site resources and facilities, and make contributions to ongoing research. They gain an understanding of the Center's programs and have the opportunity to make professional contributions to their fields. Potential areas of research for participants include digital library research, automated indexing techniques, vocabulary and thesaurus research, natural language processing, artificial intelligence research, image processing and computer-based learning. During FY 1997 the Center made one award to an undergraduate student and another to an MD/PhD post-doctoral student. A student who had just finished medical school and went on to do an internship after training for seven months at the Center was also supported. In addition, one teacher and one student from Coolidge High School, NLM's "adopted" high school were mentored.

The Medical Informatics Elective project began in 1989 as part of the NIH Computers in Clinical Medicine program for third-year and fourth-year medical students. The elective now represents a unique overview of the field involving an extensive lecture

series and individual research projects with preceptors at NLM and other NIH institutes and divisions. Five medical students from schools across the U.S. completed the elective in March-April 1997. Each student participated in a seminar series and completed an independent research project including formal presentations and a manuscript under the preceptorship of a member of the NLM professional staff.

Learning Technologies

Several projects represent work in educational and learning technologies. The Learning Center for Interactive Technology provides visitors with an opportunity to view leading edge technologies and to consult with Learning Center staff on the use of these technologies in the health sciences. Center staff regularly demonstrate NLM's products and research results, including Internet Grateful Med, PubMed, and the Visible Human project. The Distributed Learning Environments project investigates the potential of telecommunications and computer-mediated technologies to provide high-quality learning experiences at a distance. Project staff are working with other collaborating staff from CDC and NLM on issues of information access for public health professionals. The Digital Microscopy System investigates the parameters that govern the online indexing of medical photomicrographs for educational purposes.

The Learning Center for Interactive Technology

The Learning Center for Interactive Technology offers a wide range of computer-based information and educational technologies representing stand-alone, Intranet, Internet and World Wide Web applications in the health sciences. In 1997 850 individuals representing 52 distinct groups visited the Learning Center, bringing the grand total of visitors to just under 11,000. In addition to the traditional constituency of health professional teachers, students, administrators, and researchers, The Learning Center experienced a noticeable increase in the number of high school and college student visitors. This increase was due in large part to the great interest generated at the high school and college level by the Visible Human Project and the availability of a variety of Visible Human applications in the Learning Center. Another contributing factor was inclusion of the Learning Center in many of the general tours provided for visitors to the NIH.

In addition to the enormous interest in multimedia applications of the Visible Human Project, visitors were also interested in Internet and web-based instructional resources, authoring tools for network

accessible instructional resources, creating distributed learning environments in support of health professions education, and the newly introduced PubMed. Staff provided demonstrations, facilitated "hands-on" practice with new and emerging online technologies, and used the Cognitive Science Branch Web site to point visitors to multimedia educational resources in medicine throughout the world. When new training opportunities become available they are announced on the Cognitive Science Branch homepage (wwwcgsb.nlm.nih.gov).

Distributed Learning Environments

The general research and demonstration focus of the distributed learning environments project is to investigate innovative means for assisting health science institutions in their use of online distance learning technologies. An important goal of the project is to develop collaborative strategies that institutions can use to successfully integrate a wide range of technologies. A consortium of four health science institutions has been established and is working with Learning Center staff to develop the program; so far literature reviews of content important to the project have been completed. Staff have also generated a list of crucial research questions to be addressed. The literature reviews and research questions are crucial phases in the creation of a "Research-Based Handbook for Distributed Learning" that will be helpful to institutions interested in resource sharing and collaboration.

Digital Microscopy System

In 1994 project staff set about attempting to determine the parameters that govern the indexing of a library of medical photomicrographs so that eventually a library of such materials can be available to medical practitioners, researchers and students via the World Wide Web. In particular, the project intended to determine whether, as had been suggested by the Board of Scientific Counselors, such images could be described using the concepts found in the UMLS Metathesaurus.

A well-defined sub-set of the universe of such images was chosen for the tasks. DMS had in hand a good quality set of prostate photomicrographs that were recorded in analog format by another Center project (Computer-assisted Curriculum Delivery Systems) and had access to an expert consultant in Genito-Urinary (GU) cytology, also as a result of work on the previous project. During the 1994-95 period an expert panel of four GU pathologists indexed a series of 51 prostate slides. They worked independently and supplied to NLM keywords (index terms) to describe morphology,

a diagnosis, and a written slide description. The results indicated that although they always agreed on the diagnosis, they frequently used different words to describe the slides morphologically. Further, only about 60% of the terms used could be mapped to the 1995 version of the UMLS Metathesaurus. Use of the analog images (on a videodisc) proved to be cumbersome and so in mid-1995 one of the consultants came to NLM to record a few prostate images in digital format.

In 1996 the project team installed a Web server that provides indexers with the capability of having on a single screen a photomicrograph and the input screen for the Metathesaurus. Also in 1996 a collaborative effort was initiated with the Department of Pathology at the University of Pittsburgh and WWW pages were developed that permit indexers to see images stored on the project Web server and enter keywords and diagnoses. Six genito-urinary pathologists working at four sites re-indexed a set of prostate images that were recorded in 1995 at NLM. They entered up to 20 index terms for each of 11 test photomicrographs and those terms were submitted directly (via the API supplied by the UMLS group at NLM) to the UMLS Metathesaurus and the results were posted to a database. Only 30% of the terms submitted online were found in the 1996 Metathesaurus. Terms that were not found were later submitted as batch files to the UMLS Large Scale Vocabulary Test. Forty-four percent were found, meaning that the 1997 version of the Metathesaurus could recognize about 74% of terms that pathologists use to describe photomicrographs of prostatic carcinoma. These results were incorporated into a paper that was presented at the fall 1997 meeting of the American Medical Informatics Association.

Cervical Cancer Project—Screening for Life

"Screening for Life," an instructional program to promote the early detection of cervical cancer, was described in last year's report. The project team has now produced a complete Web version that includes all the text, graphics and video clips from the original program. The system is intended for general practitioners, internists, nurses and other health professionals who provide cervical cancer screening services.

Making Medical Science Visible

The "Making Medical Science Visible" videotape, which documents the graduate level training required for the production of successful visuals for medicine and science, was described in last year's report. The class of 1995 from the Department of Art As Applied to Medicine at the Johns Hopkins

University School of Medicine was closely followed during their graduate training. The final 28-minute tape has received several citations: the Videography Awards' "Award of Excellence" and the Telly Awards' "Silver Medal." Both represent the highest honors given by the respective organizations to projects that "clearly exceed industry standards."

Educational and Information Program Support Project

A specially shot videotape about the Human Genome Map on the Internet was prepared for the Human Genome Map News Conference held in the Lister Hill Auditorium, October 24, 1996. The tape was sent by satellite to U.S. television stations .

On January 30, 1997, Dr. Lindberg recorded two interviews with Dr. Michael DeBakey. Dr. DeBakey discussed his more than 50 year association with the NLM and its predecessor institution, the Army Medical Library. Also discussed was Dr. DeBakey's experience in Russia in the fall of 1996 when he was called in to consult on the diagnosis of President Boris Yeltsin. Included in the edited discussion were applicable historic photographic stills, motion picture film and videotapes.

The June 26 news conference announcing that access to MEDLINE on the Internet is now free, was videotaped and edited into a final program. The news conference, held at the Dirksen Senate Office Building, featured comments from Senator Tom Harkin, Senator Arlen Specter, NIH Director Harold Varmus, and Vice President Al Gore.

On August 5, 1997, at the Old Executive Office Building, camera crews videotaped the news conference at which Vice President Al Gore announced the availability of the Cancer Genome Anatomy Project on the Internet.

On September 30, 1997, the House Subcommittee held an NIH overview hearing. Appearing were NIH Director Harold Varmus, NLM Director Donald A.B. Lindberg, and a number of other NIH officials. A camera crew videotaped the proceedings and copies of the tape were provided to the NLM Office of Public Information, the House Commerce Committee, and several Committee members.

Resource Support and Development

System Security and Advanced Network Planning

Work during this year included the areas of system security, the NLM network, and the Next Generation Internet. System security was enhanced with the deployment of secure login and remote shell programs. The NLM network effort involved the implementation of backup routers and additional Asynchronous Transfer Mode (ATM) capabilities. The ATM network will position NLM well for participation in the Next Generation Internet project.

The system security work included development of a package for easy installation of the SSH (secure shell) set of programs on Sun and Silicon Graphics workstations. The SSH system encrypts all transactions to impede the interception of passwords and other confidential information by programs such as network sniffers. In addition the SSH system provides a means of secure access and file transfer for root accounts over the network. Additional work included preparation for deployment of a proactive password program and the implementation of PGP public key encryption. The proactive password program will force the selection of good (i.e., difficult to crack) passwords. The PGP system will be used to transfer encrypted information in email or in files or to authenticate such information.

Work has continued on the deployment of ATM within the Center. The scope of the ATM network was extended to include the eight machines of the Internet Grateful Med server farm. The ATM switches now use the ATM FORUM Private Network-Node Interface (PNNI), a protocol analogous to the Internet Protocol (IP) Open Shortest Path First (OSPF) routing protocol. The PNNI will permit the interconnection of a large number of switches in the same way the IP routing protocol permits the interconnection of a large number of workstations.

Two high-performance routers are key elements in the connection of the network to the Internet. When the primary router malfunctions, the backup secondary router takes over the routing function. Both routers now use the Hot Standby Protocol, which switches automatically to the backup when necessary in a matter of seconds.

Ultimately, NLM will be connected to a GigaPoP, a Gigabit-capacity Point of Presence, perhaps at the University of Maryland. The Next Generation Internet will use these GigaPoPs to connect networks such as NLM's and NIH's to the Next Generation Internet community. While the details of the connections have yet to be formalized, the connections are likely to be realized via ATM OC3 connections. The Center is procuring additional components to upgrade the capacity of existing ATM switches to serve a larger number of workstations. With this accomplished, LHNCBC and NLM will be in a good position to take advantage of the capabilities of the Next Generation Internet.

Engineering Laboratories

Housed in the document imaging laboratory are advanced systems to electro-optically capture the digital images of documents. Subsystems are available to perform image enhancement, segmentation, compression, OCR and storage on digital optical disk media. The laboratory also includes high-end Pentium-class workstations running under Windows 95 and NT, and Novell Netware-based networks, for performing document image processing. Equipment developed in-house includes document capture workstations, for both loose-leaf and bound volume documents. Other systems have integrated optical disk drives and high resolution softcopy display devices for documents. These are configured into systems that serve as laboratory testbeds to support research into automated document delivery, document archiving, and techniques for image enhancement, manipulation, portrait detection, skew detection, segmentation, compression for high density storage and high speed transmission, omnifont text recognition, and related areas.

The image processing laboratory supports the investigation of image processing techniques for both grayscale and color biomedical imagery at high resolution. It consists of equipment to capture, process, transmit and display such high-resolution digital images. Large-volume storage is provided by a jukebox containing 144 5.25-inch rewriteable magneto-optical platters, each formatted for Unix. Each platter has a storage capacity of 586 megabytes, for a total jukebox storage capacity of 81.5 gigabytes. In addition to the jukebox, mass storage is also provided by a Sun SPARC storage unit, a RAID system, consisting of 18 1.2 gigabyte hard drives, six independent fast buffered SCSI-2 buses, and connected to an SBUS card in a SPARC 20 via a 25 megabytes/s fiber channel connector.

Core Resource Support

Several projects provide core resource support for the Center, as well as, in some cases, for all of NLM. Support projects provide the increasingly critical planning, operational, and troubleshooting functions for the Center's shared communications resources and networks, including critical 24 hour support for access to public NLM systems. FDDI and ISDN networking are in active use and experiments with ATM are under way. The LHC server and associated Internet (BBN PLANET), Ethernet, Appletalk, and Novell communications facilities are used by nearly every research project at the Center as well as by other NLM staff and external collaborators. Equipment and services required for the videographics facility, multimedia development, and the video production capability of the NLM are also included under core support. The Center

provides photographic, graphics, and video support to all of NLM, including the operation and maintenance of the video production capability. Core support also includes maintenance and supplies, procurement assistance, financial tracking, property inventory, staff training and staff travel. Center staff regularly present papers at scientific meetings and several serve as project officers for ongoing contract activities, including the NLM's Telemedicine Initiative awards.

External Research Support

In FY 1998 Center staff will be involved in several NLM supported efforts. NLM will be a major partner in the multi-agency Digital Libraries Initiative, phase 2. The program is administered by the National Science Foundation. The sponsored research program will address the entire digital libraries life cycle from information creation to archiving and preservation. In the fall of 1997, NLM joined the MIT Media Lab News in the Future Consortium. Membership in the consortium will allow NLM to have access to innovative research being conducted at the media lab in a variety of areas, including machine understanding of content, computer modeling, and automatic collection and display of information. The Center's Office of High Performance Computing and Communications coordinates NLM's Telemedicine research awards.

Telemedicine

Research and evaluation issues included in NLM's telemedicine program include: (1) the impact of telemedicine on the health care system as a whole and on cost, quality, and access to care for specific populations; (2) the benefits of integrated access to practice guidelines, expert systems, bibliographic databases, electronic publications, and other knowledge-based information from within computer-based patient record systems and other automated systems that support research and practice; (3) the maintenance of patient confidentiality as increasing amounts of electronic health data are transmitted via telecommunications during health care and aggregated for public health and research purposes; and (4) the development of data standards and uniform practices for effective transmission, aggregation, and integration of health care, public health, and research data.

The NLM is funding 19 telemedicine projects, affecting rural, inner-city, and suburban areas, with a total budget of $42 million. The 19 multi-year projects, located in 13 states and the District of Columbia, will serve as models for: evaluating the impact of telemedicine on cost, quality, and access to health care;

assessing various approaches to ensuring the confidentiality of health data transmitted via electronic networks; and testing emerging health data standards. Each project will review and apply recommendations from two NLM-supported studies by the National Academy of Sciences' Institute of Medicine:

NAS/IOM Report: "Telemedicine: A Guide to Assessing Telecommunications for Health Care"

Telemedicine is receiving increasing attention not only in remote areas where health care access is troublesome but also in urban and suburban locations. Yet the benefits and costs of this blend of medicine and digital technologies must be better demonstrated before today's cautious decision-makers invest significant funds in its development. The telemedicine study, designed to develop a framework for evaluating the patient care applications of telemedicine, identifies managerial, technical, policy, legal, and human factors that must be taken into account in evaluating a telemedicine program. The IOM-appointed committee reviewed previous efforts to establish evaluation frameworks and reports on results from several completed studies of image transmission, consulting from remote locations, and other telemedicine programs. The committee also examined basic elements of an evaluation and considered relevant issues of quality, accessibility, and cost of health care.

NAS/CSTB Report: "For the Record: Protecting Electronic Health Information"

When visiting a doctor, information may be recorded in an office computer. Tests may be sent to a laboratory or a consulting physician. Relevant information may be transmitted to a health insurer or pharmacy. Data may be collected by the state government or by an organization that accredits health care or studies medical costs. By making information more readily available to those who need it, greater use of computerized health information can help improve the quality of health care and reduce its costs. Yet health care organizations must find ways to ensure that electronic health information is not improperly divulged. Patient privacy has been an issue since the oath of Hippocrates first called on physicians to "keep silence" on patient matters, and with highly sensitive data entering patient records, concerns over privacy and security are growing. "For the Record: Protecting Electronic Health Information" responds to the health care industry's need for greater guidance in protecting health information that increasingly flows through the national information infrastructure—from patient to provider, payer, analyst, employer, government agency,

medical product manufacturer, and beyond. This report makes practical detailed recommendations for technical and organizational solutions and national-level initiatives. The report describes two major types of privacy and security concerns: the increased potential for inappropriate release of information held by individual organizations (whether by those with access to computerized records or those who break into them) and systemic concerns derived from open and widespread sharing of data among various parties. The committee reports on the technological and organizational aspects of security management, the effectiveness of technologies for user authentication and encryption, obstacles and incentives in the adoption of new technologies, and mechanisms for training, monitoring, and enforcement. The report also reviews the growing interest in electronic medical records, the increasing value of health information to providers and others, and the current legal environment for protecting health data.

Office of the Public Health Service Historian

The PHS Historian's Office is also a part of the Center. Created in 1992, the Historian's Office was transferred to NLM as part of the 1995 reorganization of the Department. A significant amount of time of this Office in FY 1997 was devoted to preparations for the celebration of the PHS Bicentennial in 1998. In addition to participating in the planning of events, the Office worked on several specific Bicentennial projects, including an exhibit on the PHS at Ellis Island at the National Museum of Health and Medicine (scheduled to open in February, 1998), the introduction of a "Bicentennial Bits" column in the *Commissioned Corps Bulletin*, and a PHS Bicentennial calendar. Staff submitted several articles for publication and delivered a dozen lectures on various historical topics. The Office also answered numerous historical queries and cooperated with other agencies in assuring the preservation of important documents and artifacts relevant to PHS history (including materials from the now-closed St. Elizabeths Hospital Museum). In November, John Parascandola received the PHS Surgeon General's Exemplary Service Award for his work as curator of the exhibit on "Doctors at the Gate: The United States Public Health Service at Ellis Island" (displayed at the Ellis Island Immigration Museum, May 24-July 31, 1996). A major emphasis for the office in FY 1998 will be to provide support for the PHS Bicentennial commemoration in 1998: exhibits, publications, lectures, and other special events. The office is working with staff of the History of Medicine Division on a small PHS history exhibit for the NLM.

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.

NCBI consists of senior scientists, postdoctoral fellows, and support staff working in computational biology research and the application of these methods to building public information resources. 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.

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 growth rate of GenBank continues exponentially. In FY 1997, approximately 700,000 new sequences were added. The million mark in terms of number of sequences was reached in September 1996 and the two-million mark will probably be reached by January 1998. It took GenBank 16 years to accumulate the first million sequences and will take less than 16 months to acquire the second million sequences. Thus, the doubling rate continues with a 16-month time constant. For the coming year, the rate of growth shows no signs of abating. In fact, major technological breakthroughs in automated sequencing technology promise that the exponential growth rate will only increase.

Complete genome sequences are now available for representative species from three domains of life--Archaea, Bacteria, and Eucaryotes. During FY 1997, the full sequences for *Helicobacter pylori*, *Mycoplasma genitalium*, *Methanococcus jannaschii*, and *Synechocystis* PCC6803 joined the other completely sequenced genomes of *Escherichia coli* and *Saccharomyces cerevisiae* (yeast). Other organisms for which there was substantial growth in the number of GenBank records are *Mus musculus* (house mouse), *Caenorhabditis elegans* (a roundworm), and *Drosophilia melanogaster* (fruit fly). Sequences from these organisms continue to provide valuable clues for understanding the functioning of human genes.

NCBI produces GenBank from thousands of sequence records submitted directly from authors prior to publication; records submitted to our international collaborators, EMBL and DDBJ, are added daily. Indexers with specialized training in molecular biology create the GenBank records and apply rigorous quality control procedures on the data. As a final step, senior NCBI scientists review the records for accuracy of biological information. Improving the biological accuracy of submitted data and correcting existing entries are high priorities for the GenBank team. New releases of GenBank are made every two months; daily updates are made available via the Internet and the WWW.

Comprehensive coverage of all sequence data, protein as well as DNA, is provided by GenBank along with the corresponding MEDLINE bibliographic information, including abstracts. 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. 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. Data obtained through cDNA sequencing, Expressed Sequence Tags (ESTs), are critical to understanding gene function and represent a growing proportion of the database. As such, additional annotation is available for these sequences as part of the separate dbEST database.

An international collaboration with the EMBL Data Library at Hinxton Hall, UK, and the DNA Database of Japan in Mishima facilitates the exchange of sequence data worldwide through a shared automated system of daily updates. Other cooperative arrangements, such as with the U.S. Patent & Trademark Office for sequences from issued patents, augment the data collection efforts and ensure the comprehensiveness of the database.

PubMed

A major investment in NCBI's time and resources was made this year to the design, building, and maintenance of the PubMed database. Development of PubMed was initiated last year with an experimental version of Entrez on the World Wide Web that contained all of MEDLINE with links to full-text publishers' Web sites. Links to DNA and protein sequences, genome mapping data, and 3-D protein structures complete the integrated approach that Entrez is known for. In addition to the complete MEDLINE, PREMEDLINE is also included, as well as the entire journal coverage for journals linked to full-text publishers' sites (such as *Science*) that are selectively indexed for MEDLINE based on the relevancy of the article to biomedicine.

The algorithms for "neighboring" in the DNA and protein databases were adapted to MEDLINE so that links could be pre-computed to retrieve similar articles. After several months of testing within the molecular biology community, as well as the larger biomedical and public communities, the enormous level of usage and acceptance led to the announcement by Vice-President Gore in June 1997 that PubMed would be freely available.

In response to requests from both professional search specialists and experienced scientist/physician users, advanced searching modes have been refined to allow more sophisticated search methods and strategies. An evaluation was conducted with the National Network of Libraries of Medicine members, and NLM's Reference Section and MEDLARS Management Section staffs continue to provide feedback for the addition of features and standardization with the existing NLM mainframe systems.

The publishing community has responded in kind with growth from 24 to over 100 journals participating as full-text linked sites. Some publishers provide access only by subscription, and others provide free access to the current issue with subscription required for back issues.

PubMed usage has grown extraordinarily rapidly, with over one million different users accessing it in the four months since it became publicly available. Each day a quarter of a million searches of PubMed are performed.

Other Specialized Databases and Tools

NCBI continued to be active in the design, implementation, and distribution of new databases and software tools for the molecular biology community. Vice President Gore announced the NCI's Cancer Genome Anatomy Project, CGAP, at a press conference in August 1997. The CGAP Web site, jointly developed by NCI and NCBI, brings together data on gene expression in normal, precancerous, and malignant tissues, representing collaborations among many government, academic and industrial laboratories. The Gene Map of the Human Genome, available on NCBI's Web site, was produced in collaboration with eight labs in the U.S. and overseas and published in the October 25, 1996 issue of Science. This map shows the chromosome location of over 16,000 human genes with links to the underlying sequence and map data. It is updated regularly as new mapping results become available. The Database of Genome Survey Sequences (dbGSS), which is also contained in GenBank as the GSS division, was expanded this year with more comprehensive information on contributors, experimental conditions, and genetic map locations.

The BLAST sequence searching server is one of NCBI's most heavily used services and its usage continues to grow at a fast pace. 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 network to a BLAST server running on a computer at the NCBI. In a few seconds, the BLAST server compares the user's sequence with up to a million known sequences and determines the closest matches.

Each day more than 40,000 sequence searches are performed, with users submitting their requests through e-mail, server/client programs, and the World Wide Web. The e-mail service has a public key encryption option to guarantee the confidentiality of user data as it traverses the public networks. Through significant improvements to the BLAST software, which processes queries as fast as 1.6 seconds per search, and continued upgrades of hardware, NCBI has been able to accommodate the increasing database size and increasing number of sites using the service.

The BLAST suite of programs for sequence similarity searching was a particular focus for both extensive software improvements and an expansion of databases used with the software. Two new server algorithms were introduced in Version 2.0: Gapped BLAST, which allows the introduction of gaps (deletions and insertions) into alignments; and Position-Specific Iterated BLAST (PSI-BLAST), a new generation of protein database search programs that provide iterative searching using statistically significant alignments after an initial query.

Specialized BLAST services were added that search recently released microbial genomes and tentative human consensus sequences from The Institute for Genomic Research (TIGR) and the University of Oklahoma's Advanced Center for Genome Technology (ACGT). In addition, the Genome Survey Sequence (GSS), High Throughput Sequence (HTG), and Escherichia coli databases were added to the BLAST database suite for searching.

A new tool called Electronic PCR was developed as a Web page for detecting PCR-based sequence tagged sites (STSs), which have been used as landmarks for construction of various types of genomic maps. Work is nearly complete on the High Throughput Genome (HTG) Web page, which will contain information for both regular GenBank users and sequencing centers about HTG sequence submission, processing, and access. A new Web page of Clusters of Orthologous Groups (COGs) is being developed that compares protein sequences encoded in seven complete genomes representing five major phylogenetic lineages.

The members of the NCBI taxonomy group plan the overall structure of the taxonomy database and Web pages, monitor the literature, and maintain contact with off-site taxonomy advisors and the collaborating sequence databases in Europe and Japan. They add new species or do other edits to the taxonomy database when

so required by new systematic insights and guide the NCBI indexing staff on taxonomic issues.

The taxonomy database contains 35,000 organisms represented with one or more sequences in GenBank (32,500 at the species level). About 800 species a month were added to the database in FY 1997, for an overall increase of 40% for the year. The taxonomy group is currently composed of three biologists and one computer scientist. A curator for green plants and fungi joined NCBI this year when one member of the group left. A fourth biologist with primary expertise in Metazoan zoology is being sought.

In collaboration with the NCBI structural biology researchers, the taxonomy group has assigned species names and classifications to all of the entries in the PDB structure database, and have set up an internal mechanism for assigning classifications to new entries. Work is in progress to index the structure entries through the taxonomy browser, as is currently done with nucleotide and protein entries.

For the past year, the international collaborators, EMBL and DDBJ, have adopted NCBI's taxonomy database as the standard classification. EMBL and DDBJ indexers use the database and routinely consult with NCBI taxonomists on a daily basis.

Database Access

Entrez Retrieval System

The major database retrieval system at NCBI, Entrez, was originally developed for searching nucleotide and protein sequence databases and related MEDLINE citations, and has been expanded to include the integrated set of PubMed, MMDB 3-D Structure, Genomes, and Taxonomy databases. Users can search gigabytes of sequence and literature data using techniques that are fast and easy to use. 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.

Entrez's design permits incorporating additional linked databases without changes in the user interface. WWW Entrez now provides graphical views of nucleotide and protein sequences and access to the NCBI Genomes database, which contains graphical views of sequences and chromosome maps. The structure viewer, Cn3D, permits visualization of 3-dimensional protein structures.

Entrez users submit over 80,000 text searches and 35,000 sequence similarity searches daily. Nearly all of the increase in search activity has been from Web-based services. Approximately 1.2 million different IP addresses have used NCBI Web services in the past four months. Of these users, some 200,000 have accessed exclusively the molecular biology databases. Each day, 50,000 users access NCBI's databases and the number continues to grow.

Other Network Services

Overall, usage of NCBI's World Wide Web services, first introduced in December 1993, continues to expand as more services are added. In addition to PubMed, CGAP, the Human Gene Map, dbGSS, Electronic PCR, and new versions of BLAST and Sequin, NCBI staff continued to make access and usage easier with improved documentation and tutorials. Information about NCBI, its databases and services, data submission and update, and individual scientist's research projects is readily available, as well as an ever-increasing number of search tools. The WWW server provides capabilities for Entrez and BLAST searches and submission by BankIt. Many other WWW servers have links to the NCBI server to conduct searches and obtain the latest GenBank records. At the end of FY 1997, NCBI's site was averaging over 1,000,000 accesses daily.

GenBank is also distributed over the Internet through the standard File Transfer Protocol (FTP) program. Many large commercial and academic sites maintain a local copy of GenBank. Every two months, 600 sites download new full releases of GenBank. NCBI's Data Repository, with over 50 additional molecular biology databases, is also distributed via FTP; over 10 gigabytes of data, including daily GenBank and dbEST update files, are downloaded daily.

NCBI's electronic mail servers, RETRIEVE and QUERY, are used to retrieve records from several sequence databases, including GenBank, EMBL, Swiss-Prot, and PIR, by sending a mail message containing a text query to the server. 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 1,500 queries are handled daily by the RETRIEVE and QUERY servers.

The improvement of NCBI's sequence submission software continued to be a high priority. Numerous improvements were made to Sequin, the stand-alone submission tool released last year, including the integration of Network Entrez, PowerBlast, a repeat finder, an ORF finder, and a new alignment algorithm into the software. In addition, users can now update existing GenBank records with Sequin, and replace or merge sequences with ease. Complete documentation and a tutorial are maintained on the Web site. The Web submission tool, BankIt, now in its third year of use, continues to be the most popular submission tool, but submissions entering via this route has dropped from over 80% to about 65% with the growing use of Sequin.

NCBI uses a network of Unix-based symmetric multiprocessor servers to support its Internet database services. Each day NCBI generates over 60 gigabytes of data to send over the Internet in response to over half a million user transactions (queries, similarity searches, and FTP downloads of data files). During FY 1997 many of these servers were upgraded to accommodate the increased demand and the increased size of the databases they serve. In most cases, the servers were upgraded by the addition of CPUs and memory. In other cases, existing servers were upgraded to new generations of CPU architecture. Internally, network speed was increased through the installation of 100 megabit/sec Fast Ethernet interfaces for major server systems. These connections complement the existing 100 megabit/sec FDDI interfaces for older servers.

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. With ASN.1 definitions and the NCBI software toolkit, complicated analysis programs can be readily constructed from pre-existing sets of modular tools, saving considerable time and programming effort.

Basic Research

Basic research is at the core of NCBI's mission. The Computational Biology and Information Engineering Branches at the NCBI comprise 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.

There has been a variety of applied and theoretical studies of biologically important molecules and their function, as well as continued development and improvement of algorithms and statistics for their analysis. Computer program development has included PROBE for the automatic delineation of protein families; PowerBLAST, a network application for interactive sequence analysis and annotation; PSI-BLAST for iteratively searching protein databases with position-specific score matrices; and modification of BLAST to run three times faster and produce gapped alignments. Macromolecular sequence analysis programs have been applied to investigate HIV-1 haplotypes, HTLV-I subtypes, CED-3 and CED-4, the protein subunit of ribonuclease P, ubiquitin regulators, 16S-like rDNA sequences, human repetitive elements, integral membrane phosphatases, histone N-acetyltransferases, spindle assembly checkpoint proteins, and DNA damage-responsive cell cycle checkpoint proteins.

Genome-scale projects have continued to be a staple of Computational Biology Branch research. Projects have included the creation of a transcript map for the 2.8 Mb region containing the multiple endocrine neoplasia-type 1 (MEN1) locus, and the positional cloning of the MEN1 gene; the comparison of archaeal and bacterial genomes; a comparison of ESTs to human genomic sequences; and a comparative analysis of approximately 1200 orthologous human and mouse mRNA and protein sequences. Other efforts have included the development of a database of protein domains; the development of a database of histones and histone fold sequences and structures; research on a model of the HIV infectivity; description of a measure for the success of protein fold recognition; and, a study of the extent and nature of contacts between protein molecules in crystal lattices.

The intramural group is engaged in over 40 projects, many of which involve collaborations with NIH and other research laboratories. The work is reviewed by a Board of Scientific Counselors of 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, over 50 in FY 1997, and the requests for outside collaborations. The staff has made 34 presentations at major scientific meetings, academic departments, and companies engaged in molecular biology research; in addition, they made 23 presentations at NCBI's computational biology lecture series.

The Visitors Program continues to be successful in bringing members of the scientific community to the NCBI to engage in collaborative research with the Computational Biology Branch as well as joint activities in database design and implementation with the Information Engineering Branch. This program, administered in conjunction with Oak Ridge Associated Universities, facilitated visits by 43 individual senior researchers this past year.

User Support and Outreach

As part of its mandate to facilitate the use of databases and software by the biology community, NCBI maintains a user support group with broad experience in handling biology and medical information. The primary focus of this group is to support the particular services that NCBI offers by e-mail, phone, and fax. NCBI has extended its outreach to the library science community by invited presentations and workshops on biotechnology information topics.

As the number of database services and number of users has increased, the scope of user support services has also expanded. NCBI staff in the Information Resources Branch, with contractor support, provide responses to queries for information and assistance. The three main areas of user support include: information about GenBank and related molecular biology database 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. In fall 1997, assistance with general PubMed inquiries from both users and publishers was provided by staff from MEDLARS Management Section, NLM.

To increase awareness of NCBI and its programs, NCBI staff participate in exhibits, seminars, workshops, and courses, both nationally and internationally. In FY 1997 NCBI staffed exhibits at scientific society meetings, including the American Society for Human Genetics; American Society for Cell Biology; Biophysical Society; HUGO--HGM '97; International Conference on Mathematical & Computer Modelling and Scientific Computing; American Society for Microbiology; Special Libraries Association; American Crystallographic Association; and, American Society for Biochemistry and Molecular Biology. Additional exhibits were staffed at more specialized

meetings, including TIGR/Small Genomes, Genome Sequencing & Analysis Conference, and NIH Web Day.

In addition, NCBI members participated as faculty at courses sponsored by the American Association for Cancer Research; Catholic University of American; Johns Hopkins School of Medicine; George Mason University's Institute for Computational Sciences and Informatics; The Jackson Laboratory; NLM/Marine Biological Laboratory; the Medical Library Association; the International Center for Genetic Engineering & Biotechnology in Trieste, Italy; and, the Cancer Institute Hospital in Beijing, People's Republic of China.

The newsletter was distributed to a mailing list of 25,000 biologists and institutions, and new Fact Sheets on programs and services were distributed at all public forums where NCBI was represented.

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 Human Genome Research Institute and the National Cancer Institute on databases and informatics programs that impact information exchange on a national level.

Extramural Programs

Funding for extramural bioinformatics activities is the responsibility of another division of the NLM, the Extramural Program. It 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. Extramural 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 and the application of this knowledge to medical practice 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 rapidly addresses 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

The Extramural Programs Division (EP) of NLM continues to receive its budget under two different authorizing acts, the Medical Library Assistance Act (unique to NLM), and Public Health Law 301 (covers all of NIH). The funds are expended as grants-in-aid, and in some instances as contracts, to the extramural community in support of the goals of the NLM. Review and award procedures conform to NIH policies.

This chapter includes Table 10 (FY 1997 Budget By Category) and, as appendixes at the end of the report: Appendix 1: FY 1997 Grant Portfolio; Appendix 2: FY 1997 BLRC Members; Appendix 3: FY 1997 Board of Regents Subcommittee Members.

The Grant Programs

EP issues grants in several categories for which the motif in general is medical informatics:

- Resource (for Information Management)
- Training
- Research (and Research Resources)
- Publication
- Educational Technology
- SBIR/STTR
- Other

Resource

Resource Grants, authorized by the Medical Library Assistance Act, support access to information as well as promote networking, integrating, and connecting computer and communications systems. There are four types of Resource Grants which range in complexity relative to their intent as well as dollar amounts and duration of time. They are considered "seeding" grants, thereby funding the large capital costs of initiating a resource or service or program to eventually become self-sustaining. The commonalities among the four Resource Grants are that public and private, nonprofit health institutions/organizations engaged in education, research, patient care, and administration are eligible and that there must be library involvement.

Information Access Grants

Information Access Grants, aimed primarily at hospitals, clinics, community health centers, and similar small health organizations, support computers to access Internet Grateful Med/PubMed and also fund the automation of public access catalogs for health science libraries. These grants provide up to $12,000 per institution and are available to single as well as multiple institutions. In FY 1997 there were two grants to Area Health Education Centers (Rural Health Projects, Inc. in Enid, OK, and Southwest Georgia AHEC in Albany) to support Internet Grateful Med access in 22 hospitals and other health-related institutions where a variety of health students receive clinical training. Similarly, Information Access Grants to the University of Alabama in Tuscaloosa and the University of North Texas in Fort Worth are funding computers to some 19 sites where students are serving residencies, clerkships, or preceptorships. Two additional Information Access Grants have assisted in the installation of automated public access catalogs at Mercy Hospital in Pittsburgh and the Lanterman State Hospital and Development Center in Pomona, CA.

Information Systems Grants

Information Systems Grants, ranging up to $150,000 per year for up to three years, are aimed at larger hospitals and medical schools and support computer access to information on a more complex level than Information Access Grants. In FY 1997 three such grants were awarded. The first, to Tufts University Health Sciences Library, supports access to digital multimedia resources so that the content of these valuable resources can be better accessed and retrieved for health science students, educators, and researchers. The second such grant was awarded to the University of Minnesota's School of Nursing for its Rural Nurse Practitioners Programs so that nursing students training in rural areas will have computers to access information resources for their studies. Documents are retrieved through the participation of the Bio-Medical Library at the University of Minnesota. The third Information Systems Grant is to the Northern Wisconsin Area Health Education Center at Wausau and is funding the development of the Northwoods Healthnet so that health professionals and health students have access to information resources. Close to 50 sites throughout northern Wisconsin are participating including public health departments and tribal clinics. The health sciences library of the Marshfield Clinic serves as the resource center with collection and staff resources. An extensive training program is a major component of this project.

Internet Connections Grants

The Internet Connection Grants provide funding for single institutions up to $30,000 and multi-institutions up to $50,000 to initiate Internet access for health institutions by funding gateway/router equipment, Internet Service Provider fees, and line charges in the first year. Some institutions with existing Internet access extend it to other institutions. In FY 1997 15 Internet Connection Grants were awarded: Berkshire Medical Center, Berkshire, MA; La Porte Hospital, La Porte, IN; Washington University, St. Louis, MO; Putnam Memorial Hospital, Burlington, VT; Medical College of Wisconsin, Milwaukee, WI; North Carolina AHEC, Chapel Hill, NC; Ephrata Community Hospital, Ephrata, PA; Baptist Hospital, Birmingham, AL; Grande Ronde Hospital, La Grande, OR; Burke Rehabilitation Hospital, White Plains, NY; Pikesville United Methodist Hospital, Pikesville, KY; Atlanticare Health System, Egg Harbor Township, NJ; Wright State University, Dayton, OH; Hope of Lee County Hospice, Fort Myers, FL; California College of Podiatric Medicine, San Francisco, CA; and Critical Illness/Trauma Foundation, Bozeman, MT.

IAIMS Grants

Integrated Advanced Information Management Systems (IAIMS) Grants are designed to support institution-wide information systems that link a variety of individual and institutional databases and information systems for patient care, education, research, library, and administration. IAIMS Grants support two phases, planning and implementation, with the program goal being to support organizational mechanisms that foster the integration and sharing of various information systems. The planning phase funds up to $150,000 for one to two years; the operational phase up to $500,000 per year for five years or $550,000 for an IAIMS apprenticeship option. In FY 1997 five IAIMS Grants were funded to supplement previously funded Planning Grants in order to evolve to the Operational Phase.

Training And Fellowships (MLAA)

Training in Informatics

Managing the potential of computers and telecommunication for health care information requires investigators who understand fundamental problems of knowledge representation, decision support, and human-computer interface. NLM remains the principal support nationally for research training in the fields of medical informatics, including clinical and basic

science domains. NLM provides three mechanisms of support for its training activities. Five-year institutional training grants support approximately 150 trainees at predoctoral and postdoctoral levels. Institutions currently receiving such support:

1. "Boston Consortium" (Harvard Medical School; Massachusetts Institute of Technology; Brigham and Women's Hospital; Massachusetts General Hospital)

2. University of Minnesota

3. University of Pittsburgh

4. Stanford University

5. Yale University

6. Rice University; Baylor College of Medicine; University of Houston

7. Oregon Health Sciences University

8. University of North Carolina; Duke University

9. University of Missouri, Columbia

10. Columbia University

11. University of Utah

12. University of Indiana, Regenstrieff Institute

Individual informatics research fellowships are available to those who seek research training similar to offerings at the institutional training sites but at a site of their choosing. Individual applied informatics fellowships are available to individuals who want to learn informatics techniques and technology for application in their current professional specialties. NLM initiated five fellowships this year.

Planning for the Education of Health Sciences Libraries

In early 1995, as a one-time initiative, NLM issued a Request for Applications (RFA) for Planning Grants for Education and Training of Health Sciences Librarians to provide further planning for education in four areas identified by the NLM Planning Panel on the Education and Training of Health Sciences Librarians: 1) evolving role for health sciences librarians; 2) professional education programs; 3) lifelong learning programs; and 4) recruitment including minorities. Twenty-one applications were received and seven

awards were made: Welch Medical Library of The Johns Hopkins University, Baltimore, MD; Graduate School of Library and Information Science, University of Illinois, Champaign, IL; School of Library and Information Science, University of Missouri, Columbia, MO; School of Information and Library Science, University of North Carolina, Chapel Hill, NC; School of Library and Information Science, University of Pittsburgh, Pittsburgh, PA; College of Library and Information Science, University of South Carolina, Columbia, SC; and Eskind Biomedical Library, Vanderbilt University, Nashville, TN.

After reviewing options, in FY 1997 NLM elected to approach the issue with a 3-element project beginning in FY 1998:

- All existing NLM Informatics Training Programs would be strongly encouraged to develop and offer training within the curriculum suitable for those interested in health science libraries. NLM agreed to provide additional funding for any slots awarded to librarians.

- Applied Informatics Fellowships were again widely publicized to those in the library community as a means of supporting training at adequate stipend levels to those interested in obtaining informatics training but perhaps at a mid-career level or otherwise not able to move for educational purposes.

- Library Operations developed a second, optional year of training for NLM's Library Associate Program for those Associates interested in spending a second year of "internship" at an appropriate health center library capable of providing intensive experience in informatics.

Training of Minorities

NLM is participating in an NIH-wide fellowship program aimed at encouraging under-represented minorities into research careers.

Research and Research Resources (PHS 301)

Research grants are made through a variety of mechanisms, including individual research projects, cooperative agreements, research resource grants and others.

Medical Informatics

NLM's research grants sponsor the investigation of basic and applied medical knowledge issues that arise at the intersection of biomedicine, computer science, and human behavioral sciences. This year, NLM initiated 14 new individual informatics research projects and co-funded two more with another institute. In addition there was one First Independent Research Support and Transition Award and one Shannon Award for a young investigator. The "Shannon Awards" are grants for highly promising or unusually significant projects which, for one reason or another, fall just short of normal funding plans. They are modestly budgeted, but they give the investigator an opportunity to start work on which to base a subsequent, and larger, proposal.

To promote informatics research projects of particular interest to nursing, a Program Announcement (PA-95-010), Enhancing Clinical Care Through Nursing Informatics, jointly supported by NLM and the National Institute of Nursing Research, was released December 2, 1994. Although NINR has subsequently refocused its funding priorities, the PA stands. Applications received because of it are reviewed by NLM's BLRC, and if meritorious, are funded by NLM and/or NINR.

In general, an on-going EP project is targeted toward encouraging nurses to apply for relevant NLM grants, making nurses aware of both Program Announcements and Requests for Applications by wide electronic distribution .

Biotechnology Informatics (Computational Molecular Biology)

The techniques of informatics are indispensable tools for handling the complex data generated by molecular biology research. NLM continues to provide research grants for informatics projects in this area of basic medical science, as well as training grants, and grants for support of research resources. A related problem concerns the development and maintenance of electronic databases on which researchers increasingly rely, and for which no other source of support has yet been identified. This year, NLM awarded two new investigator initiated research projects, one conference grant, and one database project grant, co-funded with another institute.

NLM also participates with 15 other organizations in the Human Brain Project, which seeks innovative methods for discovering and managing increasingly complex information in the neuroscience. For example, NLM awarded a research grant to Dr. Douglas M. Bowden of Washington University who plans to integrate a wide range of data on human and primate neuroanatomy in a computerized Brain Information Management System.

Publication Grant Program

The Publication Grant Program provides short-term financial support for selected not-for-profit, biomedical scientific publications. Studies prepared or published under this NLM program include critical reviews or monographs on special areas of biomedical research and practice; research monographs in the history of medicine and the life sciences; writings on medical informatics, health information science and biotechnology information; and, in certain instances, secondary literature tools and scientifically significant symposia. Resources in recent years have been used principally for history of medicine projects, but projects in electronic publishing, video, and other media were also supported. The program has an informal self-imposed ceiling on direct costs per grant per year. In November the ceiling was raised from $25,000 to $35,000.

Educational Technology

Extramural activity in educational technology development has been limited to those projects that have included media and electronic resources in their research design, either as an independent or dependent variable. Researchers and other experts seeking funding for educational technologies have been advised to consider coupling media or electronic resources to be developed with dissemination or translation of research findings.

Staff refer inquiries about funding for educational, mediated products to other entities of NIH, DHHS, or other Federal agencies. The growth in the number of Internet-accessible sites of information about grants, contracts, and collaborative programs has greatly facilitated such referral.

SBIR/STTR (PHS 301)

All NIH research grant programs, including NLM's, by Congressional mandate must allocate a fixed percentage of available funds every year to Small Business Innovation Research (SBIR) grants. These projects may involve a Phase I grant for product design, and a Phase II grant for testing and prototyping. This year, NLM's allocation was awarded to four applicants for Phase I support. NLM also participates in the other mandated fund allocation program, Small Business Technology Transfer, but generally it contributes its small allocation to other NIH institutes, as it did this year.

Other Grants

Conference Grants

Support for conferences and workshops is intended to help scientific communities identify research needs, share results, and prepare for productive new work

Biomedical Ethics

Ethical issues in health care and research produce an enormous literature. This literature comes from law, medicine, public health, and government. The National Reference Center for Bioethics Literature at Georgetown University continues to offer invaluable resources and guidance for workers in this area. An Extramural Programs contract maintains the Center. A complementary contract from Library Operations supports an indexing activity that contributes to BIOETHICSLINE, one of NLM's MEDLINE databases. This year NLM renewed its contracts to continue these important services.

Other Extramural Programs Activities

HPCC and Outreach

The Outreach and the High Performance Computing and Communications initiatives of NLM are elements of the formal grant programs, and are being met more fully since revision of the resource programs, the continued success of the Connections program, and the Electronic Medical Record System cooperative grants.

Minority Support

The Information Access Grant awarded to the University of Texas Health Science Center at San Antonio includes several institutional members serving significant Hispanic populations.

Grants Management Highlights

The Grants Management staff reviews NLM grant applications for administrative content and compliance with guidelines and directives; prepares and disseminates grant award documents; maintains official grant files for NLM; provides consultation and assistance to grantees on appropriate business management concepts; and advises NLM officials on grants management policy and procedures.

The Grants Management staff, which consist of four employees, issued a total of 154 awards for FY 1997. Effective October 1, 1997, the Notice of Grant Award for all of NIH was redesigned from a document format into a letter format. The same information is contained in the letter format. The purpose of this change was to enable the NIH to e-mail the Notice of Grant Awards to the Grantee Institution instead of mailing hard copies. This change in mailing the Notice of Grant Award should occur sometime in the very near future. In addition, we will join most of the other institutes of NIH in using the new Information Management Planning Analysis and Coordination II (IMPAC II) system to issue our Notice of Grant Awards.

Review Committee Activities

NLM's scientific merit peer review group, the Biomedical Library Review Committee (BLRC) met three times in 1997 and reviewed 122 applications of which 75 were approved. The Committee operates as a "flexible" review group; i.e., it is composed of three standing subcommittees, consisting of nine members on the Medical Library Resource Subcommittee; seven members on the Medical Informatics Subcommittee; and four members on the Biotechnology Information Subcommittee. The subcommittees consider research applications in medical library projects, medical informatics, and biotechnology information respectively.

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; when at least two members of the scientific merit review group dissented from the majority; when a policy issue is identified, and when an application is from a foreign institution. The Extramural Programs Subcommittee makes recommendations to the full Board which votes on the applications.

Review Reform

The NIH is participating in the President's "Reinventing Government" initiative. A portion of this activity has been the establishment of a Peer Review Oversight Group (PROG) charged with coordinating, evaluating, and making policy recommendations for all peer review conducted at NIH. The Oversight Group is continuing to meet and will consider input from the applicant community. Some changes which have occurred include:

- The name of the Division of Research Grants (DRG), which is now the Center for Scientific Review (CSR), and the reorganization of a number of standing study sections.

- Five specific criteria have been chosen as a basis for assessing the merit of NIH research project grant applications. These criteria are Significance, Approach, Innovation, Investigators, and Environment with the emphasis of each criteria depending on the type of application being reviewed.

- FIRST (R29) awards for newly independent biomedical investigators will become R01(traditional research project) awards after June 1998. Applicants wishing to apply for a FIRST award will indicate this on the cover of the application and should also check off the appropriate box on the new 398 application form. The traditional FIRST award review standards and criteria will be used when reviewing these applications and this change will allow for more flexible funding options.

- New guidelines will go into effect after October 1, 1998 for the inclusion of children in NIH funded research involving human subjects.

Divisional Operations

The operating budget was adequate for indicated site visits, and continuation of the Division's on-going maintenance. EP incorporated the new NIH streamlining initiatives in timekeeping and procurement.

Administration and Personnel

There have been two significant positions filled within Extramural Programs. Dr. Sharee Pepper is the new Scientific Review Administrator and Mr. John Seachrist is the new Grants Management Officer. The operating budget was adequate for maintaining the Division's activities. Additional funds were requested and approved for the anticipated relocation of Extramural Programs to the Rockledge One Building. This move is scheduled to take place sometime in the early summer of 1998. EP continues to incorporate the

new NIH streamlining initiatives by utilizing appropriate staff in various tasks throughout the year.

Summary

NLM's EP, like almost all extramural grant divisions at NIH, regrets that not all applications of good quality can be funded, but the grants which can be made are furthering NLM goals in key areas. However, support for developing the educational technology of informatics remains uncomfortably small, and, most importantly, we have not yet expanded the Informatics research budget commensurate with the increase in Informatics scientists leaving our training programs.

Table 11

Extramural Grant and Contract Program
(dollars in thousands)

Category	FY 1995*		FY 1996*		FY 1997	
	No.	$	No.	$	No.	$
Resource projects	**53**	**5,351**	**53**	**5,754**	**40**	**5,307**
IAIMS	(7)	(2,231)	(10)	(3,011)	(8)	(3,040)
Access	(11)	(775)	(5)	(476)	(6)	(320)
Systems	(10)	(1,437)	(12)	(1,387)	(11)	(1,412)
Connections	(25)	(908)	(26)	(880)	(15)	(535)
Research	**59**	**13,652**	**68**	**14,817**	**80**	**14,994**
Medical informatics projects	(21)	(4,515)	(24)	(4,433)	(32)	(5,955)
Medical informatics resource	(1)	(1,335)	(1)	(1,419)	(1)	(230)
Biotechnology	(20)	(4,381)	(19)	(5,264)	(27)	(6,259)
Cooperative agreements	(5)	(2,099)	(7)	(1,872)	(2)	(754)
Career awards	(11)	(1,138)	(16)	(1,638)	(18)	(1,796)
Library science	(1)	(184)	(1)	(191)	(0)	…..
Training	**26**	**5,088**	**19**	**4,501**	**20**	**5,645**
Institutional	(17)	(4,731)	(10)	(4,178)	(12)	(5,290)
Fellowship	(9)	(357)	(9)	(323)	(8)	(355)
Publications	**16**	**513**	**12**	**327**	**9**	**247**
Educational technology	**1**	**27**	**…**	**…..**	**…**	**…..**
Bioethics	**1**	**416**	**1**	**458**	**1**	**498**
SBIR/STTR	**1**	**191**	**2**	**196**	**4**	**401**
Regional Medical Library	**8**	**5,545**	**8**	**6,283**	**8**	**6,611**
NIH Tap		**…..**		**849**		**830**
Totals:	**(165)**	**$30,783**	**(163)**	**$33,185**	**(162)**	**$34,533**

*Some figures have been corrected

OFFICE OF COMPUTER AND COMMUNICATIONS SYSTEMS

Fernando Burbano
Director

Systems Reinvention

In FY 97 the Office of Computer and Communications Systems continued its efforts in support of the Systems Reinvention project. Key staff members were involved with the development of the Integrated Library System (ILS), Document Delivery System (Relais), Internet Grateful Med (IGM), PubMed, and Grateful Med for Windows.

Integrated Library System

The Integrated Library System project is a joint OCCS/LO project. Staff members from both organizations met with senior management of Endeavor Information Systems, Inc. (EIS) to determine whether the Voyager ILS product was a viable option for NLM. Briefings were presented to a broad audience of NLM staff. Based on positive preliminary assessment, a more in-depth review of Voyager will be undertaken. The system was installed and a 3-month evaluation begun. OCCS initiated two ImageWorld task orders to acquire Voyager and to conduct training and orientation. In addition, EIS will perform workflow analysis of NLM's processes. OCCS and LO staff have begun the planning for conversion of legacy system data to a format acceptable by Voyager. OCCS is exploring the possible use of Voyager for data entry and maintenance of bibliographic databases.

Registration Server

A joint OCCS/NCBI/LHC/LO project was established to support ordering of material from PubMed and IGM. OCCS is to develop a registration server to support registration and verification of individuals requesting ILL requests. Completed requests will be transferred to DOCLINE for Processing.

Relais

OCCS supported the implementation of the Relais system, which was procured by Library Operations to assist with filling Interlibrary Loan and overnight photocopy requests. Network Systems Incorporated (NSI) was awarded the contract to install Relais and has primary responsibility for implementation of the system. This effort required substantial support from the Office of Computer and Communications Systems, including:

- working with NSI on devising the protocol for downloading requests from the DOCLINE and LOCATOR systems and updating DOCLINE;
- implementing the DOCLINE and LOCATOR side of the download/update;
- implementing the necessary LAN infrastructure;
- implementing an NT domain;
- installing necessary software components on staff workstations;
- installing, configuring and administering an Oracle database;
- devising and implementing a security plan; and
- general technical oversight to the implementation.

CustomerQ

OCCS Implemented and provided support for the Quintus CustomerQ product which Library Operations procured to support their Customer Service work. This included installing and configuring an Oracle database, installing several version of the CustomerQ software, and working with the Customer Service Team on using the "out of the box" CustomerQ system.

PubMed

Support for PubMed included preparation of databases and procedures for transfer from MEDLARS to PubMed; consultation on database content and searching techniques, especially as related to vocabulary and mapping techniques.

Internet Grateful Med (IGM)

Support of Free Searching. In July, the IGM started working with the clock spooler, which resulted in growth of searches of 25% per month. ELHILL Response time remains in the 60-70 millisecond ranges per program transaction. In September daily and monthly records were shattered with 1.12 million searches being performed. To put this feat in context, there were 7.4 million searches performed in 1996.

ELHILL

Development effort centered on producing a mainframe Web server and spooler built as an interim solution to allow organizations to connect via the Web without the need of a User ID and password. With the decision to provide free Internet-based searching of MEDLARS databases, this project was terminated.

Grateful Med

This year has been a pivotal year for Grateful Med. The Windows Version 1.0 was successfully released to over 20,000 subscribers and is performing as anticipated. All domestic Value Added Network (VAN) access has been transitioned over to the AT&T FTS2000 Packet Switch Service with minimal impact to our users. NLM support for both the DOS and Macintosh versions of Grateful Med has been scaled back as MEDLARS users are moving over to the Windows version, Internet Grateful Med, and PubMed. A new Windows version of Grateful Med which includes access to AIDSLINE, AIDSDRUGS, AIDSTRIALS, SDILINE, and PREMEDLINE databases will be released by the end of the year. It is anticipated that this will be the final release of Grateful Med for Windows.

MeSH Support System

In 1997 OCCS started the full-scale development of a new system for supporting the development and maintenance of NLM's Medical Subject Headings (MeSH). MeSH is the controlled vocabulary that is used for indexing and retrieval access to NLM's bibliographic databases such as MEDLINE. The MeSH support system incorporates a number of complex requirements for developing and maintaining the terms in a rigorous and highly controlled fashion. The MeSH system is being rehosted using client/server computer technology with the Oracle database management system (DBMS). Previously MeSH has been supported using mainframe-based technology and the Model 204 DBMS. This change is requiring a complete system rewrite and overhaul of all system requirements.

Currently, the MeSH/Oracle project has made substantial progress in a number of areas with the main emphasis being placed on the creation, editing, and validation of MeSH Descriptor records. The Delphi-based client workstation interface system to the Oracle server has been functioning well and new functionality is added to it on a regular basis. Portions of the MeSH/Oracle system are now being demonstrated,

tested, and evaluated frequently, as new functionality is implemented. A test workstation has been set up in the MeSH Section in order to provide for convenient access and testing by the MeSH support staff.

Oracle database tables that will be used to support linkages to the UMLS database system have also been designed and implemented. These table structures have been reviewed and approved by the Development Team and the MeSH Staff. Preliminary DBMS table designs has also been completed for the MeSH Qualifier and Chemical records.

NLM Computer Center

Hardware

The NLM network infrastructure has been redesigned and upgraded to make it faster and more efficient while increasing capacity and reliability. A major effort was undertaken to upgrade NLM production subnets. Production subnets support 10 and 100MB/s transfer speeds, equal to the main backbone speeds. Three protocols are now supported: 10baseT, 100baseT, and 56 kbs and ISDN access. Remote indexers connect to NLM through a new CUBIX system that provides 48 modems and a dedicated Pentium processor server. Additional work was undertaken to support a high availability infrastructure. Preparation for operational systems deployment included acquisition, installation, and testing of 100 gigabytes of hardware-based RAID storage and 2 terabytes of online high-speed tape backup capacity. An Automated Tape Library System was installed and supports 2000 tape cartridges and provides automated tape management for the NLM mainframe and non-mainframe systems.

Software

Development effort on the Online Registration System includes substantial software design and development to provide web access and online registration to MEDLARS databases. The legacy system was replaced by an integrated software solution. Another major initiative for this year was to develop and establish a group to assist with the conversion to 32-bit operating systems for desktop use.

Procurement activity has resulted in the transition from the prior LAN support contractor (AAC Associates) to the MIL Corporation. Considerable effort was required to transition responsibility for NLM's extensive and complex network and core application services.

Technical service support enhancements include developing support for FTP for OCLC and CONSER, developing support for Z39.50, and developing support for making preservation data available for FTP. TESS (Technical Services System) integrates various functions of the Technical Services Division with the emphasis on the cataloging processes. It is a client/server application with the major databases residing on the IBM mainframe and the editing of records on PC workstations. The communication subsystems are based on the TCP/IP protocols making TESS operable over both Local Area Networks (LANs) and the Internet.

The initial release of TESS provided for the creation and maintenance of original cataloging, followed by a framework for the integration of acquisition and cataloging activities. Authority control for the cataloging function was next integrated into TESS. This release also included the capability for creation and maintenance of the name authority file.

In 1997 programs were developed to FTP data to OCLC in support of the CONSER cooperative agreement. The subsystem that is used to import data from the Library of Congress was also fully converted to a Z39.50 based client-server application. This has provided a new capability to obtain catalog information from OCLC and any other organization that has a Z39.50 server. Preservation data was also imported from the Inquire databases, and it is now available for distribution in USMARC formats.

Z39.50

In 1997 Z39.50 access to the PREMEDLINE database was added to the Z39.50 server. This was done in order to satisfy a requirement for online PREMEDLINE access by the University of California library system. The UC system uses Z39.50 to retrieve information from its central retrieval system, MELVYL, and it can also retrieve PREMEDLINE information from NLM in a similar fashion.

Inquery Retrieval System Evaluation

In 1997 OCCS loaded MEDLINE data into the Inquery retrieval system that was developed by the Center for Intelligent Information Retrieval (CIIR) at the University of Massachusetts. A variety of tests and evaluations were done on the Inquery retrieval system, including testing with a complete copy of MEDLINE in one large Inquery database.

Information Systems Laboratory (ISL)

The Information Systems Laboratory is a core facility designed to employ such emerging technology as distributed processing, open systems, high-speed networks, and worldwide connectivity. 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 has now become a major domestic and international access pathway. The ISL has become the core team for Unix systems hardware and software support in OCCS. Several Unix-based production systems now depend upon ISL hardware and software in order to support public user access to NLM data. The Information Systems Laboratory continued to provide system support for Unix-based applications and services. Performance and operational analyses resulted in several applications being re-platformed to more appropriate hardware configurations. A major consideration was to develop system support architectures without any single points of failure, and substantive strides were made. A major support area was the development and support of NLM's World Wide Web server. Based on the Hyperdoc server developed in the Lister Hill Center, the WWW server has been restructured, reformatted, and re-implemented as an operational facility.

MEDLARS

Year2000: Approval was obtained to proceed with several recommendations that involve NLM MEDLARS II software, data and responses to public inquiries regarding year2000 compliance. It is expected that most of the MEDLARS II searchable data will be Year2000 compliant when the 1998 system becomes operational in December 1997.

OLDMEDLINE: This database, currently containing 1964-5 *Index Medicus* (IM) citations, became operational. A task order has been awarded for the keyboarding of 1962-3 IM data and it will be added to the database.

Citation Maintenance Systems (CMS): A report subsystem was created to provide statistical information. It allows authorized users to print reports consisting of full or partial records selected from CMS databases. AIDSLINE Citation Maintenance System was created to provide for maintenance of AIDSLINE data.

MEDRECS: This new system was created and became operational. MEDRECS is a Local Area Networked based system that provides for scanning of selected journal title citation (titles and abstracts) data and linking this data to other bibliographic parts of the record. The completed citations are file transferred and become input to update MEDLINE.

Billing systems: MEDLARS user information, statistics, master account data, and invoices both monthly and quarterly are now electronically transferred to NTIS.

DOCLINE and Loansome Doc: Programs and procedures to download requests from DOCLINE to Relais and update DOCLINE with data from Relais are complete and in final system testing. Major changes were made in DOCLINE to allow delivery method specification in preparation for Relais. Ability to order non-serial MEDLARS citations from IGM, WGM, and DOCLINE was also added. A plan to link PubMed with Loansome Doc was completed and approved. The coding is progressing. The link is expected to be available in January 1998.

Publications: Changes were made to the MeSH Trees publication for 1998. A program was written to allow selective abstracts to be printed in the *Bibliography of Bioethics*. Work was performed with GPO to test sending an SGML version of IM electronically. They constructed a DTD to produce IM on their PC based system. Work continues on diacritic marks and final project plans.

Other Accomplishments

- Approximately 3% or 240,000 citations in the
- MEDLINE family of 8.7 million citations were class maintained. This consisted of over 300 individual tasks of which approximately 40% were

MeSH, approximately 40% chemical and 20% other related tasks. In addition, appropriate tasks were used and class maintenance was performed for other databases, e.g.; CANCERLINE, HealthSTAR, AIDSLINE, SPACELINE, HMD. Expansion of NIH Grant Numbers was performed in thousands of records for these databases. Lastly, Publication Type (PT) data received from the Cochrane Center was applied to hundreds of citations.

- Eagle Design and Management, Inc., was awarded a contract to perform the study and prototype for "Alternate Media for Distribution of NLM Leased Databases." Their report is due early in FY 1998.

- OCCS assisted in the creation of the current client/server MARS system. This system provides for the scanning/OCR of abstracts and titles for approximately 500 citations/week. Other fields of data are keyboarded using the system and this data is linked to the abstract and title. Data are uploaded to the mainframe, put in PREMEDLINE, made available to NCBI for PubMed and placed in the mainframe Indexing system for indexing.

- OCCS established procedures and updated MEDLINE daily for several months during the keyboarding contract protest. Procedures were established to process data prepared and sent to NLM from other sources based upon collaborative studies (Carnegie Mellon) and purchase orders let to other contractors. In addition, OCCS worked with NCBI to process SGML data provided in a more accelerated way. Several significant changes were made to the AIMS system to accommodate data coming from several directions during this time of crisis.

ADMINISTRATION

Executive Officer
Donald C. Poppke

National Performance Review

NLM continued its high priority effort, the Systems Re-engineering Initiative, in support of the National Performance Review. A designated "Reinvention Laboratory" under the Administration's National Performance Review, NLM is pursuing its theme to reinvent the Library's information systems, to move to a more flexible, powerful, and maintainable computer system that will improve internal processing and provide innovative services to outside users.

The NLM System Reinvention project made progress in several areas in FY 1997:

New Retrieval System: After a careful evaluation of several software products, NLM selected the internally developed PubMed software to replace the ELHILL retrieval system which has provided access to MEDLINE and other NLM databases for many years. Although it will be a while before the ELHILL system is phased out, PubMed has already proven to be a remarkable success. Officially announced at the end of June 1997, after a 6-month public trial period, the Web-based PubMed was already handling close to 4 million MEDLINE searches a month by September. PubMed is at http://www.ncbi.nlm.nih.gov/PubMed/.

Internet Grateful Med(IGM): NLM's original Web-based front-end to MEDLINE and other databases continued to be enhanced during the year. IGM now provides access to eleven NLM databases as well as to the Loansome Doc feature for ordering copies of articles found in an IGM search.

Document Delivery System: A new system known as Relais, which is based on scanning technology and workflow management software, was installed and tailored to NLM's needs during the past year. Relais will streamline NLM's document delivery process, eliminating much of the current manual effort and making it much more efficient to transmit document images electronically. The system is scheduled to be fully operational in the fall of 1997.

Integrated Library System (ILS): Implementation of an ILS, originally scheduled to begin in FY 1997, was delayed because the selected ILS vendor failed to develop an acceptable product. NLM is now evaluating a different product that already is being used by a large number of libraries. If the results of this evaluation are positive, implementation of the new system will begin in early 1998.

Financial Resources

In FY 1997, the Library had a total appropriation of $150,376,000. Table 12 displays the FY 1997 budget authority plus reimbursements from other agencies, and the allocation of these resources by program activity.

TABLE 12

Financial Resources and Allocations, FY 1997
(dollars in thousands)

Budget Authority:
    Appropriation, NLM. $150,376
Plus: Reimbursements 7,794

TOTAL **$158,170**

Budget Allocation:
    Extramural Programs .... $ 34,533
    Intramural Programs 115,835
    Library Operations.... .... (62,243)
    Lister Hill National Center for ....
      Biomedical Communications .... (33,452)
    National Center for Biotechnology ....
      Information .... (12,252)
    Toxicology Information .... (7,888)
Research Management and Support 7,802

TOTAL **$158,170**

The 1997 appropriation language authorized the Library to use personal services contracts and provided for the availability of $4.0 million without fiscal year limitations. These authorities are key elements of NLM's system reinvention initiative.

Personnel

In January 1997, *Marjorie A. Cahn* was appointed to the position of Head, National Information Center on Health Services Research and Health Care Technology (NICHSR), within the Division of Library Operations. Ms. Cahn has an extensive background in the health services research field. She holds a B.A. from the University of Maryland and an M.A. in Health Policy/Planning from George Washington University. Ms. Cahn previously had served as Director of Administration for the NIDDK-funded U. S. Renal Data System where she administered the technical and

logical coordination for that program.

In January 1997, *Yo Matsuo, Ph.D.*, and *Erik Sonnhammer, Ph.D.* were appointed as Visiting Fellows with NCBI. Dr. Yo Matsuo received his Ph.D. in biophysics from Kyoto University. His thesis work was in the field of computational structural biology. While at the NCBI, Dr. Matsuo will conduct research on sequence-structure "threading" methods and on improvement of NCBI information retrieval services in the area of structural biology. His research goal is to construct an automated scanning system to allow large-scale comparison of new protein sequences against the known structure database. Dr. Erik Sonnhammer obtained his Ph.D. in bioinformatics from the Sanger Center, Hinxton, Cambridge, England. Prior to his work at Cambridge, he obtained a M.Sc. in biochemistry from the Royal Institute of Technology, Stockholm, Sweden. He has also obtained relevant postdoctoral experience as a visiting scientist at the European Molecular Biology Laboratory, Heidelberg, Germany and as a computer analyst at the Netherlands Cancer Institute, Amsterdam, Netherlands.

In March, *Yanli Wang, Ph.D.* transferred to NCBI as a Visiting Fellow from the Laboratory of Molecular Biology, National Cancer Institute. She joined NCI in August 1995. Dr. Wang received her Ph.D. in structural biology from Peking University, Beijing, China. While at the NCBI, Dr. Wang will conduct research in protein threading and fold recognition algorithms. She will concentrate on browser software that facilitates comparison of threading models with known members of a protein family.

Also in March, *Hughes Sicotte, Ph.D., Christopher Leonard, Ph.D.* and *Kim Pruitt, Ph.D.* were selected to receive NIH Postdoctoral Intramural Research Training Award Fellowships with the NCBI. Dr. Hughes Sicotte received his Ph.D. in Physics from Princeton University. Dr. Sicotte's research will include the continuing analysis of protein families by developing new algorithms for sequence analysis. Dr. Christopher Leonard received his Ph.D. in microbiology, immunology and pathology from Cornell University. While at the NCBI, Dr. Leonard will undertake an analysis of proteins involved in cell cycle control and apoptosis in search of common motifs within a given species, as well as comparative genomic studies of similar functions across various species. These studies could lead to the identification of as yet unknown proteins with roles in cell cycle control. Dr. Kim Pruitt received her Ph.D. in genetics from Cornell University. While at the NCBI, Dr. Pruitt's research will concentrate in the field of bioinformatics to increase her understanding of database organization and generation, and the current biological database analytic tools.

In April, *Hidemi Watanabe, Ph.D.* was appointed as an NCBI Visiting Fellow. Dr. Watanabe received his Ph.D. in biophysics from the Science University of Tokyo in 1995 and he was a Postdoctoral Research Fellow at the National Institute of Genetics, Mishima, Japan. At NCBI, Dr. Watanabe is working with Dr. Eugene Koonin studying problems of gene phylogeny using the available collection of complete genome sequences from bacteria, archaea, and eukaryotes.

In June, *Kenneth J. Addess, Ph.D.* was selected to receive an NIH Postdoctoral Intramural Research Training Award (IRTA) Fellowship with NCBI. Dr. Addess received his doctoral degree in biochemistry from the University of California, Los Angeles. Prior to joining NCBI he served as a postdoctoral fellow with the Department of Chemistry and Biochemistry of the University of Colorado. His research at NCBI will consist of improving the 3D structure visualization algorithms used in Entrez, NCBI's Internet retrieval software.

In July, *Qiuwei Xu, Ph.D.* received a Postdoctoral Intramural Research Training Award Fellowship from the NCBI. Dr. Xu received his Ph.D. in chemistry from the University of Nebraska-Lincoln. Before his appointment with NCBI, Dr. Xu served as a Research Associate at the University of Maryland. While at NCBI, Dr. Xu will develop a more efficient and reliable algorithm to predict protein conformation and improve 3D modeling software like Cn3D.

In August, the Office of Computer and Communications Systems (OCCS) filled two senior level positions: *Mr. Joseph W. Hutchins* was appointed Chief, Development Branch, OCCS, and *Mr. Roy A. Standing* was appointed as Chief, Information Management Branch, OCCS. Mr. Hutchins has been with NLM for 27 years, and Mr. Standing has been with NLM for eighteen years.

In August, *Lewis Geer, Ph.D.* received a Postdoctoral Intramural Research Training Award Fellowship from the NCBI. Dr. Geer received his Ph.D. in physics from the Washington University in St. Louis. Before his appointment with NCBI, Dr. Geer was a Program Manager at Microsoft Corp. While at NCBI, Dr. Geer will develop software for simultaneous viewing of multiple sequence and protein structure alignments.

In September, *Alexa McCray, Ph.D.* was appointed Director of the Lister Hill National Center for Biomedical Communications (LHNCBC). Dr. McCray received her masters and doctoral degrees in linguistics from Georgetown University. She joined the NLM in 1986 as a Special Expert and has served as the Chief of the Educational Technology Branch and the Cognitive Science Branch, LHNCBC, since 1986. She

is an expert in the field of informatics and in the development of communications technologies to improve access to biomedical information by health care professionals. Her unique educational and experiential backgrounds in fields relevant to the LHNCBC will allow her to provide the leadership needed in developing and executing state-of-the-art scientific and technical research and development programs in biomedical communications.

In September, *Julia Royall* joined the NLM as a Special Expert with the Office of Health Information Programs Development. Ms. Royall has 20 years of experience in various capacities involving telecommunications from content production management to devising new technology solutions for providing access to remote and underserved communities around the world. She served for five years as Deputy Director of SatelLife, a non-profit Cambridge, Massachusetts company dedicated to satellite delivery of public health and medical information in developing regions including the African countries of Uganda and Zambia. As a Special Expert with OHIPD, Ms. Royall will develop and carry out projects in support of NLM's international programs and outreach initiatives. She will plan, organize, conduct, and report on the results of efforts to enhance the communications and networking capabilities of African malaria scientists in the context of NLM's participation with NIH in the Multilateral Initiative on Malaria.

Also in September, *John Seachrist, Jr.,* was appointed as NLM's new Grants Management Officer. For the past 16 years, Mr. Seachrist was the Grants Management Specialist of the National Institute of Neurological Disorders and Stroke, NIH. As Grants Management Officer, NLM, Mr. Seachrist is responsible for the overall fiscal and administrative management of the Library's grants programs.

In September, *Anna Panchenko, Ph.D.* received a Postdoctoral Intramural Research Training Award Fellowship from the NCBI. Dr. Panchenko received her Ph.D. in biophysics from the Moscow State University. Before her appointment with NCBI, Dr. Panchenko served as a Postdoctoral Research Associate at the University of Illinois at Urbana-Champaign, School of Chemical Sciences. While at NCBI, Dr. Panchenko will conduct research on protein domain architecture and fold-recognition techniques as applied to protein domains.

In November, *Michael Galperin, Ph.D.*, was awarded an NIH Postdoctoral IRTA. Dr. Galperin received his Ph.D. in Microbiology from the Moscow State University. While with NCBI, Dr. Galperin will conduct research in computational molecular biology.

NCBI also welcomed a New Visiting Fellow

to its staff in November: *Andrei Gabrielian, Ph.D.* Dr. Gabrielian received his Ph.D. in chemistry from Moscow State University, Engelhardt Institute of Molecular Biology. He served as a postdoctoral fellow at the International Center for Genetic Engineering and Biotechnology in Trieste, Italy. Dr. Gabrielian's research with NCBI will involve the investigation of structure function relationships in protein DNA and protein complexes.

Retirements

In June 1997, *Ruth E. Bortz*, Grants Management Specialist, Division of Extramural Programs, retired from the Federal government with more than 30 years of service with the NLM. Ms. Bortz began her Federal career in 1963 and joined the NLM in September 1966.

In July 1997, *Roger W. Dahlen, Ph.D.*, retired from the Federal service after 26 years with the NLM. During Dr. Dahlen's tenure with the Library, he served as Chief, biomedical Information Support Branch, Division of Extramural Programs.

Awards

The President of the United States conferred on NLM Director *Donald A.B. Lindberg, M.D.,* the rank of Meritorious Executive in the Senior Executive Service for sustained superior accomplishment in management of programs of the United States government and for noteworthy achievement of quality and efficiency in the public service.

The Medical Library Association Board of Directors awarded *Dr. Lindberg* and *Mr. Kent A. Smith*, Deputy Director, NLM, the MLA's highest honor; the President's Award at the 97th Annual Meeting. The award recognizes exceptional contributions to the Medical Library Association and the health sciences information profession.

The NLM Board of Regents Award for Scholarship or Technical Achievement was awarded to *Frank L. Walker, M.S.E.E.*, for his design and development of DocView, an original engineering accomplishment and significant advance in facilitating access to biomedical literature over the Internet.

The Frank B. Rogers Award recognizes employees who have made significant contributions to the Library's fundamental operational programs and services. The recipient of the 1997 award was *Lou Wave Knecht,* Division of Library Operations, for her notable intellectual contributions to improving the systems by which NLM creates and maintains the quality of the MEDLARS databases.

The NIH Director's Award was presented to

Alexa T. McCray, Ph.D., for her innovative research in medical language processing and leadership in conceptualization of the Unified Medical Language System's Semantic Network and Knowledge Sources Server.

The NLM Director's Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to *Eve Marie Lacroix,* for her outstanding leadership and initiative in using information technology to increase the speed of NLM's document delivery service, to enhance access to information about NLM's programs and services, and to improve communication among the Library's employees.

The NIH Merit Award was presented to six employees: *Susan U. Levine* and *Donald C. Poppke,* Office of Administration; *Margaret M. Byrnes, Marjorie A. Cahn* and *Mina Smith*, Division of Library Operations; and *William T. Hole, M.D.,* Lister Hill National Center for Biomedical Communications. Ms. Levine was recognized for her successful leadership as Chief Budget Officer for the NLM and her consistently excellent work in support of the Library's financial management activities. Mr. Poppke was recognized for his exceptional leadership in the enhancement and streamlining of the Library's administrative services and his strong commitment to customer-responsive administrative support. Ms. Byrnes was recognized for her exceptional leadership and achievement in developing and managing the Preservation and Collection Management Program at NLM. Ms. Cahn was recognized for establishing and leading NLM's Health Services Research Information Program. Ms. Smith was recognized for her contributions to the accurate and timely production of the MEDLINE database. Dr. Hole was recognized for exemplary creativity, persistence and sustained achievement in the development, enhancement and annual release of the Unified Medical Language System Metathesaurus.

The PHS Commissioned Corps Meritorious Service Medal was awarded to *David J. Lipman, M.D.* for exceptional skills in directing the National Center for Biotechnology Information in developing innovative solutions for meeting the needs of the biological research community.

The NIH Quality of Work Life Award was presented to the following two individuals and team: *Lois Ann Colaianni,* for her outstanding and creative efforts to offer maximum workplace flexibility to division staff and thus improve quality of work life for employees; *Dr. Nancy Wright,* for actively and successfully fostering participation of her staff in the Flexiplace Program, which allows employees to work at home at least one day a week; and to the NLM Intranet Development Team consisting of *Eve-Marie*

Lacroix, Joyce Backus, Joseph Pagano, and Naomi Miller for their group efforts and the positive impact they have had on their fellow employees' work life and work environment.

The NIH Purchasing Award was presented to *Uyen H. Phuong* for demonstrating resourcefulness in resolving problems affecting the small purchases activities of the National Library of Medicine.

Table 13

Staff, FY 1997 Full-Time Equivalents

Program	*Full-Time Permanent*	*Other*
Office of the Director	4	5
Office of Public Information	6	2
Office of Administration.	50	7
Office of Computer and Communications Systems	54	8
Extramural Programs	13	4
Lister Hill National Center for Biomedical Communications	68	8
National Center for Bio-technology Information	25	11
Specialized Information Services	26	3
Library Operations	236	36
TOTAL	**492**	**84**
TOTAL FTE's	**576**	

NLM Diversity Council

Cassandra Allen, Chair
Public Services Division

This year, the NLM Diversity Council received training that better defined the concept of diversity for members of the Council. The training included a team building component that has greatly assisted the Council in functioning as a cohesive body. After the training session, the Council proceeded with the task of developing its infrastructure and developing goals. The Council also accomplished the following:

- Prepared documents defining the roles of Council officers and members; establishing governance rules; and identifying preliminary goals;
- Established working subcommittees;
- Provided feedback to the NIH Office of Equal Opportunity for the development of the Diversity Catalysts' Implementation Manual;
- Participated in the NIH Diversity Catalyst

meetings;

- Initiated development of a diversity home page on the NLM Intranet;
- Prepared a bibliography of diversity related articles, monographs, videos and web sites; and
- Provided diversity training for NLM staff

Office of Equal Employment Opportunity

David Nash
NLM EEO Manager

The major area of EEO emphasis in FY 1997 was the employment of staff with disabilities, the working conditions for these employees, and the Library's accommodations for patrons with physical disabilities. Among the advances this year:

- The NLM web site was enhanced with an updated "accessibility at the National Library of Medicine" page, which details the parking, entrance ramp, rest facilities, reading machine, speech magnification software, speech synthesizer, enlarged keyboard lettering, print enlarger, and wide stack aisles available for patrons with various physical disabilities [www.nlm.nih.gov/services/onsite/ada.html].

- The NLM EEO Manager was involved with both the Federal Employment of the Hearing Impaired Conference and the President's Committee on Employment of People with Disabilities.

- The NLM Executive Officer served on an NIH ad hoc committee designed to provide interpretive services to hearing impaired employees. NLM has arranged for Library programs to use, when needed, the services of a sign interpreter.

- The Library has ordered a TTY phone for patrons and employees in the public area of Building 38.

- To better serve patrons who are mobile via wheelchairs, push pads have been installed for entrance to the rest rooms.

- Tours of NLM were conducted to give hearing impaired patrons a demonstration of NLM's capabilities. NLM is now known as a leader in the NIH community in its commitment to patrons with disabilities.

Other EEO Areas

The NLM EEO Manager undertakes continuing education courses to keep abreast of developments in EEO and thus be able to provide professional advice to the Director, managers, supervisors, and NLM employees on matters of disability and race issues. Also, he has made special efforts to extend NLM outreach into the various communities and organizations that frequently have access to such qualified minority individuals. For example:

- The NLM has a continuing special outreach program with Largo Senior High School, focusing on career enhancement and opportunities in medical biotechnology. The staff of NLM's National Center for Biotechnology Information are of great help in maintaining this program.

- There are special efforts to reach such populations as the greater Prince George's County area (a predominantly African American community). The Library exhibited at a job fair on career opportunities sponsored by Representative A. R. Wynn (D-MD). In addition, the EEO Manager was a featured speaker at Congressman Wynn's Education and Career Enhancement Conference, where he had discussions with the governor of Maryland about NLM outreach and employment goals.

- To address the underrepresentation of Hispanics in the NLM/NIH work force, NLM exhibited and provided information at the La Raza Conference on Hispanic health issues and also worked with the Hispanic Medical Association.
- The NLM EEO Manager attended, exhibited at, and addressed the Association of American Indian Physicians. As a result, NLM has established a continuing dialog with this group representing a key underserved population.

- The Library has continuing interaction with such important organizations as the National Association of Health Services Executives, the National Medical Association, the Congressional Black Caucus, and the National Association for the Advancement of Colored People.

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

1. **MIDDLE ATLANTIC REGION**
   The New York Academy of Medicine
   1216 Fifth Avenue
   New York, NY 10029-5283
   (212) 822-7396 FAX (212) 534-7042
   States served: DE, NJ, NY, PA
   URL: http://www.nnlm.nih.gov/mar

2. **SOUTHEASTERN/ATLANTIC REGION**
   University of Maryland at Baltimore
   Health Science and Human Services Library
   601 Lombard Street
   Baltimore, MD 21201
   (410) 706-2855 FAX (410) 706-0099
   States served: AL, FL, GA, MD, MS, NC,
   SC, TN, VA, WV, DC, VI, PR
   URL: http://www.nnlm.nih.gov/sar

3. **GREATER MIDWEST REGION**
   University of Illinois at Chicago
   Library of the Health Sciences (M/C 763)
   1750 West Polk Street
   Chicago, IL 60612-7223
   (312) 996-2464 FAX (312) 996-2226
   States served: IA, IL, IN, KY, MI, MN,
   ND, OH, SD, WI
   URL: http://www.nnlm.nih.gov/gmr

4. **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
   URL: http://www.nnlm.nih.gov/mr

5. **SOUTH CENTRAL REGION**
   Houston Academy of Medicine-
   Texas Medical Center Library
   1133 M.D. Anderson Boulevard
   Houston, TX 77030-2809
   (713) 799-7880 FAX (713) 790-7030
   States served: AR, LA, NM, OK, TX
   URL: http://www.nnlm.nih.gov/scr

6. **PACIFIC NORTHWEST REGION**
   University of Washington
   Regional Medical Library, HSLIC
   Box 357155
   Seattle, WA 98195-7155
   (206) 543-8262 FAX (206) 543-2469
   States served: AK, ID, MT, OR, WA
   URL: http://www.nnlm.nih.gov/pnr

7. **PACIFIC SOUTHWEST REGION**
   University of California, Los Angeles
   Louise M. Darling Biomedical Library
   Box 951798
   Los Angeles, CA 90024-1798
   (310) 825-1200 FAX (310) 825-5389
   States served: AZ, CA, HI, NV and U.S.
   Territories in the Pacific Basin
   URL: http://www.nnlm.nih.gov/psr

8. **NEW ENGLAND REGION**
   University of Connecticut Health Center
   Lyman Maynard Stowe Library
   263 Farmington Avenue
   Farmington, CT 06030-5370
   (860) 679-4500 FAX (860) 679-1305
   States served: CT, MA, ME, NH, RI, VT
   URL: http://www.nnlm.nih.gov/ner

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:

DeBAKEY, Michael E., M.D. (Chair)
Chancellor Emeritus
Baylor College of Medicine
Houston, TX

ALBRIGHT, Tenley E., M.D.
Two Commonwealth Ave.
Boston, MA

BALL, Marion, Ed.D.
Vice President for Information Sciences
First Consulting Group
Baltimore, MD

BARUCH, Jordan, Sc.D.
President, Jordan Baruch Associates
Washington, D.C.

BOND, Enriqueta, M.D.
President, Burroughs Wellcome Fund
Durham, NC

FONSECA, Raymond J., D.M.D.
Dean, School of Dental Medicine
University of Pennsylvania
Philadelphia, PA

FULLER, Sherrilynne, Ph.D.
Acting Director, Informatics
University of Washington School of Medicine
Seattle, WA

GAGE, John
Director, Science Office
Sun Microsystems Computer Corp.
Palo Alto, CA

KLEIN, Michele
Systems Director, Library Services
Children's Hospital of Michigan
Detroit, MI

NOLAN, George H., MD
Director, Department of Obstetrics and Gynecology
Henry Ford Hospital
Detroit, MI

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:

BUCHANAN, Bruce G., Ph.D. (Chair)
Professor of Computer Science
University of Pittsburgh
Pittsburgh, PA

CLAYTON, Paul D., M.D.
Professor of Medical Informatics
Columbia Presbyterian Medical Center
New York, NY

JAFFE, Conrade C., M.D.
Director, Center for Advanced Instructional Media
Yale University School of Medicine
New Haven CT

KAHN, Michael G., M.D., Ph.D.
Rodeer Systems, Inc.
Broomfield, CO

MASYS, Daniel R., M.D.
Director of Biomedical Informatics
School of Medicine
University of California
San Diego, CA

MITRA, Sunanda, Ph.D.
Professor of Electrical Engineering
Texas Tech University
Lubbock, TX

WILKERSON, LuAnn, Ed.D.
Director, Center for Educational Development
UCLA School of Medicine
Los Angeles, CA

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:

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

BUETOW, Kenneth H., Ph.D.
Fox Chase Cancer Center
Philadelphia, PA

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

PACE, Norman R., Ph.D.
Department of Plant and Microbial Biology
University of California
Berkeley, CA

SCHLICK, Tamar, Ph.D.
Professor
Chemistry Department
New York University
New York, NY

WILLIAMS, Myra N., Ph.D.
Princeton, NJ

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:

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

ALLMAN, Robert M., M.D.
Professor of Radiology
Univ. of Maryland School of Medicine
Baltimore, MD

ASH, Joan S. Ph.D.
Associate Professor
Library and Medical Informatics
Oregon Health Sciences University
Portland, OR

BASLER, Thomas G., Ph.D.
Professor and Director
Libraries and Learning Resource Center
Medical University of South Carolina
Charleston, SC

BROADNAX, Lavonda
Automation Operations Coordinator
Library of Congress
Washington, DC

CHUEH, Henry C., M.D.
Associate Director
Laboratory of Computer Science
Massachusetts General Hospital
Boston, MA

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

FLORANCE, Valerie, Ph.D.
Director, Edward G. Miner Library
University of Rochester
Rochester, NY

FRIEDMAN, Richard B., M.D.
Medical Director
Waianae Comprehensive Health Center
Waianae, HI

FUCHS, Rainer T., Ph.D.
Vice President and CIO
ARIAD Pharmaceuticals
Cambridge, MA

HOLST, Ruth
Director, Library Services
Columbia Hospital
Milwaukee, WI

MODEL, Peter, Ph.D.
Professor of Biochemistry
Rockefeller University
New York, NY

MOLHOLT, Pat A.
Assistant Vice President
Columbia University Health Sciences
New York, NY

MOULT, John, Ph.F.
Professor
Center for Advanced Research in Biotechnology
Rockville, MD

ORTHNER, Helmuth, Ph.D.
Professor of Medical Informatics
University of Utah Health Sciences Center
Salt Lake City, UT

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

RINDFLEISCH, Thomas
Director, Lane Medical Library
Director, Center for Advanced Medical Informatics

Stanford University Medical Center
Stanford, CA

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

SEARLS, David B., Ph.D.
Group Director of Bioinformatics
SmithKline Beecham Pharmaceuticals
King of Prussia, PA

TANG, Paul C., M.D.
Associate Professor of Medicine
Medical Director, Information Systems
Northwestern Memorial Hospital
Chicago, IL

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:

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

CABELLO, Felipe C., M.D.
Dept. of Microbiology and Immunology
New York Medical College
New York, NY

CLEVER, Linda Hawes, M.D.
Chair, Dept. of Occupational Health
California Pacific Medical Center
San Francisco, CA 94120

COLLEN, Morris F., M.D.
Director Emeritus
Kaiser Permanente Division of Research
Oakland, CA

EPSTEIN, Neal, M.D.
NHLBI Cardiology Branch
National Institutes of Health
Bethesda, MD

LI, Yihong, Ph.D.
Assistant Professor
Oral Biology Department
University of Alabama
Birmingham, AL

MAKINEN, Ruth H.
Head, Technical Services
University of Minnesota
Bio-Medical Library
Minneapolis, MN

MATHIEU, Alix, M.D.
Professor of Anesthesia
University of Cincinnati
College of Medicine
Cincinnati, OH

O'DONNELL, Anne Elizabeth, M.D.
Assistant Professor
Department of Medicine
Georgetown University
Washington, D.C.

ROLETT, Ellis L., M.D.
Professor of Medicine
Dartmouth-Hitchcock Medical Center
Lebanon, NH

STRICKLAND, Ora Lea, Ph.D.
Professor, School of Nursing
Emory University
Atlanta, GA

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

Appendix 7:   Extramural Programs Grants—FY 1997

Awards listed by Program Area. Those preceded by an '*' are first year, or competing awards, all others are continuation of awards made in prior years.

Fellowships

Medical Informatics Research Fellowships

Carter, Alexandre R., $35,316
5 F31 LM00040-04
Minority Predoctoral Fellowship Program – NIGMS
National Institutes of Health

Kogelnik, Andreas M., $15,996
5 F37 LM00043-03
Mitochondria Genetics Information System
Emory University, Atlanta, GA

Soo Hoo, Kent, Jr., $19,057
5 F37 LM00044-02
Content Based Indexing for Medical Image Management
University of California, San Francisco

*Norris, Patrick R., $25,172
1 F37 LM00053-01A1
Therapeutic Goals in Intelligent ICU Monitoring
Vanderbilt University, Nashville, TN

*Brossette, Stephen E., $20,996
1 F37 LM00057-01
Knowledge Discovery/Data Mining in Epidemiology Surveys
University of Alabama, Birmingham

*Schoeflelr, Katherine M., $34,008
1 F37 LM00062-01
Measurable Attributes for Controlled Terminologies
Duke University, Durham, NC

Applied Medical Informatics Fellowships

Bradley, Johanna R., $58,600
5 F38 LM00045-02
Creation, Access and Use of Networked Health Information
National Library of Medicine

*Nigrink, Daniel J., $71,600
1 F38 LM00055-01A1
Database Independent Analysis of Endocrine Tests

Children's Hospital, Boston, MA

*Steinberger, Eileen K., $52,500
1 F38 LM00061-01
Expert System in the Diagnosis of Lymphoma and Leukemia
University of Maryland
Baltimore Professional School

**Integrated Advanced Information Management Systems
(IAIMS) Grants**

Phase I

Rubenstein, Arthur H., $75,000
3 G08 LM05657-01S1
IAIMS Planning at the University of Chicago
University of Chicago

Niland, Joyce C., $145,243
5 G08 LM05697-02
IAIMS Phase I Planning Grant for City of Hope - CRIIS
City of Hope National Medical Center
Duarte, CA

Ward, Richard E., $75,000
3 G08 LM05720-01S1
Planning a Comprehensive IAIMS Capable of Supporting CQI
Case Western Reserve University
Henry Ford Hospital, Detroit, MI

Frisse, Mark E., $75,000
3 G08 LM06002-01S1
IAIMS Planning at Washington University Medical School
Washington University, St. Louis, MO

Sokol, Robert J., $149,864
5 G08 LM06009-02
Enhancing Healthcare Effectiveness - IAIMS in Detroit
Wayne State University, Detroit, MI

Larson, Paul F., $149,922

5 G08 LM06230-02
IAIMS Planning
University of Medicine & Dentistry of New Jersey,
Newark

Flanagan, James, $75,000
3 G08 LM06234-01S1
University of Iowa Health Sciences Center IAIMS
Planning
University of Iowa, Iowa City

Gardner, Reed M., $50,000
3 G08 LM06237-01S1
IAIMS Planning at IHC
LDS Hospital, Salt Lake City, UT

Phase II

Mitchell, Joyce A., $550,000
5 G08 LM05415-04
Missouri IAIMS
University of Missouri, Columbia

Stead, William W., $550,000
5 G08 LM05443-05
Fast Track Provision of IAIMS
Vanderbilt University, Nashville, TN

Paton, John A., $545,430
5 G08 LM05583-03
IAIMS Implementation at Yale
Yale University, New Haven, CT

Fuller, Sherrilynne, $549,995
5 G08 LM05620-04
Creating the UWHSC Integrated Information
Infrastructure
University of Washington, Seattle

Internet Connection Grants

*McCue, JacK D., $47,805
1 G08 LM06401-01A1
Internet Access for Berkshire Health Systems
Berkshire Medical Center
Pittsfield, MA

*Backer, M. Barbara, $21,200
1 G08 LM06414-01A1
Provide Internet Access to La Porte Hospital and
Clinics
La Porte, IN

*Rich, E. Scott, $23,193
1 G08 LM06426-01A1

Extended LAN for Cope
Washington University, St. Louis, MO

*Heintz, Alexandra H., $30,000
1 G08 LM06564-01
Putnam Memorial Health Corp Internet Capacity
Bennington, VT

*Gottlieb, Mark S., $26,051
1 G08 LM06579-01
Inter-Site Internet Networking—Family Medicine
Medical College of Wisconsin, Milwaukee

*McDuffee, Diana C., $49,987
1 G08 LM06609-01
Community Clinics Internet Connection
University of North Carolina, Chapel Hill

*Oliphant, Gregory J., $27,265
1 G08 LM06610-01
Internet Connection for Ephrata Community Hospital
Ephrata, PA

*Montgomery, Louise L., $24,756
1 G08 LM06612-01
Baptist Internet Healthcare Links, BIHL
Baptist Health
Little Rock, AR

*Weber, Kathryn H., $29,025
1 G08 LM06620-01
Northeast Oregon Medical Internet Connection
Project
Grande Ronde Hospital
La Grande, OR

*Weiss, Nancy S., $29,521
1 G08 LM06623-01
Burke Internet Connection
M. Burke Rehabilitation Hospital
White Plains, NY

*Shanklin, Russell W., $29,924
1 G08 LM06626-01
Pikesville Methodist Hospital Internet ConnectioN
Pikesville, KY

*Sharrott, Lawrence H., $28,500
1 G08 LM06628-01
Internet Access for Atlanticare Health System
Egg Harbor Township, NJ

*Nussbaum, Noel, $42,164
1 G08 LM06632-01
A Regional Network for Medical Education

Wright State University, Dayton, OH

*Livering, Earl D., $26,627
1 G08 LM06634-01
Internet Connection for Hope Hospice
Hope of Lee County, Inc.
Fort Myers, FL

*Page, Jeffery C., $50,000
1 G08 LM06637-01
CCPM Satellite Location Internet Connection
California College of Podiatric Medicine
San Francisco, CA

*Sanddal, Nels D., $46,070
1 G08 LM06644-01
Tenkids Internet Access
Critical Illness/Trauma Foundation, Inc.
Bozeman, MT

Resource Information Systems Grants

Seltzer, Ada M., $135,083
5 G08 LM05478-04
Health Sciences Information Network
University of Mississippi Medical Center
Jackson, MS

Hagen, Michael D., $59,335
5 G08 LM05614-03
Information Systems Grant
University of Kentucky, Lexington

Dennis, Sharon E., $48,794
5 G08 LM05684-03
Model Multimedia Support Center for the Health
Sciences
University of Utah, Salt Lake City

Scott, Ursula D., $150,782
5 G08 LM05771-02
Baylorlink - An NLM Information System Grant
Baylor Research Institute
Dallas, TX

Eaton, Elizabeth K., $148,677
1 G08 LM05774-01A2
Creating and Managing Digital Multimedia
Resources
Tufts University, Boston, MA

Messerle, Judith R., $118,456
5 G08 LM05791-03
Creation of a Knowledge Laboratory
Harvard University, Boston, MA

De Bord, William J., $101,678
5 G08 LM06000-02
Health Sciences Information Data Link
Morehead State University
Morehead, KY

Jones, Michael O., $129,650
5 G08 LM06218-02
Computerized Archive of American Traditional
Medicine
University of California, Los Angeles

Lehman, Harold P., $86,085
5 G08 LM06232-02
Internet Based Image Base for Medical Education
Johns Hopkins University, Baltimore, MD

*Mueller, Christine A., $142,902
1 G08 LM06239-012A1
Health Information Access for Rural Nurse
Practitioners
University of Minnesota Twin Cities, Minneapolis

*Matthew, Suzanne L., $190,450
1 G08 LM06492-01
Northwoods Healthnet - A Northern WI Information
System
Northern Wisconsin AHEC, Inc., Wausau, WI

Resource Information Access Grants

Perry, Richard R., $64,778
3 G07 LM05447-02S1A1
Rural health projects, Inc., Med-fax network
Rural Health Projects, Inc., Enid, OK

Reynolds, Pamela M., $99,975
3 G07 LM05474-01S3
Sowega Information Delivery Electronic Access/Idea
Sites
Southwest Georgia Area Health Education Center,
Albany

*Ireland, Robert B., $9,250
1 G07 LM06380-01A1
Internet Access and Training
University of Alabama, Tuscaloosa

*Fredericks, Treva A., $12,000
1 G07 LM06469-01
Lanterman Developmental Center Medical Library
Online
Lanterman State Hospital and Develop Center
Pomona, CA

*Hogan, Linda, $12,000
1 G07 LM06474-01
Electronic Linkage—Hospital Staff to Library
Resources
Mercy Hospital of Pittsburgh, PA

*Coleridge, Samuel T., $61,208
1 G07 LM06487-01
Improve Information Access in Family Practice
Residency
University of North Texas Health Science Center,
Fort Worth

*Bowling, John R., $130,519
1 G07 LM06535-01
Information Access for Preceptors and Medical
Students
University of North Texas Health Science Center,
Fort Worth

Publication Grants

Hast, Malcolm H., $36,821
5 R01 LM05657-03
Annotated Translation of Vesalius' Fabrica
Northwestern University, Chicago

Faden, RUTH R., $33,000
5 R01 LM05700-03
Ethics and Allocation—Foundations of American
Health Care
Johns Hopkins University, Baltimore, MD

Funk, Carla J., $150
5 R01 LM05906-03
History of the Medical Library Association
Medical Library Association, Chicago

*Lunbeck, Elizabeth, $33,951
1 R01 LM05934-01A1
Practicing Psychoanalysis in the U.S., 1946-1970
Princeton University, Princeton, NJ

*Meldrum, Marcia L., $16,496
1 R01 LM05983-01A1
Randomized Clinical Trials, 1946-1970
Individual Award
Los Angeles, CA

*Bonner, Thomas N., $23,175
1 R01 LM06262-01
Biography of Abraham Flexner (1866-1959)
Wayne State University, Detroit

*Murphy, Gregory J., $39,051

1 R01 LM06265-01
Electronic Freud Textbase Prototype
Rutgers the State University
New Brunswick, NJ

*Buhler-Wilkerson, Karen A., $39,750
1 R01 LM06304-01
Endless Obligation—Home Care in America
1880-1930
University of Pennsylvania, Philadelphia

*Proctor, Robert A., $35,356
1 R01 LM06528-01
Nazi Cancer Research and Policy—A History
Pennsylvania State University
University Park, PA

Biotechnology Research Grants

*Herzenberg, Leonore A., $220,818
2 R01 LM04836-07
FACS-PENGUIN—Knowledge Base Support for
Flow Cytometry
Stanford University, Stanford, CA

Myers, Eugene W., Jr., $157,056
5 R01 LM04960-09
Efficient Software for the Analysis of Biosequences
University of Arizona, Tucson

Pearson, William R., $226,265
5 R01 LM04969-10
Comparison of Protein Sequences and Structures
University of Virginia, Charlottesville

Miller, Webb C., $382,981
5 R01 LM05110-09
Software for Analyzing Biosequence Data
Pennsylvania State University
University Park, PA

Smith, Temple F., $952,184
5 P41 LM05205-14
Biomolecular Engineering Research Center
Boston University

Brutlag, Douglas L., $214,055
5 R01 LM05716-04
Multiple Representations of Biological Sequences
Stanford University, CA

Roberts, Richard J., $150,618
5 P41 LM05800-03
REBASE - The Restriction Enzyme Database
New England Biolabs, INC.

Beverly, MA

Altman, Russ B., $325,902
5 R01 LM06244-02
Representing Biological Data for Molecular
Modeling
Stanford University, CA

Medical Informatics Research Grants

*Ellisman, Mark H., $9,550
1 R01 DC03192-01A1
Development of a3D Cell-Centered Neuronal
Database
University of California, San Diego

Friedman, Charles P., $190,378
5 R01 LM04843-11
Information and Cognition in Medical Education
University of Pittsburgh

Jaffe, Carl C., $225,982
5 R01 LM05007-08
Indexing Image Databases for Motion Similarity
Retrieval
Yale University, New Haven, CT

*Berner, Eta S., $131,845
2 R01 LM05125-06
Physician Use of Diagnostic Decision Support
System DatA
University of Alabama, Birmingham

Jelliffe, Roger W., $443,755
5 R01 LM05401-06
New Decision Supports and Databases for Drug
Dosage
University of Southern California, Los Angeles

Chute, Christopher G., $172,085
5 R01 LM05416-05
Latent Semantic Indexing in Support of Data
Retrieval
Mayo Foundation, Rochester, MN

Begg, Colin B., $92,540
5 R01 LM05524-05
Meta Analysis Methodology for Diagnostic Test
Evaluation
Solan-Kettering Institute for Cancer Research
New York, NY

Widman, Lawrence E., $179,894
5 R01 LM05530-07
Knowledge Based Interpretation of Cardiac

Arrhythmias
University of Texas Health Science Center, San
Antonio

Selker, Harry P., $478,143
5 R01 LM05607-03
New Mathematical Models for Medical Events
New England Medical Center
Boston, MA

Elstein, Arthur S., $93,166
3 R01 LM05630-04
Effect of Decision Support Systems On Clinical
Reasoning
University of Illinois, Chicago

Toga, Arthur W., $217,093
5 R01 LM05639-03
Digital Representation and Visualization of Human
Brain
University of California, Los Angeles

Musen, Mark A., $507,992
5 R01 LM057087-03
Software Architecture for Guideline Directed
Therapy
Stanford University, CA

Goldbaum, Michael H., $265,260
5 R01 LM05759-09
Structured Analysis of the Retina
University of California, San Diego

Webber, Bonnie L., $236,000
5 R01 LM05764-03
Effective Information Delivery to Clinical Personnel
University of Pennsylvania, Philadelphia

Lang, Walter P., $97,410
5 R01 LM05917-02
MEDLINE and Computer Conferencing by Dentists
University of Michigan, Ann Arbor

Sahni, Sartaj K., $267,560
5 R01 LM05944-03
Algorithms for Compression and Registration of
Brain MRI
University of Florida, Gainesville

*Marshall, Bryan E., $240,944
1 R01 LM05997-01A2
Information Integration and Virtual Therapy in the
SICU
University of Pennsylvania, Philadelphia

*Miller, Randolph A., $344,940
1 R01 LM06226-01A1
Patient Care Provider Order Entry with Tactical Support
Vanderbilt University, Nashville, TN

Good, Walter F., $161,287
7 R01 LM06236-02
Non ROC Measures for Evaluating Image Compression
Allegheny University of Health Sciences
Philadelphia, PA

Bowden, Douglas M., $171,051
5 R01 LM06243-02
Spatial/Symbolic Brain Information Management System
University of Washington, Seattle

*Brennan, Patricia F., $189,000
1 R01 LM06249-01A2
Customized Computer Support - Home Care of CABG Patients
University of Wisconsin, Madison

*Huang, H. K., $296,031
1 R01 LM06270-01
Digital Hand Atlas in Assessment of Skeletal Development
University of California, San Francisco

*Friedman, Carol, $218,581
1 R01 LM06274-01A1
Unlocking Data from Medical Records with Text Processing
Queens College, New York, NY

*Hersh, William, $183,419
1 R01 LM06311-01A1
New Model and Approach to Retrieval System Evaluation
Oregon Health Sciences University, Portland

*Brinkley, James F., $298,924
1 R01 LM06316-01
Structure Based Visual Access to Biomedical Information
University of Washington, Seattle

*Grobe, Susan J., $172,081
1 R01 LM06325-01
Automated Analysis of Intervention Narrative
University of Texas, Austin

*Luther, Paul M., $128,043

1 R01 LM06326-01A1
Three Dimensional Reconstruction of Synapses
University of Maryland
Baltimore Professional School

*Goodwin, Linda K., $279,649
1 R01 LM06488-01
Informatics Tools and Medical Perinatal Knowledge Building
Duke University, Durham, NC

*Ohno-Machado, Lucila, $50,000
1 R55 LM06538-01
Component Based Tools for Connectionist Classification
Beth Israel Deaconess Medical Center
Boston, MA

*Haug, Peter J., $189,816
1 R01 LM06539-01
Semantic Parser for Medical Free Text
LSD Hospital, Salt Lake City, UT

Letovsky, Stanley I., $87,366
5 R01 AG13743-03
Spatially, Oriented Database for Digital Brain Images
Johns Hopkins University, Baltimore, MD

Medical Informatics FIRST Awards

Cooper, Gregory F., $102,782
5 R29 LM05291-05
Structuring Medical Knowledge—Probabilistic Inference
University of Pittsburgh

Balas, E. Andrew, $34,195
5 R01 LM05545-03
Meta Analysis of Clinical Information Service Trials
University of Missouri, Columbia

*Shiffman, Richard N., $62,831
1 R29 LM05552-01A1
Yale University, New Haven, CT

Lenert, Leslie A., $112,700
5 R29 LM05626-04
Computer Interpretation of Free-Text Data
Stanford University, CA

Hripcsak, George, $113,161
5 R29 LM0527-04
Linking Knowledge-Based Systems to Clinical Databases
Columbia University, New York, NY

Lehmann, Harold P., $119,705
5 R29 LM05647-04
Formalizing the Notion of Clinical Significance
Johns Hopkins University, Baltimore, MD

Gorman, Paul Northrop, $100,263
5 R29 LM05663-03
Assessment of Information Seeking in Primary Care
Oregon Health Sciences University, Portland

Yang, Yiming, $95,259
5 R29 LM05714-03
LLSF Mapping for Indexing and Retrieval of
MEDLINE
Carnegie-Mellon University
Pittsburgh, PA

Johnson, Stephen B., $112,172
5 R29 LM05783-03
Access to Medical Information Through Natural
Language
Columbia University, New York, NY

Tong, David A., $83,300
7 R29 LM06004-02
Model Based Interpretation of Intracardiac
Electrograms
University of Texas Health Science Center
San Antonio, TX

*Wagner, Michael M., $95,423
1 R29 LM06233-01A1
Belief Network Based Reminder Systems That Learn
University of Pittsburgh

Rutledge, Geoffrey W., $122,489
5 R29 LM06235-02
Advanced Computer Methods for ICU Care
Beth Israel Deaconess Medical Center
Boston, MA

Langlotz, Curtis P., $106,335
5 R29 LM06238-02
Computer-Based Explanation Methods for Decision
Models
University of Pennsylvania, Philadelphia

Shahar, Yuval, $106,135
5 R29 LM06245-02
Knowledge Based Temporal Abstraction of Clinical
Data
Stanford University, CA

*Rolland, Jannick P., $100,037
1 R29 LM06322-01A1

3D Dynamic Anatomy—A Virtual Reality Prototype
University of Central Florida, Orlando

*Brummer, Marijn, $155,512
1 R29 LM06486-01
Interactive 4D Visualization of Congenital Heart
Disease
Emory University, Atlanta, GA

Biotechnology FIRST Awards

Karp Peter D., $113,809
5 R29 LM05413-05
Biological Knowledge-Base Management System
SKI International
Menlo Park, CA

Wu, Cathy H., $98,580
5 R29 LM05524-05
Classification Neural Networks for Genome Research
University of Texas Health Center, Tyler

Altman, Russ B., $107,963
5 R29 LM05652-04
Modeling and Computing with Uncertain Structures
Stanford University, CA

Goldstein, Richard A., $87,255
5 R29 LM05770-03
Computational Approaches to Protein Structure
Prediction
University of Michigan, Ann Arbor

Research Training in Medical Informatics

Shortliffe, Edward H., $642,509
2 T15 LM07033-14
Graduate Training in Medical Information Sciences
Stanford University, CA

Gatewood, Lael C., $414,129
2 T15 LM07041-14
Research Training in Medical Informatics
University of Minnesota , Twin Cities
Minneapolis, MN

*Miller, Perry L., $331,436
2 T15 LM07056-11
Medical Informatics Research Training at Yale
Yale University, New Haven, CT

*Friedman, Charles P., $370,249
2 T1507059-11
Pittsburgh Medical Informatics Training Program
University of Pittsburgh

*Downs, Stephen M., $488,219
2 T15 LM07071-06
Duke - UNC Training Program in Medical Informatics
University of North Carolina, Chapel Hill

*Clayton, Paul D., $517,778
2 T15 LM07079-06
Medical Informatics Research Training at Columbia
Columbia University
New York, NY

*Spackman, Kent, $169,192
2 T15 LM07088-06
Training Program in Health Informatics
Oregon Health Sciences University, Portland

*Mitchell, Joyce A., $365,215
2 T15 LM07089-06
Medical Informatics Research Training Program
University of Missouri, Columbia

*Greenes, Robert A., $1,007,939
2 T15 LM07092-06
Harvard-MIT-NEMC Research Training in Health Informatics
Harvard University, Boston, MA

*Gorry, G. Anthony, $415,451
2 T15 LM07093-06
Training Program in Computational Biology and Medicine
Rice University, Houston, TX

*Tierney, William M., $129,718
1 T15 LM07117-01
Regenstrief Medical Informatics Research Fellowships
Indiana University
Purdue University at Indianapolis

*Gardner, Reed M., $263,715
1 T15 LM07124-01
University of Utah Medical Informatics Training
University of Utah, Salt Lake City

Small Business Innovative Research (SBIR) Small Business Technology Transfer Grants (STTR)

*Marcotte, Thomas D., $99,830
1 R41 MH57593-01
Development of Simulations to Detect Impaired Drivers
Systems Technology, Inc.

Hawthorne, CA

*Delaney, Gerald T., $100,000
1 R43 LM06295-01A1
Web-Based Multimedia Access to Radiologic Information
Sudbury System, Inc.
Sudbury, MA

*Frawley, Sandra J., $99,829
1 R43 LM06330-01
Linking Web-Based Retrieval to An Online Patient Record
Medical Decision Associates, Inc.
New Haven, CT

*Marchisio, Giovanni, $99,761
1 R43 LM06520-01
Bayesian Textual and Multimedia Information Retrieval
Mathsoft, Inc. ,Seattle, WA

*Myers, Glenn A., $100,000
1 R43 LM06522-01
Customizable AAC System Using Context Information
Adaptive Computer Systems, Inc.
Iowa City, IA

Medical Informatics Resource Grants

Shortliffe, Edward H., $238,169
3 P41 LM05305-05S2
Center for Advanced Medical Informatics
Stanford University, CA

Markley, John L., $430,010
5 P41 LM05799-02
Biological Magnetic Resonance Data Bank
University of Wisconsin, Madison

Medical Informatics, Research Career Development Award

Sonnenberg, Frank A., $72,630
5 K04 LM00096-04
Knowledge Management for Clinical Decision Analysis
University of Medicine and Dentistry
R. W. Johnson Medical School
Piscataway, NJ

Biotechnology Resource Grants

Smith, Temple F., $952,184

5 P41 LM05205-14
Biomolecular Engineering Research Center
Boston University

Hardison, Ross, $220,893
5 R01 LM05773-03
Network Server for Electronic Genetic Analysis
Pennsylvania State University, University Park

Ledley, Robert S., $1,133,745
5 P41 LM05798-03
Protein Information Resource
National Biomedical Research Foundation
Washington, DC

Pagon, Roberta, $149,032
5 P41 LM06001-02
Helix—A Directory 0f Medical Genetics
Laboratories
Children's Hospital and Medical Center
Seattle, WA

*Pagon, Roberta, $292,362
1 P41 LM06029-01A1
GENLINE - An Electronic Clinical Genetics
Knowledge Base
University of Washington, Seattle

Jurka, Jerzy W., $253,972
5 P41 LM06252-02
REPBASE - A Database of Repetitive Sequences
Genetic Information Research Institute
Los Altos, CA

Cooperative Agreements

Barnett, Guy O., $459,128
5 U01 LM05854-03
Point of Care Computer Based Ambulatory Patient
Record
Massachusetts General Hospital, Boston

Hersh, William R., $249,900
5 U01 LM05879-03
Vocabulary and Text Data Extraction from the EMR
Oregon Health Sciences University, Portland

Safran, Charles, $117,300
3 U01 HS08749-03S1
Sharing Paperless Records Among Provider
Networks
Beth Israel Deaconess Medical Center
Boston, MA

Conference Grants

*Moult, John, $43,800
1 R14 LM06319-01
Assessment of Protein Structure Prediction
Maryland Biotechnology Institute
College Park, MD

*Dunker, A. Keith, $14,000
1 R13 LM06323-01A1
Pacific Symposium On Biocomputing
Washington State University, Pullman

INDEX

F

G

H

I

K

L

M

N