

1996

C.02-----SEQ: N05320000

TI: NATIONAL LIBRARY OF
MEDICINE PROGRAMS 07/08/97

NATIONAL INSTITUTES OF HEALTH NATIONAL LIBRARY OF MEDICINE PROGRAMS & SERVICES FY 1996

Internet Grateful Med

Location: <http://igm.nlm.nih.gov/>

What's New? What's Cool? Destinations Net Search People Software

Internet GRATEFUL MED

U.S. National Library of Medicine

Internet Grateful Med V2.21a

NEW RELEASE

This version of Internet Grateful Med was released September 9, 1996. It can search in HealthSTAR, AIDSLINE, and PREMEDLINE in addition to MEDLINE. The default file is MEDLINE. Select the "Search Other Files" action on the Search Screen to change to one of the other files.

INTRODUCTION

[Internet Grateful Med](#) offers assisted searching in MEDLINE and other online databases of the U.S. [National Library of Medicine](#) (NLM). It was developed through the User Access Services project of NLM's [System Reinvention](#) initiative. Internet Grateful Med can map user terms through NLM's [Unified Medical Language System®](#) (UMLS®) [Metathesaurus®](#) to help users create, submit and refine searches in MEDLINE® and other MEDLARS databases.

USER ID AND PASSWORD

Searching MEDLINE requires a valid MEDLARS user ID code and password.

User ID code: Password:

[APPLY ONLINE FOR A NEW NLM USER ACCOUNT](#) (last updated June 18, 1996)

[CHARGES TO USE INTERNET GRATEFUL MED](#) (last updated July 19, 1996)

[SEND QUESTIONS OR COMMENTS TO THE DEVELOPERS](#)

HELP TEXT

Context-sensitive help is available from many areas of the program. To invoke it, select the nearest information icon . To return from the help text, use the "Back" function of your



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

*Office of Public Information
National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894
(301) 496-6308
e-Mail: publicinfo@nlm.nih.gov*

Cover: Internet Grateful Med Home Page (<http://igm.nlm.nih.gov>)

NATIONAL INSTITUTES OF HEALTH
NATIONAL LIBRARY OF MEDICINE

Programs and Services
Fiscal Year 1996

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

National Library of Medicine Catalog in Publication

/
675 M4
U56an

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

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

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

Contents

Preface	v
High Performance Computer and Communications Program.....	1
Telemedicine	1
Visible Human Project	2
Internet Grateful Med	2
Unified Medical Language System	3
Office of Health Information Programs Development.....	4
NLM Long Range Plan	4
International Programs	4
Outreach	6
Library Operations.....	8
Planning and Management	8
Collection Development and Management.....	8
Bibliographic Control.....	10
Network Services.....	12
Reference Services	13
Document Delivery	14
National Network of Libraries of Medicine	14
Special Onsite Programs.....	15
Customer Service Policy	16
Specialized Information Services	22
TOXNET	22
Databases Under ELHILL.....	23
AIDS Services	24
Other Programs.....	24
Lister Hill National Center for Biomedical Communications	26
Computer Science Branch	26
Communications Engineering Branch.....	29
Cognitive Science Branch	34
Audiovisual Program Development Branch.....	40
National Center for Biotechnology Information	43
Database Building	43
Database Access	44
Basic Research.....	46
User Support and Outreach.....	47
Extramural Programs.....	47
Biotechnology Information in the Future	47
Extramural Programs	48
The Grant Programs	48
Other Extramural Programs Activities	51
Grant Review.....	51
Divisional Operations.....	51
Summary	52

Office of Computer and Communications Systems.....	53
Administration	57
National Performance Review.....	57
Financial Resources.....	57
Personnel	58
NLM Diversity Council.....	60
Office of Equal Employment Opportunity	60
Index	74
NLM Organization Chart.....	Inside back cover

Appendixes

1. Regional Medical Libraries	62
2. Board of Regents	63
3. Board of Scientific Counselors/LHC.....	64
4. Board of Scientific Counselors/NCBI	65
5. Biomedical Library Review Committee	66
6. Literature Selection Technical Review Committee	67
7. Extramural Programs Grants-FY 1996.....	68

Tables

Table 1. International MEDLARS Centers	7
Table 2. Growth of Collections	17
Table 3. Acquisition Statistics.....	17
Table 4. Cataloging Statistics.....	18
Table 5. Bibliographic Statistics	18
Table 6. Circulation Statistics	18
Table 7. Online Searches.....	19
Table 8. Off-line Searches.....	20
Table 9. Reference Services	20
Table 10. History of Medicine Activities.....	21
Table 11. Extramural Grants and Contracts Program.....	52
Table 12. Financial Resources and Allocations	58
Table 13. Staff	60

Preface

There were two signal events in Fiscal Year 1996 that led to an unusual amount of press attention for the National Library of Medicine: the announcement in November 1995 of the Visible Human Female and the introduction of the Internet Grateful Med in April 1996 (both in the HPCC Chapter). A number of other accomplishments in FY 1996 are worth noting, however, including:

- Publishing two major reports—"Survey of Online Customers" and "Outreach Activities of the National Library of Medicine: A Five-Year Review",
- Managing to keep MEDLINE operating during weeks of shutdown because of Federal furloughs and blizzards,
- Increasing to 140,000 the number of user codes for NLM's national online network,
- Introducing two new databases—SPACELINE, a space medicine database produced in cooperation with the National Aeronautics and Space Administration, and HealthSTAR, a merging of two existing NLM databases into one file with 2.5 million records on health care administration, planning, health services research, and clinical practice guidelines,
- Announcing a formal customer service policy that commits the Library to providing a high level of service to its customers (see page 16),
- Taking over from the Public Health Service two new programs: the PHS Historian's Office and the bimonthly journal *Public Health Reports*;
- Awarding five-year contracts to eight institutions to serve as Regional Medical Libraries in the National Network of Libraries of Medicine, and
- Three new Regents were appointed by the HHS Secretary in 1996: Enriqueta Bond, M.D., Raymond J. Fonseca, D.M.D., and John Gage. Steven J. Phillips, M.D., was elected Chairman of the Board by his colleagues.

For the success we continue to enjoy in meeting the information needs of health professionals, I would like to thank the Library's hard-working staff, our friends and collaborators in the National Network of Libraries of Medicine, the experts who serve on the Library's various advisory Boards, and the American public who continue to support the Library and the National Institutes of Health with their tax dollars.



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

HIGH PERFORMANCE COMPUTING AND COMMUNICATIONS PROGRAM

The High Performance Computing and Communications initiative is the rubric under which the Library engages in several important technology-related programs. Up until January 1995, NLM director Donald A. B. Lindberg, M.D., headed the coordinating office for the HPCC multiagency initiative. He had assumed that post in 1992 and held it concurrently with his directorship of the NLM. In FY 1996 an Office of High Performance Computing and Communications was created within the Lister Hill Center. Two of that Office's projects—telemedicine and Visible Human—are described here, along with the Internet Grateful Med and UMLS, two projects that cut across several NLM programs.

Telemedicine

NLM's telemedicine program, coordinated by the Lister Hill Center's Office of High Performance Computing and Communications, is designed (1) to evaluate the impact of the National Information Infrastructure (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.

NLM is the lead agency within DHHS for the government's High Performance Computing and Communications initiative and as such has a direct interest of the use and effects of the 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.

Major research and evaluation issues included in NLM's telemedicine program arising from the current and future impact of the NII 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 important 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 Library announced that it was funding 19 telemedicine projects, affecting rural, inner-city, and suburban areas, with a total budget of \$42 million. The 19 multi-year telemedicine 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 National Academy of Sciences studies on criteria for evaluation of telemedicine and on best practices for ensuring the confidentiality of electronic health data. NLM is the principal patron of these studies.

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. NLM sponsored a study by the Institute of Medicine (IOM) designed to develop a framework for evaluating the patient care applications of telemedicine. *Telemedicine: A Guide to Assessing Telecommunications for Health Care* identifies managerial, technical, policy, legal, and human factors that must be taken into account in evaluating a telemedicine program. The IOM appointed a committee to review 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.

M. Ackerman, Ph.D.

Visible Human Project

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 700 licensees in 26 countries. The Visible Human Project has been featured in well over 800 newspaper articles, news and science magazines, and radio and TV programs worldwide.

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 kindergarten 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. Automobile manufacturers now include passenger injury models based on Visible Human data to 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.

Phase two of the Visible Human Project is to turn 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 labeled. The relationship of each object to the other objects in its cross-section and in the adjacent cross-sections is being cataloged. The extent of a single object which spans several cross-sections must be noted. In order to accomplish this, information about building geographic databases and databases associated with computer-aided drafting systems are being used as starting points for development of this unique interactive anatomical digital atlas.

M Ackerman, Ph D

Internet Grateful Med

The National Library of Medicine launched the Internet Grateful Med on April 16, 1996, at a conference on health-care applications of the information superhighway sponsored by the Friends of the National Library of Medicine. The introduction of IGM was made by NLM Director Donald A B Lindberg, M.D., Michael E DeBakey, M.D. (a member of the NLM Board of Regents), and Senator Bill Frist (R-TN), a physician.

The new system, which is the first major product of the NLM Reinvention Initiative, allows anyone with access to the World Wide Web to search the MEDLINE database of 8 million journal article references and abstracts. It improves considerably on the existing versions of Grateful Med (used primarily with dial-up telephone networks) by affording much greater flexibility in using medical vocabulary (including terms from the UMLS Metathesaurus) and the easy use of limiting factors such as language, age group, and year of publication. The Internet Grateful Med allows users to sign up online and, because the Internet is global in nature, even health professionals in other countries can join the network.

The system has extensive context-sensitive online help that can be displayed from nearly every point in the program. There is also a 20-slide hypertext introduction and system overview. A 25-segment hypertext "New User's Survival Guide" leads the user step by step through a series of actions and searches that demonstrate the capabilities of Internet Grateful Med. An unusual "Analyze Search" function offers related terms from among many thousands of cross references.

The instant appeal of Internet Grateful Med resulted in a dramatic increase in the number of persons using the Library's online network, and online computer usage statistics are repeatedly hitting all-time highs. The system received another boost in popularity when Ann Landers printed a letter from Dr. Michael E DeBakey praising the new system. The Internet Grateful Med was improved later in FY 1996 by adding NLM's AIDS and health services

research databases (including access to clinical practice guidelines) to its searching capabilities, also added was a direct link to the 60,000 Images from the History of Medicine. More databases will be added in the future.

R. Mehnert

Unified Medical Language System

The goal of the UMLS program is to help health professionals and researchers to retrieve and integrate biomedical information emanating from a variety of electronic sources, including computer-based patient records, bibliographic databases, factual databases, and expert systems. To do this, it must overcome obstacles of language (various information sources use various vocabularies), disparate search procedures, and wide distribution of computer-based information across the electronic landscape. The UMLS seeks to build "knowledge sources" that may be used to develop interfaces by those creating medical information systems. These knowledge sources map concepts and terms from many different biomedical vocabularies and classifications and also provide machine-processable descriptions of the contents of health-related databases.

FY 1996 marked the 10th anniversary of the UMLS project, which is directed by a multi-disciplinary team of NLM staff and involves medical informatics research groups across the United States working under competitively awarded contracts and grants. More than 700 UMLS users worldwide receive annual editions of the UMLS Knowledge Sources.

Several improvements in distribution of the UMLS Knowledge Sources occurred in 1996. The release date was moved to early in the calendar year to allow UMLS applications that use the MeSH terminology to have access to the most current edition. A new Web-based front-end to the UMLS Knowledge Source Server provides easy cross-

platform access to the UMLS Knowledge Sources via the Internet. As a result, NLM ceased distribution of separate Macintosh and DOS browser applications for the UMLS Metathesaurus.

NLM continued to make major additions to the Metathesaurus, a database of concepts and terms from more than 30 vocabularies and classifications. The 1996 edition incorporates additional sections of SNOMED International, more cancer terminology from PDQ, and the German translation of MeSH produced by the International MEDLARS Center in that country. Through the collaboration of several International MEDLARS Centers, the Metathesaurus now has MeSH terminology in English, French, German, Portuguese, and Spanish.

Users in commercial companies, universities, hospitals, libraries, and other research institutions are applying the UMLS Knowledge Sources to a variety of information problems, including creation of computer-based patient records, linking patient records to related information in clinical practice guidelines and bibliographic databases, Web information retrieval, and natural language processing. NLM itself uses the UMLS Metathesaurus in the Internet Grateful Med interface to MEDLARS databases and in natural language processing, indexing, and retrieval research.

In FY 1996, NLM began a large-scale national test to determine the extent to which the 1996 Metathesaurus and several planned additions to it cover the concepts and terms needed in U.S. health data applications. The test, which is co-sponsored by the Agency for Health Care Policy and Research, uses a special Web-based interface to the UMLS Knowledge Source Server to allow participation by groups across the country. The results, which will be available in 1997, will assist NLM in setting priorities for UMLS development and should also be helpful in developing appropriate strategies for moving toward a standard set of vocabularies for U.S. health data.

B. L. Humphreys

OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT

Elliot R. Siegel, Ph.D.
Associate Director

Long Range Plan

In January 1985, NLM's Board of Regents undertook to develop a Long Range Plan to guide the Library in using its human, physical, and financial resources to fulfill its mission. This culminated in the adoption of a Plan by the Board of Regents containing goals, recommendations, and resource requirements, published in 1987. NLM continually expands and updates the Plan, recent 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 current planning effort is focusing on NLM's international programs.

The Long Range Plan is integral to the day to day management of the Library. Over the past 10 years, NLM has successfully integrated its budget and planning processes, so that the budget is "driven" by the Long Range Plan. New initiatives are presented in terms of the Plan, and progress toward achievement of recommendations is tracked. This gives NLM a rational process for setting goals and objectives and measuring program performance that is built into the fabric of its management processes.

Planning Panel on International Programs

The NLM's International Programs are integral to the Library's responsibilities in biomedical information transfer. The U.S. health community depends on NLM's information retrieval system to identify the relevant information from the world's biomedical literature and, in turn, the world health community shares the benefits of NLM's advanced information systems in meeting its own requirements for needed information.

The Board of Regents approved the formation of a Long Range Planning Panel for NLM's international programs. The purpose of the panel is 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. It is expected that the panel will meet three times in 1996-7, and will present its final report to the Board for approval and inclusion in the NLM Long Range Plan.

Dr. Donald S. Fredrickson, former NIH Director, has accepted the chairmanship of this panel. Dr. Fredrickson is a distinguished physician and scientist who has been President of the Institute of Medicine and President and Chief Executive Officer of the Howard Hughes Medical Institute, among many other accomplishments and honors.

The panel has 24 regular members and it is being advised by a number of consultants in areas such as health sciences librarianship, electronic publishing, and telecommunications. The Panel is being asked to tie its considerations to the following frame of issues:

- What work ought NLM do overseas or with other countries?
- Should NLM focus its efforts on particular countries or regions?
- Should NLM focus its efforts on particular types of biomedical information?
- Are there particular communication methods or strategies that should receive special attention?

We seek to emerge at the end of three meetings with useful maps to guide the Board of Regents in constructing its blueprints for the future international roles of the Library.

International Programs

While we plan for the future of the NLM International Programs, activities continue in the present. In FY 1996, NLM marked another year of activities with individual countries, international government organizations such as the World Health Organization (WHO) and the Pan American Health Organization (PAHO), and international nongovernmental organizations such as the International Council for Scientific and Technical Information (ICSTI). NLM international activities also included training for colleagues from abroad, the NLM publication exchange program with 168 libraries in 58 countries, and receiving numerous professional visitors from abroad.

Country-to-Country Activities and Bilateral Agreements

International MEDLARS Centers

To assist international health professionals in accessing NLM MEDLARS data bases, NLM has signed Memoranda of Understanding with partners in

18 foreign countries and with two international organizations (Table 1)

Twenty-seven representatives from 17 of these International MEDLARS Centers attended the 1996 International MEDLARS Policy Advisory Group at the NLM March 26-27, 1996 Technical programs on NLM System Reinvention, biotechnology information activities, Internet Grateful Med, and a demonstration of the Visible Human were presented to the group. Issue panels on Unified Medical Language System and MeSH translations, regional medical bibliographies, outreach initiatives, access to Internet Grateful Med and the Internet, document delivery, and telemedicine applications provided important information to the group on NLM activities. As Chair of the Long Range Planning Panel on International Programs (LRP-IP), Dr. Donald S. Fredrickson, NLM Scholar in Residence, NIH Director Emeritus, asked for advice from IMPAG concerning what should NLM be doing internationally in the next 5 to 10 years, and what contributions is NLM in a position to make to the world's health. The input received from the IMPAG provides important issues for consideration by the Long Range Planning Panel on International Programs (see above).

Hong Kong

The signing of a Memorandum of Understanding between the Chinese University of Hong Kong (CUHK) and the National Library of Medicine (NLM) on March 26, 1996, designated Hong Kong as the 20th International MEDLARS Center. The CUHK has already demonstrated successfully from many hospitals, doctors' offices, and libraries how to use Grateful Med and the Internet to search NLM MEDLARS databases. The Hong Kong center will be the second foreign center to access NLM's databases exclusively over the Internet using the Grateful Med software.

Israel

In an effort to provide foreign users with cost-savings and conveniences associated with a predictable charge for searching NLM's database, an experimental arrangement has been undertaken with the Israel MEDLARS Center to purchase a block of codes and to market these among current and prospective individual users in the region with access to the Internet. (The Egyptian MEDLARS Center has been offered a similar opportunity.) The intent is to gain needed international experience with the economics and dynamics of such competitive pricing

schemes that are proving to be very popular in the U.S.

Newly Independent States

The NLM, with assistance from the U.S. Department of State, is helping medical scientists of several newly independent states of the former Soviet Union with access to biomedical information. The medical libraries of Belarus, Ukraine, Kazakhstan, Kyrgyzstan, Uzbekistan, Lithuania, Latvia, and Estonia are being provided online access to NLM's biomedical databases as well as journal subscriptions and interlibrary loans. By 1995, all libraries were able to search NLM databases and to communicate with colleagues at NLM and other participating libraries via e-mail.

In 1996, an experiment in upgrading telecommunications infrastructure was completed in Lithuania that enabled both the Vilnius University Medical Faculty and the Lithuanian National Medical Library to use Internet services other than e-mail. Additional codes for MEDLARS access were given to branch medical libraries and researchers in certain countries receiving U.S. support for collaborative research projects. Every library has been conducting regular MEDLARS searches using the hardware and software provided, with one exception, and journals have been flowing to the central Asian republics from the American Association for the Advancement of Science. Interlibrary loans have been more problematic and uneven, and the volume has depended on each country's infrastructure and established lines of communication with its collaborating library in the U.S. Additional journal subscriptions are being provided to supplement interlibrary loans where appropriate. Equipment upgrades are being explored in cases where there may be sufficient infrastructure to support them.

Activities with International Organizations

PAHO

The Pan American Health Organization has been an important collaborator with the NLM. In 1988, Dr. Carlos Gamboa, working with colleagues at the University of Chile's School of Medicine, successfully devised BITNIS as a means to conduct MEDLINE searches over high-speed computer networks using Grateful Med. BITNIS is a gateway system that imbeds a MEDLINE search within an e-mail packet envelope, thereby providing an inexpensive and fast search service in regions of the world where the cost of access by traditional

communications systems would be prohibitive for most health professionals. During 1996, searches were performed from more than 40 countries, but the goal remains one of migrating these users over to systems such as Internet Grateful Med when full Internet capabilities are in place.

WHO

The National Library of Medicine and the World Health Organization 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

Working with ICSTI, NLM is leading the development of a demonstration project intended to reduce existing technological and tariff barriers to the flow of scientific and health-related information over the Internet to developing regions of the world. Focused initially in the eastern Caribbean, the NLM/ICSTI effort will be carried out in conjunction with a number of other international organizations, including UNESCO, PAHO, the International Telecommunications Union, and the United Nations Development Programmes.

International Meetings and 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 1996 were representatives from the following countries:

Argentina, Australia, Belarus, Brazil, Bulgaria, Canada, China, Czech Republic, Egypt, Ecuador, Estonia, France, Georgia, Germany, Hong Kong, Israel, Japan, Korea, Latvia, Lithuania, Mexico, Netherlands, Poland, Romania, Russia, South Africa, Sweden, Switzerland, Singapore, Slovakia, Slovenia, Spain, Taiwan, Thailand, Ukraine, United Kingdom, Uruguay

Outreach

In the late fall of 1995, NLM conducted a random sample survey of 2,500 online customers. The purpose of the survey was to assess the availability of end-user technology capable of supporting a transition from Grateful Med to new forms of access and retrieval, in particular Internet Grateful Med which was under development at the time. The survey results indicated that about 70 percent of the NLM's customers already had access to the Internet, although the majority of these Internet-capable users were not accessing NLM databases via the Internet. In addition, according to the survey results, Internet access is uneven, especially in rural areas and at hospitals. More than a third of the rural users and almost half of users at hospitals reported that they did not have access to the Internet. Similarly, about half of users at rural libraries did not have Internet access. These results reinforce the need for continuing special outreach efforts directed at improving access for rural and hospital-based users and rural libraries.

In FY 1996, NLM published the first comprehensive and detailed assessment of the outreach efforts undertaken in collaboration with the eight regional medical libraries and the more than 4,500 members of the National Network of Libraries of Medicine. Published as a special supplement to the April 1996 issue of the *Bulletin of the Medical Library Association*, the report highlights the priority and importance of outreach across all of NLM's programs.

A newly developed outreach project is being undertaken by the Regional Medical Library at the University of Washington. The effort is directed toward better understanding how to build sustainable and workable means for Native American health care providers (and, by extension, other underserved communities) to access NLM and other biomedical information resources. A primary goal is to develop models of the facilitators for and barriers to effective access. Model components may include geographic area served, technologies, integrated infrastructure, extent of use of computerized biomedical databases, training needs, institutional participants, cultural/language factors, and local leadership factors. The models may also provide the frameworks for evaluating progress toward sustainable and workable access.

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)
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

Library Operations (LO) provides the core NLM services that improve access to biomedical and health care information for health professionals and researchers around the world. The largest of NLM's Divisions, LO acquires and preserves the biomedical literature, develops and maintains a biomedical thesaurus and a classification that are widely used to organize biomedical information, indexes and catalogs the materials acquired by the Library using this thesaurus and classification; distributes NLM's authoritative indexing and cataloging records online, in machine-readable form, and in publications, provides document delivery, reference, and research assistance as a back-up to services available from other U.S. health sciences libraries, and coordinates the National Network of Libraries of Medicine (NN/LM) which provides primary information services to U.S. health professionals and researchers.

The multidisciplinary LO staff also maintains an active research program in the history of medicine, conducts research and evaluation related to LO programs and services, and directs or contributes actively to many NLM-wide initiatives. LO's basic services form the essential foundation for NLM's Outreach Program and for NLM's biotechnology, AIDS, and health services research information programs.

Planning and Management

Minimizing the combined effects of partial Government shut-downs, snow days, and the extended suspension of the MEDLINE data entry contract became an important management challenge during the past year. Most FY 1996 LO production and service figures are not comparable to those of previous years due to these unusual disruptions. To reduce the Library's dependence on keyboarding as a data entry mechanism, LO is collaborating with other NLM components to establish a robust OCR (optical character recognition) input stream for MEDLINE data and to obtain electronic citation and abstract data direct from publishers. Additional information about these efforts appears in this and other chapters of the report.

The NLM System Reinvention project, which is designed to replace the Library's outmoded legacy systems with more flexible and maintainable software that will help users and staff members do

their work more efficiently, continued to be a major priority. LO staff members assisted with final testing and release of Internet Grateful Med and final testing of Grateful Med for Windows and also tested potential replacements for NLM's underlying retrieval system. ELHILL LO and OCCS selected a commercial integrated library system (ILS) that will be adapted to support many of NLM's basic library functions and purchased Relais, a product originally developed for the Canadian Institute of Scientific and Technical Information, that will be used to automate internal operations related to the Library's document delivery service.

LO also continued work on re-engineering its customer service activities. A new NLM customer service policy was announced and released on the Library's Web site in early 1996. To obtain a more accurate picture of current customer service traffic, LO directed a project to collect and analyze all customer inquiries received at service points throughout NLM during a one-week period in December 1995. The data indicated that 92% of customer inquiries were handled by the combination of the Online Service Desk, the Reference Section, the Collection Access Section, and the National Center for Biotechnology Information and that 84% of the customer transactions took 5 minutes or less. To facilitate a more coordinated approach to handling customer service inquiries and obtaining customer feedback, LO has obtained the CustomerQ, commercial customer service support software, and the Teloquent automated call distribution system.

Like others engaged in organizing and providing access to biomedical information, LO is examining how to manage the challenges and opportunities offered by the rapidly growing number of electronic information sources. Specific projects are described elsewhere in this chapter.

Collection Development and Management

Collection development entails establishing and updating selection policy, acquiring and processing relevant literature in all formats and languages, and preserving the materials acquired. As of September 30, 1996, NLM owned 2,181,783 printed books, journal volumes, theses, and pamphlets and 3,004,029 nonprint items, including audiovisuals, computer software, microforms, prints, photographs, and manuscripts (Table 2).

Selection

LO staff members and their agents select literature for the NLM collection based on guidelines published in the *Collection Development Manual of*

the *National Library of Medicine*, which is revised every 5-7 years. In between major revisions, LO develops more specific operational criteria for types of material that are presenting selection challenges and for emerging subject areas. In FY 1996, LO implemented more detailed coverage guidelines for some types of serials and audiovisuals. With the assistance of Lucretia W. McClure, Librarian emerita of the University of Rochester Medical Center, coverage guidelines for annual reports, county medical society bulletins, and state journals were examined, and recommendations developed for a systematic review of this material. Testing began of draft criteria for selecting the electronic information sources that will be described in NLM's databases, including AVLINE, DIRLINE, and the UMLS Information Sources Map. NLM joined with the Library of Congress and the National Agricultural Library to revise the joint NLM/NAL/LC collecting statement for human nutrition and made the joint collecting statement for veterinary medicine available on NLM's Web site. The national libraries issue joint statements to clarify their collecting responsibilities in subject areas that are partially covered by each of them.

LO conducts periodic collection assessment studies to determine how successfully it is applying NLM collection guidelines. In FY 1996, current serial titles in sociology, social work, and counseling were reviewed, and some subscriptions were cancelled as a result.

Acquisitions

In FY 1996, LO received and processed 174,992 modern books, serial issues, audiovisuals, and software packages (Table 3). The Library's collection grew by 37,553 volumes and 39,304 other items (e.g., audiovisuals, microforms, software, pictures, manuscripts). Among other interesting historical acquisitions, NLM acquired a number of rare 16th century books including, *Dissectionis Venarum Artuarumque Commentarium* (Paris, 1526), the first separate editions of Galen's treatises on the dissection of veins and arteries, nerves, and muscles translated by Antonius Fortobus, Pierre Tolet's French translation of Galen's treatise on blood-letting, *De la Raison de Cvrer per Euacuation de Sang* (Paris, 1584), *Paedotrophiae* (Paris, 1584), a first edition of Scevole de Sainte-Marie's poem on the care and feeding of children, Joannes Gilinus' *la. Antiquar. Arx Erat hic Quondam Domus Est ubi & Hospita Virgo qua* (Milan, 1508), a first edition of what is probably the first printed book on hospital history and management, about the 15th century Ospedale Maggiore of Milan, a first edition of

Nicolas Monardes' work on blood-letting, *De Secanda Vena Impleuriti inter Grecos et Arabes* (Seville, 1539); and a first edition of Paracelsus' *De Natura Rerum IX Bucher*. The Library also obtained several manuscripts of physicians who practiced in America in the late 18th and early 19th centuries: the account book of a Revolutionary War physician, Joseph Fiske, Jr. of Lexington, Massachusetts (1784-98, 1803); the two-volume medical daybook of Nathaniel Swift, a country practitioner in rural Massachusetts in the aftermath of the War of 1812, and the journal of an anonymous physician working in the Adirondacks from 1827-33. William H. Helfand continued his generosity to NLM by donating 149 historical postcards of hospitals and a collection of pharmaceutical handbooks, catalogs, and therapeutic manuals which provide excellent source material for the study of late 19th and early 20th century medical practice and the development of the pharmaceutical industry.

As a cost-saving measure, the Library canceled additional second copy subscriptions for journals indexed in MEDLINE that are not heavily requested by onsite and offsite users. Efforts to improve acquisition channels for material from Eastern Europe and the former Soviet Republics began to bear fruit. The automated system used to create and edit literature invoice records was enhanced to improve the accuracy of invoice data. Other improvements to existing automated processing systems were held to a minimum as staff members worked on the procurement of the commercial integrated library system that will serve as the platform for replacing existing legacy systems.

Preservation and Collection Management

Preservation and collection management cover all activities that extend the life of NLM's archival collection of biomedical literature and keep that collection in good order. These activities include: binding incoming journal issues, microfilming brittle volumes, providing conservation treatment for rare and unique items, maintaining appropriate storage procedures, facilities, and conditions for all segments of the collection, and preventing and responding to emergencies that threaten library materials. NLM also distributes preservation information to assist other libraries, explores the use of new technology to preserve library materials, and promotes the use of more permanent media in new biomedical publications. In May 1996, NLM's National Program for the Preservation of the Biomedical Literature received the Louise Darling Medal for Distinguished Achievement in Collection Development in the Health Sciences from the Medical Library

Association

In FY 1996, LO bound 18,308 volumes, microfilmed 12 million brittle pages, and gave conservation treatment to 250 items in the special historical collections. Substantial progress was made on the project to ensure that there is archival quality microfilm for all brittle non-US *Index Medicus* titles (The U.S. titles were completed in FY 1995). A new PC-based automated system manages and tracks microfilming production operations, since these functions cannot be quickly addressed by the new ILS software NLM will procure as part of System Reinvention.

Much of the planned work to establish an onsite facility for minor book repairs was completed. The revision of the Disaster Recovery Plan is also proceeding. A study of collection storage space needs showed that, at current growth projections, NLM will have space for its collection through the year 2014. Since the rate of conversion to electronic publication remains difficult to predict, the Library will re-examine its space needs in 2000.

In FY 1996, LO began distributing preservation data in MARC format, thus greatly enhancing access to information about what NLM has preserved. The Preservation Section also established a home page on NLM's Web site to improve dissemination of preservation information to health sciences libraries.

Bibliographic Control

To help health professionals and researchers to identify relevant information within the world's biomedical literature, NLM develops the Medical Subject Headings (MeSH) thesaurus for use in indexing, cataloging, and online searching, links this thesaurus to other biomedical vocabularies in the Unified Medical Language System (UMLS) Metathesaurus, maintains the *NLM Classification* for subject arrangement of books on library shelves; and produces authoritative cataloging and indexing records for newly published or acquired books, serial titles, audiovisuals, computer software, manuscripts, and journal articles.

Thesaurus

The 1997 MeSH contains 18,531 main subject headings, 787 subheadings or qualifiers, 83 publication types, and about 91,000 supplementary records for chemicals and other substances. For 1997, LO staff added 349 new main headings, replaced 71 main headings with more current terminology, and created 1,019 new cross-references. There were special efforts to improve terminology for viruses and

virus diseases, dentistry, space medicine, and the history of medicine. The enhancements to dental terminology were based on recommendations prepared by Dr. Colette Hochstein, then an NLM Associate, who analyzed terms in vocabularies provided by the American Dental Association. Thirty-five new publication types and 15 new geographic subheadings were established for 1997 as part of a broader project to eliminate differences in the use of MeSH in NLM cataloging and indexing. The goal of this project is to remove barriers to effect cross-file retrieval by NLM's users.

The MeSH Section manages the editing of the content of the UMLS Metathesaurus, which is released annually in January. In FY 1996, most of the editing for the 1997 Metathesaurus was completed. More than 80,000 concepts and 124,000 terms will be added to the Metathesaurus for 1997, including the remainder of *SNOMED International*, the American Medical Association's *Current Procedural Terminology (CPT)*, *Logical Observations Identifiers Names and Codes (LOINC)*, and about one third of the *Read Thesaurus*, which is used by the National Health Service in the United Kingdom. These additions will improve the Metathesaurus's already considerable utility as a source of controlled vocabulary for clinical information systems and will improve the ability of Internet Grateful Med and other search interfaces to map users' clinical terminology to MeSH terms for better MEDLINE retrieval.

OCCS staff have begun development of a new MeSH maintenance system that will be built on the Oracle relational database management software that will also underlie NLM's new integrated library system. The new MeSH system will be based on a fundamental data model that is similar to that of the UMLS Metathesaurus to facilitate an eventual integration of the systems used to build and maintain MeSH and the Metathesaurus.

Cataloging

The LO staff catalogs the literature acquired by the Library both to alert users to what is available from NLM and to provide cataloging records that can be used by other libraries, thus reducing the level of effort required to organize and manage library collections throughout the world. Many health sciences libraries also use MeSH and the *National Library of Medicine Classification* to catalog materials of local interest for which no NLM cataloging record is produced.

In FY 1996, the Technical Services Division cataloged 20,365 modern books, serials, nonprint items, and cataloging-in-publication (CIP) galley,

using a combination of in-house staff, contractors, an interagency agreement with the Library of Congress, and assistance from the International MEDLARS Center in China. There was a net reduction of 1,663 in the working inventory of uncataloged modern books. Samples of a range of electronic information sources were cataloged as part of a project to develop NLM policy for bibliographic control of these materials. The History of Medicine Division cataloged 351 rare books and began applying MeSH headings to pre-1801 works. Previously these items did not receive subject cataloging.

Images from the History of Medicine, the Web-based image database and associated catalog for NLM's historical prints and photographs collection, is being moved from the prototype system developed in the Lister Hill Center to a new production system based on the integrated library system software that is likely to be a key part of the NLM System Reinvention project. The new environment will facilitate upgrading of the picture catalog records, as well as user access to the image database.

Significant improvements in the distribution of NLM's machine-readable cataloging records are described in the Network Services section of this chapter.

Indexing

LO indexes more than 3,800 biomedical journals to help health professionals and researchers identify articles on specific topics, using a combination of in-house staff, commercial contractors, international MEDLARS centers, and cooperating organizations. LO also annotates its indexed citations when previously indexed articles are retracted, corrected, or challenged in subsequently published notices or commentaries. LO staff members assist the National Center for Biotechnology Information (see NCBI chapter) by indexing gene sequences and editing author sequence submissions for incorporation into GenBank.

The Literature Selection Technical Review Committee (LSTRC- see Appendix 6 for members), an NIH-chartered committee, advises NLM about the journals that should be indexed in MEDLINE, *Index Medicus*, and other NLM databases. In FY 1996, the LSTRC reviewed 395 journal titles and rated 77 sufficiently highly for them to be added to MEDLINE. The LSTRC reviewed recommendations on journal coverage from professional societies in the fields of obstetrics and gynecology, psychiatry, the behavioral sciences, and hematology and also considered the results of a special review of journals in the Chinese language. As a result of these reviews,

the LSTRC advised NLM to index 38 additional journals and to discontinue indexing 29 journals in these areas.

In FY 1996, NLM added 336,000 citations to MEDLINE, an 18% decrease from FY 1995 due to the impact of a two-month suspension of a newly awarded data entry contract. The contract award made in January 1996 was protested by an unsuccessful bidder, and the Government Services Administration eventually instructed NLM to reopen the negotiations with both the winner and the protestor in the competitive range. Both firms submitted revised proposals, and the second technical review yielded the same result as the first.

To minimize the immediate impact of the suspension, staff members throughout NLM volunteered their services as emergency keyboarders and proofreaders, work began on a parallel OCR input stream that will be developed into a production operation as insurance against future disruptions to the data entry operation, the existing project to obtain machine-readable citation and abstract data from publishers was accelerated, and, as soon as legally able to do so, the Library purchased short-term data entry and keyboarding support from a range of outside sources. The Index Section handled the large and complex task of coordinating input from all these sources and ensuring that contract indexers continued to receive an adequate volume of work. The backlog that developed during the suspension and its aftermath should be cleared by early calendar 1997.

LO, the Office of the Director, the Lister Hill Center, and the Office of Computer and Communications Systems cooperated to establish a basic OCD (optical character recognition) operation at the Library that can be used to input abstracts. Additional research and development will continue in FY 1997 to produce software that more correctly interprets the Greek letters, other special characters, and formulas that appear frequently in the journals NLM indexes. In an effort coordinated by the National Center for Biotechnology Information (see NCBI chapter), NLM received 3,816 usable electronic citations and abstracts from publishers in FY 1996. The volume of data received via this mechanism is expected to increase gradually over the next several years.

LO staff members are directing or contributing to several subprojects in the Next Generation Indexing project, which will examine whether more intensive use of automation can produce comparable retrieval performance while reducing the time and human resources needed to index articles. The Lister Hill Center is directing this NLM-wide effort.

Network Services

To improve access to current biomedical information, NLM disseminates its indexing and cataloging data online, in machine-readable formats, and in publications, answers requests for reference and research assistance from onsite and remote users, delivers documents from the NLM collection as a back-up to services available from other libraries and suppliers, and coordinates the National Network of Libraries of Medicine (NN/LM) to equalize health professionals' access to information irrespective of their geographic locations. The goal of NLM's Outreach Program is to link every US health professional to these services.

Online Services

NLM's online users conducted about 7.4 million searches in FY 1996 (Table 7). This figure does not include searches performed on the computer systems and CD-ROM products of the many organizations that lease data from NLM and probably represents less than half of the total online use of the Library's databases.

At the end of FY 1996, 139,257 individuals and institutions held codes for searching the online databases available on NLM's systems, an increase of about 25,000 or 21% from FY 1995. Many new users took advantage of a new online registration system, and the number of code applications submitted on paper forms declined sharply. Individuals from outside the United States may now access NLM's system directly, without applying through an International MEDLARS Center, if they register online, use Internet Grateful Med, and pay via credit card. Increasing numbers of US users obtain online access through fixed-fee arrangements negotiated by their employing institutions. There are 38 such arrangements covering 200 different sites.

Most of NLM's online users search via one of the several versions of Grateful Med, including the new Internet Grateful Med that was released in April 1996. Any code holder with an appropriate World Wide Web browser can point to NLM's Web site and use Internet Grateful Med, which employs the UMLS Metathesaurus to help users construct more successful search strategies. A new version of Grateful Med for Windows was fully tested in FY 1996 and will be released in early FY 1997. The results of the FY 1996 survey of NLM's online users (see chapter on Health Information Programs Development) indicate that the majority of these users are likely to switch to either the Internet or the Windows version of Grateful Med, allowing the Library to phase out the DOS version as planned for

NLM System Reinvention

In FY 1996, the Library took steps to make references to both newer and older journal articles available online. PREMEDLINE, a database of in-process MEDLINE citations that have not yet been proofread or indexed with MeSH headings, became publicly available in August 1996. This file has been particularly useful while NLM works to reduce the backlog caused by the data entry contract suspension. LO and OCCS staff also completed most of the work to put 1964-65 *Index Medicus* data online at NLM, using original MEDLARS data that had been retained by DIMDI, the German MEDLARS Center. OLDMEDLINE will become available to the public in early FY 1997, and NLM will proceed to work on making earlier data available as well. Other important changes to NLM's online services included the October 1995 release of SPACELINE, a space medicine and life sciences database produced by NLM and the National Aeronautic and Space Administration, the April 1996 merger of the HEALTH and HSTAR databases to form HealthSTAR, the addition of 14,000 references to items not represented in other NLM databases to HISTLINE, and the enhancement of HSTAT with cross-file searching and many new full-text documents, including the 1995 revision of the *Guide to Clinical Preventive Services*.

FY 1996 was a transition year for NLM's online services training (online) program for health sciences librarians and other search intermediaries. The number of regional online training centers was reduced from three to one—the New York Academy of Medicine—in April 1996. Work proceeded on restructuring the "Fundamentals of MEDLARS Searching" class to rely more heavily on multimedia and computer-assisted instruction and to reduce its length to two days. The revised course will be introduced in 1997. In FY 1996, NLM and the Regional training sites taught 45 online training courses, including Fundamentals courses, specialized training modules, and MEDLARS updates, to a total of 688 students.

Machine-readable databases

To promote broad access to biomedical information, NLM leases its databases in machine-readable form to commercial database vendors, international MEDLARS centers, universities, and other interested organizations. The licensees then make the data available online or in CD-ROM products. In FY 1996, NLM distributed data from one or more of its databases to 98 different organizations. Although they pay for a proportional share of the cost of making MEDLARS data

available in machine-readable form, several licensees are now offering free access to MEDLINE via the Internet, usually in conjunction with commercial advertising or other services for which there may be a fee

LO continues to streamline its data distribution procedures. In FY 1996, all information related to licensing NLM data became available on the NLM Web Site. The Library developed a policy that permits people to download small amounts of MEDLARS data for incorporation in their products at regular online service rates. In 1997, organizations will be able to establish license agreements for MeSH data online via the Internet, and additional formats of MeSH data will be available via ftp. In FY 1996, ftp distribution of cataloging data was expanded to include weekly USMARC (MACHine-Readable Cataloging) bibliographic record shipments and the transfer of NLM name authority records to the Library of Congress. NLM completed the conversion of its USMARC distribution format to the new integrated format. The Library also began to include NLM preservation microfilm information in its MARC records.

Publications

NLM publishes some of its authoritative data in printed publications, of which *Index Medicus* is the oldest and largest. As NLM's primary clientele becomes able to use alternative electronic dissemination mechanisms, the Library reviews and modifies or eliminates specific printed publications that have outlived their usefulness in that format. In keeping with this strategy, NLM ceased publication of *Cumulated Abridged Index Medicus* with the 1994 edition, and will cease publishing *Health Sciences Serials* and the *MeSH Supplementary Chemicals* with their 1996 issues. Print remains an attractive format for some products, however. In FY 1996, NLM published *A Shelflist of Islamic Medical Manuscripts at the National Library of Medicine* in printed form. This publication reflects the extensive scholarly assistance previously provided by Emilie Savage-Smith, Ph.D., Oxford University, which allowed HMD to prepare full cataloging records for this important group of historical materials.

The Library continues to increase the number and types of publications available via the Internet. In FY 1996, LO played a key role in the reorganization and expansion of NLM's public World Wide Web site (www.nlm.nih.gov). Since March 1986, Internet users obtained more than 244,763 copies of NLM publications—from fact sheets to extensive bibliographies and reports—using the Web.

Current Bibliographies in Medicine is

available both in print and electronically. Its issues address topics of current interest that may be difficult to search in NLM's databases or appear in the literature of several multiple disciplines. The bibliographies cover subjects important to specific programs of NLM, NIH, and other Health and Human Services 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. FY 1996 additions to the series included *Critical Pathways; The Role of Dietary Supplements for Physically Active People, Confidentiality of Electronic Health Data*, produced as an NLM contribution to the Interdepartmental Health Privacy Working Group and as background for a National Academy of Sciences study funded by NLM; and *Public Health Informatics*, one outgrowth of the meeting on "Making a Powerful Connection: the Health of the Public and the National Information Infrastructure" held at the Library in April 1995.

Reference Services

LO provides reference service and research assistance to onsite and remote users as a backup to services available from other health sciences libraries in the National Network of Libraries of Medicine. In FY 1996, 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 72,510 requests for reference assistance from onsite and remote users (Tables 9, 10). The number of inquiries received via the Internet continues to grow.

LO is heavily involved with expanding and organizing the information on NLM's World Wide Web site to better serve the Library's users and employees. Notable LO additions to the Web site included a pointer page for finding information on health and medical topics, a Preservation and Collection Management home page, the online SERHOLD Manual, a report on the use of NLM databases in health services research, and extensive information for researchers visiting the Library or using NLM services from a distance. A newly established Intranet also provides NLM staff members with ready access to administrative and technical information that helps them do their jobs more efficiently. In FY 1996, NLM's CD-ROM network was upgraded and expanded to include more reference tools useful to onsite users and staff members. The Reference Section also installed additional software and equipment to help people with disabilities to use onsite services. Among the new tools provided for those with disabilities are

speech synthesizers and adjustable height workstations for wheelchair access

Document Delivery

LO provides copies of documents from NLM's comprehensive collection of biomedical literature as a back-up to services available from other NN/LM libraries and document suppliers. Staff also retrieve documents from NLM's closed stacks for use by onsite patrons. In FY 1996, the Library handled 584,738 requests for post-1913 documents (Table 6) and filled 5,289 requests for items from the historical and special collections (Table 10). Interlibrary loan requests for modern items increased 1 per cent and onsite requests increased 4 per cent, despite the unusually large number of days NLM was closed. Requests for Internet delivery of documents increased 57 per cent to a total of 20,384. NLM filled 64 per cent of the interlibrary loan requests it received, down 6 per cent from FY 1995 due to the impact of the shutdowns.

After a comparison of delivery times, the Library changed to FedEx for delivery of books and audiovisuals requested on interlibrary loan. This change has improved delivery time to west coast libraries and reduced NLM staff processing time. Based on the results of a cost study, NLM will raise its ILL charges by \$1.00 effective October 1, 1996.

As part of NLM System Reinvention, the Relais system has been purchased to provide enhanced automated support to the Library's internal document delivery operations. Relais will be customized to interact with other NLM systems and to accommodate NLM's unique requirements. NLM's goal is to deliver documents electronically within 24 hours of receipt of request. The Relais system uses scanners instead of photocopiers to capture images of the requested document that can then be sent to the requester by Ariel, electronic mail, fax, or postal mail. The images will not be retained by NLM. LO's document delivery procedures will be modified to make optimum use of Relais features.

At the close of the fiscal year, 2,875 libraries were using DOCLINE, NLM's automated document request and routing system. These libraries entered 2.9 million document requests into DOCLINE in FY 1996, which is roughly equivalent to the FY 1995 traffic. Ninety-three per cent of the requests were filled. DOCLINE routes requests based on the 1.4 million holdings statements for 3,181 health sciences libraries, primarily in the United States and Canada. During FY 1996, progress was made on improving the procedures for updating SERHOLD to reduce contention with other internal processing systems and to allow more frequent and less labor-intensive

holdings updates.

In FY 1996, NLM and the Regional Medical Libraries stepped up efforts to encourage DOCLINE participants to access the system via the Internet to reduce NLM telecommunications costs. Twenty-six percent of DOCLINE usage currently occurs via the Internet. A number of changes were made to the DOCLINE system to improve its efficiency. Users may now print request numbers in bar code format and specify additional delivery methods, such as Internet. More DOCLINE statistics are available to participants via ftp, and the Regional Medical Libraries can update DOCLINE participant information via the World Wide Web.

National Network of Libraries of Medicine

The purpose of the NN/LM is to provide U.S. health science practitioners, investigators, 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 geographic location or institutional affiliation. The network has more than 4,500 members, including health sciences libraries of every size and type located throughout the country. NLM's NN/LM Office oversees the network programs which are coordinated and administered by Regional Medical Libraries (RMLs) in eight geographical regions. New RML contracts for 1996-2001 were awarded in April 1996 following a competitive process that began in FY 1995. All of the incumbent contractors received new awards. (See Appendix 1 for a list of the RMLs.)

The National Information Infrastructure offers the network new and exciting opportunities to improve information services to health professionals in 1996-2001. The RMLs are increasing their efforts to facilitate connectivity to the Internet for member libraries and health professionals and are using the Internet to develop innovative services. The NN/LM Office and the RMLs communicate regularly via e-mail and audio-teleconference to ensure that new and enhanced programs and services are piloted in different regions and, if successful, introduced throughout the country.

The NN/LM program is a core component of NLM's outreach initiative. The RMLs and other NN/LM members have developed and implemented many projects to reach out to underserved health professionals in both rural and inner-city areas. In a recent example, the first special enhancement project under the new NN/LM contracts was awarded to the Pacific Northwest RML to develop a model for planning and evaluating information outreach in minority communities. The project will work with

evaluation specialists and groups in the northwest concerned with Native American health and will develop benchmarks and test strategies for improving the outcomes for health information outreach

During the new contract period, the RMLs will collaborate with the Centers for Disease Control and Prevention and several NLM components to expand outreach efforts to public health professionals. This effort, which will build on the public health outreach experience of the RMLs in the Middle Atlantic, Greater Midwest, South Central, and Pacific Northwest Regions, is another follow-on to the meeting on public health and the National Information Infrastructure held at NLM in April 1995.

The RMLs 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 and at some distant meetings focused on health services research. In FY 1996, NLM and NN/LM services were displayed at about 82 exhibits at national, regional, and state association meetings across the country.

Special Onsite Programs

In addition to reference and document delivery services, NLM offers a variety of special programs and services to those who visit the Library in Bethesda, including guided tours, briefings on NLM's operations and new services, and historical exhibits, seminars, and symposia. The Library also conducts a 1-year post-master's training program for librarians with potential for leadership in health information programs and services.

Public Tours and Briefings

NLM attracts many domestic and international visitors interested in biomedical communication, information technology, health sciences librarianship, and the history of medicine. In FY 1996, LO staff members conducted 127 regular daily tours for a total of 488 visitors. The Office of Public Information (Office of the Director) arranged 95 special group tours and briefings for 1,779 visitors and arranged a press conference in conjunction with the public release of Internet Grateful Med in April 1996 (see chapter on High Performance Computing and Communications).

Historical Programs

The History of Medicine Division hosted a 1-day conference on "Medicine in the Indian Princely

States" on December 1, 1995 and a symposium, "Public Health, Demography, and American Medicine", on May 22, 1996, in honor of James H. Cassedy's contributions to the NLM and to historical scholarship. The symposium was co-sponsored by the Washington Society for the History of Medicine.

HMD, the NLM EEO Office, and the NLM EEO Committee sponsored two historical lectures. Evelyn M. Hammonds, Assistant Professor of Science, Massachusetts Institute of Technology, gave a talk on "Race, Gender, and AIDS: An Historical Perspective" on February 29, 1996. This was the most recent of an annual series of NLM African American History Month lectures. Mary Fissell of the Johns Hopkins Institute of the History of Medicine discussed "Vernacular Bodies: Popular Medical Works for Women" in NLM's first Women's History Month lecture on March 14, 1996. HMD also arranged for several historical seminars, including "In Search of Dr. George E. Burch and 20th Century Cardiology" by Vivian Burch Martin, "Louis Pasteur: Scientific Method and Ethical Quandaries," by Gerald L. Geison, and "To the Book and to the Letter: Exhibiting Books, Manuscripts, and Documents" by Lou Storey.

The HMD staff installed two historical exhibits in the NLM lobby in FY 1996: "Disease and Death in the Neighborhood: Medical Maps of Washington, D.C., 1878 to 1909" and "Extraordinary Objects, Extraordinary Stories: Celebrating the NLM Collections." Staff members also published the results of their scholarship in a number of books, chapters, articles, and reviews during the past year.

NLM Associate Program

The Library has a 1-year competitive training program that 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. Three Associates participated in the program in 1995/1996, including one from the Czech Republic. Among the projects undertaken by the 1995/96 Associates were a study of potential additional sources of literature suitable for the SPACELINE database, the development of a mechanism for conducting comparative analyses of results returned by different retrieval interfaces, and work on two issues of *Current Bibliographies in Medicine*.

The international Associate returned to work in her home country, and the two 1995/96 U.S. Associates accepted jobs at the National Institutes of Health Library and the Kevric Company. Four new U.S. Associates began the 1996/97 program in September 1996. Effective in September, the NLM

Associates are part of a special educational program administered on the Library's behalf by the Oak Ridge Institute for Science and Education and are no longer directly employed by NLM. This change does not affect the scope and purpose of the Associate program.

Customer Service Policy

The Library's Mission

The National Library of Medicine collects, organizes and disseminates the biomedical literature of the world in order to advance the medical and related sciences and to improve the public health. The Library serves as a national information resource for research, health care, the education of health professionals, and service activities of Federal and private agencies, organizations, institutions, and individuals.

The Scope of the NLM Collection

NLM's collection is intended to serve the information needs of U.S. health professionals, but access to information in the collection is available world wide. The Library comprehensively collects the world's substantive biomedical publications to support its mission. Although the importance of consumer health information is recognized, and the public may use the Library's information services, the Library does not collect information that is written for the public.

NLM's Customers

Our primary customers include: health science librarians and information specialists, health care providers, researchers, scholars and students, and historians.

Customer Service Values

- We will treat all of our customers with courtesy.
- Our customers will be assisted by knowledgeable staff.
- We will provide information and instructional materials to assist our customers in using our products and services.
- We will meet or exceed our established deadlines for providing customer service.
- Periodically, we will assess our customers' needs and level of satisfaction with our service.

Customer Service Commitment

Our goal is to respond to your inquiry according to the following standards:

- If you call the Library using our 800 numbers, staff will be available to assist you during business hours. Our business hours will be prominently posted and noted in our brochures and through other information sources.
- If you mail or fax your request, you will be sent a response within ten working days after we receive your inquiry.
- If you contact us via e-mail, you will be sent a response within five working days after we receive your inquiry.

If we cannot provide a full response within these time frames, you will receive confirmation that we received your request, with an estimate of the delay before we can provide the full reply.

If you visit the Library, you will receive information on how to use our collections and services from staff in our Reading Rooms. Staff will be available to assist you during business hours, which will be prominently posted and noted in our brochures.

Table 2**Growth of Collections**

<i>Collection</i>	<i>Previous Total (9/30/95)</i>	<i>FY 1996</i>	<i>New Total (9/30/96)</i>
<i>Book Materials</i>			
<i>Monographs:</i>			
Before 1500	576	0	576
1501-1600	5,787	6	5,793
1601-1700	10,120	4	10,124
1701-1800	24,441	5	24,446
1801-1870	41,106	15	41,121
Americana	2,341	0	2,341
1870-Present	613,777	15,322	629,099
Theses (historical)	281,794	0	281,794
Pamphlets	172,021	0	172,021
Bound serial volumes	1,049,964	22,361	1,072,325
Volumes withdrawn	(57,697)	(160)	(57,857)
Total volumes	2,144,230	37,553	2,181,783
<i>Nonbook Materials</i>			
<i>Microforms:</i>			
Total microforms	394,084	37,222	431,306
Reels of microfilm	79,639	3,900	83,539
Number of microfiche	314,445	33,322	347,767
Audiovisuals	58,722	2,001	60,723
Computer software	776	81	857
Pictures	56,601	0	56,601
Manuscripts	2,454,542	0	2,454,542
Total nonbook	2,964,725	39,304	3,004,029
Total book and nonbook	5,108,955	76,857	5,185,812

Table 3**Acquisition Statistics**

<i>Acquisitions</i>	<i>FY 1994</i>	<i>FY 1995</i>	<i>FY 1996</i>
Serial titles received	23,250	22,600	22,522
Publications processed:			
Serial pieces	154,076	154,342	154,146
Other	22,569	22,743	20,846
Total	176,645	177,085	174,992
Obligations for:			
Publications	\$4,456,480	\$4,788,181	\$5,102,058
Included for rare books	(\$207,575)	(\$189,527)	(\$209,178)

Table 4

Cataloging Statistics

	<i>FY 1994</i>	<i>FY 1995</i>	<i>FY 1996</i>
Completed Cataloging	19,556	22,232	20,365

Table 5

Bibliographic Services

<i>Services</i>	<i>FY 1994</i>	<i>FY 1995</i>	<i>FY 1996</i>
Citations published in MEDLINE	367,877	392,354	322,825
For <i>Index Medicus</i>	351,958	374,907	309,038
Recurring bibliographies.....	12	6.....	6
Journals indexed for <i>Index Medicus</i>	3,127	3,093.....	3,205
Abstracts entered	274,514	293,724.....	242,544

Table 6

Circulation Statistics

<i>Activity</i>	<i>FY 1994</i>	<i>FY 1995</i>	<i>FY 1996</i>
Requests Received	539,988	573,541	584,738
Interlibrary Loan.....	324,670	345,428.....	347,992
Onsite	215,318	228,113.....	236,746
Requests Filled:	410,453	424,169.....	410,732
Interlibrary Loan.....	229,949	245,078.....	227,810
Photocopy.....	217,627	231,993.....	215,461
Original	10,864	11,384.....	10,814
Audiovisual	1,458	1,701.....	1,535
Onsite	180,504	179,091.....	182,922

Table 7**Online Searches***

DATABASES	FY 1994	FY 1995	FY 1996
AIDSDRUGS	1,288	1,453	1,240
AIDSLINE	54,596	66,200	59,507
AIDSTRIALS	2,319	3,148	2,203
ALERT	2,027	3,062	3,244
AVLINE	27,269	27,543	22,712
BIOETHICS	15,075	15,248	14,352
BIOTECHSEEK	725		
CANCERLIT	88,077	91,324	80,206
CATLINE	363,805	390,067	382,321
CCRIS	4,048	3,168	2,697
CHEMID	10,608	9,970	9,151
CHEMLINE	16,588	13,075	12,150
DART	3,061	3,861	2,556
DENTALPROJ	146	142	72
DIRLINE	17,064	33,427	27,602
DOCUSER	14,353	17,699	14,664
EMIC	3,577	3,787	3,743
EMICBACK	2,995	2,039	1,566
ETICBACK	896	1,046	666
GENETOX	1,808	1,913	2,935
HEALTH	186,701	178,295	93,183
HEALTHSTAR			76,857
HISTLINE	4,374	6,806	6,577
HSDB	35,767	30,296	24,947
HSRPROJ	93	1,266	1,756
HSTAR	27,789	41,670	27,263
IRIS	21,453	17,593	15,870
LOAN STATUS	14,495	19,939	21,304
MEDLINE	4,989,911	5,262,329	5,414,712
MeSH	37,221	38,058	39,138
PDQ	23,794	20,303	17,404
POPLINE	24,610	22,864	19,731
PREMEDLINE			3,001
REFLINE	42,518	39,984	33,541
RTECS	14,685	13,808	12,159
SDILINE	44,259	218,654	188,418
SERLINE	418,162	633,615	664,265
SPACELINE		1,507	3,119
STORED SEARCH	336	108	162
TOXLINE	69,944	71,631	66,761
TOXLIT	11,766	11,074	9,954
TRI	21,320	12,793	8,959
TRIFACTS	577	511	441
USERS	3,912	3,701	3,911
YEAR86	2	4	0
Total	6,625,774	7,334,981	7,397,020

*Beginning in FY 1995, figures include off-line searches.

Table 8

Off-line Searches*

<i>DATABASES</i>	<i>FY 1994</i>
AIDSLINE.....	1,710
AVLINE.....	75
BIOETHICS.....	5
CANCERLIT.....	3,177
CATLINE.....	504
CHEMLINE.....	0
DENTALPROJ.....	0
HEALTH.....	9,943
HISTLINE.....	6
MEDLINE.....	3,733
MeSH VOCABULARY FILE.....	0
POPLINE.....	4,692
SDILINE.....	210,612
SERLINE.....	3
TOXLINE.....	3,557
TOXLIT.....	2,234
Total	240,251

*Off-line searches are no longer being counted separately.

Table 9

Reference Services

<i>Activity</i>	<i>FY 1994</i>	<i>FY 1995</i>	<i>FY1996</i>
Reference Section:			
Offsite requests.....	22,706	25,135	24,799
Onsite requests.....	42,482	43,132	34,796
Total	65,188	68,267	59,595

Table 10

History of Medicine Activities

Activity	FY 1994	FY 1995	FY 1996
Acquisitions:			
Books	126	154.....	188
Modern manuscripts	193,725	184,000.....	163,500
Prints and photographs	313	2,046.....	183
Historical audiovisuals	61	147.....	108
Processing:			
Books cataloged	340	2,451.....	351
Modern manuscripts processed	0	0.....	0
Pictures cataloged	1	0.....	0
Citations indexed	3,697	2,510.....	1,363
Public Services:			
Reference questions answered.....	13,516	13,434.....	12,885
Onsite requests filled	7,866	7,273.....	5,289

SPECIALIZED INFORMATION SERVICES

Melvin Spann, Ph.D.
Associate Director

The Division of Specialized Information Services (SIS) has been increasingly responsive to demands for up-to-date information in toxicology and environmental health and in HIV/AIDS. The evolution of the Toxicology and Environmental Health Information Program (TEHIP) has resulted in the development of new and easier methods for information access and expanded sources. As public concern with exposure to environmental hazards has grown, SIS over the past three decades has implemented the use of state-of-the-art computer and communication technologies to meet concomitant demand for information. Use of these new technologies has allowed the Division to reach beyond its traditional user base and traditional information products.

The Division has created a World Wide Web server to distribute information more effectively, efficiently and at lower cost. The graphical nature of the WWW allows the Division to provide access to additional types of information as well as to link information facilitating its use.

We anticipate further expanding services in response to recommendations from the Institute of Medicine's panel examining ways to improve the utility of the TEHIP databases. The Division's AIDS activities also continue to expand as the outreach program in HIV/AIDS connects with more of the Historically Black Colleges and Universities and community-based organizations.

TOXNET

The Toxicology Data Network (TOXNET), the Library's networked microprocessor system, had two major developments in FY 1996: discontinuation of VAN access to TOXNET and implementation of a new version of the Remote Data Entry System (RDES). Searching TOXNET databases is mainly done now through FTS 2000 and the Internet. Members of the Scientific Review Panel now use the new RDES software to do online commentary and review of the Hazardous Substances Data Bank (HSDB). In addition:

- A feasibility study to investigate mounting the National Cancer Institute's voluminous PHS-149

data (i.e., chemical carcinogenicity studies) as an online TOXNET database was completed.

- Work continued on the further enhancement and implementation of the Windows workstation for building and updating HSDB records.
- Work also continued on researching alternative configurations and costs for possible new hardware and operating system platforms for the TOXNET system.
- More work has been done to implement a fully associative graphical user interface using the FORMS Software embedded in the experimental SIS server. This will facilitate literature searches by users who also need to search ELHILL databases such as MEDLINE and TOXLINE.

Databases under TOXNET

The **Hazardous Substances Data Bank (HSDB)** continues to be the most highly used data bank on TOXNET, averaging over 10,000 searches each month. Increased emphasis was placed on providing more data on human toxicology and clinical medicine within HSDB, in keeping with the recommendations of the Board of Regents' Subcommittee on TEHIP. Additional recommendations are anticipated from the Institute of Medicine Evaluation Study. Changes to the composition of the Scientific Review Panel (SRP) are being made to accommodate the anticipated shift in content emphasis. Newer sources of relevant data have been added to maintain the currency for the 4,500 HSDB records. Because of increased staff efforts, 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. SRP members, in collaboration with SIS staff are exploring new approaches for data presentation for the HSDB metal records.

The **Toxic Chemical Release Inventory (TRI)** series of files now includes eight online files, TRI87 through TRI94. These files remain an important resource for environmental release data and continue to attract new users. Mandated by the Emergency Planning and Community Right-to-Know Act (Title III of the Superfund Amendments and Reauthorization Act of 1986), these EPA-sponsored databases contain environmental release data to air, water, and soil for almost 400 EPA-specified chemicals. Starting with the TRI91 file, the reporting facilities were required to report source reduction and recycling activities, in addition to environmental releases. These additional reporting requirements

have considerably increased the size and complexity of the databases.

In spite of the additional data, the TRI94 file was released in June, 1996. This relatively early release date was made possible by increased TOXNET system efficiencies, and improvements in data handling at the EPA. EPA is planning to include over 280 additional chemicals in 1996 and require submissions from federal and military installations. Another major change in TRI95 will be two versions of the reporting form, a long and a short, requiring unit record changes. 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, who 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 on NLM's TOXNET system. This data bank is supported by the National Cancer Institute and has grown to over 7,300 records. The chemical-specific data covers the areas of carcinogenesis, mutagenesis, tumor promotion, and tumor inhibition studies.

The **Integrated Risk Information System (IRIS)**, EPA's official health risk assessment file, continues to experience high usage on TOXNET. Beginning with version 6.5, Grateful Med contains form screens to facilitate the searching of IRIS by inexperienced users. Creating a search mechanism with the TOXNET Concept Menus is also under consideration. 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 3,000 chemicals. GENE-TOX receives a high level of interest among users in other countries where mutagenesis is a factor in regulations.

The **Registry of Toxic Effects of Chemical Substances (RTECS)** is a data bank 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. NIOSH has announced that the full RTECS will no longer be available as a printed or microfiche product, so online access via NLM is now even more important. RTECS contains information on over 135,000 chemicals.

The **Developmental and Reproductive Toxicology (DART)** database now contains over 30,000 citations from literature published since 1989 on agents that may cause birth defects. This year, a significant effort was made to add relevant technical reports to DART. Records from DART are also added to TOXLINE on a quarterly basis. DART is a continuation of the Environmental Teratology Information Center backfile (ETICBACK) database, which contains almost 50,000 citations to literature published from 1950-1989. ETICBACK citations are also found in TOXLINE. DART is funded by the EPA, the National Institute of Environmental Health Sciences, and the FDA's National Center for Toxicological Research. NLM provides funding and management for the project.

The **Environmental Mutagen Information Center (EMIC)** database contains over 14,000 citations to literature published since 1992 on agents that have been tested for genotoxic activity. A backfile for EMIC (EMICBACK) contains 70,000 citations to the literature published from 1950-1991. 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 National Institute of Environmental Health Sciences and managed by NLM.

Databases under ELHILL

CHEMID (Chemical Identification File) is an NLM online chemical dictionary that contains 290,000 records, primarily describing chemicals of biomedical and regulatory importance. It also contains an important set of regulatory data, collectively known as SUPERLIST. More than 12,000 records are augmented with the name and an indication of source for chemicals mentioned in one or more of 33 lists, e.g., the Department of Transportation Hazardous Materials List and 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 1996 new file generation programs allowed more frequent updates to ChemID and thus more currency for users. Data such as Carcinogenic Status from regulatory sources were added to the Classification Code field and were made searchable for the first time. During FY 1997, the number of lists in SUPERLIST should increase to over 35 and the file size of ChemID will grow to 300,000 records, reflecting increasing coverage of biomedical and regulatory sources.

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 (CAS) from its Registry System, and this is augmented extensively by NLM with nonproprietary data from a variety of sources. CHEMLINE now contains 1,550,000 records on chemical substances. In 1996, efforts to use the new ChemID file generation techniques for CHEMLINE were studied, and analysis will continue into 1997.

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. The Chemical Abstracts Service requires usage royalties, therefore, information from this source is used to create two separate online bibliographic files, TOXLIT and TOXLIT65. The four databases in the TOXLINE family of services now contain over four million records. Approximately 20,000 records are added with each monthly update. During FY 1996, SIS added a new subfile from the Swedish National Chemical Inspectorate's file, RISKLINE. We continue to explore ways that the Unified Medical Language System can be utilized to accommodate vocabulary changes in TOXLINE subfiles.

DIRLINE (Directory of Information Resources Online) is NLM's online directory of organizations, databases, projects, and programs which focus on biomedicine. These resources provide information to users that may not be available from any other NLM database. A new version of *Health Hotlines* (a listing of organizations from DIRLINE with toll-free phone numbers) was produced which will also be available through a simple search interface on the SIS Home Page.

AIDS Services

AIDSTRIALS and AIDS DRUGS are the databases generated as part of the Public Health Service's AIDS Clinical Trials Information Network. This year citations to published results of clinical trials started being added to the AIDSTRIALS records. This information, along with the citation number from AIDSLINE, facilitates access to result information for users.

NLM awarded 22 AIDS Information Outreach Purchase Orders in FY 1996. Seven awards were made to previous award recipients to allow them to continue or expand successful projects. The remaining 15 awards were for new projects. NLM's collaborative project with the Wheaton Regional Library, Montgomery County, Maryland, has provided NLM with interesting information on the use of NLM's resources and health information by the general public. Very useful relationships have developed between other local government agencies and the library. The introduction to the public of access to the World Wide Web and training in how to use this capability has also been a very exciting part of this project.

A new NLM AIDS home page was developed as part of the overall development of the SIS World Wide Web service. In addition to access to NLM's AIDS-related databases and other resources, NLM staff worked with the sponsors of the XI International Conference on AIDS to quickly create a database which debuted the day the conference started.

Other Programs

Internet

The SIS World Wide Web Server began to offer a greatly expanded panoply of data to Internet users. Among the new offerings were chemical structures displays for substances in the AIDS DRUGS file, searching the abstracts of the XI Conference on AIDS, new WWW versions of the SIS demo and slide overview packages which use the latest JAVA script and frames capabilities of WWW browsers, and a conversion of all SIS Fact Sheets to a hypertext (HTML) format. While the SIS TEHIP and AIDS Internet Gophers remain open to users, increasingly data from these sources are being converted to WWW format to allow use of the more flexible graphical capabilities of the Web. In 1997, additional searching capabilities will be made available on the SIS Server, including searching HSDB chemicals by structure, substructure, and similarity, as well as searching DIRLINE records to allow public access to this useful collection of information resources.

New WWW initiatives to commence in FY 1997 include the development of an operational Web-based user interface to access the TOXNET system and a collaborative research participation program with the Artificial Intelligence Group at the University of Arizona to automatically generate a

Toxicology and Environmental Health Concept Thesaurus from the TOXLINE database

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 project, NLM has collaborated with the Agency for Toxic Substances and Disease Registry (ATSDR) to train representatives from 57 additional schools in the use of NLM's valuable online resources. During the past Fiscal Year, one of the training classes was hosted by Howard University and it included HBCUs 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. During the past year, new PC-based workstations were provided to the nine original participating HBCUs.

SIS has also implemented an AIDS Information Training component to the existing Toxicology Information Outreach Project. The goal of this project is to provide training in the use of available information resources in HIV/AIDS for faculty, staff, students, and community providers associated with the HBCUs. NLM is working with the National Association for Equal Opportunity in Higher Education (NAFEO) and the National AIDS Minority Information and Education Program to

identify participants. A number of training classes were held during the year, including one at Howard University for attendees at NAFEO's annual conference.

User Support Computer-Based Activities

The main focus of our computer-based training activities over the past year was to develop a demonstration of the TEHIP databases and a TEHIP slide presentation with access from the WWW. The TEHIP demonstration gives a brief description and several sample searches for each database. The slide presentation covers the historical development of the program and the status of current projects. Both products are currently available on the SIS WWW server and use the latest Java script and frames capabilities of Web browsers. In addition, SIS completed development of the introductory module of TOXICOLOGY TUTOR, a computer-based toxicology course. The course was developed under TOOLBOOK for use in a Windows environment. Currently, TOXICOLOGY TUTOR is being revised for release on the WWW.

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 through the SIS WWW server and gopher.

LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS

Harold M. Schoolman, M.D.
Acting Director

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

A Board of Scientific Counselors meets to review the quality and contents of the intramural research programs within the Lister Hill Center. The Board is composed of scientific and technical experts (see Appendix 5 for a list of members) who are prominent leaders in the fields of medicine, computer science, engineering, and health professions education.

In FY 1996 a new organizational element was added to the Lister Hill Center: the Office of High Performance Computing and Communications. This office includes the Library's HPPC-related programs that were originally part of the HPPC National Coordinating Office when it was located at the NLM. The major activities of that Office are in the Chapter on High Performance Computing and Communications.

The Center is now composed of an Office of the Director, the Office of High Performance Computing and Communications, and four branches:

- Computer Science Branch
- Communications Engineering Branch
- Cognitive Science Branch
- Audiovisual Program Development Branch

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

- Computer and information science as applied to the problems of the Library, of Biomedical research, and health care delivery,
- Biomedical image engineering, including image acquisition, processing, storage, retrieval, and

communications, and

- Use of computer and image technologies for health professions education

Computer Science Branch

Research projects of the Computer Science Branch concentrate on applying artificial intelligence techniques to problems in the representation, retrieval and manipulation of biomedical knowledge. CSB projects involve both basic and applied research in such areas as expert systems, intelligent database systems, multimedia hypertext information delivery, machine learning, and machine-assisted indexing for information classification and retrieval. The research addresses issues in knowledge representation, knowledge base structure, knowledge acquisition, 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 data banks, and multimedia knowledge-based systems with interactive video capability.

Branch staff members participate in individual and team research projects. Several are principals in the User Access Services project of NLM's System Reinvention program and in the development of the Metathesaurus and the Information Sources Map of NLM's Unified Medical Language System initiative. They participate in the medical informatics and information science research communities and professional specialty societies. Several staff members 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 Chief Dr. Lawrence Kingsland and computer specialist Ms. May Cheh co-direct the 8-week NIH elective in Medical Informatics for third- and fourth-year medical students each spring.

User Access Services Project

The User Access Services Project applies techniques of computer science, information science, artificial intelligence, and networked distributed computing to the problems of helping users gain access to the information in NLM's multimillion record databases. The members of this group under project leader Dr. Lawrence Kingsland have developed a new program for assisted searching in NLM's online databases over the World Wide Web: Internet Grateful Med (IGM), an important

component of NLM's System Reinvention initiative, was released to the public in April 1996. 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.

Tens of thousands of physicians have been introduced to Internet Grateful Med through outreach programs of the National Network of Libraries of Medicine and professional specialty conferences such as the annual conferences of the Radiological Society of North America, the American College of Physicians, and the American Medical Informatics Association. The program has been a considerable success. The initial version of Internet Grateful Med searched only in MEDLINE. A much-needed online self-registration process was added to the system in June 1996. By the end of FY 1996, the self-registration system had been used by more than 4,200 users from 52 countries to establish new individual accounts for searching in NLM's databases. Searching in AIDSLINE, HealthSTAR, and the new PREMEDLINE database was added to Internet Grateful Med early in September 1996. Internet Grateful Med won one of several Federal Showcase Site awards at the Federal Webmaster Workshop in August, 1996.

By the end of FY 1996, five months after its release, usage of the Internet Grateful Med server was running at a rate of 1.3 million World Wide Web connection requests from users at nearly 42,000 individual computers on the Internet each month. Weekdays in late September, the server handled about 60,000 Web connections from users at nearly 2,400 individual computers on the Internet each day. In the aggregate, Internet Grateful Med users made nearly 80,000 searches that month in MEDLINE or the AIDSLINE, HealthSTAR or PREMEDLINE databases.

During FY 1997, the Internet Grateful Med development team will add searching capability to a number of additional NLM databases to the system. Linkages to additional systems will be expanded, and means of identifying and searching in multiple databases from a single user query will be explored.

Expert Systems Program

The objective of the Expert Systems Program is to create computer systems that facilitate access to medical knowledge. The Expert Systems group has built a multimedia expert system shell

called CTX, for Criteria Table Expert. This year, progress was made in the design and development of a World Wide Web version of the system to allow Web-based delivery of expert systems based on the criteria table knowledge representation pioneered by NLM researchers.

The Expert Systems Program's flagship research vehicle is the AI/RHEUM knowledge based expert system for diagnosis in rheumatology. This system was improved with the inclusion of additional video images for physician education and new text descriptions to provide physicians with expanded information resources at the conclusion of the system interaction. Update of the system's disease knowledge base has continued with the collaboration of expert rheumatologists Dr. H. James Williams of the University of Utah and Dr. Balu Athreya of Thomas Jefferson University.

The Expert Systems Program continues to participate in exploring and defining the issues of evaluation of knowledge based systems. A 2-day Workshop on the Evaluation of Knowledge Based Systems was held at NLM December 7-8, 1995. The workshop was co-sponsored by NLM and the National Science Foundation, the Agency for Health Care Policy and Research, and Microsoft Research.

Unified Medical Language System Metathesaurus

Each year more than 85% of the UMLS Metathesaurus effort goes into creating the new release for the annual edition of the UMLS Knowledge Sources. Developing the 1997 Metathesaurus, with a very significant advance in content, was the main achievement of Dr. William Hole's Metathesaurus production group this fiscal year. The new Metathesaurus includes 331,756 concepts (78,864 more than in 1996, a 31.2% increase) and 740,170 distinct names for these concepts (197,062 more, a 36.2% increase). All are linked to 827,362 entries in 35 vocabularies, thesauri, or classifications (an increase of 237,964 or 40.37%).

Major additions for 1997 included the complete Snomed International, the complete AMA CPT96, and the complete official ICD9-CM including fully specified names for the 5th digit ICD9-CM terms. One third of the Read Clinical Classification (the standard for medical coding in the United Kingdom) was also added.

Under NLM review and supervision, most vocabulary inversion and insertion is done by contractor Lexical Technology, Inc. The majority of the editing is done on remote client systems by contractor Kevric, Inc. Dr. William Hole and Ms. Tammy Powell of the McSH Section, Library Operations, are project officers for these contractors.

Major progress was made this year on the redesigned Metathesaurus Enhancement and Maintenance Environment software system ("Meme2"). Initial complete server and client versions were created and tested. Final testing and switchover to the new system is planned for spring 1997. This system allows editors to perform editing actions at the same level at which editors think, rather than in unnecessarily complex detail. It will store actions to allow analysis and the "do" and "undo" of algorithmic and human editing steps. It will address fundamental limits of relational database management system (RDBMS) performance encountered in use of the present system, and will form the basis for integrated editing of multiple vocabularies within a common environment. The server side of Meme2 is being developed at Lexical Technology, Inc. The client design and testing are being done by NLM. Alternative RDBMS environments (including an Oracle version) are being built and tested at NLM.

Unified Medical Language System Information Sources Map

The UMLS Information Sources Map (ISM) is an attempt to develop a software environment that will accept a biomedical query, identify relevant network-accessible information resources, and (where possible and appropriate) connect to those resources and retrieve helpful information. The approach to this problem is two-tiered: identify resources, then identify and retrieve information within those resources. Although progress has been made on both tiers, the first has received more attention thus far. Identification of resources follows the classical library model: the nature and content of each resource is described by the equivalent of a card catalog entry (the ISM record).

FY 1996 was a busy year for the ISM Project team (Dr R.P.C. Rodgers, Dr Kevin Scott, and Mr Lawrence Wright). Development work was shifted into the new object-oriented Java language, which is well suited for World Wide Web-based information systems. With the collaboration of medical librarians Ms. Betsy Humphreys and Ms. Jana Bradley, the ISM record structure and content was extensively revised. ISM records were instantiated within a set of relational tables running under the Oracle database system. The Oracle database was set up in such a way as to make it available to the UMLS Metathesaurus and the UMLS Semantic Network developers as well.

A completely new version of Sourcerer, the prototype ISM user interface, was presented. This new prototype accepts natural language queries, then

employs tools developed by Dr Alan Aronson and Mr. Thomas Rindfleisch of the Cognitive Sciences Branch to identify phrases within the query and subsequently to identify UMLS Metathesaurus concepts within those phrases. Concepts highlighted by the user are then used to search the ISM database for relevant resources. The ISM team created an interface framework based on the Common Object Broker Request Architecture (CORBA) for both the ISM database and the UMLS Knowledge Source Server, and began working with the other UMLS components to make this approach available to them as well.

With the assistance of Ms. Betsy Humphreys and her colleagues in the Division of Library Operations, real-world biomedical queries from the NIH and from the Oregon Health Sciences University were assembled as test sets to assist in the design and evaluation of Sourcerer. Strategies were developed for creating larger test sets.

With the collaboration of the Corporation for National Research Initiatives, an experimental "handle server" was established at NLM. The handle server will be used for experiments in the use of the Uniform Resource Name, an identifier attached permanently to an information object, in place of the Uniform Resource Locator ubiquitous in World Wide Web documents today.

HyperDOC, the initial version of NLM's principal Web server and home page, was created several years ago by Dr Rodgers. This was one of the first Web servers in the Federal government, and one of the more complex and interesting early Web sites. In FY 1996, Dr Rodgers helped to transition this system into a full-fledged NLM service encompassing all divisions of the Library. This included recruitment of a new technical support person to work within NLM's Office of Computer and Communications Systems. Dr Rodgers also made public an online multimedia version of the Islamic Medicine exhibit that was on view at NLM in 1994. He was asked to advise the NIH-wide group that is now developing a Web-based system for disseminating NIH's consumer health information.

As a member of both the International World Wide Web Conference Committee and the Paris Program Committee, Dr Rodgers participated in the Fifth International World Wide Web Conference in Paris, France May 4-12, 1996. At that conference he organized and co-chaired an expert panel titled "Multicasting & Real-Time Applications and the Future of the Web." This was the first presentation at a major international meeting to use Internet-based teleconferencing: half the panelists participated from four different sites in California, the other half were present in Paris. Dr Rodgers

chaired the session on visualization applications at the Federal Webmaster's Workshop in August 1996 and he made a presentation during this session about NLM's Visible Human Project

Full Text-Online Reference Works (ORW) Project

Mr. Charles Goldstein, long-time Chief of the Information Technology Branch (ITB), retired from the Federal service during FY 1996. The staff and projects of ITB were integrated into the Computer Science Branch. The primary efforts in the Full Text-Online Reference Works project during FY 1996 focused on improving the Web-based client (and interface) to the retrieval system by which documents from the Health Services/Technology Assessment Text (HSTAT) collection are provided online. In addition, efforts were made to simplify the addition of new data by revising the Standard Generalized Markup Language (SGML) encoding requirements and providing Document Type Definitions to the data providers tailored to the style of report or document to be submitted.

Approximately 10 megabytes of new databases (exclusive of the graphics associated with them) were added to HSTAT, a 20% increase in the size of the available online collection. These include Agency for Health Care Policy and Research clinical guidelines and NIH Consensus Development Reports.

An agreement was made with the Center for Substance Abuse Treatment of the Substance Abuse and Mental Health Services Administration (SAMHSA/CSAT) to include their Treatment Improvement Protocols in the collection. Approximately 90% of the data received for inclusion in HSTAT prior to FY 1996 has been modified to take advantage of the features of Hypertext Markup Language 3.x, especially the ability to render tabular data in a reasonable and relatively readable manner in the Web browser.

Machine Learning Project

As reported in *NLM Programs and Services* in past years, the Lister Hill Center has been exploring the application of machine learning technology to biomedical problems in an activity known as the Machine Learning Project. The underlying motivation for this work arises from the explosion of available biomedical information and the less well acknowledged explosion of analytical tools and techniques that can be applied to the information.

During FY 1996, the Machine Learning Project has entered into collaborations with

researchers around the world to apply machine learning technology generally to problems of biomedical significance. Included in this work are applications to predicting the lifespan of HIV-infected individuals, understanding protein folding, and new research in modeling of complex assemblies of neurons.

MedIndEx Project

Work continues on the MedIndEx Project, reported in previous years, the main objective of which is to facilitate expert indexing. The MedIndEx Project develops and tests interactive knowledge-based systems for computer-assisted indexing of medical literature currently indexed in MEDLINE using terms from the Medical Subject Headings thesaurus. Another focus of this research is to develop intelligent retrieval systems utilizing the same representations and environment of the indexing system.

In addition to the Indexer Interface currently being developed, the MedIndEx System includes software designed to assist knowledge engineers in ensuring a consistent, compact, and syntactically correct knowledge base. An evaluation of the system is in progress.

Medical Informatics Training Program

The Medical Informatics Training Program was established in February 1996 at the suggestion of the LHCBC Board of Scientific Counselors. The Program develops, coordinates, and evaluates medical informatics training at LHCBC. It provides recruitment and advising of students, the matching of students with mentors and preceptors, and the administrative support needed to bring students to NLM and provide stipends where appropriate.

Opportunities are provided for students at various stages in their careers. These training opportunities benefit both the students and the NLM researchers, providing contact for NLM researchers with bright young people interested in medical informatics.

During FY 1996, the LHCBC Medical Informatics Training Program provided training to high school, college, medical school, post-M.D. and postdoctoral students and to high school teachers and university faculty members.

Communications Engineering Branch

The Communications Engineering Branch focuses on R&D in image engineering and communications engineering. Included are 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 (OCR) and man-machine interface design applied to automated data entry.

This applied R&D is 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/x-ray image databases, and future imaging applications in support of medical educational packages employing digitized radiographic, anatomic, and other imagery. Information on Branch projects appear in our Web server at <http://archive.nlm.nih.gov>

Automating Data Entry into MEDLINE

In response to a crisis in early 1996 in the data entry of citations and abstracts of medical journal articles for the MEDLINE database, the Communications Engineering Branch developed a system, temporarily code named MARS for Medical Article Record System. MARS is designed to serve as an alternative to manual keyboarding, the traditional method of data entry followed for many years. MARS combines the keyboarding of citation data which may appear in different areas of a journal issue with scanning and the automatic text conversion of abstracts which, if keyboarded, proves very labor-intensive.

Designed to enter 500 records a day, 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, we 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.

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 sent 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.

One advantage of this networked approach is that all three operator functions (scanning, reconciling and citation entry) can be done concurrently, so that a fixed sequence of operations is not necessary. For example, citations may be keyboarded before or after abstracts are converted. The network server maintains directories in which the scanned TIFF images, the text abstract files, and the citation files are all retained till they are acted upon.

Current work is focused on increasing the level of automation in the system. Some 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, (c) providing a highlight-and-click capability to capitalize or de-capitalize words or entire text lines, and (d) changing icons, error messages, and buttons for clarity.

Research toward more comprehensive automation is planned through automated field entry by page segmentation software being developed by Mr. Daniel Le and the application of field reformatting and validation rules. To identify the bottlenecks in the current system, Dr. Susan Hauser and Mr. Lew Berman are developing instrumentation code to synchronize all workstations to provide a common time-stamp origin, to create log files, extract timing data from the log files and to analyze the recorded data.

DocView

As described in last year's *NLM Programs and Services*, this program aims to investigate the role of advanced client software that provides access

to, and facilitates delivery of, bitmapped images of biomedical documents over the Internet. While Internet transmission of documents has been implemented by systems such as Research Libraries Group's Ariel, now a popular way for libraries to exchange documents among themselves, there has been less of a focus on serving the end user. Our development of DocView is centered on the end user receiving bitmapped document images over the Internet from servers (Web, Gopher or FTP), or scanned and sent over the Internet by Ariel systems. The DocView client software runs under any of the versions of Microsoft Windows.

DocView is currently being beta tested to establish the technical feasibility of easy and intuitive document delivery over the Internet. Mr. Frank Walker, the principal designer of DocView, released the seventh version of the DocView beta program for the final phase of the ongoing beta testing with over 35 organizations in nine countries.

DocView offers a number of functions, including the ability to receive documents at the end user's desktop, forward the document (with or without added text), order a document, view monochrome images, process images (zoom, scroll, etc.); insert an electronic bookmark; copy and paste parts of a document, and manage folders or collections of documents.

Investigation into DocView's performance and usage relies partly on feedback from users (beta testers). User feedback is implemented by both a built-in Comments function as well as an electronic user survey.

One new feature added in response to user feedback was an Adaptive Communications Algorithm. In this design DocView automatically maintains a database of IP addresses of the ten most recent Ariel systems that have sent the user documents. When the user powers up DocView, the system uses this database to contact each of the Ariel systems and receives documents from any Ariel that has documents in its queue for this particular user.

Three papers describing DocView were presented at the 1994, 1995, and 1996 Integrated Online Library Systems (IOLS) meetings, and published in their proceedings. The most recent references: (1) Walker FL, Thoma GR. Internet document access and delivery. *Proc. 11th IOLS Meeting*. Medford NJ. Learned Information, Inc, May 1996, pp. 107-16 (2) Thoma GR, Walker FL. Access to document images over the Internet. In: Rada R, Ghaoui C, eds. *Medical Multimedia*; Oxford UK: Intellect, 1995, 179-93.

Workstation for Interlibrary Loan (WILL)

As described in last year's *NLM Programs and Services*, the pilot System for Automated Interlibrary Loan (SAIL), after 4 1/2 years of successful operation, was discontinued. The project evolved to a different design and produced a system called WILL. WILL is a multifunctional integrated workstation that delivers documents on demand. WILL automatically retrieves ILL requests from DOCLINE, parses the request into data fields, provides a GUI for the operator to scan requested documents, and automatically delivers document images by all three delivery modes (print for mail, fax and Internet), and updates DOCLINE with status. The only process that cannot be automated is the physical handling and scanning of the documents.

To investigate the role of WILL in interlibrary loan services, a beta test was installed by Dr. Susan Hauser at the NIH Library. Between late April and the end of September 1996 over 119,500 requests were automatically retrieved, sorted, and printed out by WILL, and status of the requests automatically updated in the request routing system DOCLINE. Since the vast majority of the requests were for photocopies of articles, for which the conventional photocopy machines were used since they are faster than printers, the most fruitful role for WILL was the automatic retrieval of requests and updating of status rather than the automated fax and Internet delivery of documents. However, over this time period some documents were indeed scanned and delivered by these modes. The beta test will continue through the next fiscal year.

WILL was described in two internal technical reports issued by the Branch in early FY 1996. These are titled WILL Advances in the management of interlibrary loan, and WILL. Design of a standalone WILL unit. Dr. Thoma presented talks on WILL at the Computers in Libraries conference in Arlington, Va., in February 1996 and the 11th Computers in Healthcare Education Symposium in Philadelphia in April 1996. An extended abstract was published: Thoma GR. Automating Medical Document Delivery. *Proc. 11th Computers in Healthcare Education Symposium*, April 1996, pp 68-9.

Image Compression

This R&D activity was proposed on May 6, 1996 by Dr. Thoma to the Board of Scientific

Counselors as part of a project called Image Storage and Transmission Optimization. The objective of the Image Compression project is to conduct study communications and compression techniques to address the storage and transmission of the highly data-intensive color cryosection image files from the Visible Human Project. The investigation is directed toward both a lossy as well as a lossless technique. The strategy is to store the data losslessly while providing access to it via the Internet at a lossiness controlled by the user.

The fundamental notion in designing the lossless compression technique is to exploit the interframe redundancy expected to exist in the images, rectangulate and extract the region of interest and eliminate the background, and apply a lossless compression technique such as Huffman coding.

The approach considered for the lossy compression work is to determine the "best" wavelet transform and optimum number of taps to decompose the images followed by a suitable quantization technique, either scalar or vector. Finally a lossless technique will be applied to the quantized data.

Following the May 6 presentation to the Board of Scientific Counselors, and their strong endorsement, an initial step was completed: the estimate of data compaction resulting from removal of the blue background. This is an important issue for both lossy as well as lossless compression, but particularly for the latter since it is not expected to yield a very high compression ratio. The region of interest was defined as the biological image segments as well as the fiducial marks on each slice. The fiducial marks, which are cross-sections of two I-shaped cables and a plastic rod, appear at the same place with respect to the body on all slices, and are included because they can be used to register the slices. The extraction method used is rectangulation around the selected areas rather than wavelength discrimination as suggested by other investigators. We found wavelength discrimination to be unreliable since the pixel values representing blue occur in areas other than the background. Rectangulation was applied to every tenth slice, i.e., 180 out of the 1,870 slices in the male dataset. Since 10 slices represent 1 cm in the male body, the size of the rectangulated segments for every tenth slice should be close to the average for every slice. The average size of rectangulated body segments was combined with the size of the rectangulated fiducial marks to give an approximate CR for the entire dataset, which was calculated to be 3.545. This means that any additional compression offered by either lossy or lossless methods would be increased by this multiplicative factor.

These items were discussed by Dr. Thoma at

the NLM-sponsored Visible Human Project Conference in a paper titled "Image compression approaches for the Visible Human dataset."

Communications Engineering Research

In light of NLM's initiatives in telemedicine, we consider R&D in communications engineering crucial. This research activity has included testing of experimental methods for Internet file transmission, satellite tests conducted with the University of California at San Francisco using the NASA Advanced Communications Technology Satellite, investigation of use of the World Wide Web for dissemination of biomedical database information, and experiments using asynchronous transfer mode technology over dedicated high-speed communications links.

This work is focused on the optimization of TCP/IP performance, both for file transfer and for interactive applications, and includes experiments both on the Internet and on dedicated communications links of opportunity. An experimental optimization method (multisocket transmission) developed in-house is being investigated.

Internet TCP file transmission experiments were conducted in 1996 between NLM and the University of Arizona, using both conventional FTP and the NLM multisocket method. Visible Human digital color files (7 megabytes) were used in this work. A similar experiment was conducted on a dedicated T-1 (1.544 megabits/sec) Very Small Aperture Terminal (VSAT) link between NLM and the University of California at San Francisco; the VSAT work also included experiments with an RFC 1323 TCP implementation. Results of the work were reported at a workshop on the Visible Human project, conducted at the 1996 SPIE Medical Imaging conference in Newport Beach, CA. In both the Internet and VSAT work, a substantial transmission rate improvement was measured by use of the multisocket method and by use of RFC 1323 TCP.

For evaluating performance over our communications test-beds, we are using the in-house-developed Medical Information Retrieval System (MIRS), which allows query and retrieval of text and images from biomedical databases. MIRS is Sun-workstation based and requires the installation of MIRS client software. The MIRS concept is being extended by investigation into use of the World Wide Web and new Java software technology for biomedical information delivery on a large class of hardware platforms.

Mr. Long and Mr. Michael Gill have begun to

evaluate ATM technology for Visible Human and digital x-ray image transmission with the establishment of a DS-3 (45 megabits/sec) link from NLM to the Texas Medical Center in Houston. Experiments are beginning on a second ATM connection, which links NLM to the ATDNET, at OC-3 (155 megabits/sec) rates. This connection also makes available to NLM use of the NASA High Data Rate Terminal at Goddard Space Flight Center, over which additional satellite-mediated experiments are planned.

Selected papers on the communications activity include (1) Long LR. Transmission of medical images over wide area networks. *Proceedings of the 1st Visible Human Conference*, Bethesda, MD, October 7-9, 1996. (2) Long LR, Gill MG, Thoma GR. High-speed satellite access to bio-medical text/image databases. *Advanced Digital Libraries '96 Forum on Research and Technology Advances in Digital Libraries*, Library of Congress, Washington, DC, May 13-15, 1996. (3) Bazill TM, Huang HK, Thoma GR, Long LR, Gill MJ. High speed image transmission via a communications satellite. *Proc SPIE Medical Imaging '96*, Newport Beach CA, p 337. (4) Long LR, Thoma GR. Medical image database access via satellite. *Proc ACTS Results Conference*, NASA Lewis Research Center, Cleveland OH. Sept 1995, VSAT section.

Document Image Analysis

As described in last year's *NLM Programs and Services*, Mr. Daniel Le developed 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. NLM has applied to the U.S. Patent and Trademark Office for a patent to protect this invention.

This research is described in several papers in the literature, most recently in Le DX, Thoma GR, Wechsler H. Classification of binary document images into textual or non-textual data blocks using neural network models. *Machine Vision and Applications*, Issue 8, 1995, 429-35.

Digital X-ray Imaging Projects

These projects have been described in *NLM Programs and Services* in previous years. One of the projects named DXPNET for *Digital X-ray Prototype*

Network serves as a vehicle to address engineering goals while creating an archive of digitized radiographic images accessible over the Internet. An initial prototype Standardized Reading Workstation (SRW) was developed on a Sun 4/260 platform, with image display on both the standard Sun monitor and a high resolution Megascan monitor. Mr. Rodney Long and colleagues have completed all planned functions for the workstation. They have completed system tests and debugged the software. As a result, the basic functional capability of the system to collect readings is available, and it is well-positioned to move to the field-test level.

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 functions for zooming and histogram equalization have also been completed, but these routines are general enough to accommodate other image processing functions. Another function being developed for the SRW is an online radiological atlas that a reader can refer to while reading the cervical and lumbar images.

To increase the effective image access and delivery speed, the storage system has been expanded beyond the optical jukebox to include a Sun Sparearray RAID system controlled by a Sun Spare 20. A performance study was done to identify the RAID level optimum for this data collection, and was published. Hauser SE, Berman LE, Thoma GR. Is the bang worth the buck? A RAID performance study. *Proc 5th NASA Goddard Conference on Mass Storage Systems and Technologies*. NASA Pub No 3340, Sept 1996, 131-40.

As an extension of the DXPNET work, our National Center for Health Statistics collaborators have proposed a future project to read 10,000 hands and knees 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. We have completed the design of a workstation to identify the optimum resolution required. The idea is to use a few score images from a study on the Pima Indians as a way to establish resolution requirements for the large NHANES III collection. The Pima Indian x-ray images have been digitized at three different spatial resolutions: 2000 x 1634, 3000 x 2400 and 3000 x 4900. The software will display these images on the high resolution monitor, allow subject experts to examine these and

record interpretations. This workstation will be completed in 1997

While the SRW and related systems were 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 focused on the development of a prototype system named MIRS (Medical Information Retrieval System), designed to enable general access via Internet to the mixed text/image database consisting of NHANES II collateral data and the x-ray images. MIRS was developed on a Sun Sparc 10 platform and uses X-Windows and TAE Plus for interface development. Access to the data is by SQL query, and the data is organized in an Illustra DBMS. An anticipated beta test of MIRS will demonstrate access to the integrated database consisting of the images, the standardized readings, and the corresponding NHANES collateral data (demographic, blood chemistry lab data, medical questionnaire data, etc.). Although beta testing of MIRS will be done over the Internet, ATM, as well as over the ACTS satellite link, it was demonstrated at the Department of Health and Human Services in Washington DC and at the Maryland Governor's Technology Showcase in Baltimore in 1996. Recent references to our design of MIRS are (1) Thoma GR, Berman LE, Long LR: Internet access to a biomedical text/x-ray image databank. *Proc 19th International Online Meeting Oxford UK. Learned Information Europe Ltd, 1995, 429-35.* (2) Thoma GR, Long LR, Berman LE: A client/server system for Internet access to biomedical text/image databanks. *Computerized Medical Imaging and Graphics* (1996), 20 4, 259-68

To increase widespread acceptance of MIRS in the future, we are designing a platform-independent Web-based MIRS system using the Sun Sparc-based system design as a foundation. The new design, written entirely in Java, provides input field validation; simple query formulating capability to automatically generate sql queries, multithreaded image download and display capability enabling multiple images to be retrieved and displayed simultaneously; and support for Web-based database access using Java Database Connectivity. An early version of this system has been demonstrated, as noted earlier

Visible Embryo Project

The goal of this project, now completed, was the development of a Web-based distribution system for the embryo images and associated text from the world-famous Carnegie Embryo Collection housed at the Armed Forces Institute of Pathology (AFIP). Mr

Low Berman led the development effort and alpha and beta test phases. The beta test group consisted of 38 participants from a wide range of disciplines: medical researchers, students, NIH staff, and database suppliers. Following the test phases, we transitioned the Visible Embryo Web site in June 1996 to a Silicon Graphics machine at AFIP. The project was funded by the National Institute of Child Health and Human Development

Engineering Laboratories

The R&D conducted by the Communications Engineering Branch relies on two laboratories which are designed, equipped, and maintained by the Branch. The Document Imaging Laboratory supports the DocView, WILL, and Document Image Analysis projects. The Image Processing Laboratory supports the investigation of image processing techniques for both grayscale and color biomedical imagery at high resolution. As described in last year's *NLM Programs and Services*, both laboratories contain high-end workstations and a variety of Ethernet, Token Ring, and other network connections.

Cognitive Science Branch

The Cognitive Science Branch (CgSB) conducts research and development in learning and information technologies. CgSB staff participate in the health professions education, computer and information sciences, and medical informatics communities. The branch maintains a World Wide Web server (www.wegsb.nlm.nih.gov) that provides information relevant to health professions education and information technologies. The server features information about CgSB research projects, resources, and monographs and provides links to related Web sites.

Research staff have developed a variety of tutorials on Internet-based resources and distance learning which are also available on the server. The CgSB Web site is updated regularly and new features continue to be added.

Research in Learning Technologies

The Learning Center for Interactive Technology

The Learning Center for Interactive Technology showcases a wide range of computer-based information and educational technologies representing stand-alone, local area networked, and wide area networked applications in the health sciences. More than 10,000 individuals representing a highly diverse population of health professionals and

students have visited the Center since its inception in 1984. The fifty-plus groups visiting The Learning Center in 1996 included university presidents, international health leaders, business representatives, foreign exchange students, medical and nursing informatics students, students from local elementary and high schools, and Elder Hostel members (citizens 65 years and older).

Visitors this year were especially interested in how to use the Internet and the World Wide Web to find and provide health-related information. Staff provided demonstrations of new and emerging online technologies and used the CgSB Web site to point visitors to a variety of online multimedia educational resources in medicine. Staff also provided in-depth individual and group consultations on how to create online multimedia resources and demonstrated exemplary distance learning applications available over the Internet from various health sciences institutions.

Continuing collaborations and outreach activities by Learning Center staff during 1996 included seven professional presentations and publishing two revised technology monographs and one original tutorial. Staff also presented eight tutorials on "Introduction to the Internet and World Wide Web Radiology Resources" and nine tutorials on "Developing Resources for World Wide Web and RSNA Electronic Journal (RSNA EJ)" to more than 600 health professionals at the Radiological Society of North America Annual 1996 Meeting. In collaboration with other NLM programs, The Learning Center also served as the setting for outreach activities in which several teachers and students completed practicums to meet the requirements of their respective internships.

Distributed Learning Environments

The purpose of the Distributed Learning Environments project is to investigate, demonstrate, and evaluate various information and communication technologies for distributed learning. Emerging telecommunications and computer-mediated technologies, especially those based on the concepts of distributed computing architectures and client/server computing systems, can expand and enhance learning experiences and opportunities by making them available anyplace, anytime. These new technologies and conceptual frameworks offer a new environment for developing and delivering learning experiences, one that is decentralized and emphasizes learning over teaching. The major objective of this project is to identify the potential of these technologies and concepts to provide enriched distance learning experiences in the health

professions.

Project efforts focus on developing innovative means for delivering online distance learning integrating telecommunications, the Internet and the World Wide Web. Project staff collaborate closely with health science institutions and other government programs involved in distance learning. Applications will exemplify collaborative authorship of learning resources by developers, the collaborative use of these resources by faculty, resource sharing/exchange by both faculty and developers, and cooperative learning by students. Developers, users, and resources that reside at different sites will be interconnected by emerging communication technologies.

In 1996, a review of the literature on distributed learning environments, distance learning, and collaborative learning was completed. As a result of this review, staff made seven presentations on distributed learning in the health professions at regional and national conferences and published one journal paper. Additionally, an ongoing assessment of current and emerging telecommunications and computer-mediated technologies was initiated. Collaborations are ongoing with health science institutions and other government distance learning programs.

Digital Microscopy System Project

In 1994 the Digital Microscopy System (DMS) project 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 LHCBC Board of Scientific Counselors, such images can 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 task. DMS had in hand a set of prostate photomicrographs that were recorded in analog format by another LHCBC project 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 (resident 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.

Three significant developments came in FY 1996. First, DMS installed a WWW server on one of its computers providing potential indexers with the capability of having on a single screen a photomicrograph and the input screen for the Metathesaurus. The server may be accessed through the CgSB Web site. Access to the indexing research page is restricted but a password may be obtained by contacting the project officer via the email link on the DMS home page.

Second, 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 DMS Web server and enter keywords and diagnoses. The twelve images that were captured at NLM in 1995 have now been indexed by six pathologists operating remotely from Ann Arbor, Baltimore, Denver and Pittsburgh. The index terms entered into the Pittsburgh database have been submitted to the NLM/AHCPR Large Scale Vocabulary Test and the data are just now being analyzed. Even before the analysis is complete, cursory examination indicates that well over 90% of the terms are recognized by what will be the next edition of the Metathesaurus.

Third, DMS has developed a WWW page that will give indexers access to the images and a controlled vocabulary (Metathesaurus concepts). During early 1997 the team of indexers will re-index the test set of images using the controlled vocabulary and inter-indexer variabilities can then be studied.

Research in Information Technologies

Natural Language Systems Program

The Natural Language Systems (NLS) research team is concerned with investigating the contributions that natural language processing techniques can make to the task of mediating between the language of users and the language of the databases they attempt to access. The successful integration of these techniques with other sophisticated retrieval strategies has the potential of contributing to the resolution of some of the most difficult problems faced by health professionals as they attempt to access biomedical information in computerized form.

The focus of NLS work is the development of

SPECIALIST, an experimental natural language processing system for the biomedical domain. The SPECIALIST system includes several modules which are based on the major components of natural language—lexicon, morphology, syntax, and semantics. The morphological component is concerned with the structure of words and the rules of word formation. The syntactic component treats the constituent structure of phrases and sentences, and the semantic component is concerned with meaning, with emphasis on the domain of discourse. All three rely heavily on the lexical component which encodes the information specific to the lexical items in the language.

An ongoing and fundamental aspect of NLS research is the evaluation of the effectiveness of natural language processing in the context of information retrieval. NLS staff regularly conduct experiments using the SPECIALIST system and UMLS knowledge to assess incremental improvements in retrieval effectiveness. It is a testable hypothesis that methods such as those under development by the NLS group, which depend on extensive linguistic and domain knowledge, provide an advantage over methods which do not have access to such knowledge.

Lexical Systems Project

The Lexical Systems Project brings together several activities centered around lexical issues, including development and maintenance of the SPECIALIST lexicon and the lexical tools distributed with the UMLS as well as a project to exploit those resources for text analysis. The volume of text available in electronic form is increasing exponentially. Text processing algorithms exploiting the lexicon and lexical tools can be used in experiments comparing word-based and concept-based retrieval algorithms.

The SPECIALIST Lexicon

A major component of the Lexical Systems project is the SPECIALIST lexicon, a large syntactic lexicon of medical terminology containing over 100,000 lexical items. Lexical items are collected into unit records containing morphological, syntactic, and spelling information about each item. Morphological information includes full inflectional information for each lexical 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 sequencing

information for adjectives and complement patterns for verbs. Verb complement information is central to analysis of sentence structure. A recent experiment on the lexicon 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 lexicon has been built and maintained by NLS staff together with several expert consultants. Tools have been developed to facilitate lexicon building and to ensure the consistency of the lexical entries, Lextool, the original lexicon building tool written in Prolog, has recently been replaced by a Web-based tool called LexBuild. Both tools enforce a complete and consistent form for lexical entries, and provide 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 in 1990 when five expert consultants worked off-site using government supplied computer equipment and software to expand the lexicon. This effort brought the lexicon to over 40,000 lexical records. In 1994, when the lexicon was first released as one of the UMLS Knowledge sources it contained approximately 64,000 lexical records. Another successful lexicon building project in 1995 increased the size of the lexicon to over 80,000 items.

An ongoing lexicon building project has brought the lexicon to its current size. Two lexicon consultants are working off-site with the Web-based tool, LexBuild. New lexical records are downloaded to NLM via the Internet biweekly. This project is scheduled to continue until April 1998. Each of the lexicon building projects has emphasized quality control and correction of errors in the existing lexicon as well as growth of the lexicon. The current effort has concentrated on terms from the UMLS Metathesaurus that have been assigned lexical tags. In the future, lexical tags will be removed from the Metathesaurus and the lexicon will carry that information. Both consultants are professional linguists with broad knowledge of the issues involved in lexicon coding and previous experience as consultants in previous SPECIALIST lexicon building projects.

Lexical Tools

The lexical tools are designed to help users abstract away from certain sorts of lexical variation. Retrieval and matching tasks often require some way to ignore variation. The three tools lvg, norm, and wordind offer this sort of capability. Lvg uses the SPECIALIST lexicon, and associated databases

along with a set of heuristic rules to allow users to abstract away from alphabetic case, word order, variations in punctuation, spelling variation, inflection, and derivation. Lvg offers methods (called flows) which can combine these functions to produce indexes with varying degrees of aggressiveness with respect to the matching of terms. Essentially users use lvg to create the index keys for the terms they wish to search. Using lvg again with the same flows gives the users the capability of matching terms through the index in a way that abstracts away from the sorts of lexical variation they wish to ignore. Norm encapsulates one of the flows of lvg and is used in the UMLS Metathesaurus to create the normalized string index together with wordind to create the normalized word indexes.

Modular Text Processing System

The Lexical Systems group is developing a text analysis system based on the SPECIALIST lexicon and the lexical tools. The system is modular in design in order to allow for flexible use and continuous revision. The modules are servers which will be available to a variety of clients for a variety of uses. Interchangeable versions of particular modules should facilitate experiments to improve the system. The system will consist of several 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 preprocessor 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 structured text types. Preliminary versions of each module, with the exception of the preprocessor module, have been created in a pilot project.

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 so the lexical look-up module can 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 intended to be strongly based on the syntactic information encoded in the SPECIALIST lexicon. The structure of sentences is determined in large part by the complement patterns of the verbs that occur in them. After consideration of several grammar paradigms, Categorical 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 Categorical 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 to test the hypothesis that the SPECIALIST lexicon encodes sufficient information for parsing a good portion of medical English text.

Semantic Knowledge Representation Project

The Semantic Knowledge Representation Project was recently initiated in order to develop programs to provide usable semantic representation of biomedical free text by building on the resources currently available in the Unified Medical Language System (UMLS). Ideally, digital libraries of the future will provide "deep semantic interoperability," that is, "the ability of a user to access, consistently and coherently, similar (though autonomously defined and managed) classes of digital objects and services, distributed across heterogeneous repositories" (Lynch and Garcia-Molina 1995). A number of challenges must be met before deep semantic interoperability becomes possible, a major aspect of achieving this goal is reliable representation of the knowledge contained in text. A richer representation will be required than is currently available. This project is a first step toward providing such a representation for biomedical text. It is also hoped that the results of the project will be applicable to automatic indexing in support of enhanced information retrieval effectiveness.

As an example of the type of representation under development, (1b) contains the semantic propositions which represent some of the knowledge contained in the text in (1a).

(1a) *We used hemofiltration to treat a patient with digoxin overdose complicated by refractory hyperkalemia.*

(1b) Hemofiltration-TREATS-Overdose

Digoxin-CAUSE-Overdose
Overdose-OCCURS_IN-Patients
Hyperkalemia-COMPLICATES-Overdose

Each of the propositions in (1b) has a predicate (in upper case) which is a relation from the UMLS Semantic Network. Each of the arguments is a concept from the UMLS Metathesaurus.

A representation such as that given in (1b) could serve as a prime candidate for the automatic indexing of free text. If this could be reliably and accurately imposed on free text, other applications, such as question answering, retrieval of images, and guided browsing would be possible. Knowledge mining applications, such as gathering information from computerized patient records, would also be feasible.

The Semantic Knowledge Representation Project relies on existing resources, especially the UMLS Knowledge Sources and the SPECIALIST natural language processing tools. The Parser (supported by the SPECIALIST lexicon) provides syntactic analysis. MetaMap (supported by SPECIALIST lexical tools) maps syntactically analyzed text to Metathesaurus concepts. The program SemRep then uses the syntactic structure and the UMLS knowledge to map to the Semantic Network, thereby providing propositions as semantic conceptual structure as shown in (1b). 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 Next Generation Indexing project.

Evaluation is a central concern in this project. Effectiveness is based on the number of propositions retrieved. In this context, 100% recall indicates that all the propositions expressed in the text being processed were produced by the program, and 100% precision indicates that *only* those propositions were produced. Achieving such a level of performance is unrealistically ambitious given the current state of knowledge concerning natural language processing. Increasing precision by eliminating incorrect interpretations is currently being addressed. Preliminary error analysis indicates that incorrect semantic interpretations fall into three classes, those due to lexical considerations (ambiguity and synonymy), those due to the rules which map from text to Semantic Network relationships, and those due to the process of identifying the arguments of Semantic Network relationships. The current precision rate may be near 70%, which is quite good for natural language processing of free text, but it is hoped that the enhancements being undertaken will

significantly increase this rate

Current research focuses on enhancing two programs, MetaMap and SemRep, which are central to the semantic interpretation process. Two important aspects of enhancing MetaMap are improving the treatment of abbreviatory expressions and dealing with ambiguous mappings, since both phenomena contribute to lexical ambiguity, which is an important source of error. Increased coverage of abbreviatory expressions in medical text is needed. However, the ambiguity inherent in such expressions must be resolved. One aspect of resolving this ambiguity is to discover abbreviations expanded in the text, since, if the correct expansion is identified, these are not ambiguous. Methods for disambiguating other types of abbreviatory expressions have to be sought. Ambiguous mappings (of non-abbreviatory expressions) arise for various reasons and detract from the reliability of the final semantic interpretation. One solution to this problem is the use of a statistical technique (Hidden Markov Model) to resolve ambiguity on the basis of the semantic types associated with the mappings. Other solutions need to be explored.

In improving SemRep, the rules which map linguistic structures to relations in the Semantic Network need to be expanded. A semiautomatic method for discovering such rules, based on patterns of semantic types occurring in sentences in biomedical text, is being developed. The methods of argument identification during the semantic interpretation process need to be improved. One aspect of this improvement is a more rigorous use of information for the complement structure of verbs and nominalizations found in the lexicon. Another goal is to expand the identification of coordinate noun phrases.

Unified Medical Language System Project

UMLS Knowledge Source Server

The NLS group has designed and implemented the UMLS Knowledge Source Server for accessing information stored in the UMLS Knowledge Sources. The server provides flexible access to the UMLS data over the Internet. The system has three client interfaces: a World Wide Web interface, an application programming interface (API), and a command-line interface. The Knowledge Source Server allows the user to request information about particular concepts in the Metathesaurus, including, for example, attributes such as the concept's definition, its semantic types, the concepts that are related to it, etc. It also allows the user to request information about the attributes

themselves, for example, by asking for all the concepts that have been assigned to a particular semantic type, or by asking for all the terms that have a particular lexical tag.

The implementation of the Semantic Network module in the server computes the relationships between semantic types using the inheritance property of the network type hierarchies. Information in the network can be queried in terms of the semantic types and the relationships between them. The Information Sources Map module gives users access to descriptions of each of the databases catalogued in the Information Sources Map and also allows them to see sample records for each of the databases. The SPECIALIST lexicon can be searched, and syntactic and morphological information about the lexical item will be displayed. The lexicon also forms the basis for the lexical routines and the normalized indexes that are used in the Metathesaurus look up procedures.

The World Wide Web interface to the Knowledge Source Server allows users to browse and explore the data and to see how those data are organized in the UMLS. The command-line interface 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, and if so, what attributes the matching concepts may have. They can also filter the results, for example, by retrieving only those matching concepts that have a particular semantic type. The 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.

NLM/AHCPR Large Scale Vocabulary Test

During 1996 the NLM and the Agency for Health Care Policy and Research (AHCPR) sponsored a test to determine the extent to which a combination of existing health-related terminologies covers the vocabulary needed in health information systems. The test vocabularies are those 30 that were represented in the 1996 release of the Metathesaurus together with three planned additions: the portions of SNOMED International not already in the UMLS, the Read Clinical Classification, and the Logical Observations, Identifiers, Names, and Codes system. The NLS group designed and implemented a specialized application interface to the UMLS Knowledge Source Server for use in the Large Scale Vocabulary Test. Participants all over the country used the Web-based system over five months and submitted more than 40,000 terms to the system. The data are currently being analyzed and results are expected to be available in the spring of 1997.

Next Generation Indexing Project

The objective of the Next Generation Indexing (NGI) project 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 index terms assigned by humans.

Human indexing is an expensive and labor-intensive activity. The total costs of indexing at NLM include data entry, NLM staff indexing and revising, contract indexing, equipment, and telecommunications costs. Indexers are highly trained individuals, not only in MEDLINE indexing practice, but also in one or several of the subject domains covered by the MEDLINE database. If automated methods can be successfully developed and implemented, the project will have an important impact on NLM's ability to continue to provide high-quality services to its constituents. Secondly, but also importantly, the project should contribute significantly to information science research. As more and more documents become available in electronic form, and as more and more organizations develop "digital libraries" for their collections, automated techniques for accessing the information are required. It is not possible to index each document by hand, and new methods must be developed. The research products and publications of the NGI project should contribute positively to the information science and digital libraries research communities.

CgSB Branch Chief, Dr. Alexa McCray, is overall coordinator of the project in collaboration with Dr. Nancy Wright, head of NLM's Index Section. Project work during 1996 has involved a number of activities. Existing information retrieval (IR) research projects at the NLM were identified, project leaders were invited to join the NGI team, and existing research tools were identified for project use. These include the UMLS Knowledge Source Server, the Natural Language Systems lexical tools, the NCBI neighboring algorithms, and associative indexing algorithms developed as a joint project between SIS and OCCS. The Center's Unix computer, which is currently being upgraded to an Ultra Enterprise 4000, serves as a shared computer environment for the NGI project.

The Next Generation Indexing project will provide NLM with an opportunity to investigate the design and implementation of automated indexing methods that will partially or completely replace

current practices, as well as the cost savings that will be obtained from these automated methods. The project should also be considered an opportunity to examine some of the present assumptions about the contents of MEDLINE and existing indexing policies, and to determine their relative costs and continuing value to current and future generations of MEDLINE users.

Audiovisual Program Development Branch

The Audiovisual Program Development Branch (APDB) conducts media development activities with three specific objectives. As its most significant effort, the branch supports the Lister Hill Center's research, development, and demonstration projects with high-quality video, audio, and graphic materials. From initial project concept, through 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. In addition to technique development by the staff, the facilities and hardware systems must reflect state-of-the-art standards in a very rapidly changing field. High 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. With all of the technologies being brought together in the development and support projects of the Branch, the central core expertise remains the creative, artistic, communication skills of the staff.

Visible Human Project

A videotape, "The Visible Humans--A Step Toward Tomorrow" was completed in March 1996. Taping was done at several locations including the University of Colorado Health Sciences Center, Denver, where they videotaped actual imaging and sectioning activity over two years. In addition to telling the history of the project, the tape focuses on the applications of the datasets by NLM licensees. More than 200 copies of the tape have been distributed to date. The Branch also designed the new Visible Human Web site to provide background information on the project as well as actual sample images from the data set. It also provides links to other web sites that contain applications developed using the data set.

The Visible Male and Visible Female 35mm and 70mm films are now being stored at NLM. They are kept in a film storage environment with controlled temperature and humidity. The Visible Female 35mm original negative films have been assembled on composite reels in sequence and are ready for duplication. To produce viewable copies, an interpositive, and a duplicate negative were printed so that a positive film transparency is available for screening.

The entire 70 mm Visible Human male was scanned using the Kodak 5057 custom designed scanner. The decision was made to scan the original 70 mm for the database to provide the highest resolution images with the most accurate color reproduction. The film was scanned at a resolution of 4096 x 5568 with a pixel dimension of 144mm per pixel. There are 1878 images. The images were reviewed for quality to detect any errors or artifacts and the frames with correctable artifacts were cleaned and rescanned.

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. This was done using the 2-D and 3-D visualization tools in ANALYZE running on a Silicon Graphics workstation.

An interactive brain atlas was developed to be presented at the Visible Human Conference in October 1996. It is a SuperCard based teaching program that runs on a Macintosh platform. It includes MR and CT scans of the head as well as corresponding cryosections from the Visible Human Male data set.

Cervical Cancer Project--Screening for Life

This is an instructional program to promote the early detection of cervical cancer. Topics include risk factors, incidence and mortality charts, screening and laboratory procedures, pathology, and the Bethesda Reporting System. In collaboration with the National Cancer Institute, the program was originally designed as an interactive videodisc. In 1993, the program was redesigned for the Compact Disc-interactive format including a modem dial up to NCI's PDQ cancer information database. APDB is now in the process of beta-testing a complete Web version that includes all the text, graphics, and video clips from the original program. Intended for general practitioners, internists, nurses and other health professionals who provide cervical cancer screening services, the NCI's International Cancer Information Center will make it available as part of its web site.

Making Medical Science Visible

This videotape documents the graduate-level training required for the production of successful visuals for medicine and science. In an effort to highlight the process of training students to become biomedical visualizers, 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. During the 2-year project, a variety of activities were videotaped, including medical illustration techniques, project critiques and evaluations, academic courses, anatomical dissection and instruction, laboratory research, surgical procedures, and interviews. The final 28-minute tape is an effort to feature both the scientific and artistic training required to provide accurate and instructive visuals for science and medicine, both for the medical community and the public.

Cardiac Embryology/Visible Embryo Web Site

The desktop, interactive version of the Cardiac Embryology Project was successfully converted to a Cardiac Embryology Web site. Existing digital assets were repurposed and supplemented with additional graphics, text, and animation, taking advantage of the interactive educational possibilities of the World Wide Web. The approach to the Web site design was based on maintaining the streamlined organization of the original program, the user feedback from the Visible

Embryo alpha and beta tests, and current issues in successful Web design. The site offers a rich educational experience to the user and delivers a valuable dataset of histological human embryonic images, 3-dimensional models, instructional animations and supporting text, and bibliographic reference. The program was tested extensively by researchers, anatomists, pathologists, and others before release. The Web site has been included in the Hopkins LectureLinks Project.

Office of the Public Health Service Historian

The Office of the Public Health Service Historian became a part of APDB as of October 1, 1995. Created in 1992 in the Office of the Assistant Secretary for Health, the Historian's Office was transferred to the Library as part of the 1995 reorganization of the Department. One of the major projects of the Office in FY 1996 was the development of a 6-room exhibit titled "Doctors at the Gate: the United States Public Health Service at Ellis Island," which was on display at the Ellis Island Immigration Museum from May 24 through July 31. A laser videodisc containing a portion of the 1936 PHS film, "The Work of the Public Health Service" was prepared by APDS for the exhibit.

The office also assisted in planning commemorative activities for the 50th anniversaries of the Centers for Disease Control and Prevention, the Division of Nursing, and the National Mental Health Act of 1946, and the 25th anniversary of the National Institute on Occupational Safety and Health. Planning of the PHS Bicentennial in 1998 was also a significant effort of the Office this fiscal year. Staff carried out research, delivered lectures, and published papers on various aspects of PHS history, in addition to editing the "PHS Chronicles" historical series in *Public Health Reports*. Other responsibilities included answering historical queries from a variety of sources and working to preserve PHS documents and artifacts of historical significance such as materials from the now closed Saint Elizabeth's Hospital Museum.

Educational and Information Program Support

At the request of the MEDLARS Management Section, APDB produced two new training videotapes to be used in the MEDLARS training classes conducted at NLM and the regional centers. The tapes, which replace those made 10 years ago, are *Getting Started...a MEDLARS Overview* and *Getting Started...the MeSH Vocabulary*.

A videotape describing the major NLM programs and activities was produced by APDB. This is an update of the *Pathways* videotape and is titled *Your Key to Medical Knowledge*. The revised program was reviewed by the Board of Regents at the January 1996 meeting. The recommended content changes were made and a new narration track was recorded and a new title background was created. The finished program was also revised to include a segment on Internet Grateful Med. Open caption and foreign language versions including Spanish, Chinese, Japanese, and Russian were produced, and copies were made for distribution by NLM Public Information Office.

Mapping the Geographical History of Washington, a 15-minute videotape was designed and produced by APDB for inclusion in the History of Medicine Division lobby exhibit, "Death and Disease in the Neighborhood: Medical Maps of Washington, 1878-1909." The tape was transferred to a laser videodisc and ran in the exhibit from January through June. The program was shown more than 4,800 times during this period. Also, APDB produced two laser videodiscs for the HMD exhibit, "Emotions & Disease" which opened in the NLM lobby in October 1996.

Audiovisual Support Activities

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

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

David Lipman, M.D.
Director

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

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

There are presently 55 senior scientists, postdoctoral fellows, and support staff working at the NCBI. These scientists have backgrounds in medicine, molecular biology, biochemistry, genetics, biophysics, structural biology, computer and information science, and mathematics.

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

Database Building

GenBank - The NIH Sequence Database

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

The database contains over 1,021,000 sequences, nearly 594,000 of which are human. Over 465,000 sequences were added during the year, a 46% increase in the size of the database, with the number of human sequences representing almost half of that increase. Two divisions were added to GenBank to accommodate the past two year's growth in human sequences, which is the result of large-scale human cDNA sequencing projects. The Genome Survey Sequence (GSS) division contains random single pass read genome survey sequences and other types of sequences that will be used as a framework for the mapping and sequencing of complete genomes; the High Throughput Genome (HTG) division contains genomic sequences that are produced by high-throughput sequencing projects.

The total genome sequences are now available for representative species from three domains of life--Archaea, Bacteria, and Eucaryotes. For example, the full sequences for the bacteria *Escherichia coli*, *Haemophilus influenzae*, and *Mycoplasma pneumoniae* are in GenBank. Other organisms for which there was substantial growth in the number of GenBank records are *Mus musculus* (house mouse), *Caenorhabditis elegans* (a roundworm), and human immunodeficiency virus type 1 (HIV-1). Sequences from these organisms have provided valuable clues for understanding the functioning of human genes.

NCBI produces GenBank from thousands of sequence records submitted directly from authors prior to publication and, to a lesser degree, from extracting sequences from publications. Indexers with specialty training in molecular biology create these 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.

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.

Other Specialized Databases

NCBI was particularly active this year in the design, implementation, and distribution of new databases and software tools for the molecular biology community. Access to the Online Mendelian Inheritance in Man (OMIM), Johns Hopkins's database of genetic disorders, was moved to NCBI's WWW page with MEDLINE links via the Entrez service. A WWW page was also created for the Human/Mouse Homology Map, created by Ronald DeBry and Michael Seldin, Duke University. The *Saccharomyces cerevisiae* (yeast) database, a product of Stanford's DNA Sequencing and Technology Center, was added to the BLAST database suite for sequence similarity searching. New NCBI databases include the Molecular Modeling Database (MMDB) of 3-dimensional protein structures, incorporated into WWW Entrez with a structure-viewing program; and the Unique Human Gene Sequence Collection (UniGene), which organizes human GenBank sequences into "clusters" that represent distinct genes.

The Expressed Sequence Tags database (dbEST) and the Sequence Tagged Sites database (dbSTS) both more than doubled in size during FY 1996. dbEST, which collects the growing number of gene fragments obtained through cDNA sequencing, contains over 677,000 sequences, 70% of which are human. A significant number of the remainder are *Mus musculus* (house mouse) sequences. dbSTS

contains sequence and mapping data for over 47,000 short genomic landmark sequences. Both dbEST and dbSTS are available for public queries via the Internet.

The NCBI taxonomy group, now composed of four specialists, designed and implemented a Taxonomy Browser on the WWW that contains the names of all organisms represented in the genetic databases with at least one nucleotide or protein sequence. There are links directly from the browser to MEDLINE abstracts that describe new organisms. The source organism for each GenBank submission is automatically checked against the Taxonomy database, and if not listed, is added by one of the staff taxonomists. The international collaborators, EMBL and DDBJ, have accepted NCBI's Taxonomy database as the standard classification, and EMBL has initiated procedures for their indexers to use the database and consult with NCBI taxonomists.

The Taxonomy database contained about 25,000 organisms in FY 1996, and 500 new organisms are added monthly. Some 50 international molecular biologists and taxonomists serve as curators and provide expert review and consultation for the task of building and maintaining a current and consistent phylogenetic classification and nomenclature for the source organisms in the sequence databases.

Database Access

Entrez Retrieval System

The major database retrieval system at NCBI is Entrez, a search system for nucleotide and protein sequence databases and related MEDLINE citations. Using network or WWW versions, 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. A new structure viewer, Cn3D, was integrated into Network Entrez. In response to requests from the user community, a function that allows large numbers of

sequences to be downloaded, called Batch Entrez, was added to WWW Entrez

The CD-ROM version of Entrez was discontinued in August 1996 after results of a user survey confirmed the feasibility of network alternatives. Network Entrez usage has decreased by half, but use of the WWW version has grown by two and one-half times, reflecting the general trend of WWW usage of NCBI's services

Other Network Services

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 to several hundred thousand known sequences and determines the closest matches A new network BLAST application, called PowerBlast, was designed for automated analysis of genomic sequences

Each day more than 20,000 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 number of sites using the BLAST service Genome research is highly dependent on this resource

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 1,000 queries are handled daily by the Retrieve and Query servers.

NCBI continued to make significant improvements to the GenBank submission process in FY 1996 Sequin, a new stand-alone software submission tool, was released in May and refinements continued throughout the year. Sequin is

capable of handling simple submissions that contain a single, short, mRNA sequence, as well as complex submissions containing long sequences, multiple annotations, segmented sets of DNA, or phylogenetic and population studies. Extensive documentation is available on the World Wide Web, and versions for Macintosh, PC/Windows, and UNIX computers can be downloaded to the user's workstation. In its second year of use, the BankIt submission tool available on the World Wide Web continues to be the most popular submission tool, with nearly 80% of submissions entering GenBank via this route.

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 Sequin, the Taxonomy browser, and new versions of BLAST, Entrez, BankIt, two other WWW applications were introduced These are OrfFinder (Open Reading Frame Finder), a graphical analysis tool that finds all open reading frames in a user's sequence; and VacMan, which is software for analyzing infectivity data from animal trials Information about NCBI, its databases and services, data submission and update, and individual scientists' 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 1996, NCBI's site was averaging over 300,000 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, 500 sites download new full releases of GenBank, and daily, 120 sites download the updates. Fifty-five additional databases of molecular biology data are distributed as part of the NCBI Data Repository to over 1,000 sites daily

NCBI uses a network of Unix-based symmetric multiprocessor servers to provide Internet-based public search services such as BLAST and Entrez During FY 1996 the NCBI added a 20-processor Silicon Graphics (SGI) system dedicated to BLAST services. In addition, other SGI servers that support public services as well as basic research were upgraded to newer and more powerful CPUs. New Sun Microsystems servers were acquired to support the PubMed project, a major new experimental version of Entrez that will include all of MEDLINE with links to full-text journal sites A Sun server that supports Entrez search services and a second that supports bimonthly GenBank releases were also

upgraded to faster CPUs. Two Windows NT servers were also acquired and installed.

The servers are supplemented by Unix workstations, PCs and Apple desktop systems that are used for database and software development, research applications and data analysis, user services support, as well as conventional office and administrative applications.

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

NCBI's adoption of ASN.1 for database output has several advantages for users as well as developers. The data definitions in ASN.1 for biological objects enable the representation and structuring of complex biological data in data files without the need for a specific database management system. Manipulation of the complex objects is performed through the ASN.1 software tools that are freely distributed to the biology community. Thus, complicated analysis programs can be readily constructed from pre-existing sets of modular tools, saving considerable time and programming effort.

Basic Research

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

There have 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. This work has included a study of the infection kinetics of HIV, and the development of an infectivity assay; a modification of the BLAST programs to produce gapped alignments, and the description of refined statistics for their assessment, an analysis of the native states of random peptides; and the development of a Bayesian model and algorithm for multiple local alignment. Macromolecular sequence analysis programs have been applied to investigate histones, helicases, aconitase proteins, FAD-binding domains, lysozyme-like virulence factors, mitochondrial proteins, eukaryotic developmental proteins, and to identify a domain common to the human breast cancer gene, *BRCA1*.

This year genome-scale projects have become a staple of Computational Biology Branch research. They have involved the creation and publication of a gene map for the human genome, an integrated linkage map for *E. coli*, and a genetic map of *Salmonella typhimurium*; the molecular cloning of Hepatitis G virus, and comparative metabolic and evolutionary analyses based upon complete prokaryotic genomes. Other projects have included the development of the Molecular Modeling Database (MMDB) for the analysis of protein structure, and its integration into the WWW Entrez system; the evaluation of threading methods for sequence-structure comparison; phylogenetic analyses of pleurostomatid ciliates and trichomonads, and the theoretical and experimental study of document retrieval methods.

The intramural group is engaged in some 29 projects, many of which involve collaborations with NIH and other research laboratories. The work is reviewed by a Board of Scientific Counselors of distinguished extramural scientists (see Appendix 4 for list of members). The high caliber of the work has been evidenced by the number of peer-reviewed publications, some 60 in FY 1996, and the requests for outside collaborations and invitations for talks at international meetings.

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.

To increase awareness of NCBI and its programs, NCBI staff participate in exhibits, seminars, workshops, and courses, both nationally and internationally. In FY 1996 NCBI staffed exhibits at scientific society meetings, including the American Society for Biochemistry and Molecular Biology; American Society for Cell Biology; Biophysical Society; Experimental Biology; American Society for Microbiology; American Society of Human Genetics; and the Medical Library Association. Additional exhibits were staffed at more specialized meetings, including the Yeast Genetics and Molecular Biology Meeting, Bioinformatics and Genome Research, Tuskegee University Conference on the Human Genome Project, and Genome Project--Commercial Applications.

In addition, senior NCBI members participated as faculty at courses sponsored by the American Association for Cancer Research; the World Health Organization; Johns Hopkins School of Medicine; George Mason University's Institute for Computational Sciences and Informatics; The Jackson Laboratory; NLM/Marine Biological

Laboratory; the International Center for Genetic Engineering & Biotechnology (Trieste, Italy).

Three issues of a newsletter were 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. More than 300 NIH and extramural scientists' accounts are supported through online access to 20 databases under the IRX system.

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

Extramural Programs

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

Biotechnology Information in the Future

The explosive growth in the fields of genetics and molecular biology reinforces the need to build and maintain a strong infrastructure of information support. NCBI continues to be engaged in developing and employing new methods for disseminating knowledge to the biomedical community. Based on a core of advanced intramural research in several areas of computational biology, NCBI 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 (which covers all of NIH). The funds are expended as grants-in-aid to the extramural community in support of the goals of the National Library of Medicine. Review and award procedures conform to NIH policies.

Summary budget figures for FY 1996 (and FY 1995 and 1994 for comparison purposes) are presented in **Table 11**. Appendix 5 contains the roster of the Biomedical Library Review Committee. Appendix 7 lists grants awarded in FY 1996.

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

These categories are used to organize discussion of EP activities for this report.

Resource

The "Resource" grants described in this section are intended to strengthen the information management infrastructure of the grantee, and, therefore, provide significantly less overhead compensation to the grantee institution than do research grants. The four increasingly complex resource programs described below are designed to address needs of the broad spectrum of American health care organizations.

Internet Connections Grants (MLAA)

This program provides relatively small grants to health organizations that wish to make an initial Internet connection, or to broaden internal access to a connection already in place. To qualify, institutions must already have some local area network (LAN) capacity. The health science library usually has a key role in the Internet connection

grants, primarily in training new users. These grants are available in two forms: \$30,000 to single institutions and \$50,000 to those extending their connection to other sites. Grant funds support costs for router/gateway equipment and associated connection hardware, installation and leasing of communication circuits to connect to an Internet provider, and membership fees to a provider.

A Request for Applications issued in February 1996 resulted in 115 applications. A 2-stage review was utilized with an initial Special Review Committee recommending 45 applications for further consideration. In FY 1996 available funds permitted funding of 26 Internet Connection Grants with the majority going to hospitals; the 8 collaborative Internet Connection projects increase the total number of beneficiaries to nearly 60.

Information Access Grants (MLAA)

Information Access Grants provide \$12,000 for single institutions and \$12,000 per member for consortia to purchase basic computer and communications equipment to access NLM's Grateful Med, DOCLINE, and Loansome Doc services. Funds can also be used towards the installation of an integrated library system. Internet Access is usually incorporated into these projects. To foster NLM's ongoing Outreach Initiative, the grants support consortia in remote, rural areas where outlying institutions can network with a larger site equipped with back-up collections and library staff. FY 1996 awards were all to consortia.

Information Systems Grants (MLAA)

Information Systems Grants, which range up to \$150,000 per year for 1 to 3 years, are targeted for larger health institutions such as teaching hospitals and medical schools. Projects support computer and communications networking, integration, and connectivity and must include the library as a participating partner. Internet capabilities are again emphasized by NLM as highly desirable. Grant funds may cover personnel, equipment, software, communications expenses, travel, and supplies. Three Information Systems Grants were awarded in FY 1996

IAIMS Grants (MLAA)

Integrated Advanced Information Management Systems (IAIMS), the most ambitious of the resource grants, are institution-wide computer networks that link and relate library systems with a variety of individual and institutional databases and

information files, within and external to the institution, for patient care, research, education, and administration. Resource grants are made to assist medical centers and health science organizations in planning and implementing projects that will lead to an IAIMS. The goal of the program is to create organizational mechanisms within health science institutions to manage more effectively the knowledge of medicine, and to provide for a system of comprehensive, direct information access for health professionals.

NLM provides grant support for (1) an institution-wide planning phase where support may be for up to \$150,000 per year for 1 or 2 years, and (2) an operation phase in which IAIMS plans are implemented. Operation phase grants may be for up to \$500,000 per year for 5 years, or for \$550,000 per year if support of an IAIMS apprenticeship option is approved. There were six new planning and one implementation award in FY 1996.

Training and Fellowships (MLAA)

Training in Informatics

Realizing the potential of computers and telecommunication for managing health care information requires investigators able to address 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. Special tracks are available for those who wish to concentrate in the informatics of biotechnology, nursing, radiation oncology, or dentistry.

To complement the training of researchers, NLM now also offers training opportunities for health care professionals who want to apply the knowledge and technology of this field to some area of interest.

NLM provides three mechanisms of support for its training activities. Five-year institutional training grants, the largest type, currently support approximately 100 trainees, and produce 25-30 informatics scientists each year. During 1996 these programs and others were reviewed as part of a competitive review process which resulted in a decision to continue funding of the existing programs and to fund two more.

Individual informatics research fellowships are available to those who want research training similar to that received 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 to their current professional specialties.

Education of Health Sciences Librarians

In early 1995 NLM, as a one-time initiative, 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:

- evolving role for health sciences librarians;
- professional education programs;
- lifelong learning programs; and
- 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 Eskin Biomedical Library, Vanderbilt University, Nashville, TN.

Reports from these awardees are expected in the spring of 1997. NLM will consider follow-up options after reviewing the reports.

Training of Minorities

NLM is participating in an NIH-wide fellowship program aimed at encouraging under-represented minorities into research careers, and supported one fellow pursuing an MD/Ph.D. combined degree at Harvard Medical School in FY 1996.

Research (and Resource Resources)

Research grants are made through a variety of PHS 301 mechanisms, including individual research projects, cooperative agreements, research resource grants and others.

Medical Informatics

NLM's medical informatics 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 received permission to participate in several programs sponsored by the Office of the Director, NIH. Five

applications altogether benefited from these programs, generically called "Shannon Awards." These included three young investigators at Yale, Vanderbilt, and Massachusetts General Hospital.

NLM initiated 15 new individual informatics research projects in FY 1996. Five were Shannon Awards, five were the standard investigator-initiated research project grants, and five were First Independent Research Support Transition Awards.

A major NLM effort has been the ongoing collaboration with the Agency for Health Care Policy and Research to support research and development of Electronic Medical Record Systems (EMRS). AHCPR sponsored a general meeting for the awardees and their colleagues in April, and NLM followed with one in November. NLM staff encouraged, and attended, meetings of several smaller working groups. These groups, one in the Boston area, and the other dispersed among the western states, are addressing such issues as secure patient data transmission extraction of aggregated patient data from disparate files, and problem list managers.

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 neurosciences. 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.

Biotechnology Informatics

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 the area of biotechnology informatics. A related problem concerns the development and maintenance of the myriad electronic databases on which researchers increasingly rely, and for which no other source of support has yet been identified.

Publication Grants

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 medical 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 cases, secondary literature tools and scientifically significant symposia. Resources in recent years have been used principally for history of medicine projects, but projects involving electronic publishing, video productions, and other media were also supported. The program has an informal self-imposed ceiling on direct costs per grant of \$25,000. The three new awards made in FY 1996 are listed in Appendix 7.

Educational Technology

NLM's grant support for educational technology development is generally restricted to computer/telecommunication projects that incorporate strong research and evaluative components. Because the Library's support funds for this area are small, those who seek funding for educational technologies have been urged to focus their work on some specific area of medicine which might attract funding from one of the larger Institutes.

SBIR/SBTT (PHS 301)

All NIH research grant programs, including NLM's, by Congressional mandate 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 Dr. John Nelson of Nelson Information Systems in Chevy Chase, Maryland for work on a reprint service for internists based on 10 core journals. This award was the final one of a 2-year project period and included participation by two other NIH Institutes.

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. No conference grants were awarded in FY96.

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 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 MEDLARS databases.

Other Extramural Programs Activities

HPCC and Outreach

The Outreach and the HPCC 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.

Grant Review

First Level Peer Review

The Biomedical Library Review Committee (BLRC) is NLM's initial review group for evaluating the scientific and technical merit of grant applications. The BLRC has expertise germane to the specific program areas: library, informatics, and biotechnology. The committee met three times in FY 1996 and reviewed 87 applications, 62 of which were scored and continued in the review process. A roster of the Committee is in Appendix 5. The roster and other items relating to grant applications and, therefore, of interest to potential applicants are in NLM's World Wide Web site. (www.nlm.nih.gov/ep/extramural.html)

In FY 1996 NLM and the other NIH institutes were granted authority to establish Special Emphasis Panels (SEP). The authority permits the creation of initial review groups whose expertise is focused on a particular group of applications, for example, the Institutional Advanced Information Management

Systems (IAIMS) program. Such one-time review panels are also useful for evaluating individual fellowship applications, and applications having substantial involvement of regular BLRC members. Meetings of the Panels, like BLRC meetings, are regularly announced in the Federal Register. The utilization of Special Emphasis Panels replaces the previously used ad hoc review groups.

Second Level Review

A second level of review is performed by the Board of Regents, which meets three times a year, approximately three months after the Biomedical Library Review Committee. The Board concerns itself with the congruence of grant applications with the NLM's mission and with the policy aspects of the applications. The Extramural Programs Subcommittee of the Board reviews in depth "Special" grant applications such as those in which the recommended amount of financial support is large; when at least two members of the initial review group have dissented from the majority; where a policy issue is identified; and those from a foreign institution. The Subcommittee makes its recommendations to the full Board.

Review Reform

The National Institutes of Health 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. Items currently under consideration include the use of three specific criteria (significance, approach, and feasibility) with each criterion being scored; the use of an 8-point rating scale; and revision of the scoring system so that the better the grant, the higher the priority score number. The Oversight Group is continuing to meet and will consider input from the applicant community. Formal recommendations are expected in FY 1997.

Divisional Operations

The operating budget was adequate for indicated site visits, and maintaining the Division's activities. EP incorporated the new NIH streamlining initiatives in timekeeping and procurement. Mr. Dwight Mowery, Grants Management Specialist and Ms. Tracy Graves, Grants Clerk joined the Division.

Summary

NLM's Extramural Programs, 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 most 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 produced by our training programs.

Table 11

Extramural Grant and Contract Program

(dollars in thousands)

Category	FY 1994		FY 1995		FY 1996	
	No.	\$	No.	\$	No.	\$
Resource projects	38	5,150	54	5,421	53	5,754
IAIMS.....	(10)	(3,115)	(7)	(2,231)	(10)	(3,011)
Access.....	(2)	(80)	(11)	(772)	(5)	(476)
Systems.....	(9)	(1,283)	(10)	(1,438)	(12)	(1,387)
Connections.....	(17)	(672)	(26)	(980)	(26)	(880)
Research	53	12,436	64	13,680	64	14,817
Medical informatics projects.....	(19)	(5,681)	(25)	(5,966)	(22)	(4,433)
Medical informatics resource.....	(1)	(1,419)
Biotechnology.....	(13)	(3,776)	(21)	(4,390)	(19)	(5,264)
Cooperative agreements.....	(3)	(1,008)	(6)	(2,002)	(5)	(1,872)
Career awards.....	(16)	(1,725)	(11)	(1,138)	(16)	(1,638)
Library science.....	(2)	(246)	(1)	(184)	(1)	(191)
Training	19	4,372	25	5,072	19	4,501
Institutional.....	(10)	(3,980)	(16)	(4,676)	(10)	(4,178)
Fellowship.....	(9)	(392)	(9)	(396)	(9)	(323)
Publications	8	239	16	513	12	327
Bioethics	1	400	1	416	1	458
SBIR/STTR	2	145	1	191	1	196
Regional Medical Library	8	5,678	8	5,545	8	6,283
NIH Tap	849
Totals:	129	\$28,420	169	\$30,838	158	\$33,185

Office of Computer and Communications Systems

Fernando Burbano
Director

Grateful Med

This year was another successful year for Grateful Med. As in years past, a record number of Grateful Med searches were performed. The Internet continues to be the access method of choice, and the Grateful Med dial-up modem users were successfully transitioned from Telenet, Tymnet, and CompuServe to AT&T's FTS2000 Packet Switch Service Network. Updates for both the PC DOS version and Macintosh version of Grateful Med were distributed and a Windows version was completed.

During FY 1996 a new record of over five million Grateful Med ELHILL searches were performed. The number of Grateful Med searchers continues to increase, comprising over 80 percent of all ELHILL searchers. Also, the use of the Internet by Grateful Med patrons continues to grow. Currently, over 60 percent of all Grateful Med sessions reach NLM via the Internet. This is a 20 percent increase from last year. It is anticipated that the increased Internet usage will also facilitate the electronic distribution of future software updates and documentation. In fact, Grateful Med beta testers already use NLM's anonymous ftp server for downloading test programs and future Grateful Med for Windows users will be able to download software and a User's Guide from NLM's World Wide Web site.

A transition from the Value Added Networks (VANs) to AT&T's FTS2000 Packet Switch Service was completed this year. Grateful Med now supports only Internet, FTS2000, direct dial, or Canadian Datapac communications. Extensive testing and hardware upgrades were performed to insure a smooth move to FTS2000. A few problems did occur, but the majority of them consisted of user setup/configuration errors.

Grateful Med for Windows Version 1.0 was completed this year. The results from user response cards show that 19,000 users want the Windows version while only 700 users want to remain DOS users. The program disks and User's Guides are being produced by the National Technical Information Service and will be distributed in January 1997.

Both the PC DOS Version 6.7 update and the Macintosh Version 2.3 update were distributed in 1996. They contained the 1996 MESH files as well as

new communication scripts for FTS2000, new MEDLINE and 'older material' coverage, and HSRPROJ and SPACELINE database search screens. Over 12,000 Macintosh and 52,000 PC DOS updates were mailed out.

A contract action to provide support for Grateful Med development and maintenance was completed this year. A multiple award was made to AAC Associates, Inc., and Taj Technologies, Inc. AAC was an incumbent contractor while this will be Taj Technologies initial contract with NLM.

LAN Support

- In December 1995 a new dedicated T3 (45 Mbps) line was installed and became active. This line upgraded NLM's connections to the BBN Planet network for the Internet from the previous T1 (1.54 Mbps) circuit.
- A SONET Ring loop was installed by Bell Atlantic in the spring of 1996. This connection goes from the Bell Atlantic facilities (Rockville, Md.) through the existing NIH conduit facilities on campus to Buildings 12 and 10 and back to Bell Atlantic. This SONET Ring brings a reliable, high-speed fibre networking capability to the NLM and NIH. NLM's T3 was converted to this ring in June 1996.
- A LAN superserver (Netframe) was procured and the process of converting users from desktop servers to the superserver was started.
- GROUPWISE calendaring became operational.

TESS

TESS (Technical Services System) integrates various functions of the Technical Services Division with 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 and the Internet. The system controls the creation and maintenance of bibliographic and name authority records for NLM. NLM distributes this data to ELHILL and to subscribers of USMARC records.

Initial releases of TESS provided for the creation and maintenance of original cataloging, followed by a framework for the integration of acquisition and cataloging activities. Next, authority control for the cataloging function was integrated into TESS. This release also included the capability to create and maintain the name authority file.

Early releases of TESS included the conversion of the entire CATLINE/AVLINE files to

be supported within the TESS environment. TESS was enhanced to support all file maintenance activities of CATLINE/AVLINE, and to support all cataloging distribution activities, principally the distribution of catalog records in MARC format to MARC subscribers.

Another effort was a modification of the databases and software to comply with the USMARC standard for Format Integration Phase I. A basic Z39 client was established to permit FTP transfer of data from the Library of Congress. Software was also developed to allow the batch loading of USMARC data tapes from outside sources

During FY 1996 software was developed for the batch uploading of NLM preservation data from Inquire DBMS files. The databases and software were also modified further to comply with the USMARC standard for Format Integration Phase II. Software was also developed to output the Name Authority data in USMARC record formats.

Information Systems Laboratory (ISL)

The Information Systems Laboratory was created as a core facility to modernize and utilize the emerging technology domains of distributed processing, open systems, high speed networks, and worldwide connectivity and service provision. An early software product, TC_COMM, has been widely used for TCP/IP communications within other production systems such as TESS, Locator/CMS, and the Z39.50 server.

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

The ISL is continuing to introduce open systems computers and work-stations 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.

During FY 1996 the Information Systems Laboratory continued to provide system support for Unix-based applications and services. Several Unix-based production systems now depend upon ISL hardware and software in order to support public user access to NLM data. Performance and operational analyses resulted in several applications being

changed to more appropriate hardware configurations. An important 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 originally developed in the Lister Hill Center, the WWW server has been restructured, reformatted, and re-implemented as an operational facility.

NLM LOCATOR

OCCS developed a prototype Online Public Access Catalog (OPAC) which became a full-scale development effort. The operational system, named NLM Locator, was made available over the Internet and its usage continues to grow. NLM Locator provides direct access to the NLM collections through the ELHILL databases CATLINE (monographs), AVLINE (audiovisuals), SERLINE (serials), and DIRLINE (information resources). This project was OCCS's initial Unix-based client/server development effort, and the development team was challenged to acquire skills in new technologies. The complete project included mastering new skills in new computer hardware, system software, networking, programming languages, and a number of utility tools.

One of the important concepts of the client/server architecture is the ability to adapt to change without reengineering the application. NLM Locator utilizes a workstation client communicating with function servers which in turn communicate with a data server. The client workstations in the Reading Room are DOS PC's, while Internet users log in as VT-100 terminal sessions to a Unix client process executing in the function servers. The function servers are Sun computer systems running the Unix operating system. Having multiple function servers permits reliability as well as additional capacity, should usage require it. If necessary, another function server could be added without software modification. The data server is the NLM mainframe computer utilizing the ELHILL retrieval system. No changes to the legacy systems were required to implement Locator.

The system continues to function well and comments received via the electronic mail feature have been most complimentary. Reading Room and NLM staff usage has stabilized. Internet usage continues to grow.

The second phase of the project was to provide circulation control and collection management features for a combined system. The

LOCATOR/Collection Management Systems (Locator/CMS). It provides online patron registration, availability data, status information, requests for library materials, and a great number of management reports to the Public Services Division. The ELHILL retrieval agent subsystem was converted to C++ and it now shares much of this code with the Z39.50 server, although both systems run on separate Unix servers.

Z39.50 Server

The project to make MEDLINE available to Internet clients via the Z39.50 protocol consists of an ELHILL retrieval agent, a Z39.50 protocol engine that delivers MEDLINE information over the Internet in Z39.50 format, and a subsystem for dynamically converting ELHILL data into Z39.50 formats. The Z39.50 formats delivered include Simple Unstructured Text Record Syntax (SUTRS) and MARC, which is the most widely used Z39.50 format.

The ELHILL retrieval agent supports simple searching and full Boolean searching. It is written in C++ and it runs under the Solaris operating system on a Sun computer. The Z39.50 protocol engine is written in C and also runs on the same Sun computer.

MEDLARS Self Registration/Code Spooling Project

New development efforts combined all user authority data from the ELHILL UUSF, the INQUIRE Users database, the TOXNET User File, the PDQ system, the billing system, and the central computer into one Consolidated User Authority System (CUADS). Software and online panels were written to allow individual and mass changes to this file and to permute these changes to the above-mentioned data sources. Changes made by users of any of NLM's Information Retrieval Services (ELHILL, PDQ, and TOXNET) are now electronically transmitted to the others without the necessity of making the changes individually to each. This new software also allows for the addition of individual User ID's, definitions (and capabilities) of subscriptions, and the addition of (or mass changes to) groups of codes belonging to a subscription. WEB servers have been written for the central computer system in support of MEDLARS Self Registration, User ID/Subscription maintenance, and code spooling.

The billing system has been changed to allow transmission of data to the National Technical Information Service via FTP as an interim step to transmission via the newly developed Batch

Transmission File (BTF) which was developed for the inter-system transmission. Tools have been built for the Code Spooler which will allow institutions to come to the NLM for MEDLARS information retrieval service (Internet Grateful Med) without the need of distributing individual User ID's. Beta-testing will be conducted with one or more of these institutions in the near future.

FY 1996 saw the implementation of a new I/O Subsystem with shared caching between the ELHILL servers which have halved the response time to 80 milliseconds taking advantage of the new IBM 9672/R52 Central Computer System and allowing for the support of about 400 simultaneous users and rapid response for the growing Internet Grateful Med (IGM) system. These changes should allow for a five- to ten-fold growth of IGM.

Upgrades

The 3090/300J computer system was upgraded to an IBM 9672/R52 in March 1996. This upgrade increased the "raw" processing capacity (CPU MIPS) by 38 percent, while retaining the same newer, more reliable technology. The disk storage system was upgraded in September 1996. This upgrade replaces older, less reliable technology with newer, more reliable technology. The operating system and related components (TCP/IP, etc.) were upgraded in July 1996. This upgrade applied major maintenance and capacity improvements to the system.

MEDLARS databases

- Health and HSTAR databases were combined into HealthSTAR and a backfile (HealthStar75). Currently HealthStar75 contains 1,731,986 records with coverage from 1975-1989. HealthSTAR contains 1,174,164 records from 1976 through 9612 (entry month).
- PREMEDLINE was created as a new MEDLARS database to allow access to citations while they are being indexed i.e., prior to their entry into the MEDLINE database. The database is updated daily.
- HISTLINE was reorganized to allow better access to old records. Work has begun on an OLDMED file for citations prior to 1966. This database will contain citations with a date of publication prior to 1966.

Citation Maintenance

- Development of maintenance systems for SPACELINE and HealthSTAR databases was

completed. This involved the creation of software to permit importing and processing data from NASA and the Emergency Care Research Institute in addition to accessing the NLM online databases for maintenance purposes.

- The BIOETHICS maintenance system now allows records to be imported from other databases.
- Development of the CHEMLINE Record Maintenance System (CRMS) was completed. This system will provide for the creation and maintenance of the MEDLARS CHEMLINE data dictionary records.
- The INDX65 system for data entry and maintenance of *Index Medicus* 1965 data was developed and implemented.

Automated Indexing Management System

An online INVOICE system using a Customer Information Control System interface to the Database Management System INQUIRE was developed and installed. This system provides for data entry/maintenance of invoice data received at NLM. Programs were written to process journal issue data encoded in the Standard Generalized Markup Language (SGML). This data was provided to the NLM's National Center for Biotechnology Information by various publishers.

A Local Area Network-based system was created to provide data entry/maintenance of journal issue data.

The system provides for linking of keyed associated article data to scanned/optical character recognition abstract data. The data are then placed in the MEDLARS input streams for indexing.

DOCLINE and Loansome Doc

- Plans have begun for installing the RELAIS system at NLM. RELAIS is a fully automated intelligent document order processing and delivery system that will streamline the processing of interlibrary loans at NLM. A prototype workstation has been designed and tested and the workflow documents were finalized. DOCLINE and Locator have begun to transfer requests to RELAIS. Modifications to the LEND module in DOCLINE are complete for new delivery methods—fax, Ariel, printer and e-mail. NLM is expecting a spring 1997 installation of the system.

- A World Wide Web interface for updating DOCUSER records was installed. DOCUSER is the database containing data related to users of the DOCLINE system.

- Access to HealthSTAR, HealthSTAR75 and PREMEDLINE was provided.

- An interface to add foreign addresses, comments, and Internet addresses for electronic delivery was added to Loansome Doc and Grateful Med processing.

ADMINISTRATION

Executive Officer
Donald C. Poppke

National Performance Review

In response to the Administration's National Performance Review, NLM continued to be actively involved in streamlining efforts and the Library's systems re-engineering initiative under its role as a reinvention laboratory.

Streamlining

Overall, the NLM met its downsizing goals for FY 1996 and continues to work toward meeting future targets. The NLM closed FY 1996 with 575 full-time equivalents (FTE). Our ceiling for FY 1997 is 571 FTE.

In an NIH-wide streamlining effort to reduce support positions, several Competitive Service Centers have been established. NLM's Office of Acquisition Management became an NIH Competitive Acquisitions Service Center last year and provided procurement support for a selected group of contracts from other NIH components.

A number of important delegations of authority in the acquisitions area were redelegated to NLM's Chief Contracting Officer. Also, numerous personnel delegations of authority were granted to all NIH components by the NIH Office of Human Resources Management.

Systems Reinvention Project

Clearly, one of our most important and highly visible efforts in support of the National Performance Review is the NLM's Systems Re-engineering Initiative. The NLM is a designated "Reinvention Laboratory" and the theme of this high priority effort is 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. Making our services even more responsive to customers and to the nation's health information needs with fewer staff resources will be a

most significant accomplishment. The NLM System Reinvention project made progress in several areas in FY 1996:

- **Integrated Library System (ILS).** The ILS team completed a detailed definition of system requirements and system evaluation criteria and a survey of candidate ILS products. NLM plans to use the NIH Image World contract, early in FY 1997, to engage a library systems integrator to design and build the required system.
- **Document Delivery System.** A contract was awarded to Network Systems Inc. in June 1996 to install their document delivery system, Relais. 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 by June 1997.
- **Internet Grateful Med (IGM).** This World Wide Web service provides easy access to several of NLM's databases. The initial release in the spring of 1996 offered MEDLINE searching. Version 2.2, released in September 1996, added access to PREMEDLINE, HealthSTAR and AIDSLINE. Many more databases will be added in the coming months. IGM has proven to be very popular with the public. By August 1996, IGM was accounting for a quarter of the individuals who searched MEDLINE or other NLM databases during the month.
- **Grateful Med for Windows.** This newly designed version of Grateful Med, which provides access to MEDLINE, underwent extensive beta testing in 1996. User reaction to the product has been very favorable. A production Version 1.0 will be distributed to active users of the current DOS version of Grateful Med in January 1997.

Financial Resources

In FY 1996, the Library had a total appropriation of \$140,563,000. Table 12 displays the FY 1996 budget authority plus reimbursements from other agencies, and the allocation of these resources by program activity.

Table 12

Financial Resources and Allocations, FY 1996
(dollars in thousands)

Budget Authority:	
Appropriation, NLM	\$140,563
Plus: Reimbursements	13,283
TOTAL	\$153,846
Budget Allocation:	
Extramural Programs	\$ 33,285
Intramural Programs	112,925
Library Operations	(61,772)
Lister Hill National Center for Biomedical Communications	(34,306)
National Center for Biotechnology Information	(9,858)
Toxicology Information	(6,989)
Research Management and Support	7,636
TOTAL	\$153,846

The 1996 appropriation language authorized the Library to use personal services contracts and provided that \$4.0 million be available without fiscal year limitations. These authorities are key elements of NLM's system reinvention initiative.

Personnel

We are sad to report that the National Library of Medicine lost a long-time friend and co-worker this year. *Richard T. West*, Chief of the Office of Program Planning and Evaluation, Extramural Programs, died unexpectedly at his home on October 1, 1996. Mr. West had been with the Library for some 25 years and will be greatly missed.

Appointments

In November 1995, *Donald C. Poppke* was appointed as NLM's new Executive Officer. Mr. Poppke was formerly the Executive Officer of the National Center for Nursing Research, NIH, and before coming to NLM he was the Chief of the Public Health Branch in the Office of the HHS Assistant Secretary for Management and Budget. He received a B.S. degree in biological sciences from Washington and Lee University and an M.S. degree in the technology of management from American University. As Executive Officer, Mr. Poppke is responsible for planning and directing the

administrative management functions of the Library including financial, personnel, and contracts management, administrative services, and management analysis and legislation.

In February 1996, the Library made its first appointment under the Silvio Conte Senior Biomedical Research Service (SBRS), a new career pathway for outstanding basic science and clinical researchers at the NIH. The SBRS, comparable to the Senior Executive Service, is designed for scientists actively engaged in research. *James M. Ostell, Ph.D.*, with the National Center for Biotechnology Information (NCBI) is the first NLM employee selected for a Senior Biomedical Research Service position. Dr. Ostell has made original and important independent contributions to biomedical research in the new field of bioinformatics. He has conceived, designed, built, and delivered molecular biology systems in a broader, more unified and biologically rational way than has ever been before.

Also in February, *Kathleen Gardner Cravedi* joined the NLM as a Special Expert with the Office of Public Information. Prior to joining NLM, Ms. Cravedi worked as Director of Public Policy of the Health Foundation for Hospice and Homecare in Washington, D.C. and served as Deputy Executive Director of the Alliance for Aging, also in Washington, D.C. As a Special Expert with OPI, Ms. Cravedi will develop public information and outreach-related strategies.

In March, *Robert Ploger* was selected to receive a Fellowship under the new Technical Intramural Research Training Award (IRTA) program. This new program is designed to produce highly trained support professionals capable of performing the latest advanced techniques in a laboratory. Mr. Ploger, who will be working with NCBI staff, will gain experience in the use of UNIX operating system, programming in the computer language C, and database management using the SYBASE relational database system. He will work on biomedical sequence databases used for mapping the human and other genomes, as well as for identifying, isolating and sequencing genes of biological and medical interest.

In April, *Aaron Marchler-Bauer, Ph.D.* was appointed as an NCBI Visiting Fellow. He received his Ph.D. in biochemistry from the University of Vienna. Dr. Bauer has extensive experience in protein threading, an area that demands considerable software skills, and in applied structural analysis. At NCBI, Dr. Bauer will work on software development

for the NCBI macromolecular structure database and on independent research projects in the area of protein threading and structure comparison. His research will allow NCBI to further develop comparative analysis tools that make use of 3-D structure information.

In June, *Stewart J. Nelson, MD* was appointed head of the Medical Subject Headings Section (MeSH), Library Operations. Dr. Nelson received his medical degree from the State University of New York, Downstate Medical Center. Prior to joining the Library, Dr. Nelson practiced and taught general internal medicine at the Medical College of Georgia and at the State University of New York at Stony Brook. In addition, he worked as a consultant to Lexical Technology, Inc., where he collaborated in the development of the UMLS Metathesaurus. Dr. Nelson is a Fellow of the American College of Physicians.

In July, *Shih-Hsin Yang* was selected to receive a Predoctoral Intramural Research Training Award (IRTA) Fellowship with the NCBI. Mr. Yang will learn the techniques involved in using computers to help solve some of the information problems that arise in scientific research. He will collaborate with scientists at NCBI who are working on the macromolecular structure database and the Entrez Genomes Project.

Also in July, *Mark Alan Hershkovitz, Ph.D.* received a Postdoctoral Intramural Research Training Award (IRTA) Fellowship from the NCBI. Dr. Hershkovitz received his Ph.D. in botany from the University of California. Before his appointment with NCBI, Dr. Hershkovitz served as a Postdoctoral Scientist at the Smithsonian Institution, Laboratory of Molecular Systematics. While at the NCBI, Dr. Hershkovitz's research will include the analysis of the taxonomy of species in GenBank, the advancing of systematics of angiosperms, caryophytes, and organismal evolution at the molecular level.

Awards

The NLM Board of Regents Award for Scholarship or Technical Achievement was awarded to three individuals: *Wen-Min Kao, M.L.S.* for outstanding intellectual and editorial contributions to the fifth edition of the *National Library of Medicine Classification*, and to *Mark S. Boguski, MD, Ph.D.* and *Gregory D. Schuler, Ph.D.* for technical achievement in producing the human transcript map, a compilation and synthesis of the current state of knowledge in human gene mapping.

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 1996 award was *Joseph W. Hutchins*, Office of Computer and Communications Systems, for his exceptional contributions that have significantly improved NLM's indexing, database access, and document delivery functions.

The NIH Director's Award was presented to two employees this year: *Michael J. Ackerman, Ph.D.* and *Lawrence C. Kingsland, III, Ph.D.* Dr. Ackerman was recognized for his leadership of the Visible Human project, culminating in the release to the scientific community of two immense datasets representing a 3-dimensional male and female. Dr. Kingsland was recognized for his leadership in creating, testing, and implementing the Internet Grateful Med software that is revolutionizing how MEDLINE is searched by health professionals around the world.

The NLM Director's Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to two employees: *Joseph P. Fitzgerald*, Lister Hill National Center for Biomedical Communications, and *David J. Lipman, MD*, National Center for Biotechnology Information. Mr. Fitzgerald was recognized for his exceptional contributions to the mission of the NLM through the creative application of his artistic talent. Dr. Lipman was recognized for contributions of inestimable value to the research mission of the Library and for leading the National Center for Biotechnology Information to international prominence.

The NIH Merit Award was presented to seven employees: *Fernando Burbano*, Office of Computer and Communications Systems; *Pandora L. King*, Office of Administration; *Pamela A. Meredith*, Division of Library Operations; *Patricia S. Page*, Office of Administration; *Edmund J. Syed*, Lister Hill National Center for Biomedical Communications; *Cheryl E. White*, Division of Specialized Information Services, and posthumously, to *Richard T. West*, Division of Extramural Programs. Mr. Burbano was recognized for his efforts in sustaining the NLM computer operations at peak efficiency during furloughs, blizzards, and other exigencies. Ms. King was recognized for her exceptional dedication and excellence in providing personnel support services to the Library. Ms. Meredith was recognized for exceptional leadership and achievement in providing effective reference

services for NLM's Reading Room patrons and customers worldwide. Ms. Page was recognized for exceptional management and effective administration of the contracting function at the NLM. Mr. Syed was recognized for creative excellence while serving as software architect, principal system designer, and lead programmer for the Internet Grateful Med. Ms. White was recognized for exceptional achievement in producing documentation for SIS databases and effectively using the new medium of the Internet to distribute information. Mr. West was recognized for dedicated nurturing of NLM's Integrated Advanced Information Management System grant program to national renown as a facilitator of unified information systems at medical centers.

The NIH Purchasing Award was presented to *Liem T. Nguyen* for demonstrating outstanding leadership and resourcefulness in support of the NLM's small purchases activities.

Table 13

Staff, FY 1996 Full-Time Equivalents

<u>Program</u>	<u>Full-Time Permanent</u>	<u>Other</u>
Office of the Director.....	14	6
Office of Public Information.....	6	2
Office of Administration.....	50	8
Office of Computer and Communications Systems.....	53	6
Extramural Programs	13	5
Lister Hill National Center for Biomedical Communications.....	71	6
National Center for Biotechnology Information.....	22	11
Specialized Information Services	28	2
Library Operations.....	234	38
TOTAL	491	84
TOTAL FTEs.....	575	

NLM Diversity Council
Cassandra Allen (Chair)
Public Services Division

In July 1996, under the leadership of the NLM Director, the Library became the first component of the National Institutes of Health to establish a Diversity Council. The purpose of the Council is to provide advice to the Director to ensure the NLM, its staff and programs reflect the breadth

and richness of character of the society we serve and that everyone within the organization has the opportunity to reach his or her fullest potential.

The 11-member Council meets on a regular schedule and is mapping out a plan of action for the coming year. The Council embodies the value of bringing all voices to one table to discuss similar issues in a single advisory body.

Office of Equal Employment Opportunity

David Nash
NLM EEO Manager

In FY 1996, NLM identified areas of under-representation in the composition of its staff, especially Hispanic and Native Americans. The NLM made special efforts to cooperate with such organizations as the Association of American Indian Physicians and the InterAmerican College of Physician and Surgeons about current NLM job vacancies and other opportunities at the Library. There were exhibits and health information seminars at meetings and conferences emphasizing health issues affecting Hispanic and Native Americans. Several large groups of Hispanic high school and college students have toured the Library in the past year.

In working with African Americans, NLM's EEO Manager attended and exhibited at such conferences as the National Association for Equal Opportunity in Higher Education, composed of the 117 Historically Black Colleges and Universities. He also attended minority training seminars sponsored by the National Medical Association, National Urban League, National Association for the Advancement of Colored People, National Association of Health Services Executives, Symposium on Career Opportunities in Biomedical and Public Health Sciences, and Blacks In Government.

NLM has begun to explore developing a special outreach program with Largo Senior High School, a predominantly African American School in Maryland. This school has a four-year biotechnology medical career track program designed to train students who are seeking careers in the medical and biotechnology fields. Working with senior staff of the National Center for Biotechnology Information, a series of tours and presentations of NLM's biotechnology programs and facilities were conducted.

NLM's EEO Manager became trained and designated as a Federal mediator in the Alternative Dispute Resolution (ADR) program. The objective of the ADR program is to improve the resolution rate of EEO-related complaints, especially those that are beyond the scope of EEO laws and regulations. The

training has allowed the EEO Manager the opportunity to provide the NLM Director and staff

with expert guidance and alternatives in addressing and resolving employee-related concerns.

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
(212) 876-8763 FAX (212) 534-7042
States served: DE, NJ, NY, PA
2. **SOUTHEASTERN/ATLANTIC REGION**
University of Maryland at Baltimore
Health Sciences Library
111 South Greene Street
Baltimore, MD 21201-1583
(410) 706-2855 FAX (410) 706-0099
States served: AL, FL, GA, MD, MS, NC, SC, TN, VA, WV, DC, VI, PR
3. **GREATER MIDWEST REGION**
University of Illinois at Chicago
Library of the Health Sciences
P.O. Box 7509
Chicago, IL 60680
(312) 996-2464 FAX (312) 996-2226
States served: IA, IL, IN, KY, MI, MN, ND, OH, SD, WI
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
5. **SOUTH CENTRAL REGION**
Houston Academy of Medicine-Texas Medical Center Library
1133 M.D. Anderson Boulevard
Houston, TX 77030
(713) 790-7053 FAX (713) 790-7030
States served: AR, LA, NM, OK, TX
6. **PACIFIC NORTHWEST REGION**
University of Washington
Health Sciences Center Library
Box 357155
Seattle, WA 98195
(206) 543-8262 FAX (206) 543-2469
States served: AK, ID, MT, OR, WA
7. **PACIFIC SOUTHWEST REGION**
University of California, Los Angeles
Louise Darling Biomedical Library
10833 Le Conte Avenue
Los Angeles, CA 90024-1798
(310) 825-1200 FAX (310) 825-5389
States served: AZ, CA, HI, NV and U.S. Pacific Territories
8. **NEW ENGLAND REGION**
University of Connecticut Health Center
Lyman Maynard Stowe Library
263 Farmington Avenue
Farmington, CT 06030-4003
(203) 679-4500 FAX (203) 679-1305
States served: CT, MA, ME, NH, RI, VT

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:

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

ALBRIGHT, Tenley E., MD
Two Commonwealth Ave.
Boston, MA

BALL, Marion, Ed.D.
Vice President for Information Sciences
University of Maryland at Baltimore
Baltimore, MD

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

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

DeBAKEY, Michael E., MD
Chancellor
Baylor College of Medicine
Houston, TX

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

FULLER, Sherrilynn, Ph.D.
Director, Health Sciences Library
University of Washington
Seattle, WA

GAGE, John
Director, Science Office
Sun Microsystems Computer Corp.
Mountain View, CA

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:

MITCHELL, Joyce A., Ph.D. (Chair)
Director, Medical Informatics
University of Missouri-Columbia
Columbia, MO

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

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

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

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 Sharp & Dohme
Rahway, NJ

PACE, Norman R., Ph.D.
Distinguished Professor of Biology
Indiana University
Bloomington, IN

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

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

ZAMUDIO, Carlos
Director, Bioinformatics and Engineering
Sequana Therapeutics
La Jolla, CA 92037

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:

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

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

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

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

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

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

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

IYENGAR, S. Sitharama, Ph.D.
Professor and Chairman of Computer Science
Louisiana State University
Baton Rouge, LA

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

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

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

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

RINDFLEISCH, Thomas
Director, MCAMIS/SSRG
Stanford University
Stanford, CA

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

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

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

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

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:

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

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

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

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

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

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

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

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 1996

Awards preceded by an "*" are first year, or competing awards, all others are continuation of awards made in prior years.

Fellowships

Informatics Research Fellowships

- *Ambrosino, Richard \$43,684
University of Pittsburgh
Improving decision-support systems with domain knowledge
- Carter, Alexandre R. \$32,998
Harvard University
Minority predoctoral fellowship program--NIGMS
- Kogelnik, Andreas M. \$14,508
Emory University
Mitochondrial genetics information system
- *Soo Hoo, KENT, Jr. \$17,569
University of California San Francisco
Content based indexing for medical image management
- Zimmerman, Diane E. \$31,600
Rutgers the State University, New Brunswick
Protein NMR spectra using artificial intelligence

Applied informatics Fellowships

- Bradley, Johanna R. \$55,600
National Library of Medicine
Creation, access and use of networked health information
- Gray, Paul A., Jr. \$48,171
Duke University
Developing patient outcomes from clinical observations
- Legler, James D. \$28,000
University of Texas, San Antonio
Health Science Center
Assessment of city wide childhood immunization database
- *Weiner, Mark G. 50,696
University of Pennsylvania
Health service research workstation

Internet Connections Grants

- *Bailey, Nora J. \$28,849

- Sacred Heart Hospital of Pensacola, FL
- *Certo, Tindara J. \$50,000
Fayetteville AHEC, NC
Internet access for rural health professionals
- *Cook, Nedra J. \$29,062
Fort Sanders Regional Medical Center, TN
- *Cullander, Christopher \$22,943
University of California, San Francisco
Internet access at affiliated hospitals
- *Dujsik, Gerald \$30,000
Illinois College of Optometry
- *Dyer, David W. \$17,436
Masonic Medical Research Laboratory, Inc. NY
Full internet access for the MMRL
- *Ervin, Frank R. \$30,000
McLeod Regional Medical Center, SC
MRMC Internet access to the Pee Dee
- *Grimaldi, Deborah J. \$32,310
Doctors Community Hospital, MD
- *Hamblin, Garth M. \$49,298
Bartlett Memorial Hospital, AK
Access for rural health services in SE Alaska
- *Hirsch, Marga R. \$29,594
Pennsylvania College of Podiatric Medicine
- *Hunt, Ronald D. \$30,000
Harvard University NE Regional Primate Research Center
- *Lynch, John T. \$49,916
Connecticut Hospital Association
Connecting CT healthcare institutions to the Internet
- *Montgomery, James B. \$31,602
Central State Hospital, VA
- *Murry, Carol \$50,000
Ke Ola O Hawai'i, Inc., HI
Community partnerships healthnet
- *Norcross, Natalie \$28,924
Tuality Healthcare, Inc., OR
- *Palmore, Jonathon D. \$28,300
Community Memorial Hospital, VA
Healthcenter Internet connection
- *Romanoski, Frank T. \$19,986
Memorial Health Alliance, Mount Holly NJ
- *Roth, Juanita J. \$30,000
Southwest Missouri State University
Campus Internet expansion

- *Shatzkin, Nance \$30,000
St. Luke's Roosevelt Hospital Center, NY
- *Smith, Brian D. \$28,900
St. James Mercy Hospital, NY
- *Thomas, Karen K. \$26,850
Assn. Women's Health/Obstetric/Neonatal, DC
- *West, Dee W. \$30,000
Northern California Cancer Center
- *Wirth, Pamela R. \$50,000
Susquehanna Health System
- *Worley, Sandra \$48,104
Midwestern University, IL
- *Zajac, Rosalie V. \$49,136
Good Samaritan Hospital, OH
- *Ziegler, Judith E. \$28,481
Crozer-Keystone Health System, PA

Integrated Advanced Information Systems (IAIMS)

Phase I Grants

- *Curry, David S. \$134,487
University of Iowa
- *Frisse, Mark E. \$134,954
Washington University, MO
- *Gardner, Reed M. \$89,668
LDS Hospital, UT
- *Larson, Paul F. \$134,924
University of Medicine & Dentistry of NJ
- *Mitchell, Joyce A. \$550,000
University of Missouri at Columbia
- *Niland, Joyce C. \$149,964
City of Hope National Medical Center, CA
- *Sokol, Robert J. \$149,688
Wayne State University, MI

IAIMS, Phase II

- Fuller, Sherrilynne \$549,998
University of Washington, WA
Creating the UWHSC integrated information infrastructure
- Paton, John A. \$546,067
Yale University, CT
IAIMS implementation at Yale
- Stead, William W. \$571,500
Vanderbilt University, TN
Fast track provision of IAIMS

Library Information Science Research Grant

- Sievert, Maryellen C. \$191,460

University of Missouri, Columbia
Retrieval from full-text medical databases

Publication Grants

- Baldwin, Peter \$34,818
University of California, Los Angeles
State intervention against contagious disease
- Barr, Bernadine C. \$25,000
Stanford University, CA
Spare children--subjects in research, 1900-1945
- Carmichael, Ann G. \$8,087
Indiana University, Bloomington
History of state registration of causes of death
- Faden, Ruth R. \$33,000
Johns Hopkins University, MD
Ethics & allocation--foundations of American health care
- Funk, Carla J. \$4,050
Medical Library Association
History of the MLA
- Hast, Malcolm H. \$35,405
Northwestern University, IL
Annotated translation of Vesalius's *Fabrica*
- Lays, Ruth \$25,000
Johns Hopkins University, MD
Western concepts of psychic trauma, 1875-1995
- *Micale, Mark Stephen \$18,000
Individual award, CT
History of psychiatric classification from Hippocrates
- Rettig, Richard A. \$43,960
Rand Corporation, CA
Public policy and the treatment of kidney failure
- *Rosse, Cornelius \$33,153
University of Washington, WA
Interactive atlas of thoracic viscera
- Sievert, Maryellen C. \$26,801
University of Missouri, Columbia
Missouri thesaurus of medical informatics terminology
- *Tracy, Sarah W. \$39,701
Rutgers the State University, NJ
Medicalization of alcoholism in America, 1870 to 1920

Information Access Grants

- *Bowden, Virginia M. 120,418
University of Texas, San Antonio
Medical Internet South Texas (MIST)
- Bulgarelli, Nancy T. \$54,570
William Beaumont Hospital, MI
Tachia electronic information access for rural hospitals
- *Pittman, Jan M. \$102,046

Three Rivers AHEC, GA
Three rivers triad project
*Richwine, Margaret \$92,419
Indiana & Purdue Universities at Indianapolis
Internet library services for medical education
*White, Billie S. \$106,937
Mercy Medical Center, CA
Rural northern CA medical internet connection

Information Systems Grants

Adams, Mignon S. \$54,920
Philadelphia College of Pharmacy, PA
Electronic pharmacy library
*De Bord, William J. \$52,625
Morehead State University, KY
Health sciences information data link
Dennis, Sharon E. \$166,939
University of Utah
Model multimedia support center for the health science
*Jones, Michael O. \$146,544
University of California, Los Angeles
Computerized archive of American traditional medicine
Hagen, Michael D. \$67,132
University of Kentucky
Information for medical students in rural settings
*Kingsbury, David T. \$95,544
Johns Hopkins University, MD
Internet based imagebase for medical education
Messerle, Judith R. \$116,922
Harvard University, MA
Creation of a knowledge laboratory
Oppenheim, Roberta A. \$149,969
Forsyth Dental Center, MA
Oral health sciences information system (OHSIS)
*Scott, Ursula D. \$134,725
Baylor Research Institute, TX
Baylorlink
Self, Phyllis C. \$143,022
Virginia Commonwealth University
Drug information automated library (DIAL)
Seltzer, Ada M. \$83,410
University of Mississippi Medical Center
Health sciences information network
Troutman, William G. \$174,661
University of New Mexico, Albuquerque
Interactive medical information system in NM

Research Training in Medical Informatics

Clayton, Paul D. \$219,022
Columbia University, NY
Cooper, Gregory F. \$317,840
University of Pittsburgh
Downs, Stephen M. \$392,388
University of North Carolina, Chapel Hill
Gatewood, Lael C. \$502,337
University of Minnesota
Gorry, G. Anthony \$281,180
Rice University, TX
Greenes, Robert A. \$1,191,502
Harvard university, MA
Miller, Perry L. \$259,458
Yale University, CT
Mitchell, Joyce A. \$317,213
University of Missouri, Columbia
Shortliffe, Edward H. \$537,196
Stanford University, CA
Spackman, Kent \$159,672
Oregon Health Sciences University

Biotechnology FIRST Awards

Altman, Russ B. \$104,141
Stanford University, CA
Modeling and computing with uncertain structures
Goldstein, Richard A. \$102,770
University of Michigan, Ann Arbor
Computational approaches to protein structure prediction
Karp, Peter D. \$102,286
SRI International, CA
Biological knowledge-base management system
Wu, Cathy H. \$95,089
University of Texas, Tyler
Classification neural networks for genome research

Biotechnology Research Grants

*Altman, Russ B. \$324,757
Stanford University, CA
Representing biological data for molecular modeling
Brutlag, Douglas L. \$210,310
Stanford University, CA
Multiple representations of biological sequences
Davison, Daniel B. \$175,339
University of Houston, TX
Multiple sequence alignment server and algorithm development
Hardison, Ross C. \$212,398
Pennsylvania State University

- Network server for electronic genetic analysis
- *Miller, Webb C. \$393,494
 Pennsylvania State University
 Software for analyzing biosequence data
- *Myers, Eugene W. Jr. \$155,214
 University of Arizona
 Efficient software for the analysis of
 biosequences
- Pearson, William R. \$255,637
 University of Virginia, Charlottesville
 Comparison of protein sequences and structures

Biotechnology Resource Grants

- *Jurka, Jerzy W. \$244,203
 Genetic Information Research Institute, CA
 Repbase--a database of repetitive sequences
- Ledley, Robert S. \$1,090,179
 National Biomedical Research Foundation, DC
 Protein information resource
- *Markley, John L. \$446,305
 University of Wisconsin, Madison
 Biological magnetic resonance data bank
- *Pagon, Roberta A. \$150,330
 Children's Hospital and Medical Center, WA
 Helix--a directory of medical genetics
 laboratories
- Roberts, Richard J. \$162,032
 New England Biolabs, Inc., MA
 Rebase--the restriction enzyme database
- Smith, Temple F. \$840,262
 Boston University
 Biomolecular engineering research center

Electronic Medical Records Systems Cooperative Agreements

- Barnett, Guy O. \$444,053
 Massachusetts General Hospital
 Point of care computer-based ambulatory
 patient record
- Chute, Christopher G. \$100,000
 Mayo Foundation, MN
 Multi-institutional testbed for clinical
 vocabulary
 (co-funded with another agency)
- Cimino, James J. \$400,934
 Columbia University, NY
 Controlled vocabulary in primary care
 electronic records
- Hersh, William R. \$283,557
 Oregon Health Sciences University
 Vocabulary and text data extraction from the
 EMR

- Kahn, Michael G. \$233,246
 Washington University, MO
 Automated tools to support health services
 research
- Kohane, Isaac S. \$287,176
 Children's Hospital, MA
 Multiplatform internet access to multimedia EMRs
- Safran, Charles \$123,448
 Beth Israel Hospital, MA
 Sharing paperless records among provider
 networks
 (co-funded with another agency)

Medical Informatics FIRST/Career Awards

- Cooper, Gregory F. \$99,157
 University of Pittsburgh
 Structuring medical knowledge--probabilistic
 inference
- Friedman, Carol \$109,392
 Queens College, NY
 Integrated medical text processing system
- Gorman, Paul Northrop \$99,107
 Oregon Health Sciences University
 Assessment of information seeking in primary
 care
- Hripcsak, George \$106,829
 Columbia University, NY
 Linking knowledge-based systems to clinical
 databases
- Johnson, Stephen B. \$107,858
 Columbia University, NY
 Access to medical information through natural
 language
- Kahn, Michael G. \$116,557
 Washington University, MO
 Time and context in medical databases
- *Langlotz, Curtis P. \$109,512
 University of Pennsylvania
 Computer-based explanation methods for
 decision models
- Lehmann, Harold P. \$109,527
 Johns Hopkins University, MD
 formalizing the notion of clinical significance
- Lenert, Leslie A. \$112,700
 Stanford University, CA
 Computer interpretation of free-text data
- *Rutledge, Geoffrey W. \$124,384
 Beth Israel Hospital, MA
 Advanced computer methods for ICU care
- *Shahar, Yuval \$104,035
 Stanford University, CA
 Knowledge based temporal abstraction of
 clinical data
- Sittig, Dean F. \$96,608
 Brigham and Women's Hospital, MA
 Intelligent real time information synthesis for

monitors
 Sonnenberg, Frank A. \$65,880
 University of Medicine & Dentistry of NJ
 Knowledge management for clinical decision
 analysis
 *Tong, David A. \$94,532
 University of Oklahoma
 Model-based interpretation of intracardiac
 electrograms
 Yang, Yiming \$131,529
 Carnegie-Mellon University, PA
 LLSF mapping for indexing and retrieval of
 MEDLINE

Medical Informatics Research Grants

Begg, Colin B. \$90,614
 Sloan-Kettering Institute for Cancer Research,
 NY
 Meta-analysis methodology for diagnostic test
 evaluation
 *Bowden, Douglas M. \$178,795
 University of Washington, WA
 Spatial/symbolic brain information management
 system
 Brinkley, James F. \$95,456
 University of Washington, WA
 Structural information framework for brain
 mapping
 (co-funded with another NIH institute)
 *Chute, Christopher G. \$176,155
 Mayo Foundation, MN
 Latent semantic indexing in support of data
 retrieval
 Elstein, Arthur S. \$187,807
 University of Illinois at Chicago
 Effect of decision support systems on clinical
 reasoning
 Ezquerra, Norberto F. \$239,847
 Georgia Tech Research Corporation
 Knowledge-based system for cardiac image
 interpretation
 Friedman, Charles P. \$178,690
 University of North Carolina, Chapel Hill
 Information and cognition in medical education
 Goldbaum, Michael H. \$245,750
 University of California, San Diego
 Structured analysis of the retina
 Hayward, Robert S. \$115,370
 McMaster University, Canada
 Evaluation of a database of practice guideline
 abstracts
 *Jaffe, Carl C. \$231,240
 Yale university, CT
 Indexing image databases for motion similarity

retrieval
 Jelliffe, Roger W. \$431,187
 University of California, Los Angeles
 New decision supports and databases for drug
 dosage
 *Lang, Walter P. \$151,529
 University of Michigan, Ann Arbor
 MEDLINE and computer conferencing by
 dentists
 Letovsky, Stanley I. \$83,590
 Johns Hopkins University, MD
 Spatially oriented database for digital brain
 images
 (co-funded with another NIH institute)
 Musen, Mark A. \$479,180
 Stanford University, CA
 Software architecture for guideline directed
 therapy
 Patil, Ramesh S. \$254,544
 University of Southern California
 Clinical management of critical illness using AI
 Sahni, Sartaj K. \$279,223
 University of Florida
 Algorithms for compression and registration of
 brain MRI
 Selker, Harry P. \$283,005
 New England Medical Center, MA
 New mathematical models for medical events
 *Shiffman, Richard N. \$100,000
 Yale University, CT
 Knowledge processing for clinical practice
 guidelines
 Toga, Arthur W. \$208,743
 University of California, Los Angeles
 Digital representation and visualization of
 human brain
 Webber, Bonnie L. \$224,960
 University of Pennsylvania
 Effective information delivery to clinical
 personnel
 Widman, Lawrence E. \$172,974
 University of Oklahoma Health Sciences Center
 Knowledge-based interpretation of cardiac
 arrhythmias

Medical Informatics Resource Grants

Balas, E. Andrew \$94,897
 University of Missouri, Columbia
 Meta-analysis of clinical information service
 trials
 *Giuse, Dario A. \$100,000
 Vanderbilt University, TN
 Improving knowledge acquisition for medical
 diagnosis

*Ho, Bruce K. \$100,000
University of California, Los Angeles
fast browsing of pacs image archive via network
Shortliffe, Edward H. \$1,419,125
Stanford University, CA
Center for advanced medical informatics

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

Nelson, John M. \$380,198
Nelson Information Systems, Inc., MD
Advanced pilot study of a journal management
system
Jimison, Holly B. \$14,000
Levernois & Associates, Inc., OR
Multimedia tool enhancing informed consent
(co-funded with another NIH institute)

Index

A

Agency for Health Care Policy and Research	3, 39, 50
Agency for Toxic Substances and Disease Registry	23, 25
AI/RHEUM	27
AIDS Clinical Trials Information Network	24
AIDS Services	24
AIDSLINE	27
Animal Testing	25
Ann Landers	2
Applied informatics fellowships	49
Appropriation	57
Ariel	14
ASN I	46
Associate Program	15
Audiovisual Program Development Branch	40
AVLINE	9, 53, 54
Awards	59

B

BankIt	45
Bibliographic Control	10
BIOETHICSLINE	51
Biomedical ethics	51
Biomedical Library Review Committee	51
Biotechnology informatics	50
BITNIS	5
BLAST	44
Board of Regents	4, 51
Board of Scientific Counselors	26, 46
Bulletin of the Medical Library Association	6

C

Cardiac Embryology Project	41
Cataloging	10
Categorical Grammar Parser	38
CATLINE	53, 54
Centers for Disease Control and Prevention	15
Cervical Cancer Project	41
Chemical Carcinogenesis Research Information System	23
CHEMID	23
CHEMLINE	24, 56
China	11
Chinese University of Hong Kong	5
Conference Grants	50
Consensus Development Conferences	13
Consolidated User Authority System	55
Corporation for National Research Initiatives	28
Cumulated Abridged Index Medicus	13
Current Bibliographies in Medicine	13, 15
Customer Service	8, 16

D

Department of Transportation	23
------------------------------	----

Developmental and Reproductive Toxicology	23
Digital Microscopy System	35
DIRLINE	9, 24, 54
Disaster Recovery Plan	10
Distributed Learning Environments	35
Diversity Council	60
DOCLINE	14, 56
Document Delivery	14
Document Imaging Laboratory	34
DocView	31
DXPNET	33

E

Educational Technology Grants	50
Electronic Medical Record Systems	50
Electronic Medical Record Systems grants	50
ELHILL	53
EMBL	44
Entrez	44
Environmental Mutagen Information Center	23
Equal Employment Opportunity	60
Exhibits	15
Expert Systems Program	27
Expressed Sequence Tags database	44
Extramural Programs Division	48
Extramural Programs Subcommittee	51

F

First Independent Research Support Transition Awards	50
Fixed-fee arrangements	12
Frank B. Rogers Award	59
Friends of the National Library of Medicine	2
FTS2000	53
Full Text-Online Reference Works	29

G

GenBank	43, 45
GENE-TOX	23
Genome Survey Sequence	43
Gopher	54
Grants for Education and Training of Health Sciences Librarians	49
Grateful Med	6, 12, 53
Guide to Clinical Preventive Services	12

H

Hazardous Substances Data Bank	22
Health Hotlines	24
Health Sciences Serials	13
HealthSTAR	12, 27, 55
Holland	9
High Performance Computing and Communications	26
High Throughput Genome	43
HISTLINE	12, 55

Historical acquisitions	9
Historical lectures	15
Historically Black Colleges and Universities	25
History of Medicine Division	11, 15
Hong Kong	5
HSRPROJ	53
HSTAT	29
Human Brain Project	50
HyperDOC	28, 54

I

IAIMS Grants	48
Image Compression	32
Image Processing Laboratory	34
Images from the History of Medicine	3, 11
Index Medicus	11, 13, 56
Indexing	11, 40
Informatics research fellowships	49
Information Access Grants	48
Information Sources Map	26, 28, 39
Information Systems Grants	48
Information Systems Laboratory	54
Institute of Medicine	1
Institutional training grants	49
Integrated Advanced Information Management Systems	48
Integrated library system	8, 10, 11, 57
Integrated Risk Information System	23
Interlibrary loan	14, 31
International Council for Scientific and Technical Information	6
International MEDLARS Centers	3, 4
International Programs	4
Internet	6, 12, 13, 14, 24, 31, 34, 35, 39, 53
Internet Connections Grants	48
Internet Grateful Med	2, 12, 26, 55, 57
Intranet	13
Islamic Medical Manuscripts	13
Israel MEDLARS Center	5

J

Johns Hopkins University	41
--------------------------	----

L

Large Scale Vocabulary Test	36, 39
LexBuild	37
Lexical Systems Project	36
Library of Congress	9, 11, 13, 54
Library Operations	8
Lister Hill National Center for Biomedical Communications	26
Literature Selection Technical Review Committee	11
Lithuania	5
Loansome Doc	56
Locator	54
Long Range Plan	4
Louise Darling Medal	9

M

Machine Learning Project	29
Medical Article Record System	30
Medical informatics research grants	49
Medical Informatics Training Program	29
Medical Information Retrieval System	32, 34
Medical Library Assistance Act	48
Medical Library Association	10
Medical Subject Headings	10, 59
MEDLINE	2, 5, 8, 11, 27, 30, 40
MeSH	3
MeSH Supplementary Chemicals	13
MetaMap	38
Metathesaurus	3, 10, 26, 27, 35, 37, 38, 39
Minority Support	51
Mississippi Delta Project	25
Molecular Modeling Database	44, 46

N

National Academy of Sciences	1
National Aeronautic and Space Administration	12
National Agricultural Library	9
National Cancer Institute	22
National Center for Bioethics Literature	51
National Center for Biotechnology Information	43
National Center for Health Statistics	33
National Information Center on Health Services Research and Health Care Technology	13
National Information Infrastructure	1
National Institute of Environmental Health Sciences	23
National Institute of Occupational Safety and Health	23
National Network of Libraries of Medicine	14, 27
National Performance Review	57
National Program for the Preservation of the Biomedical Literature	9
National Technical Information Service	53, 55
Native American health	6
Native Americans	15
Natural Language Systems Program	36
Newly Independent States	5
Next Generation Indexing	11, 40
NHANES	34
NIH Director's Award	59
NIH Merit Award	59
NLM Classification	10
NLM Director's Award	59

O

Oak Ridge Institute for Science and Education	16
OLDMED	55
OLDMEDLINE	12
Online Mendelian Inheritance in Man	44
Online Public Access Catalog	54
Online Services	12
Optical character recognition	11
OrfFinder	45
Outreach	6, 14, 25

P

Pan American Health Organization	5
PDQ	55
Peer Review Oversight Group	51
Personnel	58
Planning Panel on the Education and Training of Health Sciences Librarians	49
PREMEDLINE	12, 27 55
Preservation	9
Public Health Service Historian	42
Public Services Division	55
Publication Grant Program	50
PubMed	45

R

Reference services	13
Regional Medical Libraries	14
Registry of Toxic Effects of Chemical Substances	23
Reinvention	26
Relais	8, 14, 56, 57
RELAIS	56
Resource grants	48

S

Semantic Knowledge Representation Project	38
Semantic Network	38
SemRep	38
Sequence Tagged Sites	44
Sequin	45
SERHOLD	14
SERLINE	54
Small Business Innovation Research	50
Small Business Technology Transfer	50
Sourcerer	28, 38
SPACELINE	12, 15, 53, 55
Special Emphasis Panels	51
SPECIALIST	36, 38, 39
Specialized Information Services	22
Substance Abuse and Mental Health Services Administration	29
Superfund Amendments	22
SUPERLIST	23
System Reinvention	8, 11, 14, 57
Systems Re-engineering Initiative	57

T

Taxonomy Browser	44
------------------	----

Technical Services Division	10, 53
Telemedicine	1, 32
TESS (Technical Services System)	53
The Learning Center for Interactive Technology	34
Tokenizer	37
Tours	15
Toxic Chemical Release Inventory	22
Toxicology and Environmental Health Information Program	22
Toxicology Information Outreach Project	25
TOXLINE	24
TOXLIT	24
TOXNET	22, 55
Training, online	12
Training in Informatics	49
Training of minorities	49

U

UMLS Information Sources Map	9
UMLS Knowledge Source Server	39
UMLS Knowledge Sources	3, 38
UMLS Metathesaurus	2, 3, 12
Unified Medical Language System	3, 10, 26, 38, 39
Unique Human Gene Sequence Collection	44
User Access Services Project	26
User survey	6
USMARC	53

V

VacMan	45
Visible Embryo Project	34
Visible Human Project	2, 32, 41
Visitors	6, 15, 46

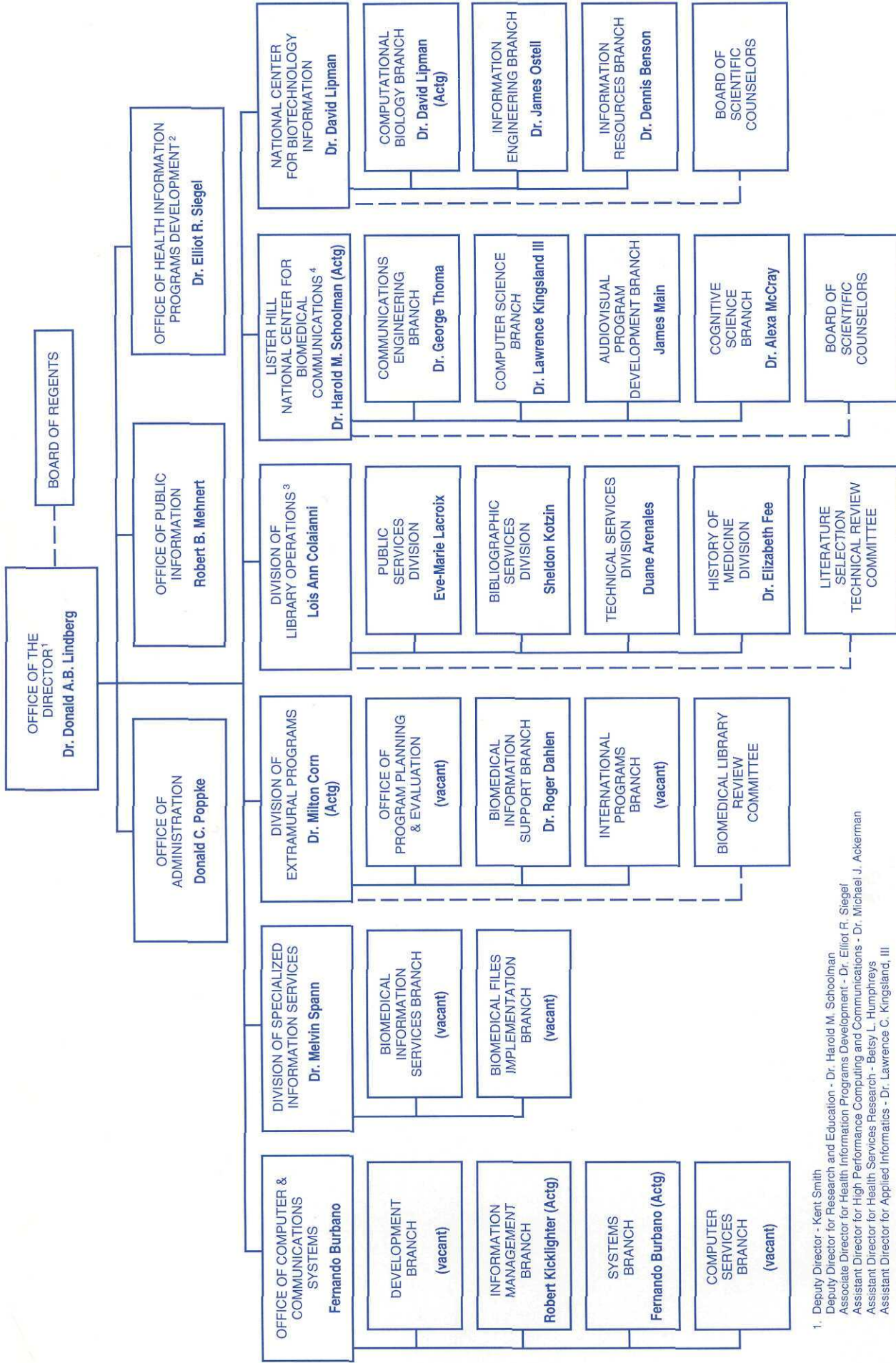
W

Wheaton Regional Library	24
WILL	31
Windows Grateful Med	12
Workstation for Interlibrary Loan (WILL)	31
World Health Organization	6
World Wide Web	13, 22, 24, 28, 35, 39, 41, 45, 54

Z

Z39.50	54, 55
--------	--------

National Library of Medicine



1. Deputy Director - Kent Smith
 Deputy Director for Research and Education - Dr. Harold M. Schoolman
 Associate Director for Health Information Programs Development - Dr. Elliot R. Siegel
 Assistant Director for High Performance Computing and Communications - Dr. Michael J. Ackerman
 Assistant Director for Health Services Research - Betsy L. Humphreys
 Assistant Director for Applied Informatics - Dr. Lawrence C. Kingsland, III

2. Includes International Programs

3. Includes:
 National Network of Libraries of Medicine Program, Program Coordinator - Becky J. Lyon
 Medical Subject Headings Section - Stuart Nelson, M.D.
 National Information Center on Health Services Research and Health Care Technology - Marjorie A. Cain

4. Office of High Performance Computing and Communications, Head - Dr. Michael J. Ackerman

NIH Publication No. 97- 256A