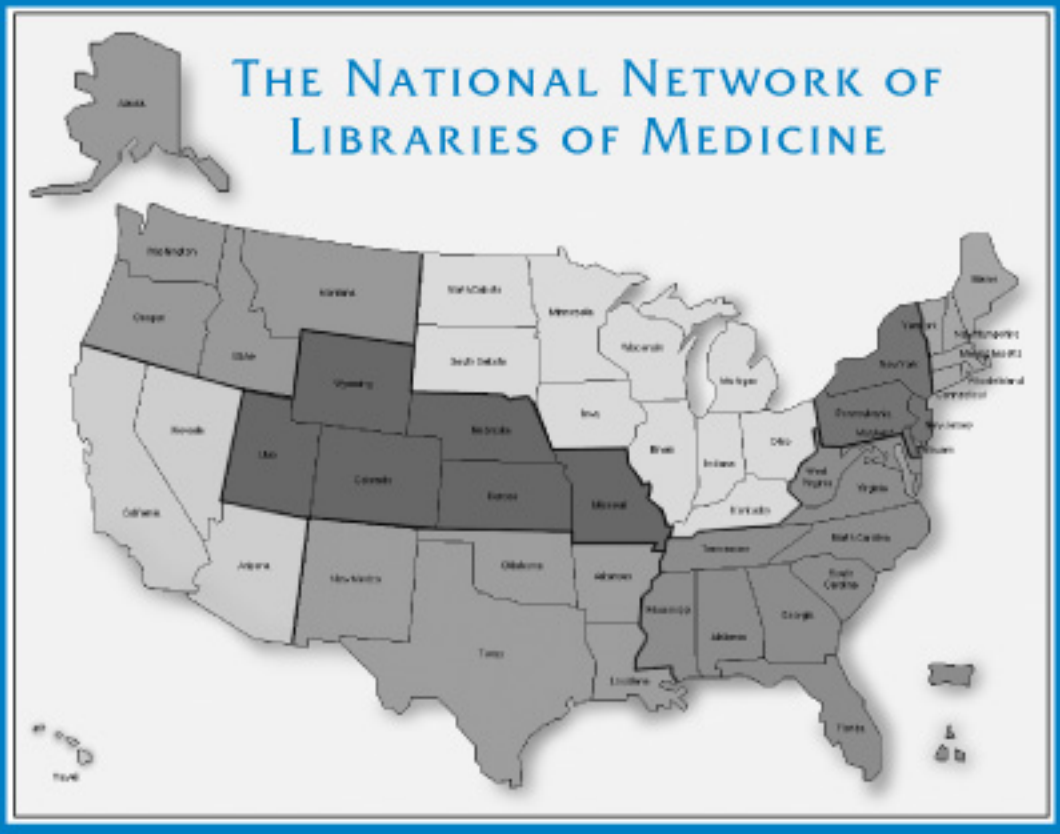
NATIONAL INSTITUTES OF HEALTH
NATIONAL LIBRARY OF MEDICINE
PROGRAMS & SERVICES 2001



DEPARTMENT OF HEALTH AND HUMAN SERVICES

1: Middle Atlantic Region: New York Academy of Medicine

2: Southeastern/Atlantic Region: University of Maryland at Baltimore

3: Greater Midwest Region: University of Illinois at Chicago

4: Midcontinental Region: University of Utah

5: South Central Region: Houston Academy of Medicine

6: Pacific Northwest Region: University of Washington

7: Pacific Southwest Region: University of California, Los Angeles

8: New England Region: University of Massachusetts

NATIONAL INSTITUTES OF HEALTH

NATIONAL LIBRARY OF MEDICINE

PROGRAMS AND SERVICES

FISCAL YEAR 2001

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
BETHESDA, MARYLAND

National Library of Medicine Catalog in Publication

Z
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 DHEW publication ; no. (NIH) 78-256, etc.; for 1979-80 as NIH publication ; no. 80-256, etc.
Vols. for 1981-available from the National Technical Information Service, Springfield, Va.
ISSN 0163-4569 = National Library of Medicine programs and services.

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

CONTENTS

Appendixes

Tables

PREFACE

The pages of this year's report reveal that the National Library of Medicine continues to make progress in creating ever-more useful information services for the health professions and the public. To name just a few of the highlights: free access to the DNA sequence of the human genome was made available by NLM's National Center for Biotechnology Information; a new database, ArcticHealth, was introduced by the Specialized Information Services Division; MEDLINEplus was improved by the addition of a daily news feed from the public media and some 30 interactive health education modules; two prominent scientists were added to the increasingly popular *Profiles in Science*; new 5-year contracts for the Regional Medical Libraries were awarded; web-based information resources were introduced on bioterrorism, anthrax, etc., that responded to the September 11 attacks; and the History of Medicine Division has created a wonderful new exhibit in the main rotunda, "The Once and Future Web." This list only scratches the surface of the progress we made in the last 12 months.

These accomplishments prompt me to reflect on the tremendous range of talent required to operate an institution as complex as the National Library of Medicine. We have librarians, of course—the people who acquire the diverse materials for the collection, who bring order to it by cataloging and indexing, and who help others have access to it. We have historians and preservationists. We have systems analysts and computer scientists. As befits an organization that is part of the National Institutes of Health, we have scientists and technical experts of many and various disciplines. And, of course, there are the many support staff whose efforts enable the work of others.

The design for expanding NLM's existing facilities is moving forward. On July 24, 2001, President Bush signed the 2001 Supplemental Appropriations Act. The Conference Report accompanying the bill directed that $7,115,000 be transferred to the National Institutes of Health "for purposes of the design of a National Library of Medicine facility." This transfer of funds, along with funds previously transferred to the NIH Buildings and Facilities account, clears the way financially for the completion of the design for the new and expanded facilities.

Finally, I would like to acknowledge the work of the many health professionals and information specialists who serve as advisors on the Board of Regents, Board of Scientific Counselors, and other advisory groups. Their perspective, as leaders in their fields, keeps us pointed in the right direction.

_____

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

OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT

Elliot R. Siegel, Ph.D.
Associate Director

The Office of Health Information Programs Development is responsible for three major functions:

- establishing, planning, and implementing the NLM Long Range Plan and related planning and analysis activities;
- planning, developing, and evaluating a nationwide NLM outreach and consumer health program to improve access to NLM information services by all, including minority, rural, and other underserved populations; and
- conducting NLM's international programs.

Planning and Analysis

NLM Long Range Plan

The NLM Long Range Plan 2000–2005, published in 2000, remains at the heart of NLM's planning and budget activities. Its four goals form the basis for NLM operating budgets each year:

- Organize health-related information and provide access to it;
- Encourage use of high-quality information by health professionals and the public;
- Strengthen the informatics infrastructure of biomedicine and health ; and
- Conduct and support informatics research.

Additionally, the NLM Board of Regents has identified in the Plan its highest priority new initiatives for special emphasis in the next five years:

- Health Information for the Public
- Molecular Biology Information Systems
- Training for Computational Biology
- Definition of the Research Publication of the Future
- Permanent Access to Electronic Information
- Fundamental Informatics Research
- Global Health Partnerships

All NLM Long Range Plan documents are available on the NLM web site.

Other Planning Activities

OHIPD maintains involvement in many NIH-related planning and evaluation activities, including the preparation of Science Advances and other materials required by NIH for the Government Performance and Results Act (GPRA) and appropriations hearings, and answering queries about NLM's involvement in a variety of disease and policy-related areas.

In addition to specific outreach and consumer health projects outlined below, OHIPD has overall responsibility for developing and coordinating the NLM Health Disparities Plan. This plan outlines NLM strategies and activities undertaken in support of NIH efforts to understand and eliminate health disparities between minority and majority populations.

This office has convened and is chairing the NLM Coordinating Committee on Outreach, Consumer Health and Health Disparities. This Committee plans, develops, and coordinates NLM outreach and consumer health activities. It is charged with:

- Articulating NLM's separate and overlapping goals for Outreach, Consumer Health and Health Disparities;
- Recommending program and funding priorities in each of these areas for the NLM budget for a 3–5 year period;
- Identifying target populations, collaborators and strategies for developing and undertaking new project initiatives within the context on NLM's Long Range Plan and Health Disparities Plan;

- Documenting specific plans for evaluating new and current activities; and
- Identifying and addressing specific implementation challenges that have proven difficult in the past.

It is important for NLM to be able to describe and analyze its outreach, consumer health, and health disparities projects in order to identify areas of opportunity, report on their progress, and plan for new initiatives. A major activity of the Committee in FY2002 will be to develop a database of NLM outreach and consumer health projects. NLM's Office of Computer and Communications Systems will develop, host, and support this database with assistance from OHIPD staff and the committee.

Outreach and Consumer Health

NLM carries out a diverse set of activities directed at building awareness and use of its products and services by health professionals in general and by particular communities of interest. Considerable emphasis has been placed on reducing health disparities by targeting health professionals who serve rural and inner city areas. Additionally, starting in 1998, NLM has undertaken new initiatives specifically devoted to addressing the health information needs of the public. These projects build on long experience with addressing the needs of health professionals and on targeted efforts aimed at making consumers aware of medical resources, particularly in the HIV/AIDS area.

Tribal Connections

NLM has recently focused on improving Internet connectivity and access to health information services in American Indian and Alaskan Native communities. Phase I (Pacific Northwest) of tribal connections is complete, with final project evaluation now under way. Phase 2 (Pacific Southwest) sites have been selected, and implementation is well along. Also, NLM has funded a Phase 3, in which more intensive community-based outreach and training will be implemented at select Phase 1 and 2 sites to assess if these community-based approaches significantly enhance the project impacts on health information, behavior, and outcomes.

NLM/OHIPD continues to support a special tribal connections project in the Southeast, with the American Indian Cultural Center/Piscataway Indian Museum in Waldorf, Maryland. The computer lab and computer learning center have been fully implemented, and the initial round of training has been completed. NLM is now evaluating the first year results, and discussing possible follow-on training and community-based activities of interest to the Center.

Also, in 2001 NLM/OHIPD partnered with the NIH and NLM EEO Offices to participate in the Acting NIH Deputy Director's American Indian Powwow Initiative. This included exhibiting at five powwows in the Mid-Atlantic area. An estimated 5,000 persons visited the NLM booth over the course of these powwows. These activities proved to be another viable way to bring NLM's health information to the attention to segments of the Native American community and the general public.

Outreach to Seniors

CyberSeniors/CyberTeens was initiated in 2001 and is intended to train computer savvy teenagers to help senior citizens learn how to use the Internet to access health information. The project has a strong evaluation component, intended to help measure the extent to which the health information seeking behavior and actual health decisions of the participating seniors are actually changed.

Outreach to Hispanics

The Lower Rio Grande Valley Hispanic Outreach Project is a collaboration with the University of Texas Health Sciences Center to conduct a needs assessment and various health information outreach projects with Hispanic-serving community, faith-based, and educational institutions. This is the beginning of an intensified NLM effort to meet the health information needs of the Hispanic population in Texas and elsewhere.

Web Evaluation

The Internet and World Wide Web now play a dominant role in dissemination of NLM information services. The web environment in which NLM operates is rapidly changing and intensely competitive. These two factors suggested the need for a more comprehensive and dynamic NLM web planning and evaluation process. Accordingly, the NLM Director established a Web Evaluation Work Group that has been operating now for about 18 months. The Work Group is chaired by the NLM Associate Director for Health Information Programs Development, and staffed by the OHIPD. The priorities of the Work Group include: quantitative and qualitative metrics of web usage, and measures of customer perception and use of NLM web sites.

During FY2001, the Work Group pursued an integrated approach intended to encourage exchange of information and learning within NLM, and help better inform NLM management decision-making on web site research, development, and implementation. The initial round of activities included:

- an online survey of a random sample of MEDLINEplus users;
- comparison of MEDLINEplus with other health information web sites;
- access to a syndicated telephone survey of the U.S. public's online and offline health information seeking behavior;
- analysis of NLM web site log data; and
- access to Internet audience measurement estimates based on web usage by user panels organized by private sector companies.

The Work Group and OHIPD continue to explore and test a range of internal and external web evaluation methods and applications.

International Programs

Malaria Research Network for Africa

The first electronic malaria research network has been created by NLM, working in partnership with organizations in Africa, U.S., U.K., and Europe. The network enables scientists working in Africa to have full access to the Internet and the resources of the World Wide Web as well as access to medical literature. Research sites involved are a) recognized as being of high quality by the malaria research community, b) have work they are trying to accomplish but can't due to limited communication, and c) have access to the necessary resources for purchasing equipment and sustaining the system.

MIMCom, the malaria research network, comprises telecommunications, information access, new tools for research, training, and evaluation. In collaboration with partners around the world, NLM designs and operates the network and meets all costs associated with: determination of requirements; site surveys; negotiating with African telecommunications regulatory authorities; assistance with equipment purchase and installation; monitoring of the system; ongoing technical assistance, training and support; handling of monies and agreements; establishing document delivery systems and information portals; and promotion of malaria research agendas. Individual sites and their funding partners are responsible for all equipment and operating costs.

The network has its technical hub in the U.K. at Redwing Satellite Solutions, Ltd. where a large satellite dish, focused on a geostationary satellite 37,000 km above the Atlantic, is connected directly to the high-speed Internet backbone on the ground. At research sites where there is no local telecommunications service to meet the needs of the researchers, a smaller ground station in the form of a VSAT is installed. The VSAT dish antenna connects to a radio unit which in turn is connected to an existing local area network, serving the computers used by the researchers. Some sites on the network operate a wireless connection—using a long distance radio link—either to a local Internet service provider or to another MIMCom site nearby.

The system provides a permanently open link, operating reliably 24 hours a day 7 days a week. Any time of day or night, a researcher can send and receive emails, search for literature, search databases, and share files

and images with colleagues. This access to communication and information is moving researchers toward a new and more efficient way of doing collaborative research.

Satellite systems are highly reliable as they are not subject to the problems and limitations of telephone wires and other more traditional means of obtaining an Internet connection. However, satellite bandwidth is very expensive. The system design allows hundreds of researchers in Africa to share bandwidth, thereby maximizing the usage of the satellite capacity and minimizing cost per site.

Evaluation of the network has just begun, led by NLM and Mbita-ICIPE research site in Kenya. The evaluation will cover various aspects of network performance and efficient use of bandwidth as well as information use and site growth, proposals funded, papers published, and collaborations carried out. Baselines were created before the network was installed, and interviews and questionnaires are currently being conducted.

In connection with the malaria research network, NLM has launched two experimental programs to increase access to medical literature for malaria researchers in Africa. The first is a pilot document delivery system for malaria researchers through the Medical Library at University of Zimbabwe, the Medical Research Center (MRC) in South Africa, and NLM. The second, to provide access to full text, is a joint 2-year project of the NLM and the American Association for the Advancement of Science that will allow malaria researchers to receive full text articles of journals free of charge.

The project's website (www.nlm.nih.gov/mimcom/) comprises links to MEDLINE, a variety of free online journals, databases, malaria-related sites, and general information. An NLM reference librarian serves as the webmaster and will be expanding the site to include special news releases and articles of interest to researchers.

NLM has provided technical support and training to IT personnel at each site on an ongoing basis in addition to holding a comprehensive workshop for all. Additional training opportunities will build capacity among African IT specialists at the research sites, from course work for individuals to regular conference calls of the group. Follow-up visits to each site also allow for updating and trouble shooting.

Special training is being planned for researchers in the use of IT related to specific research agendas they are trying to carry out. This may include various wireless communication devices as well as personal software agents.

The Network as of October 2001:

Kenya: Kenya Medical Research Institute (KEMRI)/Centers for Disease Control and Prevention (CDC) site in Kisian; KEMRI/Wellcome Trust site in Kilifi; KEMRI/CDC/Walter Reed Army Institute for Research (WRAIR) site in Nairobi with microwave links to the U.S. Library of Congress and Wellcome Trust sites: and International Centre of Insect Physiology and Ecology (ICIPE) site at Mbita Point with the NIAID/NIH.

Ghana: Noguchi Memorial Institute in Accra with NIH, U.S. Agency for International Development (U.S.AID), U.S. Naval Institute of Medical Research (U.S.NIMR); and Navrongo Health Research Center in Navrongo (with connection to district health office), with NIH, U.S.AID, and U.S.NIMR.

Tanzania: National Institute of Medical Research (NIMR) Center in Amani with NIH; NIMR Center in Ifakara with NIH; Microwave link from NIMR headquarters to local ISP in Dar es Salaam with NIH; and Local ISP/Mobitel/CyberTwiga link for KCMC in Moshi.

Uganda: Uganda Viral Research Institute in Entebbe with CDC.

International Network Partnerships

The report of the International Planning Panel stated that there is a need to strengthen and expand efforts in global health information networking. The panel favored the development of a loosely arrayed network of international centers for medical information.

In response to the International Long Range Plan, OHIPD proposes to pursue strategies to develop these international network developments. Two initial areas for exploration are international DOCLINE libraries and library-to-library partnerships (or a combination of both areas). The purpose is to see how NLM can plan a new role internationally that strengthens our relationships with foreign libraries, particularly in underdeveloped areas.

In addition to supporting international libraries, international network partnerships can support the international research community through programs such as the Multilateral Initiative on Malaria. NLM can share its expertise in designing and implementing telecommunications capacity with scientists in developing countries, enabling researchers to communicate in a timely manner, access biomedical information resources and databases, and collaborate on proposal preparation and research implementation with colleagues in industrialized countries.

Global Internet Connectivity

End-to-end performance of the Internet, on both national and global scales, continues to be important to NLM in part because the Internet is the primary vehicle for promoting access to and dissemination of health information. This includes the further exploration of the methods and metrics needed to better understand the quality of Internet performance from the end user perspective. During 2001, NLM built on the earlier phases of end-to-end connectivity testing by conducting outreach to other researchers and organizations working in this area. The intent is to lay the groundwork for development of an NLM plan for future activities on Internet connectivity, including the use of very high bandwidth networks for health-related applications. NLM is developing partnerships with organizations such as the Cooperative Association for Internet Data Analysis at the San Diego Supercomputer Center, University of California, and is now extending these discussions to possible collaborations with the Internet 2 and related organizations.

International MEDLARS Centers

Bilateral agreements between the Library and more than 20 public institutions in foreign countries allow them to serve as International MEDLARS Centers. As such, they assist health professionals in accessing MEDLINE and other NLM databases, offer search training, provide document delivery, and perform other functions as biomedical information resource centers.

NLM's Long Range Plan 2000-2005 emphasizes the need to establish new international partnerships to leverage its resources. The establishment of future Centers will be guided by the opportunity to benefit from the international initiatives of others. On August 23, 2001 in Boston, a Memorandum of Understanding was inked between the United States and Norway. The agreement made the University of Oslo Library of Medicine and Health Sciences the newest International MEDLARS Center.

Through that collaborative arrangement with NLM, the Oslo library will provide online search assistance, training, and document delivery to health professionals and libraries in Norway and in the Baltic Countries. Library staff there will also translate NLM's vocabulary, known as Medical Subject Headings (MeSH), into Norwegian. It is particularly fitting that the University of Oslo become the newest MEDLARS Center as their outreach initiatives mirror in many important ways the objectives of NLM's own programs. The International MEDLARS Centers are:

Australia: National Library of Australia
Canada: Canada Institute for Scientific and Technical Information (CISTI)
CHINA: Institute of Medical Information Chinese Academy of Medical Sciences
Egypt: ENSTINET Academy of Scientific Research and Technology
France: French Institute of Medical and Health Research **(**INSERM)
Germany: German Institute for Medical Documentation and Information (DIMDI)
Hong Kong: The Chinese University of Hong Kong

India: National Informatics Center Ministry of Information Technology
Israel: Hebrew University
Italy: Istituto Superiore di Sanita
Japan: Japan Science and Technology Corporation (JST)
Korea: Seoul National University
Kuwait: Kuwait Institute for Medical Specialization
Mexico: Centro Nacional de Informacion y Documentacion sobre Salud (CENIDS)
Norway: University of Oslo
Russia: The State Central Scientific Medical Library
South Africa: South African Medical Research Council
Sweden: Karolinska Institute Library
Switzerland: Documentation Service of the Swiss Academy of Medical Sciences
United Kingdom: The British Library
Pan American Health Organization (BIREME/PAHO): Centro Latino Americano e de Caribe Informcao em Ciencias da Saude
Intergovernmental Organization: Science and Technology InformationCenter Taipei, Taiwan

International Visitors

In FY2001 the Office of Communications and Public Liaison arranged for 273 tours—142 regular daily (1:30 pm) tours and 131 specially arranged tours. There were 2915 visitors in total. They came from the following 55 countries:

Argentina, Bosnia, Brazil, Cameroon, Canada, Chile, China, Colombia, Costa Rica, Croatia, Denmark, Ecuador, England, France, Georgia, Germany, Ghana, Guatemala, Holland, India, Indonesia, Ireland, Israel, Italy, Ivory Coast, Japan, Jordan, Kazakhstan, Kenya, Korea, Lebanon, Mali, Mexico, Moldova, Nepal, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Russia, Singapore, Sweden, Switzerland, Taiwan, Tanzania, Thailand, Turkey, Ukraine, United States, Uzbekistan, Venezuela and Zambia.

LIBRARY OPERATIONS

Betsy L. Humphreys
Associate Director

The Library Operations (LO) Division of NLM is responsible for basic library services that ensure access to the published record of biomedical science and the health professions. LO selects, acquires, preserves, and organizes the world's biomedical literature in whatever format it is produced; maintains a subject thesaurus and a library classification scheme used by institutions worldwide to organize biomedical information; produces authoritative indexing and cataloging records; builds and disseminates bibliographic, directory, and full-text databases; provides national back-up document delivery, reference, and research assistance; helps health professionals, researchers, librarians, and the general public to make effective use of NLM's services; and coordinates the 4,700 member National Network of Libraries of Medicine®, which enhances health information services throughout the country. The services provided and coordinated by LO are the essential foundation for NLM's outreach programs to health professionals and the general public and also support the Library's programs in molecular biology, AIDS, and health services research information.

LO is the largest of NLM's Divisions, employing a multidisciplinary staff of librarians, technical information specialists, subject experts, health professionals, historians, and technical and administrative support personnel and relying on the services of a range of contractors. In addition to its basic services, LO directs the National Center for Health Services Research and Health Care Technology (NICHSR); carries out an active program in the history of medicine; works with other NLM program areas to develop new and enhanced products and services; conducts research and evaluation related to current programs and services as well as advanced information storage and retrieval; directs and post-graduate training program for medical librarians; and contributes to the development of standards for health data and knowledge-based information. LO staff members are active participants in Library-wide efforts to improve the quality of work-life at NLM, including the Diversity Council and the NLM Intranet.

Program Planning and Management

In FY2001, LO devoted considerable management attention to three key elements of the infrastructure for basic services:

- automated systems that support basic operations and services;
- contracts that support the National Network of Libraries of Medicine (NN/LM); and
- space needed for the NLM collection, onsite users, and staff.

LO continued to work closely with the Office of Computer and Communications Systems (OCCS), other NLM program areas, and outside collaborating institutions to complete the replacement of NLM's legacy systems, to transfer all the unique data to the new and more integrated databases, and to end the Library's reliance on mainframe computers as of September 28, 2001. Internet Grateful Med was retired on the same day, having admirably achieved its goal of providing an easy Internet interface to many NLM databases during the System Reinvention period.

Major system reinvention accomplishments are described throughout this report. After a lengthy recompetition process, the eight contracts for basic NN/LM services for 2001–2006 and subcontracts for a National Training Center and Clearinghouse and a National Outreach Evaluation Center were awarded on May 1, 2001. More information about the new contracts appears in the Outreach section of this chapter. NLM received authorization to proceed with the development of architectural and engineering plans for a third building late in FY2001. LO is working with the Office of Administration and other NLM components to define requirements and review plans for storing the collection, providing onsite services, and accommodating LO staff. In FY2001, LO moved the serials bibliographic unit to renovated space on Level 1 to alleviate serious overcrowding on the B-1 level and

converted a portion of the public space in the Learning Resource Center into reference staff work stations. The Learning Resource Center has relatively few simultaneous users so this change did not have a negative effect on service.

LO plans its programs to support the goals and objectives in the *NLM Long Range Plan, 2000–2005* and the closely related *NLM Strategic Plan to Reduce Racial and Ethnic Health Disparities, 2000–2005*. Most LO activities directly address the first two goals in the NLM Long Range Plan: "Organize health-related information and provide access to it" and "Promote use of health information by health professionals and the public." LO contributes to the third goal: "Strengthen the informatics infrastructure for biomedicine and health," through training and education for medical librarians and activities related to standards and information policy. LO's work on the Unified Medical Language System® and a new gene indexing initiative address the fourth goal, "Conduct and support informatics research." Two major LO priorities—enhancing the public's access to health information and developing strategies for managing information published in digital form—are designated as important areas for new emphasis in the NLM Plan.

Collection Development and Management

NLM's comprehensive collection of biomedical literature is essential to many of the Library's services. LO's goal is to build and maintain a collection that serves the current and future needs of health professionals and researchers. To accomplish this, the LO staff develops and updates a formal literature selection policy, acquires and processes literature that meets its selection guidelines in all languages and formats; organizes and maintains the collection for efficient current use; and preserves materials it acquires for use by future generations. At the close of FY2001, the NLM collection included 2.4 million volumes and 3.8 million other items, including electronic publications, audiovisuals, microforms, pictures, and manuscripts.

Selection

Literature is selected for the NLM collection by LO staff and agents who apply the guidelines in the *Collection Development Manual of the National Library of Medicine*, which typically undergoes a major review and revision every 5 to 8 years. In FY2001, LO developed plans for another such review, which will address the subject boundaries of the NLM collection; the audiences to which NLM's collection should be addressed; and the preferred formats when publications are issued in multiple media. Since the current edition of the manual was published in 1993, NLM has added the general public to its user groups, and electronic publishing has increased dramatically.

In FY2001, the Technical Services Division (TSD) expanded selection of complementary and alternative medicine serials, Asian and Pacific Islander research literature, and gray literature. TSD, NICHSR, and the National Network of Libraries of Medicine Office worked together to set up a contract arrangement with the New York Academy of Medicine to identify and catalog gray literature on topics related to health policy and public health. Visiting scholar Walter Lear, M.D., reviewed NLM's historical holdings in social medicine and community health and identified priorities for additional acquisitions by the History of Medicine Division (HMD).

Acquisitions

TSD received and processed 163,980 contemporary books, serial issues, audiovisuals, and electronic publications. (Table 2) Net totals of 46,369 volumes and 218,472 other items (e.g., manuscripts, pictures, microforms, audiovisuals) were added to the NLM collection in FY2001. With full implementation of the new Indexing Data Creation and Maintenance System (DCMS), the Serial Records Section was able to cease updating the legacy Journal Authority File in September 2001. This ended an almost three-year period in which some serials data had to be maintained in both the legacy system and the Voyager Integrated Library System implemented in late 1998. LO installed

new releases of Voyager in February and September 2001. The first of these included many new features in the acquisitions module and a new version of the online public access catalog. The second improved online catalog response time, among other bug fixes.

After consultation with the NIH Library Branch Chief and NIH legal counsel, TSD determined that NLM may use many of the NIH Library's licenses for electronic journals to provide onsite access to Reading Room patrons. By the end of FY2001, more than 600 journals were available for use in the NLM Reading Rooms, many through the NIH-wide licenses and others licensed directly by NLM for onsite access and interlibrary loan only. LO employs a number of book dealers and subscription agents to acquire literature published around the world. In FY2001, TSD expanded vendor coverage of materials published in Eastern Europe and the Baltic states.

HMD continued to add materials to the Library's outstanding collection of early printed books, manuscripts, pictures, and historical audiovisuals. Important individual items acquired in FY2001 include: Dioscorides' [*De materia medica*] (Basle, 1529), a seminal work on herbal, mineral, and animal drugs composed in Greek during the 1st century AD; Eucharius Rosslin's *Der Schwanngeren Frawen und Hebammen Rosegartten* (Augsburg, 1528), a textbook for midwives first published in 1513 with a lying-in woodcut on the title page; Paulus Juliarus' *De lepra et eius curatione* (Venice, 1545), an early treatise on leprosy; Ezio Cleti's *Animadversiones circa Partem Affectem Pleuritidis* (Rome, 1643), a treatise on lung disease; and two 17th century works on the medicinal aspects of food, *Diaeteticon* (1682) by J.S. Elsholtz and *Freywillig-auffgesprunger Granat-Apffel dess christlichen Samaritans* (Vienna, 1695).

Notable accessions to the manuscripts collection included the papers of Herbert Ley, FDA Director during the Nixon administration; the papers of Paul Cornely, first African-American President of the American Public Health Association; records from the American College of Nurse-Midwives; documents relating to "contraband hospitals" which treated freed slaves in the occupied South during the Civil War; and documents from the Department of Health and Human Services on early discussions regarding AIDS and discussions about alien excludability for medical reasons. HMD also received additions to the papers of several Nobel scientists, the Victor Whitten dermatology archives, and the NLM archives.

NLM's picture collection was enriched by the addition of many public health posters from across the U.S. and around the world, including posters related to health and the war effort in World War I and World War II and a large and splendid poster depicting the consequences of cocaine addiction on Paris and Parisians during the Jazz Age, which was transferred to NLM from the Smithsonian Institution. Other additions to the picture collections included AIDS etchings by Sue Coe, a medical caricature by Thomas Rowlandson (1756-1827), an 1888 patent medicine map of the United States, and a stereoscope card documenting a Civil War amputation scene at a field hospital. William Helfand continued to be a generous donor to the historical picture collection.

Among the historical audiovisuals acquired were videos of a series on black physicians, several films made by the epidemiologist Telford Work, a World War II film made by William Roberts, MD, a selection of films produced by the National Institute of Mental Health, and more than 1,000 films and videos of public service announcements and interviews from the Food and Drug Administration.

Preservation and Collection Management

To preserve the NLM collection and keep it readily accessible for current use, LO binds, microfilms, conserves rare and unique items, maintains appropriate storage facilities and conditions for all types of library materials, and works to prevent and respond to emergencies that could damage these materials. LO distributes data about what NLM has preserved to avoid duplication by other libraries and provides preservation information useful to other health sciences libraries on the NLM web site. NLM conducts experiments with new preservation techniques as warranted and

continues to promote the use of more permanent media in new biomedical publications.

In FY2001, LO bound 31,625 volumes, microfilmed 5,131 volumes, repaired 1,403 items in the onsite book repair and conservation laboratory, and conserved 128 items from the historical collections. New contracts were awarded for microfilm preparation and microfilming. NLM received permission to acquire library binding services independently rather than under the multi-library contract managed by the Government Printing Office. This should make it easier to obtain binding services that meet the Library's requirements. As part of the NLM System Reinvention initiative, the Preservation and Collection Management Section worked with OCCS to develop an interim binding module that is not dependent on a mainframe computer, pending the delivery of the binding module for the Voyager Integrated Library System. New procedures were implemented for systematic review and copying of historical films in need of preservation. Post–1970 motion pictures were transferred from the general collection to HMD. The onsite Book Repair and Conservation Laboratory was expanded to allow for chemical treatments and conservation of oversize and photographic materials.

The Preservation and Collection Management Section arranged for a test of mass deacidification as a potential preservation treatment for materials in the general collection, following a survey last year that indicated that nearly 800,000 volumes in the NLM collection might benefit from this treatment. A consultant will be engaged to review the scientific literature and advise NLM on the long-term effectiveness of the process and whether the Library should establish a mass deacidification program. Fortunately most current materials received by NLM are published on acid-free paper (93% of current paper-based materials received by NLM and 98.5% of journals indexed for *Index Medicus*®). In FY2001, the Preservation and Collection Management analyzed the characteristics of those current journals that are not acid-free so that NLM can develop a strategy for persuading additional publishers to use more permanent paper.

LO has initiated a pilot project to identify medical monographs published between 1830 and 1950 that are held at the Countway Library of Medicine in Boston, the New York Academy of Medicine, or the College of Physicians of Philadelphia, but not at NLM. With funds provided by NLM under its NN/LM contract, the New York Academy of Medicine will use OCLC's Automated Collection and Analysis Services to compare NLM's holdings with those of the other three libraries. The goal is to determine the extent to which there are important medical monographs in need of preservation that are not held by NLM.

NLM's most visible project related to permanent access to digital information is the PubMedCentral electronic journal archive, which is described in National Center for Biotechnology Information chapter. The NLM Director is a member of the Library of Congress's National Digital Strategy Advisory Board, which is providing advice on the development of a national plan for preservation of digital materials. In January 2001, an NLM-wide Electronic Permanence Test Working Group was appointed to develop plans for an operational test of the permanence ratings that the Library developed last year for its own electronic publications. After a review of several systems and approaches to creating and maintaining these metadata, the Group concluded that TeamSite, a web document management system that NLM recently acquired, should be suitable for this purpose. The operational test is now planned for early 2002 when TeamSite implementation has been completed.

Bibliographic Control

To facilitate access to the biomedical literature, LO creates authoritative indexing and cataloging records for journal articles, books, films, picture**s**, manuscripts, and electronic media. As the number of biomedical publications issued in electronic form increases, LO is adapting its standard indexing and cataloging practices to enhance access to electronic resources. LO also maintains the Medical Subject Headings (MeSH®), a subject thesaurus used by NLM and many other

institutions to describe the subject content of biomedical information; collaborates with Lister Hill Center to produce the Unified Medical Language System (UMLS®) Metathesaurus®, of which MeSH is an important component; and maintains the *National Library of Medicine Classification*, a scheme for arranging physical library collections by subject that is used by health sciences libraries around the world.

Thesaurus Development

The 2002 edition of MeSH contains 20,742 main headings, 82 subheadings or qualifiers, 130 publication types, and more than 125,000 supplementary records for chemicals and other substances. All MeSH supplementary records were reviewed and reorganized into the concept-oriented structure previously implemented for MeSH main headings. The virus terminology was revised to conform to the 7th Report of the International Taxonomy of Viruses, including revision of the names of all genera in the family Retroviridae and reorganization of most species and strains of HTLV viruses. Terminology related to complementary and alternative medicine was restructured into physical, sensory, mind-body, and spiritual subgroups to facilitate searching those divisions. Vocabulary related to plant families and genera was greatly expanded with more use of specific Latin binomial names as preferred terms, and instructions for indexing the use of plants for therapy were modified. Through a joint effort with the Kennedy Institute of Bioethics, terminology in the area of bioethics was enlarged and enhanced in preparation for the addition to PubMed of unique journal citations from the former BIOETHICSLINE® file. There were also many changes and expansions to terms for transport and carrier proteins.

The majority of the content editing for the 2002 version of the UMLS Metathesaurus was completed under MeSH Section supervision. Although many vocabularies were updated and a few were added for the 2002 Metathesaurus, the review of the MeSH supplementary records caused many "undiscovered" synonyms to be merged. As a result, the number of Metathesaurus concepts will decline from the 2001 figure. In August 2001, staff from the MeSH Section and the Lister Hill Center taught a one-week "Introduction to the Metathesaurus and the UMLS" class for NLM and contractor staff who review and edit Metathesaurus data.

The MeSH Section has worked with the Department of Veterans Affairs and the Food and Drug Administration to develop a plan for creating a semantic normal form for pharmaceuticals and over-the-counter medications. The first step, which was underway at the close of FY2001, is to convert the VA National Drug Formulary into that form and include it in the UMLS Metathesaurus.

Cataloging

LO catalogs the biomedical literature acquired by NLM both to document what is available from the NLM collection and to provide cataloging records that can be used by other libraries to reduce the level of effort required to organize their own collections. LO also catalogs or otherwise organizes information resources published on the World Wide Web, both to expand existing services, such as MEDLINEplus, and to contribute to the development of practical strategies for organizing credible web-based health information. In FY2001, an NLM-wide group chaired by the Cataloging Section completed the development and initial testing of a minimum set of metadata (to include the permanence levels previously mentioned) for NLM's electronic publications. An operational test will begin in early 2002 when NLM has implemented the TeamSite web management software.

In FY2001, the Cataloging Section cataloged 19,024 contemporary books, serials, nonprint items, and cataloging-in-publication (CIP) galleys, using a combination of in-house staff and contractors. NLM is encouraging the participation of biomedical publishers in the Library of Congress's electronic CIP program since this speeds availability of cataloging copy. In support of NLM System Reinvention, cataloging staff added to the Voyager Integrated Library system almost 7,500 indexed serial titles from various specialized databases (e.g., HISTLINE®, BIOETHICSLINE®, POPLINE®) to

support their indexing in the new Data Creation and Maintenance System (DCMS) and online retrieval in PubMed. A new web-based version of the *National Library of Medicine Classification* became publicly available in May 2001. The Cataloging Section and OCCS also developed a web-based editor for use in annual updates to the *Classification*.

HMD made excellent progress on cataloging rare and unique items in the historical collections. About 109 linear feet of contemporary manuscripts were cataloged, more than twice the amount done last year. Three new *Profiles in Science* debuted: Christian Anfinsen, Marshall Nirenberg, and Barbara McClintock. The McClintock site was the first Profile featuring papers held by another institution, the American Philosophical Society of Philadelphia. All existing finding guides for other manuscript collections held by NLM and for the NLM archives were converted to the electronic archival description (EAD) format and mounted on the web. A total of 510 early monographs was cataloged, more than a 10-fold increase from the previous year. The project to catalog Dorothy Schullian's "bathtub" collection of fragments found in early bindings was completed. A Short-Title List of NLM's 90 Western manuscripts written before 1601 was made available on the NLM website, pending the addition of cataloging records for these manuscripts to the Voyager Integrated Library System. HMD also prepared reports estimating the extent of the remaining unpublished collections of pamphlets, theses, and health department reports, as an initial step in planning projects to create electronic records for them.

Staff from TSD, HMD, PSD, and the Lister Hill Center undertook a special project to ensure that future Surgeon General's reports contain standard bibliographic data elements and to identify and provide electronic access to all historical reports. As part of this effort, the Cataloging Section developed a set of minimum requirements for the title pages of the reports and also established procedures for supplying cataloging in publication data to the Office of the Surgeon General.

Indexing

The Index Section in BSD has responsibility for indexing newly published articles from about 3,700 biomedical journals so that users of the MEDLINE/PubMed® database and the products generated from it can locate articles on specific biomedical topics. Existing MEDLINE® records are annotated when the articles to which they refer have been retracted, corrected, or challenged in subsequently published notices or commentaries. A combination of inhouse staff, contractors, and co-operating U.S. and international organizations perform the indexing and annotation. New 5-year indexing contracts were awarded to three companies in FY2001.

The Literature Selection Technical Review Committee (LSTRC) (Appendix 6), an NIH-chartered committee of outside experts, advises NLM about which journals should be indexed for MEDLINE and *Index Medicus*. In late FY2001, the size of the Committee was increased from 12 to 15 members to allow more rapid review of new journals. During the year, the LSTRC reviewed 425 titles and rated 131 sufficiently highly for immediate inclusion in MEDLINE; another 76 titles were accepted provisionally, pending receipt of acceptable electronic citation and abstract data from their publishers. A subject review of journals in public health was conducted with assistance from the Public Health and Health Administration Section of the Medical Library Association, the American Public Health Association, the Association of Schools of Public Health, the National Association of State and Territorial Health Officials, the National Association of County and City Health Officials, and the Public Health Foundation. It led to the addition of another three titles. There was increased emphasis on reviewing journals from developing countries that reflect indigenous public health problems with potential impact on global health.

In FY2001, NLM added 463,014 citations to MEDLINE, about 5% more than in FY2000. During FY2001, the Index Section

assumed direct responsibility for indexing dental journals previously indexed by the American Dental Association, AIDS newsletters previously indexed under a contract managed by the Specialized Information Services (SIS), and 40 toxicology journals previously cited in TOXLINE®.

Of the citations added this year, 46% were received electronically from publishers, 27% were entered via scanning and optical character recognition, and 26% were double-keyboarded. The number received electronically, the fastest and most economical method, increased about 18%. At the end of the year, NLM was receiving XML-tagged electronic citations and abstracts from more than 275 publishers for 1,609 journals, a 42% increase from the end of FY2000. During FY2001, the scanning and keyboarding systems were modified to output XML-formatted data also.

After initial data entry, all MEDLINE citations are transferred to the Indexing Data Creation and Maintenance System (DCMS) which indexers use to add subject headings and other data needed to complete the citations. The DCMS retains a maintenance copy of all indexed citations and also produces XML output to update the PubMed retrieval database. In FY2001, BSD and other LO staff completed implementation of this reinvented system, including the addition of citation maintenance capabilities, extensions required to accommodate citations for specialized subject areas, such as bioethics and space life sciences, and enhancements to improve response time and functionality. A major effort was required to obtain reliable high speed communications for the many NLM and contract indexers who work at home.

In FY2001, at the request of the NLM Director, the Index Section worked with the National Center for Biotechnology Information (NCBI) to determine whether indexers could enhance gene/protein databanks with information from, and links to, relevant MEDLINE citations. A pilot test was conducted in which six indexers added information for five organisms (human, mouse, rat, fruit fly, zebra fish) to LocusLink as a by-product of indexing articles in selected journals. The test showed that the work was feasible, and the resulting additions to LocusLink were considered highly useful by NCBI, the National Human Genome Research Institute, and other NIH components. Once necessary system and contract changes are in place, the Index Section will do gene indexing for relevant articles appearing in any journal indexed for MEDLINE. The benefits of NLM System Reinvention are readily apparent in the relative ease and speed with which the new DCMS can be modified to handle new requirements. In addition to extensions needed to accommodate gene indexing, the DCMS is also being modified to allow operational testing of algorithms to provide automated assistance to indexers that have been developed as part of the Indexing Initiative research project.

Information Products

LO collaborates with other NLM components to produce some of the world's most heavily used medical information resources, including online databases, other electronic resources, and print publications that incorporate LO's authoritative indexing, cataloging and thesaurus data.

Databases and Web Information Resources

Users conducted about 331 million searches of MEDLINE in FY2001, about 18 million via Internet Grateful Med or the NLM Gateway and the rest via PubMed. Staff from BSD and TSD worked with OCCS, the Lister Hill Center, and NCBI to complete the complex and time-consuming task of identifying and transferring unique records from the former specialized AIDSLINE®, BIOETHICSLINE, HealthSTAR, HISTLINE®, POPLINE® and SPACELINE™ files to either Locatorplus (monograph and chapter citations), PubMed (journal citations), or the NLM Gateway (meeting abstracts). The subject-oriented searching formerly provided by these databases will be replaced by a combination of subject subsets in PubMed and enhanced NLM Gateway capabilities. At the request of Johns Hopkins University, the agreement between its Population Information Program and NLM**,** which has supported online access to POPLINE on NLM's system**,** ceased at the end of July

2001. NLM will begin indexing some of the journal titles previously indexed for POPLINE. The Population Information Program will maintain its own POPLINE website.

Completion of database transition allowed NLM to phase out Internet Grateful Med, the last in the series of Grateful Med® interfaces which provided user-friendly access to a range of NLM databases since 1986. BSD assisted NCBI in developing and testing many enhancements to PubMed functionality during FY2001, including changes to search and display features occasioned by changes in NLM's XML format for journal citation data. The search strategy for the PubMed subset for AIDS was revised, and new subsets were added for bioethics, history of medicine, space life sciences, toxicology, and complementary and alternative medicine. This last subset was a joint effort with NIH's National Center on Complementary and Alternative Medicine (NCCAM). BSD and OCCS also assisted the Center in setting up access to this subset from the NCCAM website through a special "CAM on PubMed" graphic.

Other new PubMed features include LinkOut for Libraries, which supports customized links to those sources of full-text journals that have been licensed by the searcher's institution, and the ability to link to large document supplier systems, including university library systems. At the close of FY2001, 129 libraries were using the LinkOut for Libraries, and the University of California and University of Washington were included as large document delivery services. The National Information Center of Health Services Research and Health Care Technology (NICHSR) provided funding to McMaster University to update the search strategies that underlie the "Clinical Queries" feature of PubMed and to expand the queries to include health services research topics. BSD and NICHSR worked together to develop a query strategy that would retrieve systematic reviews. The results of these efforts should appear in PubMed in FY2002.

BSD staff also assisted the Lister Hill Center with the initial public implementation of the NLM Gateway in October 2000 and with subsequent enhancements, including the addition of AIDS, space life sciences, and health services research meeting abstracts and access to DIRLINE. The NLM Gateway is now the only interface to OLDMEDLINE. LO is making steady progress in converting its retrospective indexing data to electronic form. Data from the 1956 and 1957 *Current List of Medical Literature* were received from the keyboarding contractor and will be added to OLDMEDLINE in FY2002. Contracts were awarded to convert the 1946–1955 data, including one issued to Lakota Technologies, Inc., a tribal firm that is attempting to increase the number of jobs available on its reservation in South Dakota. The initial conversion of all series of the *Index-Catalogue of the Library of Surgeon-General's Office* was finished and quality review of the data is nearing completion.

During FY2001, usage of MEDLINEplus®, NLM's web information service for the general public**,** tripled—to 62 million page views. Nearly 600,000 unique visitors access the site each month. PSD and OCCS made many well-received improvements to MEDLINEplus during the year, including more than 70 new health topics, a health news feature, the interactive Patient Education Institute tutorials, redesigned and streamlined access to the adam.com medical encyclopedia, use of a spell-checker, and interface improvements based on usability testing with seniors. PSD and OCCS continued to work with the National Institute on Aging to develop an NIHSeniorHealth website, which will be linked to MEDLINEplus. The technical issues associated with providing video content in a format that can be easily accessed by seniors in their homes have delayed public release of this site. Substantial effort was also devoted to developing and implementing a new editing system for the database that underlies MEDLINEplus. PSD continued to work with the University of North Carolina to develop technical and procedural approaches to linking to and from MEDLINEplus from complementary web services covering state and local health information.

In February 2001, NLM conducted a voluntary MEDLINEplus visitor profile survey to obtain general information about who is using the site, what they are looking for, how they make use of what they find, and their degree of

satisfaction with the site. The results (a summary is available in MEDLINEplus) indicated a high degree of satisfaction with MEDLINEplus. Many of those who completed the survey provided their email addresses and indicated a willingness to provide additional information. This favorable view of MEDLINEplus was corroborated in comparative analyses of MEDLINEplus and other health website conducted by Cyber Dialogue, Inc. at the request of NLM's Office of Health Information Programs Development.

NICHSR completed development and testing of two new web information resources that will become publicly available in FY2002. The *Database of Health Services Research Resources*, which is produced with assistance from the Cecil Sheps Center at the University of North Carolina, describes data sets, survey instruments, and other tools useful in health services research with links to fuller documentation and published studies that made use of the tools. The Healthy People 2010 Information Access Project is a joint effort of NLM and the Public Health Foundation to provide ready access to information that can assist in developing strategies to meet public health goals. Staff from NICHSR, PSD, and the Specialized Information Services Division have developed evidence-based PubMed strategies that retrieve citations to articles relevant to selected Healthy People 2010 objectives, which are available along with the full-text of the chapters of Healthy People 2010 and other relevant web resources including MEDLINEplus health topic pages. Staff from NICHSR, BSD, and the Lister Hill Center also completed development and usability testing of a new version of Health Services/Technology Assessment Text (HSTAT) that will become publicly available in FY2002.

During FY2001, NLM obtained simplified URLs for many of its heavily used services**,** including pubmed.gov, docline.gov, and medlineplus.gov, and for the National Network of Libraries of Medicine (NN/LM) website.

Machine-Readable Data

NLM disseminates its databases in machine-readable form to promote the broadest possible used of its authoritative bibliographic and thesaurus data. There is no charge for any NLM database, but recipients must sign license agreements or memoranda of understanding that impose conditions on the use of the specific databases involved. The commercial companies, international MEDLARS® centers, universities, and other interested organizations that license the data make them available online or in CD-ROM products or use them to improve the functionality of a variety of biomedical information systems.

In FY2001, BSD worked with OCCS to implement major changes in the distribution of MEDLINE data. As part of System Reinvention, a new XML distribution format replaced the old ELHILL unit record format. Version 2 of the MEDLINE XML format, released in the fourth quarter of FY2001, defined data elements needed for journal citations transferred to PubMed from the specialized ELHILL databases that were eliminated as part of NLM System Reinvention. NLM also moved to daily release of MEDLINE in-process records via ftp and the use of DLT tape for distribution of retrospective data. Because the MEDLINE data are now easier to use, a record 53 additional organizations licensed MEDLINE this year, primarily for research or data mining. This brought the total number of licensees of NLM's bibliographic or toxicology data to 123 at the close of FY2001. NLM has a longstanding policy of limiting the number of non-U.S. institutions that provide public access to MEDLINE data. Given the upsurge in international requests for research use of the MEDLINE file, the Library is developing a research-only license for non-U.S. users.

NLM is moving to XML as the distribution format for other databases as well. The MeSH Section completed work on an XML format for distribution of 2002 MeSH data. BSD assisted SIS in releasing XML DTDs and sample

records for several TOXNET databases. The Cataloging Section is working on an XML format for distribution of NLM's cataloging records.

CATFILEplus, another by-product of System Reinvention, is a new member of NLM's suite of bibliographic files available in the MARC 21 format. The new data distribution includes NLM cataloging records, as well as monographs and monograph chapter records created by contributing special producers in the fields of bioethics, health technology assessment, history of medicine, population studies, and space. The latter were previously distributed in the separate BIOETHICSLINE, HealthSTAR, HISTLINE, POPLINE and SPACELINE files, which now have been integrated into Locatorplus, PubMed, and the NLM Gateway's meeting abstract file.

A total of 1,511 individuals and organizations license the UMLS Knowledge Sources for a wide variety of research, educational, and commercial purposes. In addition to ftp distribution, the UMLS files are available on CD-ROM and through the applications programming interface or interactive use of the UMLS Knowledge Source Server developed and maintained by the Lister Hill Center. BSD and NICHSR staff members are assisting with testing a new version of the Knowledge Source Server.

Print and Web Publications

NLM publishes some of its authoritative data in print publications, including *Index Medicus,* the *List of Journals Indexed in Index Medicus*, and several MeSH publications, but its electronic databases are considered the primary means of making these data available. The Library and the organizations with which it collaborates continue to review and modify or eliminate specific print publications that have outlived their usefulness, given increasing user access to electronic data. In FY2001, NLM made the decision to cease publication of *Cumulated Index Medicus*, effective with the 2000 edition, due to steadily declining sales. At the request of the American Dental Association and the American Journal of Nursing Company respectively, NLM also ceased production of the *Index of Dental Literature* and the *International Nursing Index* after completing the 2000 editions.

NLM System Reinvention necessitated complete replacement of the programs that produce the monthly *Index Medicus* and the MeSH publications. In the case of *Index Medicus*, an interim bridge approach to producing the publication also had to be developed to handle the transition period during calendar 2001. Problems associated with this approach caused significant publication delays for the monthly *Index Medicus* in 2001.

NLM's World Wide Web site is the primary vehicle for distributing a wide range of publications, including recurring newsletters and bulletins, fact sheets, technical reports, and multimedia catalogs. There were 2.6 million hits to its publication pages in FY2001, up 24% from the previous year. Issues of the *Current Bibliographies in Medicine* series continue to be extremely popular. Each issue of this series, which is edited by the Reference and Customer Services Section, addresses a topic of current interest to NLM, NIH, or other federal agencies and may be produced in conjunction with an NIH consensus development conference, a White House conference, or another meeting. Reference and sometimes NICHSR staff members collaborate with outside experts to produce each bibliography. FY2001 additions to the series included: Health Communication and Follow-Through Related to Early Identification of Deafness and Hearing Loss in Newborns, Public Health Informatics, Diagnosis and Management of Dental Caries, Youth Violence Prevention Resources, and Health Risk Communication.

The web-based *NLM Technical Bulletin*, edited by the MEDLARS Management Section, provides timely detailed information about changes and additions to NLM services that is particularly valuable for librarians and other information professionals. Individual articles are published as they are completed, which has allowed rapid dissemination of information about the many changes resulting from NLM System Reinvention. In FY2001, the scope of *Technical Bulletin* was expanded to include information about the UMLS Knowledge Sources.

HMD redesigned the *Directory of History of Medicine Collections* for 2001 to take better advantage of web search and design capabilities and also published a new web-based *Guide to Collections Relating to the History of Artificial Organs*, which includes museums and commercial companies, as well as libraries and archives. The Guide is a product of Project Bionics: Artificial Organs from Discovery to Clinical Use, a joint project of The American Society of Artificial Internal Organs (ASAIO), The Smithsonian's National Museum of American History and the National Library of Medicine, History of Medicine Division, launched in January 2000.

Core Health Policy Library Recommendations, a selection guide for libraries that support health policy programs or those working on health policy issues, was published on the NICHSR web pages. The guide was prepared by the Academy for Health Services Research and Health Policy. NICHSR is editing a multi-authored electronic text on Health Technology Assessment Information Resources for publication on NLM's website in FY2002.

PSD's Web Management Group serves as the web master for NLM's main web site. The NLM Home Page design was modified slightly to include additional information under each main category, to minimize downloading time, and to provide space for a news item or link to the current exhibit. Secondary pages, e.g., Library Services, also have improved navigation features. The Web Management Group worked with OCCS to implement FunnelWeb as the statistical package for the main web site and to set up a new Intranet statistical page for NLM staff. Many LO-managed discussion and announcement lists were converted to the Listserv software. NLM purchased the TeamSite web content management system and began its configuration. Implementation will occur in FY2002.

Direct User Services

In addition to its electronic information services, LO provides document delivery, reference, and customer service as a national and international backup to services available from other health sciences libraries and information suppliers. LO also serves a large onsite clientele in the NLM Reading Rooms.

Document Delivery

LO provides copies of documents to other U.S. and international libraries to fill requests for materials that are not readily available from other members of the National Network of Libraries of Medicine or other research libraries or document suppliers. LO also retrieves documents from the Library's closed stacks for use by onsite patrons.

In FY2001, PSD's Collection Access Section processed a total of 682,777 document requests, a 9% drop from last year. Onsite users requested 344,150 contemporary documents from NLM's closed stacks, 4% less than last year, and 4,844 items from the historical and special collections, an 8% increase from the previous year. In anticipation of increased use of electronic journals in NLM's Reading Rooms, the Collection Access Section expanded the contract photocopier arrangement to include paid printing from quiet high-speed printers at all workstations. The same card can be used to pay for either photocopying or printing. In late August 2001, the Reference and Customer Service Section implemented use of PubMed's LinkOut feature to provide access to electronic journals in the Reading Room. A new handout was prepared for patrons that includes information regarding copyright and other matters related to use of electronic journals. A baseline study of the current occupancy rates of workstations in the main Reading Room was conducted to determine whether additional workstations would be needed soon. Staff identified a software package that will assist in tracking patterns of use of Reading Room computers as an aid to identifying potential navigation enhancements or other changes that would help users find relevant information more quickly.

Remote libraries requested 338,627 contemporary documents from NLM, 13% fewer than in FY2000. The number of interlibrary loan requests sent to NLM fell sharply after the introduction of the new DOCLINE® system in July 2000, but appears to

be rebounding. NLM received 8.5% more requests in the 4th quarter of FY2001 than it received in the same quarter in FY2000. In FY2001, NLM delivered 52% of its ILL requests electronically, up from 36% last year. On July 1, 2001, NLM dropped the two-dollar international surcharge for requests delivered to international libraries by Ariel or email to their encourage use. PSD and OCCS implemented a new release of Relais, the system used to manage the delivery of documents from NLM. New features included: a "post to web" delivery method, a graphical user interface and new monitoring tools for the system administrator, the ability to send a document via any delivery method, and improved usability for the scanners.

A total of 3,223 libraries now use DOCLINE: 2,891 in the U.S., 288 in Canada, and 44 outside of North America. DOCLINE users entered 2.92 million requests into the system in FY2001, down 2% from last year; 92% of the requests were filled. The requests are routed automatically based on automated serials holdings data in the SERHOLD® database. At the end of FY2001, SERHOLD contained 1.37 million holdings statements for 50,416 serial titles held by 2,986 libraries. In close consultation with the Regional Medical Libraries, the LO/OCCS DOCLINE development team released three updated versions of DOCLINE in FY2001, each with a wide range of highly requested features. Functionality added included: additional SERHOLD reports and union lists, forms to send messages and problem reports, the ability to resubmit retired requests without re-entering data, the web as a delivery mechanism for requests, and compliance with priority 1 Section 509 standards of the Rehabilitation Act of 1998. NLM continues to test use of the ISO ILL protocol with outside organizations and hopes to implement it an option for sending requests to and from DOCLINE in early FY2002.

Loansome Doc® is a system that allows individual users of MEDLINE/PubMed and the NLM Gateway to route requests automatically for articles to a specific library that has agreed to serve them. There were 854,728 Loansome Doc requests from individuals to DOCLINE libraries, a 4% increase from FY2001. Twenty-two international libraries now offer this service to individuals and other libraries in their regions. New FY2001 participants include the Danish National Library of Science and Medicine, Copenhagen University, the University of Oslo Library of Medicine and Sciences, University Hospital, Reykjavik, Iceland, and Tel Aviv University, Israel.

Reference and Customer Service Inquiries

PSD and HMD provide reference and research assistance to onsite and remote users as a backup to services available from other health sciences libraries. PSD's Reference and Customer Service Section also has primary responsibility for responding to inquiries about NLM programs, services, and products and how they can be used effectively. Staff throughout LO and NLM provide second-level service for the questions that cannot be answered by first-line customer service staff.

In FY2001, Reference and Customer Service staff handled 110,921 legitimate inquiries from onsite users, email messages, and telephone calls. An additional 26,209 "junk" messages, including viruses sent via email, were received by the customer service email address, which added to the workload. (Prior to this year the number of junk messages received was small enough that it was included in the statistics for the number of inquiries handled.) After adjusting for the junk, both the number of requests received onsite and the number of requests received from offsite requesters were essentially equal to last year's numbers. Close to 60% of the requests come from offsite, and more than 70% of the offsite requests are received via email.

During FY2001, LO and OCCS evaluated potential replacements for the CustQ software which is currently used to record and track customer inquiries. NLM expects to implement a replacement system in FY2002 to obtain improved request analysis capabilities, to enable referral of requests to and from the Regional Medical Libraries, and to ensure more stable support for this critical service. At the suggestion of OCCS, NLM will also make use of the Native Minds software to create a proof of concept prototype for a virtual customer service representative. Reference and Customer Service

staff will convert the current knowledge base of answers to frequently asked questions into a format appropriate for the new "virtual rep."

Outreach

Many LO programs are designed to increase awareness and use of NLM's services by librarians and other information providers, health professionals, researchers, and the general public. LO coordinates the National Network of Libraries of Medicine (NN/LM) which attempts to equalize access to health information services and technology for health sciences librarians, health professionals, and the general public throughout the United States; participates in NLM-wide efforts to develop and evaluate outreach programs designed to improve health information access for underserved minority populations and the general public; develops major exhibitions and other special programs in the history of medicine; and conducts a range of training programs for health sciences librarians. Many LO staff members give presentations and demonstrations at professional meetings and write articles to highlight NLM programs and services.

National Network of Libraries of Medicine

The goal of the NN/LM is to provide timely, convenient access to biomedical and health information resources to U.S. health professionals, researchers, educators, administrators, and members of the general public. The NN/LM strives to ensure that accurate and up-to-date information is available irrespective of the user's geographic location. The network has more than 4,700 health sciences libraries, including hospital and academic medical center libraries. LO's NN/LM Office oversees the network programs that are coordinated and administered by eight Regional Medical Libraries (RMLs), under contract to NLM. (See Appendix 1 for a list of the current RMLs.)

On May 1, 2001, NLM awarded new five-year contracts for NN/LM services in all eight regions, following a selection process that included site visits to competing institutions in all regions where there was more than one bidder. A number of academic health sciences librarians, hospital librarians, and health professionals participated on the technical review and site visit teams. The 2001–2006 NN/LM program continues the focus on coordination and support for network members and outreach to health professionals, particularly those serving minority groups and working in rural areas and inner cities, but it also increases the emphasis on outreach to the general public. The goal is to increase partnerships between the NN/LM and a range of organizations, including public libraries, state libraries, health departments, tribal colleges and HBCUs, schools, churches, and other community-based organizations as a means of improving the public's access to health information. A new category of Affiliate membership has been defined for organizations that deliver health information, but do not have extensive collections of paper-based health information.

In addition to awarding the basic NN/LM contracts, NLM also funded subcontracts for two Centers that will serve the entire network: the National Training Center and Clearinghouse at the New York Academy of Medicine and the Outreach Evaluation Resource Center at the University of Washington. The Training Center will continue to provide training in the use of NLM's online systems and will also collect and link to electronic training materials and classes developed by NN/LM members in a variety of contexts. The Outreach Evaluation Center will provide training and consultation to assist network members in incorporating appropriate evaluation techniques into their outreach initiatives. A third National Center, which will provide mapping services to the NLM and the NN/LM, will be established in early FY2002. NLM is also working with the University of Connecticut on plans for upgrading and expanding that institution's Electronic Funds Transfer System, now used by about 600 libraries in four NN/LM regions to handle billing for interlibrary loan.

The NN/LM program is the core component of NLM's outreach program and its efforts to reduce health disparities. The RMLs and other network members develop and conduct many special projects to reach underserved health care professionals and to

improve the public's ability to find high quality health information. Most of these projects involve partnerships between health sciences libraries and other organizations, including public health departments, professional associations, public libraries, schools, and community-based organizations and have as one of their objectives to increase awareness and use of NLM services including MEDLINEplus, ClinicalTrials.gov, and PubMed. There is a strong emphasis on evaluation, and those who receive funding are strongly encouraged to apply the techniques described in *Measuring the Difference: Guide to Planning and Evaluating Health Information Outreach.* Due to the transition to the new NN/LM contracts, fewer new projects were started in FY2001 than in recent years. Two large projects were funded in cooperation with NLM's Office of Health Information Programs Development:

- The University of Texas Health Science Center at San Antonio is conducting a multi-faceted program of outreach and evaluation focused on the Hispanic community in the Lower Rio Grande Valley. The project includes a baseline community needs assessment, several pilot projects with a range of community organizations, a disease-specific pilot project, curriculum/faculty development at the new medical school, and assistance with the development of the MEDLINEplus Spanish-version web site.
- Phase III of the Tribal Connections Project was awarded to the University of Washington. This phase will also focus on defining a model of community-based outreach by applying appropriate evaluation criteria for health information outreach with the American Indian/Alaska Native communities. This model will serve as a framework to be applied to similar efforts with other communities with health disparities.

Several outreach projects proposed in the contract recompetition were also initiated:

- Personnel at the South Cove Community Health Center will be the focus of a project awarded to Tufts University Health Sciences Library. The Center, which has six locations, provides complete primary care services to the Asian community in the greater Boston area. In addition to introductory classes on the Internet, PubMed and HTML classes will also be conducted.
- The Yale School of Medicine, the Cushing/Whitney Medical Library, the Epidemiology and Public Health Library, and the New Haven Free Public Library will collaborate to develop a Consumer Health Information Center (CHC) at the New Haven Free Public Library. In addition to serving the New Haven community, the Free Public Library also provides information to 24 key public health officials employed by the City of New Haven Health Department in 12 distinct functional units. The CHC will serve the New Haven region as well as surrounding towns. It will provide access and training in the use of health related resources to local citizens.
- The Alumni Medical Library, Boston University Medical Center, is collaborating with the Boston Public Health Commission to facilitate access to health information on HIV/AIDS to citizens in the Boston area who use the services of the 60 Title I and City of Boston Prevention, Education, and Care funded programs. Training will be provided and the Library will expand and enhance its AIDS-focused web pages to provide quality-filtered web links and other information useful to Eastern Massachusetts and Southern New Hampshire HIV/AIDS consumers.

NLM and the NN/LM collaborate with the Centers for Disease Control and Prevention, several public health associations, and other Federal agencies in the Partners in Information Access for Public Health Professionals. As part of this initiative, NN/LM members worked with NICHSR and the NN/LM Office to organize the "Public Health Outreach Forum: What Do We Know?" at NLM on April 4-5, 2001. The meeting brought together representatives from

more than 20 NLM-funded public health outreach projects, staff from each of the RMLs, and representatives from many public health agencies and organizations to discuss what had been learned from work done to date and to make recommendations regarding future NN/LM outreach to the public health community. Papers summarizing the meeting, the lessons learned, and recommendations for the future were subsequently published in the *Bulletin of the Medical Library Association.*

The RMLs and other NN/LM members conduct most of the exhibits and demonstrations of NLM products and services at health professional, consumer health, and general library association meetings around the country. LO handles exhibits at the annual meeting of the Medical Library Association, some of the health professional and library meetings held in the D.C. area, and some distant meetings focused on health services research, public health, and the history of medicine. In FY2001, NLM and NN/LM services were displayed at 226 exhibits at national, regional, and state association meetings across the U.S. BSD staff assisted the NN/LM Office in the development of new exhibit structures for each RML.

Special NLM Outreach Initiatives

LO contributes to many NLM-wide efforts to expand outreach and services to the general public and to address racial and ethnic health disparities. In FY2001, staff members from the Office of Associate Director, the NN/LM Office, PSD, and BSD were active in NLM-wide consumer health and web evaluation committees. LO led the effort to design and test the survey instrument used in the MEDLINEplus visitor survey, organized a test of semi-structured interviews with public library users as a method for obtaining information on their health information seeking behavior in conjunction with the Region 6 and 7 RMLs, and also organized a test of the use of MEDLINEplus information prescription pads in 22 communities in Maryland in collaboration with the Region 2 RML. At LO's instigation, NLM is developing plans to conduct focus groups with consumers from different minority groups to learn more about their health information needs. The first group will be Spanish-speaking, to obtain information to guide the development of a Spanish version of MEDLINEplus.

The Office of the Associate Director and the NN/LM Office worked with the Public Library Association, a division of the American Library Association (ALA), and the Medical Library Association to organize "The Public Library and Consumer Health: Meeting Community Needs Through Resource Identification and Collaboration," a highly successful colloquium held in Washington, D.C. on January 10-11, 2001 in conjunction with the ALA Midwinter meeting. The NLM Director and a number of other NLM staff members gave presentations or tutorials at the meeting. Attendees included both public librarians and health sciences librarians. The interaction between the two groups was one of the many positive aspects of the meeting. NLM and the Public Library Association also cooperated to send a mailing about MEDLINEplus and other NLM and NN/LM services first to all public library branches in 8 states. The cover letter was signed by Dr. Lindberg and the President of the Public Library Association, and the mailing included bookmarks and a poster about MEDLINEplus. Recipients were invited to request additional supplies of bookmarks as needed. A similar package was sent to all NN/LM members in each state with a cover letter from Dr. Lindberg apprising them of the mailing to public libraries in their state and inviting their comments about MEDLINEplus or any other NLM services. NLM received a response from about 5% of each group, which is considered a good rate of return for this type of mailing, and therefore followed up with a mailing to public libraries and NN/LM members in the remaining 42 states.

LO staff members were heavily involved in NLM's partnership with Wilson High School in the District of Columbia. They helped to set up access to health information in the Parent Resource Center, provided a MEDLINEplus demonstration to parents and teachers at the opening of the Center, and taught students in 10th grade health classes how to find information in MEDLINEplus.

Historical Exhibitions and Programs

The History of Medicine Division periodically mounts major exhibitions in the NLM lobby and rotunda, with assistance from the Lister Hill Center, the Office of Communications and Public Liaison, the Office of Administration, and the Office of the Director. Designed for the interested public as well as the specialist, these exhibitions are part of NLM's outreach program. On May 21, 2001, NLM opened a new exhibition entitled *The Once and Future Web: Worlds Woven by the Telegraph and Internet*, which explores the parallel histories of the telegraph and the Internet as two electronic communications technologies that transformed the world. At the opening, the objects in the exhibitions included the first telegram ever sent (on loan from the Library of Congress), and the preserved body of Balto, the lead dog of the team that delivered diphtheria antitoxin to Nome, Alaska in 1925 after an urgent telegraphic message was sent about an outbreak of the disease (on loan from the Cleveland Museum of Natural History). In addition to physical objects, *The Once and Future Web* features touch-screen interactive stations that deliver text, images, music, videos and a searchable exhibition library for subjects ranging from Samuel F.B. Morse's original invention to the role that the Internet plays today in delivering medical information to the public. The opening of the exhibition included the premier of a humorous play depicting the development of telecommunications technology and an evening reception. An electronic version of the exhibition is available on NLM's website, which includes a "Learning Station" for high school teachers. LO provided tours of the exhibition to groups totaling more than 1,000 people during the last quarter of FY2001.

As the new exhibition opened at NLM, previous major exhibitions found new life in different formats. The DVD version of *Breath of Life*, the previous exhibition on asthma, produced by the Lister Hill Center was distributed by the National Heart, Lung, and Blood Institute to attendees at the Third Triennial World Asthma Meeting on July 13, 2001, in Chicago. The American Library Association received grant funding from the National Endowment for the Humanities to develop a touring version of *Frankenstein: Penetrating the Secrets of Nature*, an earlier NLM exhibition. NLM will provide supplemental funding as well as technical assistance. Public, academic, and health sciences libraries may apply to be one of the 40 sites to which the exhibition will tour over the next several years. Planning and research for the next major exhibition, which will focus on American women physicians, are well underway. An Ad Hoc Committee of distinguished women and men, chaired by Dr. Tenley Albright, will advise NLM on the content of this exhibition. In addition to being a distinguished surgeon and Olympic gold-medal winning skater, Dr. Albright is a former chair of the NLM Board of Regents.

In FY2001, staff from HMD and the Office of the Associate Director worked with the Office of the Director, the Lister Hill Center, and the British Library to develop and launch a "Turning the Pages" electronic exhibit of the Elizabeth Blackwell's *A Curious Herbal* (1737-1739), a beautiful book that is in both the NLM and the British Library collections. The British Library developed "Turning the Pages," a remarkable program that uses computer animation, high-quality digitized images, and touch screen technology to simulate the action of turning the pages of a book. NLM approached the British Library about adding scientific and medical works to the group of books and manuscripts available in this form. *A Curious Herbal* was launched simultaneously at NLM and the British Library on March 16 in a live transatlantic broadcast featuring NLM Director Donald A.B. Lindberg, M.D., NLM Deputy Director Kent Smith, and officials of the British Library. Work is proceeding on the next book, which will be Vesalius's *Humani corporis fabrica*, the first truly modern anatomical text. NLM will have permanent displays of "Turning the Pages" in three locations: the Visitors Center, the front lobby of the Library building, and inside HMD. In FY2001, NLM displayed "Turning the Pages" at the annual meetings of the Medical Library Association and the American Library Association where it was extremely popular.

HMD also installs "mini-exhibits" in the exhibit cases at the entrance to the HMD Reading Room. At the start of FY2001, the exhibit on display was *Joshua Lederberg: Biomedical Science and the Public Interest,* which highlighted the career of the Nobel Prize winning scientist in conjunction with his 75th birthday. Dr. Lederberg is a member of the NLM Board of Regents and chairs the PubMedCentral Advisory Committee. His papers are also featured in *Profiles in Science*. On April 13, a mini-exhibit entitled *Tempest in a Teapot* opened. Produced with assistance from Susan Junod of the Food and Drug Administration, it included a wide array of artifacts and printed materials relating to the history of tea, focusing especially on tea in medicine and the regulation of tea in United States. In FY2001, web versions of two of last year's exhibits, *Medieval Manuscripts in the NLM* and *Classics to Traditional Chinese Medicine,* were made available.

NICHSR has distributed 384 copies of the video *Health Services Research: A Historical Perspective* since its release in late FY2000. The video drew heavily on a series of oral and video interviews with health services research leaders which NICHSR had commissioned. In January 2001, Dr. Joseph Newhouse, a distinguished health services researcher and a member of the NLM Board of Regents, chaired an Ad Hoc Committee of historians and health services researchers convened to advise NLM on additional steps it should take to document the history of the field. As a result of the committee's recommendations, NICHSR is arranging for additional interviews. One with John Eisenberg, M.D., Director of the Agency for Healthcare Research and Quality has been edited and sent to him for review. Dr. Eisenberg has also donated to the Library papers relating to the founding of the Society for Medical Decision-Making.

During FY2001, HMD organized a symposium on *Frontiers in Biomedical Research, 1945-1980,* to be held in October 2001. The symposium will feature both renowned scientists and noted historians of medicine and allow considerable time for audience participation. HMD also arranged a special celebration to recognize donors to NLM's historical collections, which was planned for the night of September 11, 2001. It was cancelled due to the attacks on the World Trade Center and the Pentagon and will be rescheduled in FY2002. HMD routinely sponsors a series of seminars by historical scholars as well as special public lectures in conjunction with the NLM Diversity Council. In celebration of African American History Month, Michael Blakey (Howard University) presented a lecture entitled "New York's African Burial Ground and the Struggle for Human Rights," on February 14. On March 29th, as part of Women's History Month, Susan Wells (Temple University) spoke on "Mary Putnam Jacobi and the Speaking Picture."

HMD staff members presented historical papers and lectures at professional meetings throughout the year and also published the results to their scholarship in books, chapters, articles, and reviews. HMD continued to play a lead role in preparing the "Images from the History of Public Health" feature in the *American Journal of Public Health.* This series occasionally features picture from NLM's collection. Simon Baatz, Ph.D., a temporary NLM employee working on an updated history of NLM, conducted 45 oral history interviews on the history of the Library, and completed drafts of several chapters. Theodore Brown, Ph.D., Professor and Chair of the History Department, University of Rochester, spent several months at NLM as a visiting historical scholar.

Training, Recruitment, and Evaluation of Health Sciences Librarianship

LO develops online services training programs for health sciences librarians and other search intermediaries; oversees the activities of the NN/LM-funded National Training Center and Clearinghouse at the New York Academy of Medicine; directs the NLM Associate Fellowship program for post-masters librarians; and develops and presents continuing education programs for librarians in health services research, public health, and other topics. LO also collaborates with the Medical Library Association (MLA) and other library associations to increase the diversity of those

entering the profession, to promote multi-institution evaluation of library services, and to explore specialist roles for health sciences librarians.

In FY2001, the Medlars Management Section (MMS) and the National Training Center taught 1,189 students in 72 traditional face-to-face classes, a 6% increase in students from last year. MMS released an interactive web-based PubMed tutorial that allows anyone with web access to learn about searching PubMed. With this tutorial, NLM now offers PubMed training worldwide 24 hours a day, 7 days a week. Accessible from PubMed's sidebar, the web-based PubMed tutorial averaged about 1525 visits daily since its March debut. The tutorial was updated late in FY2001 to improve navigation, add a module to cover the Cubby feature, achieve section 508 accessibility compliance, and optimize the Flash animations to minimize download time. MMS staff assisted an NLM Associate Fellow in testing the use of Qarbon's Viewlet technology for short animated demonstrations. MMS and NICHSR have decided to use this technology for short animated demonstrations to be delivered via the web.

There were eight first-year and four second-year participants in the Associate Fellowship program in FY2001. Of the eight who finished the first year at NLM in August 2001, two elected to continue on to the optional second year at the University of California and the University of North Carolina; three accepted jobs in the United States—at NLM, the Library of the Association of American Medical Colleges, and the Louisiana State University; one returned to Canada to pursue opportunities there; and the international Associate from China accepted an internship at Vanderbilt. Of the four second-year participants, one completed the program at Vanderbilt and accepted a permanent position there; one spent a year at McMaster and then joined the Cochrane Center at Oxford. Two returned in the spring from the University of Pittsburgh and Johns Hopkins University respectively to complete the second year of the program at NLM. Five new first-year Associate Fellows entered the program in September 2001, including an international participant from Kenya.

The *NLM Long Range Plan, 2000–2005* recommends that NLM look into expanding the supply of specialist librarians in clinical informatics, bioinformatics, and health policy. In FY2001, LO provided funding to the Medical Library Association for an April 2002 conference, to be held at NLM, on the potential role of librarians as "informationists" in clinical care and clinical research. (The term "informationist" was introduced by Frank Davidoff, M.D., and Valerie Florance, Ph.D., in an editorial in *Annals of Internal Medicine* in June 2000. Most people view it as very related to, but still different from, a clinical medical librarian.) NCBI consulted with LO in developing plans for a course to train health sciences librarians and others to train biology faculty and students in the use of advanced genomic information resources. NCBI is enlisting a number of NN/LM members to help develop and test the course materials.

NICHSR continues to develop continuing education programs to increase health sciences librarians' understanding of health policy, health services research, public health, and related fields. In FY2001, NICHSR worked with MLA to sponsor a post-conference symposium on *Library Partnerships—Powerful Connections.* The agenda and slides are available on the NICHSR website. NICHSR is developing a continuing education course on health economics for the 2002 MLA meeting, which will join the other courses available on the NICHSR website shortly after it is presented. Current web course offerings include: *Introduction to Health Services Research, Introduction to Health Care Technology Assessment,* and *Finding and Using Health Statistics: A Self-Study Course*. Originally commissioned by NICHSR, this last is currently being updated with funding from the National Center for Health Statistics. The updated version will be published on the NICHSR website.

In FY2001, NLM provided funding to the Association of Research Libraries (ARL) to support the participation of health sciences librarians from minority groups in ARL leadership development programs. The Library also funded the Association of Academic Health Sciences Libraries to allow broader participation from health sciences librarians in a study of the

quality of library services using ARL's LibQual study instruments and methodology. LO continues to mount special web pages highlighting important projects undertaken by health sciences librarians in October, which MLA has designated as National Medical Librarians Month.

Health Informatics Activities

In addition to providing the Library's basic services, LO represents NLM in several initiatives designed to promote more effective health applications of advanced computing and communications technologies. In FY2001, LO continued to serve on the Department's Health Data Standards Committee that is overseeing the implementation of the administrative simplification provisions of the Health Insurance Portability and Accountability Act of 1996 (HIPAA). In this capacity, LO staff assisted in drafting a proposed revision to the language related to codes for drugs that appeared in the final HIPAA Transactions rule. LO also provides staff support to the Subcommittee on Standards and Security for the National Committee on Vital and Health Statistics. In FY2001, LO briefed the committee several times on matters related to administrative codes and classifications, clinical vocabulary, and the UMLS project and assisted in organizing the committee's hearings.

On behalf of several Federal agencies, LO initiated and now manages a contract with the Regenstrief Institute that supports the continued development and free distribution of LOINC (Logical Observations: Identifiers, Names, Codes), a detailed clinical nomenclature that is increasingly used in the automated exchange of test results. During FY2001, LO continued to represent the Department and other Federal agencies in negotiations with the College of American Pathologists for a U.S.-wide arrangement for use of the SNOMED clinical terminology. LO also worked with the Department of Veterans Affairs, the FDA, and the HL7 clinical standards development organization to develop and test a standard form for representing an orderable drug, to which the forms used in a variety of commercial drug information services and institutional drug files can be mapped.

LO served on the organizing committee for the 2001 of the American Medical Informatics Association Spring Congress on developing a national agenda for public health informatics. LO staff had primary responsibility for organizing the sections of the meeting dealing with standards and vocabulary, prepared a bibliography on public health informatics that was issued in conjunction with the meeting, and co-authored the papers summarizing the recommendations from the meeting.

Table 1

Growth of Collections

Collection	*Previous Total (9/30/00)*	*Added FY 2001*	*New Total (9/30/01)*
Book Materials			
Monographs:			
Before 1500	578	5	583
1501-1600	5818	56	5,874
1601-1700	10,139	33	10,172
1701-1800	24,483	77	24,560
1801-1870	41,168	118	41,286
Americana	2,341	0	2,341
1870-Present	682,150	13,930	696,080
Theses (historical)	281,794	0	281,794
Pamphlets	172,021	0	172,021
Bound serial volumes	1,190,381	34,717	1,225,098
Volumes withdrawn	(72,560)	(2,567)	(75,127)
Total volumes	2,338,313	46,369	2,384,682
Nonbook Materials			
Microforms:			
Reels of microfilm	109,008	7,423	116,431
Number of microfiche	420,431	18,253	438,684
Total microforms	529,439	25,676	555,115
Audiovisuals	66,476	1,814	68,290
Computer software	1,780	212	1,992
Pictures	56,940	20	56,960
Manuscripts	2,946,107	190,750	3,136,857*
Total nonbook	3,600,742	218,472	3,819,214
Total book and nonbook	**5,939,055**	**264,841**	**6,203,896**

*Equivalent to 1,792 linear feet.

Table 2

Acquisition Statistics

Acquisitions	*FY 1999*	*FY 2000*	*FY 2001*
Serial titles received	22,433	23,141	20,314
Publications processed:			
Serial pieces	123,823	143,636	142,642
Other	14,418	22,384	21,338
Total	138,241	166,020	163,980
Obligations for:			
Publications	$5,370,797	$4,895,999	$5,155,054
(For rare books)	($292,603)	($267,300)	($279,710)

Table 3

Cataloging Statistics

	FY 1999	*FY 2000*	*FY 2001*
Completed Cataloging	14,396	20,067	19,024

Table 4

Bibliographic Services

Services	*FY 1999*	*FY 2000*	*FY 2001*
Citations published in MEDLINE	434,525	442,168	463,014
For *Index Medicus*	421,423	434,813	445,041
Journals indexed for *Index Medicus*	3,394	3,472	3,707
Abstracts entered	338,435	341,682	345,624

Table 5

Web Services

Services	*FY 2000*	*FY 2001*
NLM Web Home Page		
Page Views	25,936,000	36,248,000
Unique Visitors	3,572,000	4,490,000
MEDLINEplus		
Page Views	18,437,000	62,069,000
Unique Visitors	2,098,000	4,409,000

Table 6

Circulation Statistics

Activity	*FY 1999*	*FY 2000*	*FY 2001*
Requests Received	751,732	749,869	682,777
Interlibrary Loan	396,516	390,574	338,627
Onsite	355,216	359,295	344,150
Requests Filled:	570,966	589,516	535,594
Interlibrary Loan*	301,073	299,182	251,525
Onsite	269,893	292,664	284,069

**Statistics on photocopy versus original loans filled are no longer kept.*

Table 7

Online Searches—All Databases

	FY 1999	*FY 2000*	*FY 2001*
Total online searches	191,000,000	244,000,000	313,000,000

Table 8

Reference and Customer Services

Activity FY 1997	*FY 1999*	*FY 2000*	*FY 2001*
Offsite requests	54,542	62,971	59,634
Onsite requests	56,737	51,456	51,287
Total	111,279	114,427	110,921

Table 9

Preservation Activities

Activity	*FY 2000*	*2001*
Volumes bound	31,874	31,625
Volumes microfilmed	4,513	5,131
Volumes repaired onsite	2,000	1,403
Audiovisuals preserved	46	225
Historical volumes conserved	385	128

Table 10

History of Medicine Activities

Activity	*FY 1999*	*FY 2000*	*FY 2001*
Acquisitions:			
Books	170	226	314
Modern manuscripts	129,885	1,915,550	1,340,150*
Prints and photographs	1,773	1,391	3,324
Historical audiovisuals	114	37	1593
Processing:			
Books cataloged	58	49	510
Modern manuscripts cataloged	0	87,150	190,750**
Pictures cataloged	83	256	20
Citations indexed	1,022	1,066	285
Public Services:			
Reference questions answered	14,050	15,143	15,718
Onsite requests filled	3,672	4,485	4,844

*Equivalent to 765.8 linear feet
**Equivalent to 109 linear feet

SPECIALIZED INFORMATION SERVICES

Martha Szczur
Acting Associate Director

The Toxicology and Environmental Health Information Program (TEHIP), known originally as the Toxicology Information Program, was established more than 30 years ago within the National Library of Medicine in the Division of Specialized Information Services (SIS). Over the years TEHIP has provided for the increasing need for toxicological and environmental health information by taking advantage of new computer and communication technologies to provide more rapid access to a wider audience. We have moved beyond the bounds of the physical NLM, exploring ways to point and link users to relevant sources of toxicological and environmental health information wherever these sources may reside. This is being accomplished primarily through the TEHIP and AIDS Web sites developed and maintained by SIS. Development of HIV/AIDS information resources became a focus of the Division several years ago, and now includes several collaborative efforts in information resource development and deployment, including a focus on the information needs of other special populations. This past year the Office on Outreach and Special Populations was established to coordinate activities in this area. Continuous refinements and additions to our Web-based systems are made to allow easy access to the wide range of information collected by this Division. Our usage has continued to increase over the past year with access to all toxicology and HIV/AIDS data free over the Internet.

In FY2001 SIS selected several projects for significant re-engineering, proposing new opportunities to enhance SIS information resources and provide new services in emerging areas. Prototypes are underway utilizing graphical display of data from our information resources, innovative access and interfaces for consumers, and geographical information systems. Program direction has been guided in the past by two Institute of Medicine (IOM) reports focusing on the TEHIP Program: *Toxicology and Environmental Health Information Resources: the Role of the National Library of Medicine*, released in the spring of 1997, and a follow-on report, *Internet Access to the NLM's Toxicology and Environmental Health Databases*, published in 1999. Both reports have been instrumental in our re-engineering efforts, and are used as reference for internal staff discussions at annual strategic planning retreats.

Resource Building

The wide range of resources related to toxicology and environmental health information, HIV/AIDS information, and special populations information include many databases that are created or acquired as well as other services and projects.

The **Hazardous Substances Data Bank (HSDB)** continues to be a highly used resource, averaging over 40,000 searches each month (a 30% increase over FY2000). Increased emphasis continues to be placed on providing more data on human toxicology and clinical medicine within HSDB, in keeping with past recommendations of the Board of Regents' Subcommittee on TEHIP. The selection of new members of the Scientific Review Panel for HSDB reflects this shift in content emphasis. Newer sources of relevant data are being examined for incorporation into new and existing data fields within the current 4,550 HSDB records. Because of increased staff efforts, more records are being processed through special enhancements, including source updates from various peer-reviewed files. Special summary information is being prepared to allow easier presentation of information at a health consumer level. The process of developing a new Web-based system for HSDB creation, review, and maintenance is continuing. An initial workshop to define some of the issues related to this re-engineering effort was held in October 2000, and needs analysis is well under way.

CHEMID*plus* (Chemical Identification File) is an NLM online chemical dictionary, which contains over 350,000 records, primarily

describing chemicals of biomedical and regulatory importance, and available to users on the Web. ChemID*plus* features include chemical structure search and display for 100,000 chemicals, and hyperlinked fields that retrieve data for a given chemical from other resources such as MEDLINE or HSDB. Over 15,000 records of regulatory interest collectively known as SUPERLIST are also available and hyperlinked in ChemID*plus*. During FY2001, new software enhancements and a new server provided easier access to structure display and a more robust system for ChemID*plus*.

TOXLINE (Toxicology Information On-line) is a large bibliographic database traditionally produced by merging "toxicology" subsets from secondary sources. By the end of FY2001, the database included over 3 million citations to toxicology literature going back to 1965. In FY2000, we began the transition to a next generation TOXLINE, reducing the components needed to produce the database by creating a toxicology subset on NLM's PubMed so that users can access standard journal literature in toxicology and environmental health as part of an enlarging MEDLINE database. NLM added additional journals in the area of toxicology and environmental health to MEDLINE to cover some of the literature formerly provided by outside sources. For the nonstandard journal literature in this area we created a Web-based system on TOXNET that allows efficient acquisition and updating of these components. Easy access to this TOXLINE Special database and to TOXLINE Core, the standard journal literature on PubMed, is available from the TOXNET web site.

DIRLINE (Directory of Information Resources On-line) is NLM's online directory of resources including organizations, databases, bulletin boards, as well as projects and programs with special biomedical subject focus. These resources provide information to users which may not be available from one of the other NLM bibliographic or factual databases. DIRLINE continues to receive a high level of use through a new interface, which became public in October 1999. This new interface supports direct links to the Web sites of the organizations listed in the database, as well as direct e-mail connections. Providing direct links for users facilitates ease of access for consumers as well as for health professionals. The quality and utility of the database continues to improve as duplicates have been eliminated through changes in policy and streamlining of maintenance. *Health Hotlines*, the always popular publication of health-related toll-free telephone numbers, has a Web version which also indicates the availability of Spanish speaking customer service representatives and Spanish language publications from the resources listed.

The **Toxic Chemical Release Inventory (TRI)** series of files now includes five online files, TRI95 through TRI99. These files remain an important resource for environmental release data and are a useful complement to our other databases. Mandated by the Emergency Planning and Community Right-to-Know Act (Title III of the Superfund Amendments and Reauthorization Act of 1986), these EPA databases contain data on environmental release data to air, water, and soil for over 600 EPA-specified chemicals. These files will be an important component of planned projects using geographical information systems.

The **Chemical Carcinogenesis Research Information System** (CCRIS) continues to be built, maintained, and made publicly accessible at NLM. This data bank is supported by the National Cancer Institute and has grown to over 8,000 records. The chemical-specific data covers the areas of carcinogenesis, mutagenesis, tumor promotion and tumor inhibition.

The **Integrated Risk Information System** (IRIS), EPA's official health risk assessment file, continues to experience high usage and be very popular with the user community. EPA has had a version of IRIS on the agency's Web page since 1996, and as we move to Web access we will consider how best to integrate our Web service with what EPA provides. IRIS now contains 538 chemicals.

The **GENE-TOX** file is built directly on TOXNET by EPA scientific staff. This file contains peer-reviewed genetic toxicology (mutagenicity) studies for about 3,200 chemicals. GENE-TOX receives a high level of interest among users in other countries.

The **Registry of Toxic Effects of Chemical Substances** (RTECS) is a data bank

based upon a National Institute for Occupational Safety and Health (NIOSH) file by the same name which NLM restructured and made available for online searching. With our move to free Internet access to all databases, NIOSH requested that we no longer include RTECS on our system. We continue to use RTECS in the creation of the Hazardous Substance Data Bank.

The **Developmental and Reproductive Toxicology** (DART) database now contains over 49,000 citations from literature published since 1989 on agents that may cause birth defects. DART is a continuation of the Environmental Teratology Information Center backfile (ETICBACK) database, which contains almost 50,000 citations to literature published from 1950 to 1989. DART is funded by NLM, the Environmental Protection Agency, the National Institute of Environmental Health Sciences and the FDA's National Center for Toxicological Research, and is managed by NLM.

The **Environmental Mutagen Information Center** (EMIC) database contains over 24,000 citations to literature on agents that have been tested for genotoxic activity. A backfile for EMIC (EMICBACK) contains over 75,000 citations to the literature published from 1950 to 1991. The Environmental Protection Agency, the National Institute of Environmental Health Sciences and NLM, collaborating partners in this effort, stopped compiling this special collection as of December 1999, but we will keep the collections as part of the TOXLINE Special database on TOXNET.

Resource Access

The SIS Web server provides a central point of access for the varied programs, activities, and services of the Division. Through this server users can access interactive retrieval services in toxicology and environmental health, HIV/AIDS information, or special population health information; find program descriptions and documentation; or be connected to outside related resources. During FY2001, we completed a redesign of the SIS Web site which now incorporates information about SIS in general, as well as toxicology and environmental health and AIDS information. Both the toxicology and environmental health and AIDS Web pages provide links to NLM outreach activities in these subjects, access to NLM databases, links to selected Web sites in these subjects, as well as tutorials, fact sheets, and other publications produced by SIS.

Toxicology Data Network (TOXNET)

The **Toxicology Data Network (TOXNET)**, NLM's information system providing database management for many of its toxicology files, has moved from a networked microprocessor environment to a UNIX-based platform (Solaris Version 2.6) on a SUN Enterprise 3000 computer. Integration of this configuration with other SIS database creation systems and the Web access to them is currently underway.

In FY2001, SIS continued the development of a new search interface to allow integrated access to the SIS toxicology and environmental health databases. This new search interface allows users to easily search HSDB, TOXLINE, CCRIS, Gene-Tox, DART, EMIC, IRIS, and TRI. Based on recommendations from the IOM, users are presented with a basic search screen with just a single input box for searching, with customized screens for more sophisticated users. These advanced features include Boolean searching and the ability to limit search terms to specific fields. A TOXNET user online survey is planned for the fall of 2001. New search screen designs were begun in 2001, and research and development projects such as a chemical spellchecker, automatic indexing, and a toxicology gateway system were carried out. Plans are underway to link the new NLM Gateway to the TOXNET search system, making it easier for new users to learn about our resources.

Chemical Structure Server

The chemical structure server has evolved from a mechanism to provide structure searching for chemicals covered by SIS databases to a system for integrating chemical dictionary record building and structure searching. This system uses special molecular searching programs and includes a prototype database for construction of ChemID records.

The chemical information resources continue to be consolidated on a server that meets the requirements for chemical structure creation and access.

AIDS Information Services

NLM has expanded its HIV/AIDS information services by expanding the number of relevant topic pages on MEDLINE*plus* as well as completing an overhaul and major expansion to the AIDS Web site (http://aids.nlm.nih.gov). This Web site not only contains links to NLM's programs and services, but also a well-organized and expansive set of links to many HIV/AIDS resources more technical in nature than appropriate for MEDLINE*plus*.

NLM has continued its successful AIDS Community Information Outreach Program with 16 new awards in FY2001, bringing the total number of awards made to 140.

NLM remains as the project manager for the multi-agency AIDS Clinical Trials Information Service (ACTIS) and the HIV/AIDS Treatment Information Service (ATIS). A new contract for support of NLM Clinical Information Services has been awarded that includes these services as well as certain support work for ClinicalTrials.gov and outreach programs.

Outreach / User Support

SIS has initiated a project developing a set of population-specific mini Web sites that focus on the issues of particular populations or geographic areas. These Web sites include relevant policy, legislative, and organizational information as well as organized links to health and environmental issues of that particular population. The arctic health Web site is the first of these to be released. The plan for these Web sites is for NLM to develop them and then work with a local university or agency more directly involved in the subject for continued maintenance.

NLM funded four outreach projects targeting minority populations and involving minority community-based organizations. These projects are intended to enable organizations to design local programs for improving access to consumer health information. The following organizations received funding for two-year projects:

- Northern Wisconsin AHEC (Wausau, WI)
- University of Rochester Health Sciences Library (Rochester, NY)
- Harbor View Medical Center (Seattle, WA)
- Virginia Commonwealth University (Richmond, VA)

SIS initiated a collaborative project with the DHHS Office of Minority Health (OMH). As part of their AIDS initiative, OMH conducted a needs assessment of community organizations in six major cities. Among the top needs identified by these community-based organizations was training in the use of the Internet to find health information resources. NLM is collaborating on this effort and will be providing the training in searching Internet resources.

SIS continues its support of the Toxicology Information Outreach Project (TIOP). 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 waste information resources. This year TIOP celebrated its tenth anniversary at its annual meeting at the NLM. TIOP also expanded by adding representation from the Oglala Lakota College, a tribal college, and from the University of Puerto Rico Medical School. Training was conducted at both of these new participating schools. An assessment of the program was conducted and the results will be used to formulate additional activities.

A more recent addition to NLM's outreach programs is one to improve access to health-related disaster information in three disaster-prone Central American countries: Nicaragua, Honduras, and El Salvador. NLM is funding the Regional Disaster Information Center for Latin America and the Caribbean (CRID) to strengthen the capacity of these countries to collect, index, manage, store, and disseminate

public health and medical information related to disasters.

SIS exhibited at over 30 conferences in this fiscal year. Several of these provided opportunities for presentations or workshops about NLM's information resources. In addition, SIS provided support for some conferences, including the Symposium on Career Opportunities in Biomedical Sciences sponsored by the Association of Minority Health Professions Schools. NLM also sponsored the e-health track at Expo2000 organized by Clark-Atlanta University for faculty and administrators from HBCUs, minority business leaders, and leaders of community organizations.

User Support Computer-Based Activities

SIS has developed a set of internet tutorials, *Toxicology Tutors*, which are introductory level toxicology courses available on the SIS Web server. We are considering appropriate additions to this collection for development in the future.

Other new avenues of user support are being focused at the consumer level, with a collaborative development of MEDLINE*plus* topics and addition of other special topics of concern to the general public to the SIS Web site. Our topics on Chemical Warfare Agents and Pesticide Control of West Nile Virus have been on the Web for over a year. New topics, including one on Lingering Airborne Hazards of the World Trade Center Attacks, was released in the fall of 2001.

Alternatives to Animal Testing

SIS continued to compile and publish references from the MEDLARS files that were identified as relevant to methods or procedures which could be used to reduce, refine, or replace animals in biomedical research and toxicological testing. Requests for these quarterly bibliographies have increased, as has the number of articles deemed relevant to the field. Bibliographies issued during the past four years are available on the Internet through the SIS Web Server, and the primary distribution mechanism for this project is now the Internet.

Other Specialized Services

In addition to toxicologic data files, SIS is evaluating other areas for creating specialized factual and bibliographic databases. Resource allocations are being made to determine the feasibility of initiating more clinical medicine information products for public, health professional, and scientific audiences. SIS has begun a critical review of its role in organizing and disseminating drug information in various formats, exploring a role in the assessment of the integrity and validity of such information. Another new project is developing a symptom and occupation based clinical medicine resource appropriate for use on the Web. Yet another initiative is preparing a Web resource for consumers that links brand name household products with their ingredient chemicals and potential adverse health effects. Both of these products are ready for beta testing, and are expected to be made available on the SIS Web site in 2002.

In these and other new initiatives, SIS continues to search for new ways to be responsive to user needs in acquiring and using toxicology and environmental health, HIV/AIDS, and other specialized information resources.

LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS

Alexa T. McCray, Ph.D.
Director

The Lister Hill National Center for Biomedical Communications conducts informatics research and development in support of the National Library of Medicine's mission. NLM's updated Long Range Plan (2000–2005) enumerates four broad goals for the Library as follows:

1. Organize health-related information and provide access to it;
2. Promote use of health information by health professionals and the public;
3. Strengthen the informatics infrastructure for biomedicine and health; and
4. Conduct and support informatics research.

As an R&D division, all Lister Hill Center activities are in direct support of Goal 4, and many of our research programs are discussed in that section of the Long Range Plan. In addition, however, our research is strongly motivated by the first three goals, and our activities often result in research products that are heavily used by NLM's broad constituency.

This report is organized to reflect our work in support of each of the first three goals. In some cases our work results in methods, techniques, or tools that contribute to furthering a goal, while in other cases our work leads to fully operational systems that continue to be improved on the basis of further research and experimentation.

The most current information about Lister Hill Center programs and research activities can be found at http://lhncbc.nlm.nih.gov/.

Goal 1: Organize Health-Related Information and Provide Access to It

Acquire, Organize, and Preserve Biomedical Information

Digital Library Research

The **Digital Library Research** project involves all aspects of creating and disseminating digital collections, including standards, emerging technologies and formats, copyright and legal issues, effects on previously established processes, protection of original materials, and permanent archiving of digital surrogates. Research issues currently in focus are long-term preservation of digital archives, innovative methods for creating and accessing digital library collections, and the development of modular and open information environments. Investigations concerning interoperability among digital library systems, the role of well-structured metadata, and varying "points of view" on the same underlying data set are also being pursued.

The ***Profiles in Science*** Web site uses innovative digital technology to make available the manuscript collections of biomedical scientists of the twentieth century. The content of the database is created in collaboration with the History of Medicine Division, which processes and stores the physical collections. The documents have been donated to NLM and contain published and unpublished materials, including books, journal volumes, pamphlets, diaries, letters, manuscripts, photographs, audio tapes and other audiovisual resources. Presently the database features the archives of seven prominent American biomedical scientists: Oswald Avery, Joshua Lederberg, Martin Rodbell, Julius Axelrod, Christian Anfinsen, Marshall Nirenberg, and, most recently, Barbara McClintock.

Several research projects this year continued to enhance the effectiveness of the *Profiles in Science* site. One study sought to improve the search system by analyzing user queries. Considerable effort was directed at metadata concerns, including the identification and prevention of common input errors as well as the addition of metadata elements required for permanent archival of digital objects and audio and video items. Certain metadata elements were restructured to eliminate duplicate information.

Finally, accessibility of digital objects to the visually and hearing impaired was enhanced.

During 2002 we will launch a new project in **Digital Preservation Research** as part of our overall digital library research program. Digital information in any form is at risk. Software and hardware become obsolete, and versions and file formats change, making data inaccessible. Data stored in even the simplest form is in danger due to computer media degradation and obsolescence. Online information such as e-journals and databases are susceptible; they may become partially or entirely unreadable, and may not be recoverable by the time the problem is detected. Strategies such as emulation (keeping alive the software, hardware and applications needed to access a digital object) and migration (converting the digital object to current versions and formats, and making copies to new media) will be tested and evaluated. We will conduct research into these strategies and possible alternatives to them.

Document Image Analysis and Understanding

Document image analysis and understanding research combined with database design, graphical user interface design for workstations, image processing, speech recognition and related areas underlie the development of **MARS** (Medical Article Records System), a system to automate the production of MEDLINE records from biomedical journals. MARS-1, primarily an optical character recognition (OCR) centered system designed to extract only the article abstracts while all other fields were manually entered, was supplanted by a second generation system (MARS-2) designed to extract the author names, affiliations and article title automatically. Performance data showed that while MARS-1 was a considerable improvement over the traditional keyboarding method, MARS-2 reduces the required labor effort to 25% of the manual approach.

After pages are scanned, page segmentation algorithms block out regions (zones) of contiguous text on the bitmapped image. A four-step process, combining both top-down and bottom-up strategies, is followed. First, the OCR output zones are disassembled into individual text lines. Then, the lines are split horizontally into fragments when word spaces exceed an empirically determined threshold. Third, the lines and line fragments are combined vertically into initial zones using as criteria vertical distance, line edge alignment and similarity of line features. Last, these zones are combined into final zones using as criteria horizontal distance between initial zones, zone edge alignment and similarity of zone features. This method was evaluated on 295 page images with 1180 zoned regions, and yielded an accuracy of 97.9%

Identifying or labeling the zones of interest as *authors, title, affiliation* and *abstract* requires a family of autolabeling algorithms developed on the basis of a comprehensive set of 120 rules derived from both geometric as well as non-geometric (i.e., textual or numeric data) features from the OCR output. The algorithms were tested against the images of articles from the journal titles indexed in MEDLINE excluding the approximately 1,000 titles for which publishers supply records in SGML form. The remaining 3,000+ titles are therefore candidates for the automatic processes in MARS. Errors encountered in testing the baseline algorithm were largely in labeling affiliation zones, and these were due to incorrect font attributes in the output of the commercial OCR system. To date, 2,028 journal titles can be processed automatically, but for 580 of these the publishers are delivering citations via XML tagged format, leaving 1,448 titles suitable for MARS-2 processing. This effort will continue until all the scanned journals are tested and the rules are tailored to allow automated processing of the largest possible number.

The **Indexing Initiative** project investigates methods whereby automated indexing may partially or completely substitute for expert indexing of the biomedical literature by humans. The project is pursuing concept-based indexing methods that go beyond automatic word-based indexing and will be considered a success if retrieval performance is equal to or better than that of systems using humanly assigned index terms.

Project members have developed a system, Medical Text Indexer (MTI), based on

three core indexing methodologies. The first of these calls on the MetaMap program to map citation text to concepts in the UMLS Metathesaurus. The second approach, the trigram phrase algorithm, uses character trigrams to match text to Metathesaurus concepts, while the third uses a variant of the PubMed related citations algorithm to find MeSH headings related to input text. Results from the three methods are restricted to MeSH and combined into a ranked list of recommended indexing terms.

Experiments to evaluate the efficacy of MTI indexing recommendations to NLM indexers, a semiautomatic application of MTI, are being conducted. In addition, results of the MTI system are being evaluated for use in a fully automatic indexing environment for collections of documents that will not be indexed by humans. Research into the system's indexing methods continues. In particular, a major word sense disambiguation effort based on statistical methods such as journal descriptor indexing is being undertaken to resolve ambiguities encountered during the automatic indexing process. Finally, the Indexing Initiative team plans to extend its research to address the full text documents that are becoming increasingly available.

Visible Human Project

The **Visible Human Project** data sets are designed to serve as a common reference for the study of human anatomy, as a set of common public domain data for testing medical imaging algorithms, and as a testbed 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 for a duplication fee. The data sets are being applied to a wide range of educational, diagnostic, treatment planning, virtual reality, artistic, mathematical and industrial uses by over 1700 licensees in 44 countries.

The University of Colorado Health Science Center, Center for Human Simulation is readying an alpha version of a head and neck atlas to be released as a web site in 2002. The atlas, based on the Visible Human data set, is designed to serve numerous functions. In addition to being simply an educational resource, it is to be a test platform for the development of methods and standards for digital image libraries for educational applications, and as a catalyst for the development of methods for linking images and symbolic knowledge.

We recently awarded two one-year contracts to study anatomical methods that will improve the data acquisition techniques used to obtain the original Visible Human Project data set. The first, awarded to Brigham and Women's Hospital, will attempt to overcome the problem of expanding soft tissue during the freezing process required for cryosectioning. This group will also attempt to increase the spatial voxel resolution from the original 0.33 mm^3 to 0.15 mm^3. In the second award, the University of Colorado Health Sciences Center will examine techniques to save structures damaged (e.g. teeth) or missing (e.g. ossicles of the ear) in the original data set. In addition, they will attempt to improve image contrast to aid in discriminating between anatomical structures.

Another Visible Human Project inspired initiative, the Insight Toolkit (ITK), began alpha testing this past year. The ITK makes available a variety of open source image processing algorithms for computing segmentation and registration on a variety of hardware platforms. Platforms currently supported are PCs running Visual C++, Sun Workstations running the GNU C++ compiler, SGI workstations and Linux. This work is being conducted by a consortium of universities and companies.

Three additional contracts, currently in their second of three years, involve using the Visible Human Project data set. The University of Colorado Health Sciences Center, Center for Human Simulation is exploring use of the World Wide Web to do 3D anatomical explorations for teaching. These undergraduate and postgraduate applications include audio, graphic and haptic interfaces. They have demonstrated a module for the knee, and are working on converting it to HTML for dissemination on the web. At the University of Michigan, studies are underway to develop user controllable 2D and 3D browsers

that allow manipulating arbitrarily cut planes. Stanford University researchers are experimenting with haptics in order to enable surgeons to feel as well as see their way through surgical simulations based on the Visible Human Project data set over the Internet 2 network.

This past year saw the continued maintenance of two databases to record information about Visible Human Project use. The first database logs information about the now over 1700 Visible Human Project license holders and records their intent for using the images; the second records information about the products the licensees are providing NLM in compliance with the license agreement.

We hosted the Third Visible Human Project Conference this past year. Thirty-two license holders presented papers detailing outcomes of their work with the image data set. In addition, a panel of five renowned anatomists discussed anatomy in the 21st Century, and the keynote address, "Volumetric Imaging for the Media," was presented by Alexander Tsiaras, President and CEO of Anatomical Travelogue, Inc. A full proceedings of the conference was published on CD-ROM.

With the goal of providing widespread access to the Visible Human images, to users with low speed connections as well, we are developing a new web interface to Visible Human data. This system, called AnatQuest, allows the user to quickly download selected parts of high resolution images, and then zoom and navigate over these. All the cross-sections as well as 195 rendered images (some of these from outside sources) may be accessed. Images are converted to tiled TIFF; selective downloading and display of these tiles is implemented by a servlet engine based on the Java Advanced Imaging API and the Java2D API; anatomical labels are displayed by cursor activation on regions defined by byte-masks and label tables. Research is proceeding toward improving performance, e.g., by trading off displayed tile size vs. lossy image compression.

Phase II of the high resolution scanning project, which involves the scanning of all the Visible Female 70mm film images, continued during the year. This includes the process of digitizing the complete set of 5189 film images, at 4500 ppi and 16 bits per color channel. The resulting file size of these images will be approximately 450 Mbytes per file. The scanning group digitizing the film images has developed custom software for the Windows NT platform to rapidly open and display these large files. Multiple derivative images will be provided at lower resolutions. The final images acquired from the scanning process have begun to be delivered and loaded from tapes onto a local server. They are then downloaded to a local PC workstation for viewing and quality control review. These images are being reviewed for resolution, color balance, focus, and artifacts.

Provide Access to Biomedical Information

NLM Gateway

The NLM offers an increasing number of Internet-based information resources, each with its own user interface. Lister Hill Center staff created the NLM Gateway to let users initiate searches in multiple retrieval systems from a single interface. The target audience for the new system is the Internet user who comes to NLM not knowing exactly what is available or how best to search for it. The **NLM Gateway** (http://gateway.nlm.nih.gov/), released in October 2000, now provides simultaneous searches of 11 document collections using 5 retrieval methods on different systems.

The current version of the NLM Gateway offers access to the following online resources:

- MEDLINE journal citations, 1966-present
- OLDMEDLINE journal citations, 1958-65
- LOCATORplus online catalog information for books, serial titles, audiovisuals
- MEDLINEplus consumer health information
- DIRLINE directory of health organizations
- AIDS meeting abstracts
- Health Services Research meeting abstracts
- Space Life Sciences meeting abstracts
- HSRProj information on health services

research projects
- Document delivery through NLM's Loansome Doc system
- UMLS Metathesaurus

Gateway users enter a query which is then reformulated and sent automatically to multiple retrieval systems having different characteristics but potentially useful results. Results from the target systems are presented in categories (for instance, journal article citations; books, serials and audiovisuals; consumer health information; meeting abstracts; other collections) rather than by database. In most categories, multiple document collections are searched.

Online visitors are invited to use the Gateway for an overview of some of NLM's resources. Some users will find what they need immediately. Others may find that one resource such as PubMed or MEDLINEplus has information they would like to know more about. They may then choose to go directly to that resource for a focused search using the native interface of that resource. Direct links to other major NLM resources are provided from the Gateway's search screen. This combination of a single point of access for an overview coupled with focused searches available for a second phase of inquiry should help improve user access to information offered at NLM's expanding series of Web sites.

Document Delivery over the Internet

This research area has the goal of applying document image processing to document delivery via the Internet. The two active projects in this area are **DocView** and **DocMorph**. DocView facilitates the delivery of library documents directly to the patron via the Internet. Because DocView is compatible with the Research Library Group's Ariel software, many biomedical libraries encourage their patrons to use it to receive, display, print and manage scanned images of journal articles and other documents. While Ariel is used by libraries and document suppliers routinely to send documents via the Internet to similar organizations, there are few options for end users to directly receive them. The DocView client software, which runs under any version of Microsoft Windows, enables an end user to receive documents over the Internet at the desktop, retain them in electronic form, view the images, organize the received documents into folders and file cabinets, electronically bookmark selected pages, manipulate the images (zoom, pan, scroll), copy and paste images, and print them if desired. DocView also serves as a TIFF viewer for compressed images received through the Internet by other means, such as Web browsers. Users may receive document images either via Ariel FTP or Multipurpose Internet Mail Extensions (MIME) protocols. With DocView, users may also forward documents to colleagues for collaborative work.

The DocMorph system serves as an important resource for librarians to convert library information from one form to another, often making it easier to exchange information. For instance, it is widely used to convert more than 50 different file formats to PDF for multi-platform delivery to patrons. By combining OCR with speech synthesis, DocMorph also enables the visually impaired to use library information. Dr. Richard Smith, director of the Wolfner Library for the Blind and Physically Handicapped, reported using it to convert documents to synthetic speech recorded onto audio tapes for his blind patrons. To date, more than 29,000 jobs have been submitted to DocMorph, representing 335,000 pages of information consisting of 29 Gbytes of data.

Language and Information Processing

The **Unified Medical Language System** (UMLS) project develops and distributes multi-purpose, electronic knowledge sources and associated lexical programs. System developers can use the UMLS products to enhance their applications-in systems focused on patient data, digital libraries, web and bibliographic retrieval, natural language processing, and decision support. Researchers find the UMLS products useful in investigating knowledge representation and retrieval questions. The UMLS currently comprises three knowledge sources, the Metathesaurus, the Semantic Network, and the SPECIALIST lexicon, with its associated lexical tools.

The UMLS data are made available over the Internet through the **UMLS Knowledge Source Server**, which provides direct access to each component of the UMLS. For example, users can request information about a particular concept in the Metathesaurus, including definition, semantic type, and synonyms as well as other concepts that are related to the input term. The Knowledge Source Server also accommodates navigation in the Semantic Network, allowing users to investigate relationships among semantic types and relations or to retrieve a list of Metathesaurus concepts assigned to a particular semantic type. The data in the SPECIALIST lexicon is also made available, providing the user with the syntactic and morphologic information about each lexical item it contains.

During this past year, we developed a new version of the Knowledge Source Server, which is based on a three-tier architecture. At the back end is a relational database management system that contains the UMLS data, while the middle layer consists of application logic to handle requests from clients, either web browsers or command line clients. There is also an API available for users who write their own applications. The new application relies on a reconfigured object model that dynamically populates object attributes upon request to reduce transmission traffic; multiple views of the object model are made available through a series of abstractions provided by helper methods. The redesigned server takes advantage of several Java facilities (for example, Remote Method Invocation, a server registry, and Java Database Connectivity) to provide a computationally efficient delivery mechanism for UMLS data. Alternative servlets, along with XML-encoded data and XSL style sheets, allow flexible, user-defined output capabilities.

The **Metathesaurus** is a knowledge source representing multiple biomedical vocabularies organized as concepts in a common format. It thus provides a rich terminology resource in which terms and vocabularies are linked by meaning. During this past year, the Metathesaurus group continued its two main tasks—producing increasingly comprehensive annual editions of the Metathesaurus with new and updated vocabulary sources, and developing and deploying new software systems for work on unified concept-oriented terminologies. During this past year, project staff reviewed all MeSH supplementary concepts. There is, thus, no longer any Metathesaurus content that has not received human concept-oriented review. New content for the 2002 release includes MedDRA, the FDA-mandated "Medical Dictionary for Regulatory Activities Terminology," the NCBI Taxonomy of organisms; and the first portions of the Department of Veterans Affairs National Drug Formulary, which will pilot a new standard normal form for clinical drug naming.

In 2001 we collaborated with research staff of the University of Amsterdam to develop an interactive editing interface for their International Classification of Primary Care medical vocabulary, which has been incorporated into the UMLS Metathesaurus. The project has developed a stand-alone Java-based tool for examining the vocabulary (which has concepts in 18 different languages, their character sets represented in Unicode), as well as a platform-independent web-based system using the open source tools Apache/PHP/MySQL. Because the Unicode-based work is done at the server end, and through the use of a Unicode-capable Java applet, the system can be used on clients that do not support Unicode. The work exposed and helped to remedy problems with the underlying datasets used for the non-English languages in this vocabulary.

A tutorial titled "Customizing the UMLS Metathesaurus" was presented at the Annual Symposium of the American Medical Informatics Association in November 2001. An updated, more user-friendly "MetamorphoSys" subsetting package has been created to assist users in selecting appropriate content from the Metathesaurus. Further efforts continue to provide online training materials and individual support for UMLS users.

While existing knowledge sources in the biomedical domain may be sufficient for information retrieval purposes, the organization of information in these resources is generally not suitable for reasoning. Automated inferencing requires the principled and consistent organization provided by ontologies. The

objective of the **Medical Ontology Research** project is to develop methods whereby ontologies can be acquired from existing resources and validated against other knowledge sources. Although the UMLS Metathesaurus and Semantic Network are used as the primary source of medical knowledge, OpenGALEN, CYC, and WordNet are being explored as well. During the past year, research focused on the taxonomic relation. The principles used to produce taxonomies are either intrinsic (properties of the partial ordering relation) or added to make knowledge more manageable (opposition of siblings and economy). The applicability of these principles in the UMLS as well as the theoretical issues raised by the application of these principles were addressed. The knowledge representation structure of the UMLS was also compared to general ontologies such as CYC and WordNet. Preliminary results suggest that these resources, used as a source of both lay terminology and lay knowledge, may be of interest in consumer health applications.

Effective access to biomedical information depends on reliable representation of the knowledge contained in text. The **Semantic Knowledge Representation** project develops programs that extract usable semantic information from biomedical text by building on existing resources, including the UMLS Metathesaurus, the Semantic Network, and the SPECIALIST lexical tools. Two programs in particular, MetaMap and SemRep, have been developed and are being enhanced and applied to a variety of problems in biomedical informatics. MetaMap maps noun phrases in free text to concepts in the UMLS Metathesaurus, while SemRep uses the Semantic Network to determine the relationship asserted between those concepts.

During the past year, the MetaMap Technology Transfer program (an exportable, Java-based version of MetaMap that runs under Windows or Unix/Linux) was released to the informatics community. A bug-tracking system is included to ensure that problems reported by users are addressed. SemRep was applied to the task of extracting semantic relationships regarding diagnosis and treatment from gastrointestinal endoscopy reports. SemRep was also used in research aimed at extracting molecular biology information from the research literature. One such project seeks to identify protein similarity based on functional interactions, while another extracts information supporting investigations into the genetic basis of disease.

Current research focuses on evaluating the accuracy and effectiveness of MetaMap and SemRep programs. Algorithms are being devised in MetaMap for accommodating higher-level tokens, which are semantically based groupings of lower-level lexical tokens and include mathematical formulas, bibliographic references, and locally defined acronyms and abbreviations. Effective handling of these phenomena will enhance the accuracy of MetaMap processing and the programs it supports. Other research is aimed at automatically illustrating the semantic content of anatomically oriented text. A pilot project uses our resources and an anatomical meronomy to suggest, for example, that an image of the heart highlighting the right side of that organ would be an appropriate illustration for text discussing the tricuspid valve.

The **SPECIALIST lexicon** is a large syntactic lexicon of medical and general English, and new lexical items are continually added using a lexicon-building tool developed and maintained by the group. The lexicon is released annually with the UMLS Knowledge Sources. Lexical access tools, including LVG, wordind, and norm, are also distributed with the UMLS, and a pure Java version of these tools, which is platform independent and easier to maintain, will be included with the 2002 release. Documentation and other educational materials have been revised and enhanced. The lexicon records the spelling variation inherent in English orthography; however, it cannot directly correct spelling errors. An effort is under way to investigate spelling suggestion techniques for use in terminology servers, and the most effective of these are being incorporated into the lexical access tools. Structural chemical terms pose a particular challenge to lexical tools because they do not have the characteristics of ordinary English terms. The Lexical Systems team is currently removing chemical names from the SPECIALIST lexicon using previously developed chemical identification tools under

human review. Terms removed are retained in a separate database.

Many of the most exciting discoveries in medicine are being made as investigators begin to understand the molecular basis of a host of diseases. Making the links between diseases (phenotypes) and the genes (genotypes) that trigger them is of great interest to researchers and patients alike. In the **Biomedical Knowledge Discovery** project we are exploring the information that already exists at NLM's National Center for Biotechnology Information, as well as at other sites, with the goal of developing a system that makes the link between the phenotype and the genotype. Examples of questions that such a system might answer, given a particular disease of interest, are shown below.

- What gene causes this disease?
- Is there a DNA test for this gene?
- Are there clinical trials for this disease?
- What is the function of this gene?
- What mutations have been found in this gene?
- On which chromosome is this gene located?
- Is this gene associated with any other conditions?

Preliminary investigations have shown that there is a need for standard naming of concepts, for new methods for indexing and annotating the data, and for improved algorithms for extracting knowledge. We will explore a variety of approaches to mining data in genetics databases, including enhancing the UMLS knowledge sources for this domain.

WebMIRS Project

The **WebMIRS Project** addresses fundamental issues in the handling, organization, storage, access and transmission of very large electronic files in general and digitized x-rays in particular. A special focus is research into these topics as applied to heterogeneous multimedia databases consisting of both images and text. This work has evolved from a previous project named DXPNET, conducted in collaboration with two other agencies, the National Center for Health Statistics and the National Institute of Arthritis, Musculoskeletal and Skin Diseases.

The web-based Medical Information Retrieval System (WebMIRS) is a Java applet that allows remote users to access data from two surveys conducted by the National Center for Health Statistics: the second and third National Health and Nutrition Examination Surveys (NHANES II and III), carried out during the years 1976–1980 and 1988–1994, respectively. The NHANES II database accessible through WebMIRS contains records for about 20,000 individuals, with about 2,000 fields per record; the NHANES III database contains records for about 30,000 individuals, with more than 3,000 fields per record. In addition, a user query may retrieve any of the 17,000 x-ray images collected in NHANES II, and display it in low-resolution form.

This year vertebral boundary data was added to the WebMIRS NHANES II database and made available for public use. The vertebral boundary data, produced by a board-certified radiologist for 550 of the 17,000 x-ray images in WebMIRS, consists of (x,y) coordinates for approximately 20,000 points on the vertebral boundaries in the cervical and lumbar spine images.

WebMIRS allows a user to control a graphical user interface to construct a query of the NHANES II or NHANES III data. A sample query might be equivalent to the English statements

Findrecordsforallindividualswhoreportedchronic backpain.Returntheirage,sex,race,agewhenthepainbegan,andlongestdurationofpain.Also,returntherecorddatarequiredforstatisticalanalysisanddisplaytheirx-rayimages.

WebMIRS allows the user to save the returned data to the local disk drive, where it may be analyzed with appropriate statistical tools such as the commercially available SAS and SUDAAN software. In effect, WebMIRS goes beyond data access and retrieval to data analysis. Beta testing began this year and is ongoing, with testers not only in the United States, but also in Korea, Sweden, and Mexico. WebMIRS was used in two semesters of a graduate course in public health statistics at Columbia University in 1999–2000 to

demonstrate new technological data access methods, and a real time data acquisition and analysis was demonstrated using WebMIRS/SAS/SUDAAN at the CDC Data Users Conference in Bethesda in July 2000

The Digital Atlas of the Spine is a dataset of cervical and lumbar spine images with interpretations validated by a consensus of medical experts, along with software to display and manipulate the images. The images in the Atlas were chosen from the 17,000 images in the NHANES II survey. We convened two workshops in collaboration with other NIH researchers to seek expert advice and consensus on a wide set of technical and biomedical issues related to the radiological interpretation of this set of images. Among the issues covered were the exact features to be interpreted. The selection of features, based on the consensus of experts at the workshop, took into account published studies relating to the likelihood of obtaining consistent readings for the features considered. The features identified by the workshop as consistently readable were those chosen for the Atlas. Version 2.0 of the Atlas is now being distributed for beta testing.

Goal 2: Promote Use of Health Information by Health Professionals and the Public

Increase Awareness and Use of NLM Services among Health Professionals

HSTAT (Health Services/Technology Assessment Text)

The **HSTAT** system is being used to create a model for technology transfer of a system developed through the Lister Hill Center R&D process to production status in NLM's operations division. A transfer plan was developed and discussed with NLM's Office of Computers and Communications Systems and NLM's National Information Center on Health Services Research and Health Care Technology. The plan is being modified and updated as the transfer progresses. The transfer will involve a new version of HSTAT with enhanced capabilities.

Usability testing performed on the new version of HSTAT provided valuable feedback that resulted in several modifications to the user interface. Testing for compliance with Section 508 of the Americans with Disabilities Act was also conducted, with minor modifications made as a result. Twenty-five new documents were released in HSTAT during the fiscal year, including reports from the Surgeon General that constitute a new collection of information. A major portion of the AHRQ (Agency for Healthcare Research and Quality, formerly the Agency for Health Care Policy and Research) Guideline collection was moved to archived status. The ability to include a search of the National Guideline Clearinghouse when searching HSTAT was also added. Other new features and enhancements include the organization of document titles by subject (in addition to organizing them alphabetically or by sponsoring organization), and the use of software agents to expand queries with terms from the UMLS and to check users' spelling in queries entered.

Office of the Public Health Service Historian

The **Office of the Public Health Service Historian** provides information about the history of Federal efforts devoted to public health, preserves and interprets the history of PHS, and promotes historically oriented activities across the U.S. Department of Health and Human Services, in partnership with the History Office of the Food and Drug Administration and the National Institutes of Health Historical Office.

During this past year the PHS Historian worked with other Center staff to develop a media-enhanced presentation on the history of NLM and the origins of the Lister Hill Center, for the "Getting to Know NLM" series. The Office was involved in the development of exhibits this year on "The Public Health Service Half a Century Ago" and "A 100-Year Quest for Health in the Americas (1902-2002)," and also began planning an exhibit on the history of the PHS Commissioned Corps. The Office also began to work with other NLM units on projects involving digitizing Surgeons General Reports from 1964 to the present. The Office contributed significantly to a project of the NIH History Office involving the construction of a database

of past and present NIH employees. The PHS Historian has been serving as an advisor to the Save Ellis Island foundation in its efforts to restore the historic PHS hospital buildings on the Island. The Office continued to answer numerous queries on PHS history from both within and outside the Federal Government, as well as continuing its efforts to preserve documents and artifacts related to PHS History.

Just-in-Time Information

The **Just-In-Time** (JIT) project is an attempt to build upon NLM's biomedical information databases and construct a real-time, Internet-based information system that provides succinct, highly relevant information to clinicians at the point of care. It will incorporate the literature found in MEDLINE with NLM databases that contain clinical guidelines and ongoing clinical trials. The components of the JIT research agenda include study of the structure of physician questions, improving database search strategies, and developing appropriate ranking hierarchies for medical information.

The project is currently in the process of modeling questions of clinicians. In a collaborative process with several academic medical centers, a database of clinician questions has been constructed. These questions come from actual clinician encounters and have been categorized with a novel taxonomy to facilitate future analysis and querying. Flexibility has been incorporated into this design so that other researchers will find this database a transparent and impartial repository of clinician questions. As more questions are added to the database, research will center upon testing the applicability of previously developed generic questions to real-world use. In an effort to increase the relevance of the search results, a dynamic real-time ranking program is being devised. This algorithm is critical to ensure that clinicians are not overwhelmed with information but rather have access to highly relevant information. The combination of this ranking algorithm with the knowledge developed by researching physician questions, developing generic queries, and constructing unique "hedges," offers an opportunity to build a robust JIT system.

Proteus Project

With the goal of developing a system for medical decision making, data entry and data storage in a clinical setting, the **Proteus project** investigates system architecture for using medical knowledge in the form of executable distributed components to construct clinical protocols and thereby to represent the clinical process. In this approach, called Proteus (PROTocols Editable by USers), clinical processes are represented by three types of "knowledge components": actions, processes, and events. Each such knowledge component has a mechanism to infer its own value and to determine the next action to be launched. One benefit of the clinical knowledge components is that new uses, which depend on clinical semantics, can be incorporated with relatively little effort. To demonstrate this aspect of the Proteus approach, some just-in-time features were introduced. If the user selects any transaction knowledge component, a window opens and shows in a tree structure all the possible questions pertaining to the situation represented by the knowledge component, organized into different categories. When the user selects the questions of interest and clicks on the "answer" button, a browser is opened with PubMed responses to a query string representing the question.

Increase Awareness and Use of NLM Services among the Public

ClinicalTrials.gov

ClinicalTrials.gov is a consumer health informatics application developed by Lister Hill Center staff on behalf of the NIH in response to legislation requiring NIH to create a database of clinical trials information. Increasingly, people are turning to the Internet to look for answers to their health questions, and this raises a number of research questions, including the type of content that should be created and how that content can be put into the appropriate medical context. The structure of the *ClinicalTrials.gov*

application was designed to accommodate these concerns.

ClinicalTrials.gov provides patients, families, and members of the public easy web-based access to extensive information about clinical research studies. An important feature of the system is that it offers links to other online health resources such as MEDLINEplus, which can help place clinical trials in the context of a patient's overall medical care. Currently *ClinicalTrials.gov* contains over 5,700 trials, representing some 62,000 locations, sponsored by the NIH and other Federal agencies as well as the pharmaceutical industry. Studies listed in the database are conducted primarily in the United States and Canada but include locations in approximately 70 countries.

This past year, development work was completed on a robust data entry tool to facilitate the submission of information by the pharmaceutical industry and other data providers. Lister Hill staff worked with the Food and Drug Administration to develop draft guidelines for data submission from the pharmaceutical industry. Ongoing research includes work on new search and browse facilities and an interactive map. In addition, contracts were awarded to four academic health sciences libraries to focus on further development of *ClinicalTrials.gov* training materials and outreach activities.

Terminology Server

The goal of the **Terminology Server project** is to allow biomedical information applications to customize heterogeneous medical vocabularies for various purposes. Such a service is needed to support diverse medical specialties, application domains, and user groups. For example, the terminology server could mediate information access among health consumers and medical professionals. The lack of communication resulting from a misalignment of specialized and technical terms has long been recognized as a problem in medical informatics. This research project seeks to address the problem by providing tools to help client applications bridge disparate vocabularies.

During the past year, the project focused on defining specifications for providing terminology services to clients. Case studies of existing systems, *ClinicalTrials.gov* and *Profiles in Science*, provided reference points for the requirements analysis. For example, the capability to filter vocabularies for specific characteristics will be an important feature of the terminology server. Clients may also wish to limit vocabulary terms by domain or semantic type. Additional filters can be applied as ongoing research results in new techniques. The first filter, one for natural language processing, is based on work comparing UMLS terms with text from MEDLINE citation titles and abstracts. Ongoing research also includes developing a mechanism for maintaining the currency and accuracy of the terminology server as vocabularies evolve over time.

Exhibits

Lister Hill Center staff collaborate with several other NLM divisions, other NIH institutes, and academic centers in the development of exhibits and other educational materials. The **Breath of Life Asthma** traveling exhibition structure was installed in the NLM Visitor Center on October 13, 2000. The Breath of Life Virtual Tour DVD is now available each day as part of the NLM Library Tour. The Virtual Tour DVD was presented at the National Asthma Education and Prevention Program and at the 3rd Triennial World Asthma meeting. Lister Hill Center staff helped prepare QuickTime movies of selected segments of the DVD program for the NLM Director's opening remarks, which accompanied the welcoming remarks of the National Heart, Lung, and Blood Institute (NHLBI) Director. The exhibit was an integral part of the NHLBI presence at the conference and NHLBI distributed 1500 copies of the DVD to conference attendees. Additional copies were delivered to the Chicago Asthma Coalition, which is one of several of the asthma education programs sponsored by the National Asthma Education and Prevention Program. We coordinated with NLM's Office of Communications and Public Liaison to produce a video documentary of the event.

The **Movement Disorders Video Database Project** is a collaborative project with Yale University School of Medicine's

Movement Disorders and Neurodegenerative Diseases Clinic, the Center for Advanced Instructional Media and the Biomedical Communications Department. This pilot effort established a digital video database of high-quality, full-motion video of medical significance. Neurologically based movement disorders were selected as subject matter which would be best characterized by video and audio. The video database of patients with a variety of clinically diagnosed movement disorders is undergoing updated editing and compression processes to capitalize on advances in digitization and compression schemes. This is the first step in a larger, ongoing effort to investigate the preparation of high quality, compressed video for distribution on the World Wide Web, and the delivery methodology of a medically important multimedia database

In mid March, NLM unveiled the **Turning the Pages** interactive exhibit featuring *A Curious Herbal*, written and illustrated by Elizabeth Blackwell in 1737–39. The Turning the Pages computer program simulates the turning of the pages of the digitized volume on the touch sensitive screen as well as the capability to zoom in for close-ups and hear audio commentary. The event in the NLM Visitor Center was video teleconferenced to the British Library in London and live coverage of the companion ceremony at the British Library was simultaneously fed to the audience in the NLM Visitor Center. Prior to the actual opening, a Video News Release featuring a complete edited and scripted news report had been prepared for satellite transmission to television stations throughout the United States and a DVD featuring the Curious Herbal video news release was also produced for use at the NLM exhibit at the Medical Library Association annual meeting.

The History of Medicine exhibit, **The Once and Future Web**, opened during the third week in May. Immediately prior to the opening, a program including an original play written especially for the opening was held in the Lister Hill Auditorium. The play covered the development of communications from the invention of the optical telegraph in the 18th century, tracing developments such as the electric telegraph of Samuel F. B. Morse to the Internet. Lister Hill Center staff prepared an edited composite of the event for the NLM archives. Additionally, to supplement a portion of the exhibit, the taxidermied Siberian Husky, Balto, was loaned to the exhibit by the Cleveland Museum of Natural History. Balto was the lead dog in the team of dogs that delivered life-saving diphtheria vaccine to the ice-bound city of Nome, Alaska in 1925. The vaccine had been requested by telegraph to help stem the epidemic. Rare 1925 newsreel film that featured Balto's team arriving in Nome was acquired and edited by Lister Hill Center staff. The newsreel film was mastered onto a DVD and is being used in a kiosk adjacent to Balto in the NLM lobby.

Lister Hill Center staff completed a series of videos for the Public Services Division designed to assist NLM patrons onsite and at home with detailed information ranging from directions to NLM from within the metro area to the specifics of accessing various resources of the NLM once here. We subsequently designed a Web page that delivered these videos. It will require frequent updating to reflect the physical and procedural changes that occur over time.

Goal 3: Strengthen the Informatics Infrastructure for Biomedicine and Health

Encourage Health Applications for Current and Future Internet Environments

Next Generation Internet

NLM is working to define **Next Generation Internet** (NGI) capabilities that will allow the NGI to be used routinely in health care, public health and health education, as well as biomedical, clinical and health services research. These capabilities include: quality of service, security and medical data privacy, nomadic computing, network management, and infrastructure technology as a means for collaboration.

We are supporting 15 NGI projects designed to improve our understanding of the impact of NGI technology on the nation's health care, health education, and health research systems in such areas as cost, quality, usability, efficacy and security.

We assisted the Uniformed Services University and its Medical Simulation Center in connecting to the Abilene network. NLM contracted for the dark fiber to connect the three institutions and arranged for connectivity to Abilene through the router at NLM. Test systems installed and used in-house included: Multi-Router Traffic Grapher which monitors the traffic load on network links, and generates HTML pages containing GIF live visual representations of this data; Iperf network performance measuring tool. For a cross country test for the Visible Embryo project, NetIQ's Qcheck software was employed to measure memory-to-memory tests between the Armed Forces Institute of Pathology through NLM to the San Diego Supercomputing Center.

Work on the Lister Hill Center network has continued with the development of a gigabit backbone. The existing Cisco Catalyst switches will be replaced by Extreme switches with significantly larger bandwidth capacity. The Extreme switches can handle gigabit connections to the desktop. These switches will be connected to two core gigabit switches (Extreme Black Diamond) that will provide a redundant connection between the local switches, the NGI networks, and the Internet. The result will include fully redundant paths from NLM to the Internet. Last year we connected to two NGI networks, vBNS (very high speed Backbone Network Services) and Abilene. The current connections are to Abilene and the Federal NGI network DREN, the Department of Defense Research Network. Connection to the NASA Research Network (NREN) is expected next year. The NGI networks are being used for multimedia applications involving voice and video. The Abilene network supports full IP (Internet Protocol) multicast. We use that mode to receive and transmit multicast voice and video sessions.

In an effort to increase bandwidth from the current shared 256 Kbps satellite channel among the malaria research sites in Africa, engineering staff participated in reviewing a technical proposal from Intelsat. In addition, to demonstrate video quality that may be expected unless bandwidth is increased, tests were conducted with an ISDN gateway in London on simulating a 2-hop satellite link from Africa. Bandwidths used were 128 and 256 kbps, and video quality was found to be marginal. We conducted a review on teleconferencing services for this project, and staff contributed to an NSF proposal by the National Center for Supercomputing Applications to create a center in Nairobi that is similar to others in Kenya. Specifications were generated for VCON Cruiser 384, a PC-based teleconferencing unit (384 Kbps, H.320, H.323 quality and operation).

Experiments with video conferencing and collaboration tools are being conducted with NASA, Trinity University in Dublin, Ireland, and Johnson and Johnson. An experiment in remote conferencing was completed that combined multipoint videoconferencing and streaming technology. The conference was webcast live to sites recruited from the American Association of Medical Colleges Med-Ed mailing list. Another interactive demonstration was presented over the Abilene Network from the Lister Hill Center to the Internet 2 Semiannual Conference in Washington, DC. A prototype streaming patient simulation was created working with the Simulation Center at the Uniformed Services University. Live webcasts of selected meetings of the Washington Area Computed Assisted Surgery special interest group meetings were inaugurated this past year.

In 2001, the Lister Hill Center continued to serve as a Federal representative to the Maryland Governor's Task Force on High Speed Networks and the Engineering Advisory Group. The Task Force developed a comprehensive plan for bringing the state's network infrastructure in line with the needs of the 21st century. This plan, completed and presented to the legislature, contains recommendations to combine existing state resources to maximize the state's return on investment; use existing state-owned fiber where available; and use current right-of-ways the state possesses to add additional fiber in underserved regions such as the Eastern Shore, Western and Southern Maryland. The plan also provides for equity of access to all regions of the state, and support multiple segments of our society and promotes collaboration among businesses, educational institutions, governmental bodies and research institutions. The project intends to conduct a select number of high priority pilot

projects in health care, business infrastructure development, and state government functions. A major contribution by the Lister Hill Center was made in the development of pilot projects in health care involving remote oncology treatment planning and remote intensive care support.

Telemedicine

The **Telemedicine program** was designed to evaluate the impact of advanced networking on health care, research, and public health and 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. The program also assesses the utility of emerging health data standards in health applications of advanced communications and computing technologies.

As a means of evaluating the results of the telemedicine initiatives begun in 1996 and concluded in 2000, the Lister Hill Center conducted a two-day symposium titled "Telemedicine and Telecommunications: Options for the New Century." Representatives from 19 funded telemedicine projects discussed the results of their work with an emphasis on lessons learned. Conference proceedings, including contract final reports, have been posted on the World Wide Web.

The Telemedicine Information Exchange is a web-based resource of information about telemedicine maintained by the Telemedicine Research Center, Portland, Oregon, and funded in part by NLM. During this past year approximately 5,000 non-NLM bibliographic citations, and 131 HSRPROJ-type records were received at NLM.

Ubiquitous Computing

This year we will begin a new project in **Ubiquitous Computing**. Embedded intelligence in smaller, handier forms closer to the point of use is becoming increasingly widespread. Ubiquitous computing includes wireless networking, speech technology, personal digital assistants (PDAs), radio tags, and eye-tracking technology. The first phase of this project will investigate PDAs and speech technology. One of the greatest constraints on PDAs is their input interface. These devices are too small to allow a reasonable keyboard-like interface, and the handwriting recognition they use is relatively slow and ineffective. This project will investigate the use of speech-driven interfaces to selected NLM resources. The research will be undertaken initially on desk-top units, though the output could be created using a PDA or WML simulator, to realistically constrain the visual output possible. During this past year, project staff began to develop and investigate tools for enhancing collaborative biomedical computing, and explored a variety of tools including the newly released DARPA-funded open source speaker-independent continuous speech recognition engine, sphinx 2.0.

Smart Cards

We explored several applications of **smart card technology** this past year. A smart card is a credit-card-sized plastic card with an embedded circuit chip. The chip can be a microprocessor with internal memory capable of running small programs, or simply a non-programmable memory chip. The cards can be used both for authentication and for data storage. Recent applications sometimes involve biometrics, the storage of information such as a thumbprint or an iris scan for more positive authentication than is possible with just a password. For several years we have co-sponsored the Western Governors' Association Health Passport Project, one of the largest health-oriented smart card pilot programs in this country. This project involves the storage of data from multiple Federal, state and local agencies on cards used by clients receiving benefits such as well child care, checkups, immunizations and food benefits. The mother and each child have individual cards. Health Passport cards are currently in use by 12,000 clients in three western states. Kiosks in public places allow clients to check and print information from the card.

Further Training in Medical Informatics and Librarianship

Medical Informatics Training Program

The **Medical Informatics Training Program** (MITP) provides training for students at various stages in their careers and brings talented people to the Lister Hill Center. The NLM believes that providing training benefits both students and Center scientists. The MITP recruits talented, promising students into careers in medical informatics, playing a role in developing researchers and leaders for the field. This past year, we provided training to 43 participants from 15 states and 9 countries. The participants included two high school students and teachers, 13 undergraduate students, 13 graduate or medical students, 10 postdoctoral or post-MD fellows, and five visiting faculty scholars. Students during the year worked on projects in the following areas: biomedical knowledge discovery, the clinical trials project, database systems, digital library research, image database research, information retrieval research, just-in-time medical information, knowledge based systems, natural language processing, palm technology, telemedicine, document processing and analysis, UMLS research, visualization research and Web design.

We continue to support the NIH Clinical Elective in Medical Informatics for third and fourth year medical students in March and April and continue to participate in programs supporting minority students including the Hispanic Association of Colleges and Universities and the National Association for Equal Opportunity in Higher Education summer Internship programs.

This year, we initiated a rotation for NLM Medical Informatics Trainees to provide an opportunity for fellows to learn about NLM programs and about research being conducted at the Lister Hill Center. The rotation includes a series of lectures and an opportunity for students to work closely with established scientists conducting research at the Center. The program provides participants with an opportunity to meet fellows from other NLM-funded programs and could lead to possible future collaborations with our research staff or with researchers in other NLM Training Grant Programs. This summer rotation was held at NLM in June and July with five students participating.

Lister Hill Center Organizational Structure

Lister Hill Center research is conducted by drawing on a diverse set of scientific fields and methods. Researchers have backgrounds in medicine, computer science, library and information science, linguistics, engineering, and education. The Center's research activities are regularly reviewed by an outside advisory group, the Board of Scientific Counselors, whose members are drawn from the medical informatics community (see Appendix 3).

The Center is organized into five components, together with a number of research laboratories shared by all components. Many research projects involve collaborations across organizational units. Each component has its own Web site listed below, but may also be reached through the Lister Hill Center's main Web page at http://lhncbc.nlm.nih.gov/.

The **Audiovisual Program Development Branch** conducts media development activities with three specific objectives. As its most significant effort, the branch supports the Center's research, development, and demonstration projects with high-quality video, audio, imaging, and graphics materials. From initial project concept through final project implementation and evaluation, a variety of forms and formats of visual materials are supported and staff activities include content creation, editing, enhancement, transfer and display. 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 expanded to include effective outreach activities, the range and quantity of support that the branch provides to these programs continues to increase. From applications of optical media technologies and teleconferencing to support for World Wide Web design, the requirement for graphics, video, and audio materials has increased in quantity and diversified in format. The third area of concentration is the engineering of technical improvements applied to media issues such as image quality and resolution, color fidelity, transportability, storage, and visual information communication. In addition to the development

of new methods and processes, the facilities and hardware infrastructure must reflect state-of-the-art standards in a very rapidly changing field. Current information about Audiovisual Program Development Branch activities appears at http://lhncbc.nlm.nih.gov/apdb/.

The **Cognitive Science Branch** conducts research and development in information systems informed by research in the mechanisms underlying human cognition. This involves the investigation of a variety of techniques, including linguistic, statistical, and knowledge-based methods, for improving access to biomedical information. Branch staff have developed SPECIALIST, an experimental natural language processing system for the biomedical domain. The SPECIALIST system includes several modules based on the major components of natural language: the lexicon, morphology, syntax, and semantics. The lexicon and morphological component are concerned with the structure of words and the rules of word formation. The syntactic component treats the constituent structure of phrases and sentences, while the semantic component seeks to extract biomedical content from text. Branch members actively participate in the Unified Medical Language System project and lead NLM's Indexing Initiative, whose goal is to develop automated and semi-automated techniques for indexing the biomedical literature. The Branch conducts research in digital libraries and collaborates with NLM's History of Medicine Division on *Profiles in Science*, a project to digitize collections of prominent biomedical scientists. Several Branch projects address the challenges involved in providing health information to consumers. Branch staff developed and continue to enhance ClinicalTrials.gov on behalf of the NIH. Current information about Cognitive Science Branch activities appears at http://lhncbc.nlm.nih.gov/cgsb/.

The focus of the **Communications Engineering Branch** is applied research and development in image engineering and communications engineering motivated by NLM's mission-critical tasks such as document delivery, archiving, automated data entry for the creation of MEDLINE records, Internet access to biomedical multimedia databases, and imaging applications in support of medical educational packages employing digitized radiographic, anatomic, and other imagery. Areas of active investigation center on document image analysis and understanding techniques, image compression, image enhancement, image feature identification and extraction, image segmentation toward query by image content research, image transmission and video conferencing over networks implemented via asynchronous transfer mode and satellite technologies, optical character recognition and man-machine interface design applied to automated data entry. The Branch also maintains a database of large numbers of digitized spine x-rays and bit-mapped document images that are used for intramural and collaborative research projects. The Branch hosted the 14th Annual IEEE Symposium on Computer-Based Medical Systems in July 2001. Ninety peer-reviewed papers were presented, five of them by Branch staff. In addition, special sessions on receiver operator characteristics analysis and NIH grants funding were included. This symposium was planned in cooperation with faculty at Texas Tech University, University of Connecticut, and Mt. Sinai Hospital, among others. Current information about Communications Engineering Branch activities appears at http://lhncbc.nlm.nih.gov/ceb/.

The **Computer Science Branch** applies techniques of computer science and information science to problems in the representation, retrieval and manipulation of biomedical knowledge. Branch projects involve both basic and applied research in such areas as intelligent gateway systems for simultaneous searching in multiple databases, intelligent agent technology, knowledge management, the merging of thesauri and controlled vocabularies, data mining, and machine-assisted indexing for information classification and retrieval. Research issues include knowledge representation, knowledge base structure, knowledge acquisition, and the human-machine interface for complex systems. Important components of the research include embedded intelligence systems that combine local reasoning with access to large-scale online databanks. Branch staff include the teams that developed NLM's Gateway, Internet Grateful Med and HSTAT (Health Services/Technology

Assessment Text) programs and the team that annually produces the UMLS Metathesaurus. Staff members participate actively in the medical informatics and information science research communities and other professional specialty societies. They participate in the meetings of the Internet Engineering Task Force. Branch staff coordinate a variety of training programs, including the eight-week NIH elective in medical informatics for third- and fourth-year medical students held each spring. Current information about Computer Science Branch activities appears at http://lhncbc.nlm.nih.gov/csb/.

The **Office of High Performance Computing and Communications** serves as the focal point for NLM's High Performance Computing and Communications planning and research and development activities with Federal, industrial, academic, and commercial organizations. The major activities of the office include NLM's Visible Human project, the Telemedicine Program, the Next Generation Internet, the Collaboratory for High Performance Computing and Communications, and imaging research. Staff presented tutorials on Internet technologies at the annual meeting of the Radiological Society of North America. Staff members also helped organize and present tutorials and workshops for volume graphics at the IEEE Visualization 2000 conference, and the SIGGRAPH 2001 conference on computer graphics. The office continued its sponsorship of the bimonthly meetings of the Washington Area Computer Assisted Surgery Special Interest Group. Staff members also participate in the Large Scale Networking Committee and the Joint Engineering Task Force of the interagency Information Technology Research and Development program, as well as the multi-agency Joint Telemedicine Working Group. OHPCC staff have testified before the President's Information Technology Advisory Committee on matters of the need for high speed networking by the healthcare community. Current information about the Office of High Performance Computing and Communication activities appears at http://lhncbc.nlm.nih.gov/ohpcc/.

Lister Hill Center Laboratories

The **Document Imaging Laboratory** supports DocView, MARS and other research and design projects involving document imaging. Housed in this laboratory are advanced systems to electro-optically capture the digital images of documents and subsystems to perform image enhancement, segmentation, compression, OCR and storage on high density magnetic and optical disk media. The laboratory also includes high-end Pentium-class workstations running under Windows 2000, all connected by 100 Mb/s Ethernet, for performing document image processing. Both in-house developed and commercial systems are integrated and configured to serve as laboratory testbeds to support research into automated document delivery, document archiving, and techniques for image enhancement, manipulation, portrait vs. landscape mode detection, skew detection, segmentation, compression for high density storage and high speed transmission, omnifont text recognition, and related areas.

The **Document Image Analysis Test Facility** is an off-campus facility that houses high-end Pentium workstations and servers that constitute MARS-1 and MARS-2 production systems. While routinely used to produce bibliographic citations for MEDLINE, this facility also serves as a laboratory for research into techniques for autozoning, autolabeling, autoreformatting, intelligent spellcheck and other key elements of MARS. Besides real-time performance data, also collected and archived are large numbers of bitmapped document images, zoned images, labeled zones, and corresponding OCR output data. This collection serves as ground truth data for research in document image analysis and understanding.

The **Image Processing Laboratory** is equipped with a variety of high end servers, workstations and storage devices connected by 100 Mb/s Ethernet. Most machines are equipped with multiple networking ports (FDDI, ATM, Ethernet, fast Ethernet) which allow, in addition to standard networking capabilities on the local Ethernet, the capability of alternate physical communications channels with these machines.

This capability has been used in communications engineering experiments for point-to-point satellite channels connecting these machines with remote sites. ATM switches connect the Ethernet and FDDI networks to other local area networks throughout the building, to the Internet, and to experimental ATM networks such as ATDnet and MCI's research network, in addition to vBNS, the infrastructure for the Next Generation Internet and Internet 2 initiatives. The Image Processing Laboratory supports the investigation of image processing techniques for both grayscale and color biomedical imagery at high resolution. In addition to computer and communications resources and image processing equipment to capture, process, transmit and display such high-resolution digital images, the laboratory also has a variety of image content.

The Collaboratory for High Performance Computing and Communications (Collab) was established to investigate innovative means for assisting health science institutions in their use of online distance learning technologies, to explore Next Generation Internet technologies for distance interactivity, virtual reality research, and imaging technology. A major upgrade in AC power to the Collab was completed in order to support the technologies being investigated. Innovative means for assisting health science institutions in their use of online distance learning technologies continued to be explored. The Collab Web server was put online, as were streaming video servers and multipoint video conferencing servers. Through collaborations with colleagues at the University of Utah, UCLA, and the University of Oklahoma, the EtherMed database of Web accessible health professions educational materials was expanded. The University of Alabama at Birmingham began a research collaboration with the NLM using the database.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

David Lipman, M.D.
Director

The National Center for Biotechnology Information (NCBI), established in November 1988 by Public Law 100-607, is a division of the National Library of Medicine. The establishment of the NCBI by Congress reflected the important role information science and computer technology play in helping to elucidate and understand the molecular processes that control health and disease. Since the Center's inception in 1988, NCBI has established itself as a leading resource, both nationally and internationally, for molecular biology information.

NCBI is charged with providing access to public data and analysis tools for studying molecular biology information. Over the past 13 years, the ability to integrate vast amounts of complex and diverse biological information created a new scientific discipline—bioinformatics. It is now almost impossible to think of an experimental strategy in biomedicine that does not involve some dependence on bioinformatics. At the core of this shift is the recent flood of genomic data, most notably gene sequence and mapping information. As NCBI enters into the new millennium, the horizon is ever-expanding—an explosion of scientific data that must be collected, organized, stored, analyzed, and disseminated. Through the next decade and beyond, NCBI will meet this challenge by designing, developing, and distributing the tools, databases and technologies that will enable the gene discoveries of the 21st century.

The Center meets these goals by:

- Creating automated systems for storing and analyzing information about molecular biology and genetics;
- Performing research into advanced methods of computer-based information processing for analyzing the structure and function of biologically important molecules and compounds;
- Facilitating the use of databases and software by researchers and health care personnel; and,
- Coordinating efforts to gather biotechnology information worldwide.

NCBI supports a multidisciplinary staff of senior scientists, postdoctoral fellows, and support personnel. NCBI scientists have backgrounds in medicine, molecular biology, biochemistry, genetics, biophysics, structural biology, computer and information science, and mathematics. These multidisciplinary researchers conduct studies in computational biology as well as the application of this research to the development of public information resources.

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 designed to facilitate scientific discovery.

GenBank—The NIH Sequence Database

GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. NCBI is responsible for all phases of GenBank production, support, and distribution, including timely and accurate processing of sequence records and biological review of both new sequence entries and updates to existing entries. Integrated retrieval tools have been built to search the sequence data housed in GenBank and to link the results of a search to other related sequences, as well as to bibliographic citations. Such features allow GenBank to serve as a critical research tool in the analysis and discovery of gene function. In FY2001, approximately 2 million sequences were added to GenBank, and the base count rose from 9.5 billion in August 2000 to 13.5 billion in August 2001. This rate of growth far exceeds estimated projections, and was fueled by several

genome sequencing projects and the automatic submission of large-scale batched data into GenBank.

Another important source of data for GenBank is direct sequence submissions from individual scientists. NCBI produces GenBank from thousands of sequence records submitted directly from researchers prior to publication. Records submitted to NCBI's international collaborators, EMBL (European Molecular Biology Laboratory) at Hinxton Hall, UK and DDBJ (DNA Data Bank of Japan) at Mishima, are shared through an automated system of daily updates. Other cooperative arrangements, such as with the U.S. Patent and Trademark Office for sequences from issued patents, augment the data collection effort and ensure the comprehensiveness of the database. Sequence data submitted in advance of publication is maintained as confidential, if requested.

When scientists submit their sequence data to GenBank, they receive an "accession number." This number serves as a tracking device and allows the scientist to reference the sequence in a subsequent journal article. In eight years of processing direct submissions, NCBI has issued over 560,000 accession numbers, with approximately 28% of these assigned in FY2001. There are now over 464,000 direct submission accession numbers that are publicly available and approximately 35,000 accession numbers pending release.

GenBank indexers with specialized training in molecular biology create the GenBank records and apply rigorous quality control procedures to the data. NCBI taxonomists consult on taxonomic issues, and, as a final step, senior NCBI scientists review the records for accuracy of biological information. Improving the biological accuracy of submitted data as well as updating 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 World Wide Web.

NCBI is continuously developing new tools, and enhancing existing ones, to improve access to, and the utility of, the enormous amount of data stored in GenBank. Sequence data, protein as well as DNA, is supplemented by pointers to the corresponding MEDLINE bibliographic information, including abstracts and publishers' full-text documents. GenBank provides links to textbooks, as well as outside sources, when direct links to publishers are not available. This latter service, called LinkOut, also points to other external resources that may be useful in data analysis, such as biological databases and sequencing centers. The availability of such links allows GenBank to serve as a key component in an integrated database system that offers researchers the capability to perform comprehensive and seamless searching across all available data.

GenBank has evolved to contain several types of DNA sequences, from relatively short Expressed Sequence Tags (ESTs) to assembled genomic sequences that are several hundred kilobases in length. EST data obtained through cDNA sequencing are critical to understanding gene function and therefore continue to be heavily represented in GenBank. As such, additional annotation is available for these sequences as part of a separate EST database (dbEST). NCBI continued to expand dbEST throughout the year. As of October 2001 there were 9,283,262 public EST entries stored in dbEST.

Another rapidly increasing segment of GenBank is the GSS (Genome Survey Sequences) division. The GSS division of GenBank is similar to the EST division, except that its sequences are genomic in origin, rather than cDNA. Additional data on each sequence is stored in a separate database (dbGSS) and includes detailed information about the contributors, experimental conditions, and genetic map locations. Currently, over 2,747,000 public records are stored in the dbGSS.

The STS (Sequence Tagged Site) division of GenBank also experienced significant growth in the past year. Sequence tagged sites are short sequences that are operationally unique in the genome and used to generate mapping reagents. The recently created UniSTS database reflects an expansion of the contents and information provided in the general dbSTS record and reports information about markers collected from public resources. Each marker report contains primer information, mapping data, and cross-references to other

NCBI resources, such as Map Viewer and LocusLink.

The whole genomes of over 800 organisms can now be found in Entrez Genomes. The genomes represent both completely sequenced organisms and those for which sequencing is in progress. All three main domains of life—bacteria, archaea, and eukaryote—are represented, as well as many viruses and mitochondria. New organisms added in FY2001 include: *Escherichia coli O157:H7* strain EDL93 and sub-strain "RIMD 050995", *Pasteurella multocida, Lactococcus lactis*, *Mesorhizobium leprae* strain TN, *Caulobacter crescentus, Thermoplasma volcanium, Streptococcus pyogenes* strain M1, *Staphylococcus aureus* strain N315, *Guillardia theta nucleomorph, Sulfolobus sulfataricus*, *Mycobacterium tuberculosis* CDC1551, *Mycoplasma pulmonis,* and many others. The Genomes group also installed *Zea mays* (corn) map data in the Entrez Genomes Map Viewer. Sequencing efforts for additional plants are underway. Sequences from these organisms will provide valuable clues for understanding the functioning of human genes.

The Human Genome

NCBI is responsible for collecting, managing, and analyzing the growing body of human genomic data generated from the sequencing and genome mapping initiatives of the public Human Genome Project. NCBI also plays a key role in assembling and annotating the human genome sequence. For example, NCBI recently released its first assembled view of the human genomic sequence. This assembly is based not only on the finished and draft sequences deposited by the Human Genome sequencing centers in GenBank, but also on sequences contributed to GenBank by individual scientists from around the world. Hence, this resource is truly an international public sequencing effort. Assembling the sequences is an ongoing process that involves many different steps before the data may be merged into segments of contiguous DNA. NCBI continues to improve the genome assembly by incorporating new data, filling in existing gaps, and increasing overall accuracy.

Assembling and Annotating the Human Genome

A team of NCBI scientists is also engaged in annotating, or labeling the biologically important areas of the genome. Annotation permits researchers to analyze the data in a systematic, comprehensive, and consistent manner. There are two tasks involved in annotation. The first is the correct placement of known genes into the proper genomic context and the second is the prediction of previously unknown genes based on the assembled genomic sequence. In the first task, messenger RNAs (mRNA) from the NCBI RefSeq collection—a non-redundant set of reference sequences, including genomic contigs, mRNAs of known genes, and proteins—are placed on the genome primarily by sequence alignment using tools developed at NCBI. Computer modeling is used to compensate for and overcome various problems associated with aligning the genomic and mRNA sequences.

The human genome is also being annotated with additional biological features. Examples include markers for sequence variation such as SNPs, or single nucleotide polymorphisms, and genomic position landmarks such as sequenced tagged sites. These features may be viewed using the NCBI Map Viewer, an online tool that allows you to view an organism's complete genome, as well as integrated maps for each chromosome.

Various computational approaches are also being used by NCBI investigators to accomplish the second task, that of predicting novel genes. Alignment with small snippets of expressed genes called Expressed Sequence Tags (ESTs) identifies new genes to be placed on the DNA sequence and also provides information on alternative gene splicing. Use of protein similarity analyses and gene prediction programs developed at NCBI identifies additional predicted genes.

NCBI Resources Designed to Support Analysis of the Human Genome

With the publication of the "working draft" of the human genome, the research focus is turning from analysis of specific genes or gene regions to whole genomes. To

accommodate this shift in research focus, NCBI has developed a suite of genomic resources to support comprehensive analysis of the human genome, as well as the complete genomes of several model organisms. Specialized tools and databases have also been designed to facilitate the use of this data.

NCBI's new Web page, "The Human Genome, A Guide to Online Information Resources," was released in February 2001 and is designed to serve as a nexus for the collection and storage of diverse data. This online guide provides centralized access to a full range of genome resources, including links to BLAST, dbSNP, LocusLink, RefSeq, Map Viewer, Homology Maps, UniGene, HomoloGene, and GEO. NCBI's Human Genome Sequencing site displays up-to-date information on sequencing efforts and access to various other types of resources, such as chromosome-specific BLAST searches and data relative to specific genomic contigs.

NCBI's Map Viewer provides graphical displays of features on NCBI's assembly of human genomic sequence data as well as cytogenetic, genetic, physical, and radiation hybrid maps. Map features that can be seen along the sequence include NCBI contigs (the "Contig" map), the BAC tiling path (the "GenBank" map), and the location of genes, STSs, FISH mapped clones, ESTs, GenomeScan models, and sequence variation. You can find genes or markers of interest by submitting a query against the whole genome, or by querying a chromosome at a time. Results are indicated both graphically and in a tabular format.

In FY2001, NCBI released multiple versions of the Map Viewer. These build increased functionality for users and improved query response time. The capability to view more connections between objects on the maps and between new maps was added. Users are now able to create a report of mapped objects from what is displayed on the screen. Special documentation accompanies the release of each new version and serves to report changes in Map View displays or modifications in algorithms used to make the assembly and its annotation.

New in September 2001 was NCBI's Human-Mouse Homology Map Web page. From this site one may navigate between the human and mouse genomes using NCBI's new FLASH homology browser. Links to numerous mapping resources as well as a view of various sequence alignments is also provided.

Also of interest to the scientific and academic communities is the Gene Map Web page. From Gene Map, one can display a gene map of the human genome generated by the International RH Mapping Consortium. This map includes the locations of more than 30,000 genes and provides an early glimpse of some of the most important pieces of the genome. Even more important, the map can be immediately applied by scientists to the identification and isolation of genes that either directly cause human ailments or increase our susceptibility to disease.

The Genes and Disease Web page is designed to educate the lay public and students on how sequencing of the human genome will lead to the identification of disease-causing genes; how these genes are inherited and cause disease; and, most important, how an understanding of the human genome will contribute to improving diagnosis and treatment of disease. This site was expanded in FY2001 to include a number of additional diseases and now contains descriptions for nearly 70 genetic diseases and provides links to databases and organizations that can supply additional information. For each disease-causing gene there is also a link to the PubMed literature, the Online Mendelian Inheritance in Man database (OMIM), and LocusLink.

OMIM is an electronic version of Dr. Victor McKusick's catalog of human genes and genetic disorders. The database, produced at Johns Hopkins School of Medicine, contains over 13,000 records and usage exceeds 8,600 users per day, up significantly from last year. OMIM was recently integrated into Entrez, NCBI's unique search and retrieval system, which, in turn, is linked to several other databases. This feature resulted in greater flexibility in field searching and increased relevance of retrieved information.

LocusLink, launched in FY1999, is a single-query interface to curated sequence and descriptive information about genes. LocusLink presents information on official nomenclature, aliases, sequence accession numbers,

phenotypes, EC numbers, OMIM numbers, UniGene clusters, map information, and relevant Web resources. LocusLink has rapidly expanded over the past year from 23,000 records to 88,560 records. An array of new LocusLink features include annotation based on title lines from Proteome, Inc., for human genes; gene ontology terms for human, mouse, and *Drosophila* genomes; and domain names from CDD-based analysis of RefSeq proteins. Access to protein-specific information has also been enhanced by BLink (BLAST Link) and an explicit link to Entrez proteins. LocusLink also provides one of the windows into NCBI's annotation of the human genome, with connections to Map Viewer, the graphical sequence viewer. Also, more links to HomoloGene as well as links from mouse or human genes to NCBI's computed Human-Mouse Homology Maps have recently been added.

The Reference Sequence (RefSeq) database, which also began in FY1999, provides a non-redundant set of reference standards for various molecules—from chromosomes to mRNAs to proteins. These standards furnish a foundation for the functional annotation of the human genome and a stable reference point for mutational analysis, gene expression, and polymorphism discovery. The database has grown substantially over the last year and now holds over 24,000 reference sequence records for man, mouse and rat. In addition, there are over 112,000 corresponding RefSeq protein records. The first curated genomic annotations were added to RefSeq in May 2001 and can now be retrieved in LocusLink and Map Viewer.

The most common forms of sequence variations are single nucleotide polymorphisms, or SNPs. There has been an increasing interest in SNPs detection and discovery over the last few years, as they are expected to facilitate large-scale association genetic studies. To accommodate this explosion of data, NCBI, in collaboration with NHGRI, launched the database of single nucleotide polymorphisms (dbSNP) in late FY 1998. To facilitate research efforts, dbSNP links directly to a number of software tools designed to aid in SNP analysis. Each SNP record also contains links to additional NCBI resources, including GenBank, LocusLink, dbSTS, human genome sequencing data, and PubMed.

In FY2001, an XML-based common data exchange format was initiated to integrate dbSNP with other NCBI resources, the Ensemble Annotation Project, the UCSC Genome Assembly, and smaller SNP databases such as HGBASE. The dbSNP group also added several enhancements to the dbSNP Web site, including a batch-query service that provides email-based reports for user-selected subsets of data. Additional new query services were also introduced, including enhanced locus-based queries; queries for SNPs between two STS markers on the human genome; and a free-text Entrez-like query where users can query by gene name, validation status, map location, mapping quality, SNP heterozygosity, functional class, or organism. The complete set of 2.99 million submissions were processed and reduced to a non-redundant set of 1.68 million refSNP clusters. Of this set, 1.41 million were successfully mapped and annotated on the human genome sequence.

The dbSNP sister database, dbHLA, is currently defining the SNPs in all known DRA and DRB alleles as a first step to defining molecular haplotypes for the common human tissue-typing alleles. Additionally, the dbHLA group is working with external collaborators to define reference gene sequences through the HLA region for allele-specific annotation of the reference human genome sequence. The combination of reference HLA alleles and dbSNP mapping functions is currently being used to define HLA serological alleles at the genomic level as sets of molecular haplotypes. These data are being developed as a service to the HLA research community and serve as a prototype for developing common data exchange standards.

From Human to Mouse: Model Organisms for Research

The public mouse sequencing effort has formally begun and is making rapid progress rapidly. The ultimate goals of the project include the construction of a robust physical map and a high quality, finished sequence of the mouse, as these data will provide an essential tool to

identify and study the function of human genes. The mouse genome sequence will also increase the ability of scientists to use the mouse as a model system to study and understand human disease.

All sequence data generated from this project are rapidly deposited in GenBank. To date, an initial set of whole genome shotgun data, comprising over 17 million reads, has been generated and the data are available from NCBI's Trace Archive database, established in FY2001. The mouse reads are currently being compared to the human genome, and homologous reads have been laid out along the human draft sequence. Mouse data is also being accumulated in both the RefSeq and LocusLink databases and investigators have begun to assemble the data set in order to generate larger contigs. The mouse reads are of immediate use for both human and mouse genetics and there are already examples of mouse genes that have been cloned using the available public information.

The mapping and sequencing of the genomes of all model organisms are critical to the effort to characterize, sequence, and interpret the human genome. Therefore, NCBI is also working towards the development and expansion of resources to facilitate biomedical research using other model organisms, including the rat, *S. cerevisiae* (budding yeast), *C. elegans* (round worm), *D. melanogaster* (fruitfly), and *Ararbidopsis thaliana* (a small flowering plant).

Literature Databases

PubMed is an innovative, Web-based literature retrieval system, based on NLM's MEDLINE database, that contains citations, abstracts, and indexing terms for journal articles in the biomedical sciences. It also includes URLs to full-text articles from the publishers' Web site. Early last year, a new version of PubMed was released that incorporated many new capabilities requested by the medical librarian community. At this time, functions were added or improved for limiting queries by common search filters. For example, the new version has a pull-down menu that displays search field limits, indexes, search history, and a clipboard for gathering selected articles. Context-specific help and a "frequently asked questions" section provide guidance in making the transition to the new system.

PubMed was recently assigned an additional easy-to-remember URL: pubmed.gov. Recent changes were also made to the links from PubMed. A new link to the NLM Gateway was added as well as a link to PubMed Central.

LinkOut for Libraries was released in April of this fiscal year and provides biomedical libraries the ability to link patrons from a PubMed citation directly to the full-text of an article, after the library has submitted its electronic holdings data to NCBI. As of August 2001, 298 providers had supplied links to their Web sites and allied resources based on specific citations or biological data found in PubMed and other Entrez databases.

A new Web-based interactive tutorial is now available from the PubMed sidebar. Additional system enhancements were made to PubMed throughout the year, including the addition of space life sciences-related journal citations from the former SPACELINE database; AIDS and HIV-related journal citations from the former AIDSLINE database; the addition of the sort button to all search result pages; and the addition of Complementary Medicine to the Limits Subset pull-down menu.

PubMed services have expanded in all aspects. Full-text journals that link to PubMed have doubled this year, from 1,138 in August 2000 to 2,285 in October 2001. Usage of PubMed by the scientific and lay communities has also grown considerably since its introduction in 1997. Currently, approximately 20 million searches are conducted per month and as many as 180,000 different users seek information daily via PubMed.

In collaboration with book publishers, the NCBI is also adapting textbooks for the Web and linking them to PubMed. The idea is that the textbook will serve to provide accessible background material that users can explore in order to better understand unfamiliar concepts found in a PubMed search result. The textbook, *Molecular Biology of the Cell*, 3rd ed., by Alberts et al. was the first book to be included in its entirety online. FY2001 additions include *C. elegans II* by Riddle et al. and *Retroviruses* by Coffin et al.

A collaboration between NCBI and NIH has led to the establishment of a Web-based repository for barrier-free access to primary reports in the life sciences. This repository, called PubMed Central (PMC), is based on a natural integration with the existing PubMed biomedical literature database of abstracts. PMC, a search system and archive of full-text journal literature in the life sciences, was launched in January 2001 and offers a new model for electronic scientific communication and data retrieval. The value of PubMed Central, in addition to its role as an archive, lies in the retrieval power and ease of access when data from diverse sources are stored in a common format in a single repository. PMC currently provides free and unrestricted access to the full text of forty-nine life sciences journals, with more forthcoming.

The BLAST Suite of Sequence Comparison Programs

Comparison, whether of morphological features or protein sequences, lies at the heart of biology. The introduction of BLAST in 1990 made it easier to rapidly scan huge sequence databases for overt homologies and to statistically evaluate the resulting matches. BLAST compares a user's unknown sequence against the database of all known sequences to determine likely matches. Hundreds of major sequencing centers and research institutions around the country use this software to directly query a sequence from their local computer to a BLAST server at the NCBI via the Internet. In a matter of seconds, the BLAST server compares the user's sequence with up to a million known sequences and determines the closest matches.

Not all significant homologies are overt, however. Some of the most interesting are subtle and do not rise to statistical significance during a standard BLAST search. NCBI has extended the statistical methodology in BLAST to address the problem of detecting weak, yet significant sequence similarities. Position-Specific Iterated BLAST (PSI-BLAST) searches sequence databases with a profile constructed using BLAST alignments, from which it constructs a position-specific score matrix. Several enhancements were made to the PSI-BLAST program this year, including the use of more accurately estimated statistical parameters; the filtering of database sequences, as opposed to query sequences, to prevent segments with highly restricted or biased amino acid composition from participating in the construction of profiles; and improved treatment of gaps within alignments when estimating position-specific amino acid frequencies. Many other enhancements were made to the BLAST suite of programs throughout FY2001.

New BLAST Web pages were made public early in FY2001. These pages are grouped by type of search, for example, protein-protein or nucleotide-nucleotide; allow restriction of BLAST searches through Entrez queries; provide formatting and uploading of the Position Specific Score Matrix; and generate XML output. The new BLAST site also features search-specific forms, the ability to construct a custom database on the fly, new output options, and a stable BLAST URL syntax that allows users to create custom search pages. MegaBLAST permits searching with batches of ESTs or with large cDNA or genomic sequences. Human Genome BLAST now searches the NCBI assembly of draft human genome and displays the genomic-context BLAST hits in Map Viewer.

The BLAST sequence searching server is one of NCBI's most heavily used services and its usage continues to grow at a pace reflecting the growth of GenBank. Each day more than 70,000 sequence searches are performed, with users submitting their requests through e-mail, server/client programs, and the World Wide Web. The popularity of BLAST has resulted in regular expansion of computing capacity to accommodate the growing volume of users. For example, QBLAST—a new system that obviates the need for persistent connections while users are waiting for results and allows for better distribution of the query load.

Other Specialized Databases and Tools

NCBI is well on its way to creating a new database called ProtSet, a comprehensive and stable set of protein sequences with explicit links to mRNAs and gene sequences. ProtSet will be minimally annotated because it will be

the basis for a diverse array of annotation, classification, and curation efforts, with explicit links to NCBI DNA sequences through NCBI identifiers and well-defined coordinate systems. NLM will play a unique role in this project by linking the ProtSet to the literature using the MeSH indexers to connect gene/protein sequence records with MEDLINE records, along with extracted phrases or MeSH terms to explain the basis of the link. A stable, comprehensive, and up-to-date ProtSet, along with links to the experimental evidence on function within the biomedical literature will be a valuable resource to the user community as well as for other more focused information resources. Several specialized Web services were recently released or substantially updated throughout the year.

A recent release is the SKY and CGH database combining digital imaging with cytogenetics. Digital imaging is the processing of pictures in a computer. Cytogenetics is the study of the genetic makeup of cells, and is often used in genetic diagnosis and cancer research. Breakthroughs in one field have now led to advances in the other; both spectral karyotyping (SKY) and comparative genomic hybridization (CGH), complimentary fluorescent molecular cytogenetic techniques, have benefited from the interaction of these two fields. SKY permits the simultaneous visualization of all human, or mouse, chromosomes in a different color, facilitating the identification of chromosomal aberrations. CGH utilizes the hybridization of differentially labeled tumor and reference DNA to generate a map of DNA copy number changes in tumor genomes. Together, these powerful tools are providing a means for detecting and mapping chromosomal breakpoints; detecting previously unknown chromosomal translocations; characterizing complex chromosomal rearrangements; and identifying marker chromosomes for genome mapping.

Microarray technology—a method for generating gene expression data—is another recent and important experimental breakthrough in the field of molecular genetics. As is the case with SKY and CGH, proficiency in generating data is fast overcoming the capacity for storing and analyzing it. In order to support the public use and dissemination of gene expression data, the NCBI has developed and launched the Gene Expression Omnibus, or GEO. GEO represents NCBI's effort to build an expression data repository and online resource for the storage and retrieval of gene expression data from any organism or artificial source. Many types of gene expression data will be accepted and archived as a public data set. ProbeSet, a new Entrez interface for GEO, was recently developed and released by the GEO and Entrez groups. ProbeSet is deeply indexed and is reciprocally linked to Entrez Nucleotide, PubMed, and Taxonomy. In addition, data from the Stanford Microarray Database, one of the largest collections of public microarray data and consisting of 36 sets of microarray data from 883 hybridizations across 6 species, has been acquired and deposited into GEO.

The NCBI is participating in the Mammalian Gene Collection (MGC), a new effort sponsored by the NIH. The goal of the MGC is to provide a complete set of full-length (open reading frame) sequences and cDNA clones of expressed genes for human and mouse. As of October 2001, there were 7,657 human and 2,773 mouse full-length clones stored in the database. This project will make all of the cDNA resources generated accessible to the biomedical research community. The MGC project involves the production of cDNA libraries and sequences, database and repository development, and support of research efforts leading to improved library construction, sequencing, and analytic technologies.

NCBI's Molecular Modeling DataBase (MMDB), an integral part of our Entrez information retrieval system, is a compilation of all the Protein Data Bank (PDB) three-dimensional structures of biomolecules. PDB is a collection of all publicly available three-dimensional protein structures, nucleic acids, carbohydrates and a variety of other complexes experimentally determined by X-ray crystallography and NMR and is maintained by the Research Collaboratory for Structural Bioinformatics (RCSB). The difference between the two databases is that the MMDB records reorganize and validate the information stored in the database in a way that enables cross-referencing between the chemistry and the three-dimensional structure of macromolecules. By integrating chemical, sequence, and structure

information, MMDB is designed to serve as a resource for structure-based homology modeling and protein structure prediction. MMDB currently contains over 15,000 structures, up from approximately 12,000 from last July. This figure represents an incremental increase of approximately 3,000 structures a year—a growth rate that has been constant for several years.

NCBI has developed a three-dimensional structure viewer, called Cn3D, for easy interactive visualization of molecular structures from Entrez. Cn3D also serves as visualization tool for sequences and sequence alignments. What sets Cn3D apart from other software is its ability to correlate structure and sequence information. For example, using Cn3D, a scientist can quickly locate the residues in a crystal structure that correspond to known disease mutations or conserved active site residues from a family of sequence homologs, or sequences that share a common ancestor. Cn3D displays structure-structure alignments along with the corresponding structure-based sequence alignments in order to emphasize those regions within a group of related proteins that are most conserved in structure and sequence. Cn3D also features custom labeling options, high-quality graphics, and a variety of file export formats that together make Cn3D a powerful tool for structural analysis. In FY2001, NCBI released a new version of Cn3D which contains improved quality graphics, more user display options, improved sequence and alignment viewers, coloring by alignment conservation, and the ability to save display settings and alignments.

The Conserved Domain Database (CDD) is a collection of sequence alignments and profiles representing protein domains conserved in molecular evolution. It includes domains from *Smart* and *Pfam*—two popular Web-based tools for studying sequence domains—as well as domains contributed by NCBI researchers. CD Search, another NCBI search service, can be used to identify conserved domains in a protein query sequence. CD-Search uses PSI-BLAST to compare a query sequence against specific matrices that have been prepared from conserved domain alignments present in CDD and receives several thousand queries per day. Alignments are also mapped to known three-dimensional structures, and can be displayed using Cn3D (see above).

A recently released resource displays the functional domains that make up a protein and lists other proteins with similar domain architectures. This protein architecture retrieval tool determines the domain architecture of a query protein sequence by comparing it to the CDD using RPS-BLAST. It then compares the protein's domain architecture to that of other proteins in NCBI's non-redundant sequence database. Related sequences are identified as those proteins that share one or more similar domains. The system displays these sequences using a graphical summary that depicts the types and locations of domains identified within each sequence. Links to the individual sequences, as well as to further information on their domain architectures, are also provided. As protein domains may be considered elementary units of molecular function, and proteins related by domain architecture may play similar roles in cellular processes, this resource serves as a useful tool in comparative sequence analysis.

VAST, or the Vector Alignment Search Tool, is a computer algorithm developed at NCBI for identifying similar three-dimensional protein structures. VAST is capable of detecting structural similarities between proteins stored in MMDB, even when no sequence similarity is detected. There are currently about 30 million structure-structure alignments recorded in VAST. VAST Search is NCBI's structure-structure similarity search service that compares three-dimensional coordinates of newly determined protein structures to those in the MMDB. VAST Search creates a list of structure neighbors, or related structures, that a user can then browse interactively.

The database of Clusters of Orthologous Groups of proteins (COGs) represents an attempt at the phylogenetic classification of proteins—a scheme that indicates the evolutionary relationships between organisms—from complete genomes. Each COG includes proteins that are thought to be orthologous, or connected through vertical evolutionary descent. COGs may be used to detect similarities and differences between species; for identifying protein families and predicting new protein functions; and to point to potential drug targets

in disease-causing species. The database is accompanied by the COGNITOR program, which assigns new proteins, typically from newly sequenced genomes, to pre-existing COGs.

A new Web page containing additional structural and functional information is now associated with each COG. These hyperlinked information pages include: systematic classification of COG members under the different classification systems; indications as to which COG member (if any) has been characterized genetically and biochemically; information on the domain architecture of the proteins comprising the COG and the three-dimensional structure of the domains if known or predictable; a succinct summary of the common structural and functional features of the COG members as well as peculiarities of individual members; and key references. In addition, a supplement to the COGs was made available in which proteins encoded in the genomes of two multicellular eukaryotes—the nematode *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster*—and shared with bacteria and/or archaea were included.

The purpose of NCBI's Taxonomy project is to build a consistent phylogenetic taxonomy for the NCBI sequence databases. The Taxonomy database, one component of the taxonomy project, provides general information on taxonomic resources as well as a list of outside curators currently collaborating with NCBI taxonomists. The database contains the names and lineages of the greater than 85,000 organisms represented by at least one nucleotide or protein sequence in the NCBI genetic databases. The database is recognized as the standard reference by the international sequence database collaboration. During FY2001, members of the taxonomy group maintained the overall structure of the Taxonomy database and Web pages, monitored the literature for new classifications, and maintained contact with off-site taxonomy advisors. NCBI taxonomists also provided consultation to staff of the EMBL Data Library and the DNA Database of Japan, collaborating sequence databases in Europe and Japan. Members continued to add new species or perform other edits to the database as required. Members also guided the NCBI indexing staff on taxonomic issues.

The first large set of Taxonomy LinkOut links went public early in FY2001. Links were made to four specific providers--Fishbase; HerbMed; UniGene; and COGs—as well as to two generic LinkOut providers that serve as useful site for a variety of organisms and taxonomic groups. Additional links are planned for the future.

The Taxonomy browser is an NCBI search tool that allows an individual to search the database. Using the browser, information may be retrieved on available nucleotide, protein, and structure records for a particular species or higher taxon. The NCBI Taxonomy database indexes over 85,000 organisms. The Taxonomy browser can be used to view the taxonomic position or retrieve sequence and structural data for a particular organism or group of organisms. Searches of the NCBI Taxonomy database may be made on the basis of whole, partial, or phonetically-spelled organism names, and direct links to organisms commonly used in biological research are also provided. The new Entrez Taxonomy system adds the ability to display custom taxonomic trees representing user-defined subsets of the full NCBI taxonomy.

TaxPlot, a new component of the Taxonomy project, is a research tool for conducting three-way comparisons of different genomes. Comparisons are based on the sequences of the proteins encoded in that organism's genome. To use TaxPlot, one selects a reference genome to which two other genomes will be compared. The TaxPlot tool then uses a pre-computed BLAST result to plot a point for each protein predicted to be included in the reference genome.

The Structure Group, in collaboration with NCBI taxonomists, has undertaken taxonomy annotation for the structure data stored in MMDB. A semi-automated approach has been implemented, in which a human expert checks, corrects, and validates automatic taxonomic assignments. The PDBeast software tool was developed by NCBI for this purpose. It pulls text-descriptions of "Source Organisms" from either the original PDB-Entries or user-specified information, and looks for matches in

the NCBI Taxonomy database to record taxonomy assignments.

UniGene (Unique Human Gene Sequence Collection) is NCBI's system for automatically partitioning GenBank sequences into a non-redundant set of gene-oriented clusters. Each UniGene cluster contains sequences that represent a unique gene, as well as related information such as the tissue types in which the gene has been expressed and map location. In addition to sequences of well-characterized genes, hundreds of thousands of novel expressed sequence tag (EST) sequences have been included. During FY2001 several new organisms were added to the UniGene database, including zebrafish, cow, frog, rice, wheat, barley, corn, and the plant *Arabidopsis*. As of October 2001, approximately 2,999,000 sequences were included in UniGene, with the final number of clusters (sets) totaling 96,327.

HomoloGene is a database of both curated and calculated orthologs and homologs for the human, mouse, rat, and zebrafish genes represented in NCBI's UniGene and LocusLink databases. Curated orthologs include gene pairs from the Mouse Genome Database (MGD) at the Jackson Laboratory, the Zebrafish Information (ZFIN) database at the University of Oregon and from published reports. Computed orthologs and homologs are identified from BLAST nucleotide sequence comparisons between all UniGene clusters for each pair of organisms. HomoloGene also contains a set of triplet clusters in which orthologous clusters in two organisms are both orthologous to the same cluster in a third organism. A new version of HomoloGene was released in May 2001 containing the sequences for two additional organisms—the fly and the cow. Software improvements were also made throughout the year and included tab-delimited output and calculated percent ID for affine gapping cases; selection of mRNA pairs over EST pairs to link homologous clusters; and decreased run time for HomoloGene queries.

Database Access

Entrez Retrieval System

The major database retrieval system at NCBI, Entrez, was originally developed for searching nucleotide and protein sequence databases and related MEDLINE citations. It was later expanded to include the integrated set of PubMed, Structure, Genomes, and Taxonomy databases. This year, additional databases were added to the Entrez retrieval system, including OMIM, BLINK—through Entrez Proteins, ProbeSet—a new database for gene expression data, and Books—a growing collection of biomedical books that can be searched directly. A new version of the Entrez software was made public earlier in FY2001. The major change involved the links between the databases, which are now maintained in a new system in order to enhance efficiency of use. Entrez's new design permits incorporating new linked databases without changes in the user interface, as well as additional sorting capabilities.

With Entrez, users can search gigabytes of sequence and literature data with 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 and intuitive way of accessing the data. Over 180,000 Entrez DNA and protein queries are handled per weekday and the number continues to rise.

Other Network Services

Usage of NCBI's Web services, first introduced in December 1993, continues to expand as more services are added. NCBI staff continued to make access and usage easier with improved documentation and tutorials. General information about NCBI, its databases and services, data submissions and updates, and NCBI investigator projects, as well as an ever-increasing number of search tools, are readily available via the Web. The Web server also provides capabilities for Entrez and BLAST searches and data submission through BankIt. Many other Web servers have links to the NCBI server in order to conduct searches and obtain the latest GenBank records. At the end of

FY2001, NCBI's site was averaging over 18,000,000 hits daily. Because of the mission-critical nature of NCBI's computing platforms for PubMed, Entrez, BLAST, and other services, extensive system monitoring is performed. Based on measurements taken every 15 minutes from 50 sites across the U.S. and overseas, the average time to load the entire NCBI home page is now under 1.5 seconds, an average PubMed search takes less than 3 seconds and availability has been better than 99 percent.

The improvement of NCBI's sequence submission software continued to be a high priority. A new version of Sequin, NCBI's stand-alone submission tool, was released in FY2000 and additional updates were made throughout FY2001. For example, Sequin version 3.70 for Macintosh, PC/MS Windows, and Unix computers was released in May. This new version has improved functions for updating sequences based on an alignment indexing code. BankIt, another sequence submission software tool, is now in its seventh year of use. A number of improvements designed to increase user utility were also made to BankIt throughout the year.

During FY2001, NCBI upgraded a number of its key systems to keep pace with the increase in demand for public services, such as BLAST and PubMed, as well as to accommodate the dramatic increase in the growth of GenBank. These include:

Core services—PubMed: 24 CPUs and 12 gigabytes (GB) of memory were added to the front-end Web servers that support PubMed and Entrez. Over 500 GB of storage was added to the four database servers that support PubMed and Entrez. Approximately 400 GB of storage was added to the NCBI FTP server.

Core services—BLAST: 20 8-way servers, with a total of 160 CPUs, were added to support the BLAST sequence similarity search service.

Core services—Compute Farm: Three 8-way servers, with a total of 24 CPUs, were added to the NCBI "Compute Farm," which supports genome-scale computing for basic research and for the development of new services. The total CPU count for the Compute farm is 136.

Internal network—Database support: Approximately 900 GB of FibreChannel storage was added to a Compaq SAN unit that supports a number of production and R & D database servers. Several hundred GB of directly attached SCSI storage was added to other database servers.

Internal network—Network-based storage: A single NFS server was upgraded to a 2-node cluster. The amount of storage supported by the cluster was increased from 2 terabyte (TB) to 5 TB. This system provides highly-available network-based storage for basic research and production databases.

Internal network—Security: Major upgrades and expansions were undertaken to ensure security for NCBI's public services and internal research and development computers.

Internal network—Centralized Unix desktop computing: FY2001 saw the rapid expansion of a pilot project begun in FY2000 to evaluate the use of low-cost X-Windows terminals to replace expensive workstations. In FY2001, approximately 100 terminals were deployed, in place of disk-full workstations.

Internal network—Network infrastructure: The transition to an all-switched 100/1000 Mbps network begun in FY00 was completed in FY2001. Thirty-two Gigabit Ethernet ports were added to the two core routers and approximately 250 FastEthernet ports were added in stackable switches, replacing the last of the 10 Mbps switches and hubs.

Research

Research is at the core of NCBI's mission. The Computational Biology and Information Engineering Branches are the main research branches of NCBI. Each Branch comprises a multidisciplinary team of scientists that carries out research on fundamental molecular biomedical questions by developing

and applying mathematical, statistical, and other computational methods to the life sciences. The research approach taken relies on both the theoretical and applied sciences, as, in the field of bioinformatics, these two lines of research prove mutually reinforcing and complementary. Research conducted by NCBI investigators has led to development of many new theoretical and practical models and the application of these methods to the life sciences has opened the doors to new areas of research. For example, the development and application of novel or improved algorithms to biologically important molecules has led to the identification of many previously unknown molecular structures. Structure identification, in turn, provides important clues as to how a molecule functions. From an understanding of molecular function, one can begin to elucidate its natural role in a particular molecular pathway, and from here, you can study what happens in a diseased state.

NCBI's basic research group is within the Computational biology Branch and consists of 53 senior scientists, staff scientists, research fellows, and postdoctoral fellows. Research projects include new computer methods to accommodate the analysis of genome sequences and molecular sequence databases due to the rapid growth of large-scale sequencing efforts. Other projects focus on such techniques as the analysis of particular disease genes as well as the analysis of the genomes of several pathogenic bacteria, viruses, and other parasitic organisms. Another focus is the development of computer methods for analyzing and predicting macromolecular structure and function. New areas of research include evolutionary genetics, the analysis of gene regulatory pathways, and the development of new modeling tools for tumor DNA data.

Currently, the intramural group is engaged in over 20 projects, many of which involve collaborations with other NIH institutes as well as with academia and private industry. A Board of Scientific Counselors, comprised of extramural scientists, meets twice a year to review the research activities of the Center. The high caliber of the work of this group is evidenced by the number of peer-reviewed publications, approximately 135 publications with an additional 5 in press. The staff participated in numerous oral presentations and mounted posters at various scientific meetings. Presentations were also made to visiting delegations, oversight groups, steering committees, and senior personnel from the Department of Health and Human Services. NCBI also hosted numerous outside speakers throughout the year.

The Visitors' Program continues to be successful in recruiting members of the external scientific community to engage in collaborative research with members of the NCBI Computational Biology Branch. Members of the Visitors' Program also participated in joint activities of database design and implementation with the Information Engineering Branch. NCBI researchers also continued active collaboration with the National Human Genome Research Institute on various projects, including sequence analysis, gene identification, and the analysis of experiments on gene expression. Various collaborations with other Institutes are also ongoing, including collaborations with the National Cancer Institute and the National Institute of Allergy and Infectious Diseases.

The NCBI GenBank Postdoctoral Fellow program, designed to provide for concentrated efforts on improving and strengthening GenBank, is currently filled. The NCBI uses the NIH Intramural Research Training Award Program and the Fogarty Visiting Fellow mechanisms to recruit for this program.

Outreach and Education

In FY2000, NCBI expanded its outreach and education programs to increase awareness of its myriad of public databases and specialized tools and services. NCBI staff presented at numerous scientific exhibits, seminars and workshops; sponsored a number of training courses--both lecture courses and "hands-on" courses; and published and distributed various forms of printed information.

Education: Mini-Courses and Lecture Presentations

Three new mini-courses, "Unmasking Genes in Human DNA," "Making Sense of

DNA and Protein Sequences," and "GenBank and PubMed Searching," are now offered to NIH scientists on a monthly basis. In addition, a series of presentations on Human Genome Resources has been developed for NIH Laboratory Chiefs.

Education: Bioinformatics Training

As computational capabilities and resources continue to develop, the use of computer science and technology by the biomedical community is increasing. The fusion of biomedicine and computer technology offers substantial benefits to all NIH Institutes and Centers in support of their general mission of improving the quality of the nation's health by increasing biological knowledge. In order to help NIH researchers make optimal use of computer science and technology to address problems in biology and medicine, the NCBI recently established an intramural Core Bioinformatics Facility—a network of bioinformatics specialists serving individual Institutes within the NIH. The Institutes and Centers select participants and NCBI trains these candidates on how to use the bioinformatics research tools disseminated by NCBI. In turn, core members advise researchers within their Institutes as to the best methods for conducting individual bioinformatics analyses. Information exchange among core facility members via institute-specific Web pages and a core-bio listserv allows the expertise of the entire group to focus on the diverse array of problems encountered by researchers at the NIH. Currently, the training program lasts nine weeks, with each week dedicated to exploring a major topic over a period of four days. On each of the four days, members meet for about two hours. Each two-hour session consists of an hour of lecture followed by an hour of hands-on work.

Education: Extramural Educational Collaborations

NCBI, in close collaboration with members of the biomedical community, is developing new materials for an advanced educational curriculum that would cover training issues relating to the use of various NCBI molecular biology resources. In addition, NCBI is planning to expand its existing basic training course. This course, which is now taught in a single day, will be offered over a period of either two or three days. NCBI is currently working with members of medical libraries and the academic community to determine appropriate course lengths; resources that should be covered in both the advanced and basic courses; and the development of training materials.

Outreach: User Guides for NCBI Resources

NCBI has developed a comprehensive list of fact sheets that outline the services and databases offered by NCBI, and highlight where to find them on the World Wide Web. NCBI also develops and distributes individual fact sheets that focus on a particular service or database. In addition, a number of other informational and educational resources are available on the NCBI Web site. "Articles of Interest" provides the user with a brief introduction to the field of bioinformatics and links to articles describing different NCBI resources. Another link discusses the fundamental principles underlying sequence similarity search tools. Interactive tutorials may also be found for a number of databases and search and retrieval tools. For example, "How to BLAST" is an interactive tutorial designed to help the first-time BLAST user employ this tool in their research. Tutorials for Entrez, PubMed, and OMIM were recently revised to incorporate the many new features added to these systems during the past year.

Primers have also been designed that provide a basic introduction to the science underlying NCBI resources. Topics include SNPs, ESTs, bioinformatics, molecular modeling, microarray technology, genome mapping, pharmacogenomics, SKY and CGH technologies, and phylogenetics. A basic genetics primer provides in depth information on topics such as what is a cell and how does it make DNA, RNA and proteins; what is a gene and how are genes expressed; mechanisms of genetic variation and heredity; and tools and technologies for organizing and studying genetic information.

NCBI News is a quarterly newsletter designed to inform the scientific community about NCBI's current research activities, as well as the availability of new database and software services. The newsletter contains information on user services; announcements of new or updated tutorials; a section on frequently asked questions; NCBI investigator profiles; and a bibliography of recent staff publications. In FY2001, over 39,000 printed copies of the *NCBI News* were distributed quarterly. The newsletter is also available to the general public via the NCBI Web site.

"Coffee Break," a recent educational resource at NCBI, is a collection of short reports on recent biological discoveries. Each report incorporates interactive tutorials demonstrating how bioinformatics tools are used as part of the research process. Each report is approximately 400 words and is usually based on a novel discovery reported in one or more recent articles from the peer-reviewed literature. The topics change every few months and public suggestions for future topics may be submitted to NCBI directly through this site.

NCBI in the News is a selective, annotated compilation of articles that reference NCBI programs and staff members and includes articles from the mass media as well as from the scientific and technical publications. In FY2001, NCBI was referenced in over 100 articles.

Extramural Programs

Funding for extramural bioinformatics activities is the responsibility of NLM's Extramural Programs Division. NLM funds research projects in areas defined as important to its mission. As the nation's premier repository of biomedical information, NLM has a vital interest in information management and in the enormous utility of computers and telecommunication for improving the storage, retrieval, access, and use of biomedical information. In this context, a wide variety of research in computational biology has been supported through the program, including methods and algorithms for sequence analysis, structure and function prediction, new machine architectures and specialized databases. Extramural postdoctoral training in the cross-disciplinary areas of biology, medicine, and computer science is also funded through the NLM informatics fellowship program.

Biotechnology Information in the Future

Over the past few years, there has been an explosion in the volume of genomic data produced by the scientific community, most notably in the amount of protein and gene sequence and mapping information. This is due in a large part to the recent release of the human genome, as well as the release of whole-genome sequences from other model organisms. The commitment to providing the scientific community with both the resources and tools needed to fully explore this data as quickly as possible, as well as recent advances in molecular analysis technologies, promises that the exponential growth in genomic data will only increase. This reinforces the need to build and maintain a strong infrastructure of information support. NCBI, a leader in the fields of computational biology and bioinformatics, will play an active and collaborative role in deciphering the human genome and in developing state-of-the-art software and databases for the storage, analysis, and dissemination of data. The genomic information resources developed and disseminated thus far by NCBI investigators have contributed significantly to the advancement of the basic sciences and serve as a wellspring of new methods and approaches for applied research activities. The value of these resources will continue to grow, as NCBI is committed to the challenge of designing, developing, disseminating, and managing the tools and technologies enabling the gene discoveries that will significantly impact health in the 21st century.

EXTRAMURAL PROGRAMS

Milton Corn, M.D.
Associate Director

The Extramural Programs Division (EP) continues to receive its budget under two different authorizing acts: the Medical Library Assistance Act (MLAA, unique to NLM), and Public Health Law 301 (covers all of NIH). The funds are expended mainly as grants-in-aid, and in some instances as contracts, to the extramural community in support of the goals of the Library. Review and award procedures conform to NIH policies. For a list of grants awarded in FY 2001, see http://www.nlm.nih.gov/ep/extramural.html.

EP issues grants in a broad variety of programs, all of which pertain to informatics and information management with the exception of the Publications Grant program.

- Resource Grants for information management; usually involve medical libraries
- Training and fellowship grants in support of informatics research training
- Research Grants in informatics, information science, and biomedical computing
- Research Resource grants to support informatics and bioinformatics research
- Publication grants to support preparation of scholarly manuscripts
- SBIR/STTR
- Special Projects

Resource Grants (MLAA)

Resource Grants, authorized by the Medical Library Assistance Act, support access to information as well as promote networking, integrating, and connecting computer and communications systems. There are four types of Resource Grants, which range in complexity as well as in dollar amounts and duration. They are considered "seed" grants designed to initiate a resource or service or program that is expected to become self-sustaining. All four Resource Grants are open to public and private, nonprofit health organizations engaged in health education, research, patient care, and administration, and all four strongly encourage some health science library involvement in the project.

Information Access Grants

Information Access Grants, aimed primarily at hospitals, clinics, community health centers and similar small health organizations, support installation of computers and other information technology as well as training to facilitate access to NLM's Pub Med and other databases and/or improve efficient distribution of the library resources within a region. These grants provide up to $12,000 per participating institution and are available to single as well as multiple institutions working together.

Information Systems Grants

Information Systems Grants, ranging up to $150,000 per year for up to three years, are intended for more complex projects and organizations than are the Access grants, and are suitable for a broad variety of information management projects at larger hospitals, medical schools and other health-care related institutions. These grants can be used to support both personnel and information technology, and have been widely useful in a number of areas. Planning grants are also available for those who are not quite ready to request the standard Information Systems grant.

Internet Connections Grants

The Internet Connection Grants provide grants up to $30,000 to single institutions and up to $50,000 to multi-institution conglomerates to initiate Internet access. Funds are usually used to pay for gateway/router equipment, Internet Service Provider fees, and line charges in the first year. Some institutions with existing Internet access can use these grants to improve distribution of Internet access internally, or to extend access to other institutions.

Interest in these grants diminished somewhat during the middle 1990s but has been

steady in recent years at a level of $400,000–500,000 per year. Applications in FY2001 increased markedly in response to publication of a Request For Applications, and increased emphasis on notifying potential applicants about the existence of the program.

IAIMS Grants

Integrated Advanced Information Management Systems (IAIMS) Grants are designed to facilitate institution-wide information systems that link a variety of individual and organizational databases and information systems for patient care, education, research, library, and administration. IAIMS Grants support two phases, planning and implementation, with the program goal being to support organizational mechanisms that foster the integration and sharing of various information systems, and the organization's short- and long-term planning for optimal use of information technology. The planning phase funds up to $150,000 for one to two years; the operational phase up to $500,000 per year for five years or $550,000 with an IAIMS apprenticeship option.

Although the program was initially intended to support a minimal set of models that could then serve as templates for others, experience with the grants demonstrated that the problems and therefore the solutions were parochial. It became clear that an IAIMS climate required much more emphasis on people and organizational issues than on information technology, which meant that a chief value of the grant was for smoothing the managerial interactions essential to the IAIMS goals. Although the large increase in dollar outlay for information technology by medical centers in recent years dwarfs the value of the grants, interest in these grants remains high, perhaps because of the impact of an NIH grant on otherwise stubborn management and organizational problems. In FY2001 none of the institutions with planning grants were able to compete successfully for a Phase 2 Implementation grant.

At the close of FY2001, NLM received the final report of a study of IAIMS for which the Association of American Medical Colleges had received a contract in FY2000. The study was commissioned because of a perception, shared by NLM, that so much has changed in information technology and health care organizations that the program needed to be examined and adjusted to fit the current environment.

Training And Fellowships (MLAA)

Exploiting the potential of computers and telecommunication for health care information requires investigators who understand biomedicine as well as fundamental problems of knowledge representation, decision support, and human-computer interface. NLM remains the principal support nationally for research training in the fields of medical informatics, including clinical and basic science domains. NLM provides both institutional and individual mechanisms of support for its training activities.

NLM-Supported Training Programs

Five-year institutional training grants support approximately 150 trainees at pre-doctoral and postdoctoral levels. Twelve institutions currently receive such support, but because a number of these share support with other universities and teaching hospitals, there are over 20 training sites. For the past few years, NCI and NIDR contributed funds to NLM to help support slots at these training sites for applicants interested in radiation therapy and dental informatics respectively. Following some staffing changes, NCI discontinued its support in FY1999.

BISTI and NLM's Training Programs

Interest in biomedical computing exploded at NIH after the June 1999 publication of the Biomedical Information Science Technology Initiative (BISTI) report on biomedical computing. Because the BISTI report stimulated a new set of pan-NIH grant programs to begin for the most part in FY2001, NLM provided each of the twelve institutional training programs with "BISTI" administrative supplements of $200,000 during FY2000 and

again in FY2001 as a means of initiating or enhancing training tracks in bioinformatics, i.e., the informatics of research data. The BISTI report recognized that biomedical researchers need much more training in the tools of informatics, but in addition there is now and will continue to be a marked need for informaticians with sufficient domain knowledge to develop these tools as data and data interpretation become increasingly complex. Because of its long history of supporting informatics research training, NLM is well poised to make a significant contribution to the BISTI effort.

Health Services Research

To promote research training in health services research, EP distributed $50,000 to each of the ten NLM-supported training programs that requested such supplements in FY 2000 and again in FY2001.

Individual Fellowships

Individual informatics research fellowships are available to those who seek research training similar to offerings at the institutional training sites but at a site of their choosing. Individual applied informatics fellowships are also available to individuals who want to learn informatics techniques and technology for application in their current professional specialties. To encourage mid-career applicants, the applied fellowships permit stipends of up to $58,000 per annum as substitute for salary lost during each training year. Applications for these fellowships have been predominantly for the applied fellowship, probably because of the larger stipend. Because applications have been dwindling, the program will be reevaluated in FY2002. The difficulty of informing potential mentors and candidates about the existence of these opportunities may be playing a significant role.

Education of Health Sciences Librarians in Informatics

All existing NLM Informatics Training Programs have been encouraged to develop and offer training within the curriculum suitable for those interested in health science libraries. NLM agrees to provide additional funding for any slots awarded to librarians. Response has been gratifying and is growing. Librarians are now in place at the University of Pittsburgh, Oregon Health Sciences University, University of Missouri and the University of North Carolina at Chapel Hill.

Publication Grant Program

The Publication Grant Program provides short-term financial support for selected not-for-profit, biomedical scientific publications. Studies prepared or published under this NLM program include critical reviews or research monographs in the history of medicine and life sciences; on special areas of biomedical research and practice; on medical informatics, health information science and biotechnology information; and in certain instances, secondary literature tools and scientifically significant symposia. Resources in recent years have been used principally for history of medicine projects. Standard print publication has been the most common format, but projects in electronic publishing, video, and. and other media have also been supported. The program has an informal self-imposed ceiling of $50,000 on direct costs per grant per year.

Minority Support From MLAA Funds

NLM continues its support of the NIH ongoing program for “Research Supplements for Underrepresented Minorities.” In FY2001 computer science doctorate candidate, Jonathan Allen, received a minority supplemental award as part of the program. This award to Johns Hopkins University, with mentored support by Dr. Steven Salzberg as doctoral advisor, is to develop new algorithms to significantly improve the process of automated biological protein sequence analysis. This supplemental grant award is consistent with the mission of the Biomedical Information Science and Technology Initiative (BISTI), promoting the scientific field of computational biology.

Other Minority Support

Internet Connection Grants were awarded to a broad variety of institutions serving African-American, Amerindian, and native Hawaiian populations in inner cities and rural areas. Similarly, a number of Information Access Grants were awarded to organizations serving rural and inner city populations.

Research Support (PHS 301)

Research support is provided through a variety of mechanisms, including individual research grants and contracts, cooperative agreements, research resource grants and others. NLM's research grants support both basic and applied projects involving the applications of computers and telecommunication technology to health-related issues in clinical medicine and in research.

Medical Informatics

Since inception of the grant program, the majority of NLM's research support in informatics has focused on the informatics of health care delivery with support both to applied projects (e.g. the electronic medical record, telemedicine) and related basic problems (e.g. natural language processing, data-mining, knowledge representation). Although there has been marked expansion in research support for informatics issues related to biological and medical research in recent years. NLM plans to continue its support for clinically relevant informatics.

Biotechnology Informatics (Bioinformatics)

NLM has been aware for a decade that biomedical computing is indispensable for handling the complex data and large datasets generated by research, most notably in molecular biology research and neuroscience, but also in clinically relevant areas such as outcomes research and public health issues. To facilitate this form of biomedical computing, EP has maintained a separate grant program (originally called "biotechnology" and latterly changed to "bioinformatics." NLM continues to provide research grants for informatics projects in bioinformatics, as well as training grants, and grants for support of research resources.

The BISTI report of 1999 on biomedical computing markedly increased NIH interest in potential of computing for biomedical research. In FY2000, NLM together with a number of other Institutes began a series of discussions about the various ways in which NIH intends to address national needs for training and research in biomedical computing. With participation by NLM and numerous other Institutes, NIH announced a battery of new programs responsive to BISTI in late FY2000 with the first awards to be made in FY2001. Award categories included Planning Grants for National Programs of Excellence in Biomedical Computing, specific research projects, and relevant SBIR applications. Management of BISTI is through a trans-NIH BISTI committee to which NLM sends a representative.

BISTI awards are not different in general domain from NLM's existing Bioinformatics grant program. However, EP has maintained a separate budget category for BISTI grants because new funds were specifically allocated for BISTI projects, and because both review and grant mechanisms differ from NLM's customary processes. Of the Planning Grant applications received by NIH, NLM was particularly interested in those that incorporated existing NLM-supported Informatics Research Training Programs into the plans for the Centers. In FY2001, NLM funded Planning Grants for Yale and Columbia. How the implementation grants for these centers will be handled, and when the requests for applications will be issued remains to be determined.

Databases

An issue related to BISTI concerns the development and maintenance of electronic databases on which researchers increasingly rely, and for which no other source of support has yet been identified. NLM is an important, but not the only, NIH source of support for such databases. Most of the databases funded by NLM support genomics and proteomics research. However, some awards in recent years

have been for databases relevant to clinical applications of genomics.

NLM and the Human Brain Project

NLM also participates with 15 other NIH and federal organizations in the Human Brain Project, which is led by the NIMH and seeks innovative methods for discovering and managing increasingly complex information in the neurosciences. Each participant selects grants within the project for full or shared funding. NLM participation has been steady but is rarely more than one new grant each year, and in some years none are funded.

NLM and Other Pan-NIH Projects

NLM also participates in a number of other multi-institute projects including bioengineering, pharmacogenetics, imaging, and nanotechnology. In FY2001 NLM provided co-funding to NIGMS for a pharmacogenetics database under development at Stanford. EP has not funded any bioengineering projects. Nanotechnology is still in very early stages of development for biological purposes.

NIBIB

In FY2001 Congress created a new Institute, the National Institute for Biomedical Imaging and Bioengineering. Although its research interests have not yet been fully defined, and it will not begin operating until FY2002, NIH asked all Institutes to identify currently funded image-related research projects for transfer to NIBIB. EP had five such, all of which were transferred along with the budget necessary to complete funding for the grant, ($965,909 for FY2001). Bioengineering (BECON), up to now a trans-NIH operation, will become an integral part of NIBIB.

Other Support

Conference Grants

Support for conferences and workshops is intended to help scientific communities identify research needs, share results, and prepare for productive new work. Requests for such grants are increasing. At present EP generally caps such awards at $20,000, although exceptions are made on an ad hoc basis. To expedite processing of these grants, NIH permits a two-level review to be done by NLM staff.

Biomedical Ethics

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

HPCC and Outreach

The outreach and the High Performance Computing and Communications initiatives of NLM are elements of the formal grant programs.

Special Projects

In addition to its standing grant programs, EP participates in a number of special projects often involving cooperation with another NIH institute or other Federal agency. Some examples of such activities in FY2001 follow.

The Digital Libraries Initiative-Phase 2 (DLI-2)

This initiative explores innovative digital libraries research and applications. The program extends the previously sponsored "Research on Digital Libraries Initiative." The term "digital libraries" is used to denote the vast distributed collections of text and images available through the Internet. Much research and development will be needed before these new electronic libraries can be used easily and efficiently to obtain reliable information. DLI-2 is administered by the National Science Foundation and is jointly sponsored by the NSF,

the Defense Advanced Research Projects Agency, the NLM, the Library of Congress, the National Aeronautics and Space Administration, the National Endowment for the Humanities, and others.

The project is interested in electronic information in a broad spectrum of fields in arts and science. Improving network-based information access for health care consumers is an important goal of the project for NLM, although all aspects of digital libraries as applied to health domains may compete for funding. NLM, as have the other sponsors, contributed funds to NSF, which will manage the project. NLM's commitment for FY2001 was $1,000,000 as it had been in the previous year. The DL-2 project is an arm of the HPCC initiative. Target for total project budget from all sources is $50 million over 5 years. The last installment of NLM's commitment to this program will be in FY2002.

NLM made available to interested applicants the Unified Medical Language System Knowledge Sources and the Visible Human datasets. Applicants were also free to use resources of their own choosing. Although awards were not made to fill predetermined domain quotas, the review and awards process resulted in a gratifying number of projects with health themes, and several others whose informatics component concerned issues with considerable potential benefit for health concerns. All of the contributed funds are now being used to support the out years of grants awarded during the first round of competition. As of now, no surplus funds were available to support applicants who sent in proposals during round two.

Informatics for the National Heart Attack Alert Program (Research Contracts)

This program receives approximately 2/3 of its funding from NHLBI, and the remainder from NLM. The program offered a Phase 1 feasibility contract for up to $100,000 for one year. Phase 2 called for implementation in a test population or a larger group over a period of several years. After the initial Phase 1 RFP in FY1998 which focused on 'main-line' informatics and supported 14 investigators, a second Phase 1 RFP was published in FY1999 to obtain feasibility proposals using more innovative, high-risk, high-payoff technology. Five Phase 1 contracts for nine-month planning phases were awarded in this "high-tech" group. Technologies to be explored include wearable devices, portable computing devices, games, and wireless communications devices.

In response to a Phase 2 RFP for the "main-line" Phase 1 projects, five Phase 2 contracts were awarded during late FY1999 and FY2000. A Phase 2 RFP for the "high-tech" projects was issued in late FY2000. Awards were made in FY2001 to two of the Phase 1 "high-tech" applicants. Although the original RFP contemplated the possibility of a Phase 3 for this program, neither NHLBI nor NLM is planning to proceed with another Phase.

Miscellaneous Special Projects

NLM continues its collaborative extramural funding with other agencies in support of projects broad in scope and utility and directly related to biomedical research. The agencies that received NLM funds in FY2001 were the National Center for Research Resources, National Institute of Deafness and Other Communication Disorders, National Institute of Mental Health, Agency for Healthcare Research and Quality and the National Science Foundation. NLM received co-funding for NLM grants from other organizations, including Department of the Army and the Centers for Disease Control.

SBIR/STTR (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.

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.

Grants Management Highlights

The Grants Management staff reviews NLM grant applications for compliance with guidelines and directives; prepares and disseminates grant awards; maintains official grant files for NLM; provides consultation and assistance to grantees on appropriate business management concepts; and advises NLM officials on grants management policy and procedures. The Grants Management staff, which consists of three employees, issued a total of 169 awards for FY2001, as well as supplemental awards for the Biomedical Information Science and Technology Initiative and Health Services Research.

Review Committee Activities

NLM's initial review group, the Biomedical Library Review Committee (BLRC), evaluates grant applications for scientific merit. BLRC met three times in FY2001 and reviewed 90 applications. The Committee (see Appendix 2 for roster of members) operates as a "flexible" review group; i.e., it is composed of 3 standing subcommittees: 8 members on the Medical Library Resource Subcommittee, 9 members on the Medical Informatics Subcommittee; and 4 members on the Biomedical Information Subcommittee. The subcommittees consider research applications in medical library projects, medical informatics, and biotechnology information respectively.

Thirteen Special Emphasis Panels (SEPs) were also coordinated which reviewed 250 applications. Such panels are convened on a one-time basis to review applications for which the regularly constituted review groups lack appropriate expertise, or there exists a conflict between applicant and a member of the BLRC. Use of SEPs by EP increased significantly since FY2000 because of new regulations requiring EP to supervise SEP for review of contracts as well as of grants. The significant increase in applications for SEP Panels in FY2001 (more than double those reviewed in FY2000) were in response to RFA's for Internet Connections (102) and Publications of Scholarly Documents (99). One site visit to evaluate an IAIMS application was also carried out by an ad hoc panel.

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

Appeals of review

In FY2001 EP received an appeal of a review performed by an ad hoc committee convened to consider history of medicine applications. Because staff did not sustain the appeal, the applicant requested that the matter be brought before the BOR for second and final appeal. The BOR affirmed the review, in effect denying the appeal.

Personnel Activities

EP has three Program Officers, each with an emphasis in one of the three areas of Library Resource, Informatics, and Publications. These staff members work with grant applicants during all phases of the application and review process, and subsequently monitor the work done on the awarded grants. They are an important interface of NLM with the academic community. Two new program officers were recruited and appointed in FY2001 to fill vacancies in the Library Resource and Informatics areas. EP also appointed a new Committee Management Officer to fill a retirement vacancy. A Committee Management Assistant was also hired during the year. Remaining to be filled is a second Scientific Review Administrator. IT support for EP continues to be provided by a contractor on site,

an approach necessitated when EP moved to Rockledge.

Summary

EP's grant activities in FY2001 were in conformity with previous years with the exception of the significant new emphasis on biomedical computing as stimulated by BISTI. NLM's extramural grant division, like similar divisions elsewhere at NIH, cannot fund all applications of good quality. The situation was particularly acute in FY2001 because a larger than average demand on the available budget came from the budget requirements of existing grants. Funds available for new applications were further reduced because early in the year EP funded a number of worthy FY2000 informatics applications held over for funding in FY2001 because FY2000 funds ran out. As in the previous year, a number of excellent grants, mainly in informatics research, were held over for funding in FY2002.

Figure 1—Expenditures for EP's Main Program Categories

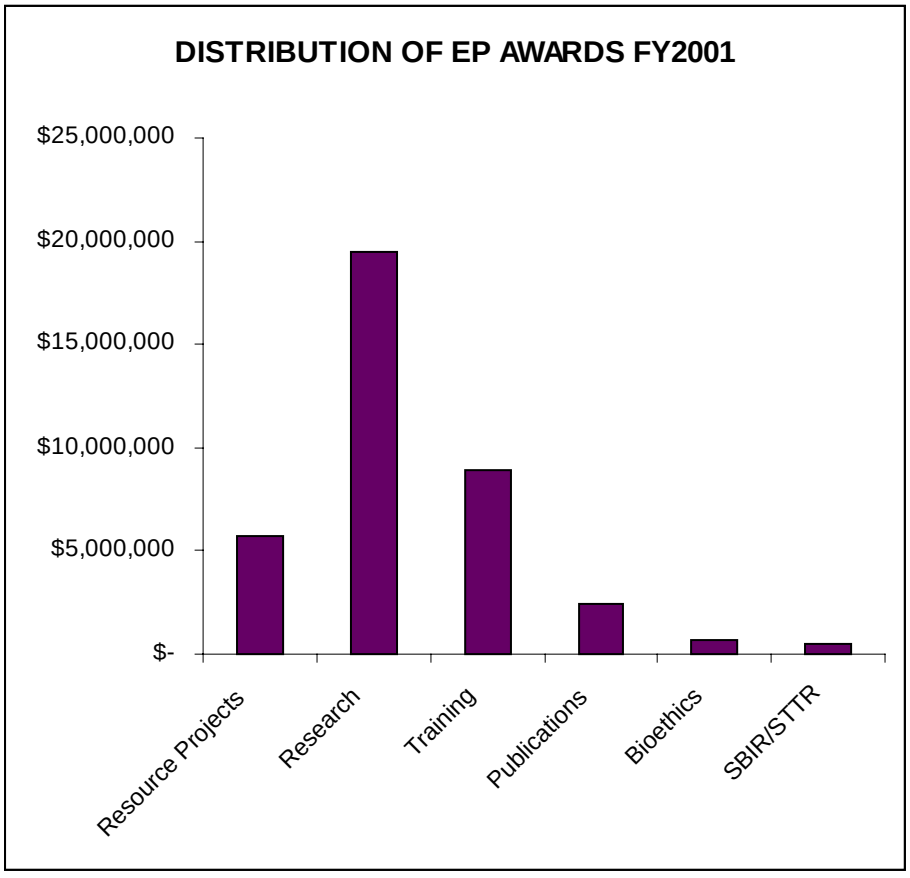


Table 11

Extramural Grants
(Dollars in thousands)

	FY 1999		*FY 2000*		*FY 2001*	
	No.	*$*	*No.*	*$*	*No.*	*$*
MLAA	82	21,408	92	25,508	174	29,551
PHS	77	18,425	68	19,325	82	22,848
Total	**159**	**39,833**	**160**	**44,833**	**256**	**52,399**

Table 12

Grants Awarded with
Medical Library Assistance Act Funds
(Dollars in Thousands)

		FY 1999		*FY 2000*		*FY 2001*	
Category	*Program*	*No.*	*$*	*No.*	*$*	*No.*	*$*
IAIMS	IAIMS Ph. I	6	872	8	1,175	5	745
	IAIMS Ph. II	5	2,742	3	1,650	1	550
	Total IAIMS	11	3,614	11	2,825	6	1,295
Training	T15	12	5,730	12	7,919	12	6,250
	BISTI Supp.	---	------	12	2,000	12	1,948
	Fellowship	9	482	13	799	12	705
	Total Training	21	6,212	37	10,718	36	8,903
Publications		5	219	6	268	39	2,406
Resource	Inf. Sys. G08	11	1,077	13	1,879	20	2,119
	Access G07	5	389	9	696	13	760
	Connect. G08	20	657	5	207	47	1,572
	Total Resource	36	2,123	27	2,782	80	4,451
Bioethics		1	529	1	530	1	697
Other	Distance Ed.	---	------	2	199	---	------
	AMI Alert*	---	------	---	------	3	1,758
	AIDS	---	------	---	------	1	74
NN/LM	Contracts	8	8,711	8	8,186	11	9,967
Total MLAA		**82**	**21,408**	**92**	**25,508**	**174**	**29,551**

*Contracts (includes $858 from NHLBI)

Table 13

Grants Awarded with PHS 301 Funds
(Dollars in Thousands)

		FY 1999		*FY 2000*		*FY 2001*	
Program		*No.*	*$*	*No.*	*$*	*No.*	*$*
Med. Informatics	R01	47	10,499	34	8,590	46	8,770
	DL2	10	1,109	1	1,000	1	1,000
	Total Med. Info.	57	11,608	35	9,590	47	9,770
Bioinformatics	R01	9	2,255	19	4,757	16	3,988
	BISTI	1	94	1	300	4	2,852
	Resource P41	4	1,776	7	2,974	9	2,765
	PDB	---	------	1	150	1	150
	Total Bioinfor.	14	4,125	28	8,181	30	9,755
DL2		1	1,000	---	------	---	------
SBIR/STTR		4	562	4	424	4	502
Bioethics		0	0	---	------	---	------
NIH Taps		0	1,030	0	1,030	0	2,671
Chairman's Grant		1	100	1	100	1	150
Total PHS		**77**	**18,425**	**68**	**19,325**	**82**	**22,848**

OFFICE OF COMPUTER AND COMMUNICATIONS SYSTEMS

Simon Y. Liu, Ph.D.
Director

The Office of Computer and Communications Systems (OCCS) provides efficient, cost-effective computing and networking services, application development, technical advice, and collaboration in informational sciences in support of the research and management programs offered through the NLM.

OCCS develops and provides the NLM backbone computer networking facilities, and supports, guides, and assists other NLM components in local area networking. The Division provides professional programming services and computational and data processing facilities to meet NLM program needs; operates and maintains the NLM Computer Center; designs and develops software; and provides extensive customer support, training courses and seminars, and documentation for computer and network users.

OCCS helps to coordinate, integrate, and standardize the vast array of computer services available throughout all NLM components. The Division also serves as a technological resource for other parts of the NLM and for other Federal organizations with biomedical, statistical, and administrative computing needs. The Division promotes the application of High Performance Computing and Communication to biomedical problems, including image processing and information security.

Executive Summary

This year, OCCS marked the completion of the NLM System Reinvention Initiative on September 30, 2001. Working jointly with other NLM organizations, the transition from legacy systems to an open systems architecture was completed. The IBM mainframe, in use for over 25 years, retired as a result of the completion of the NLM System Reinvention. A formal mainframe shutdown, executed by Dr. Donald A.B. Lindberg, was celebrated in the Lister Hill Auditorium. Other major OCCS milestones and successes this year included:

System Reinvention

The activities of FY2001 focused on completing the transition from the legacy AIMS system to the Data Creation and Maintenance System (DCMS). The DCMS was a dramatic change from the legacy system with improvements ranging from software changes to increased throughput via high-speed Cable and DSL lines whenever possible.

Data conversion deserves particular recognition this year. Conversion from the legacy format to a format suitable for loading in the appropriate target system was a monumental task. Thousands of hours were spent analyzing and reviewing data from HISTLINE, SPACELINE, AIDSLINE, BIOETHICSLINE, and POPLINE.

NLM uses the Voyager ILS for acquisitions, serials control, cataloging, collection management, circulation, and preservation. Most activities relating to Voyager this year involved the loading of legacy data produced by NLM's data conversion activities, distribution of data to licensees in various formats, and upgrading to Voyager Gold 2000 which was released in the first quarter of 2001. The major ILS accomplishments this year were:

- Accommodating the updated MARC 21 standard, which included modified character set conversion mappings.
- Data created and maintained in Voyager was extracted in XML and distributed to NCBI for use in the journal browser.
- CATFILE*plus* was developed to support US MARC distribution system. The system supports customized for recipients of NLM data.

Critical tasks for Year-end-processing (YEP) implementation were completed, including production of the M2000 Descriptor, Qualifier, and Chemical transactions for the M2000/YEP processing. Computer processing

time under the new DCMS/MeSH system was reduced dramatically from 9 months to 1 day.

The List of Journals Indexed (LJI) and List of Serials Indexed (LSI) publications are now available online in PDF and DOS text format. The production process from data extraction to publisher delivery has decreased from several weeks to just 1-2 weeks.

Three versions of DOCLINE were released this year. The most recent version, Version 1.3, was released in September 2001. DOCLINE provides document delivery service to more than 4,000 U.S. and Canadian medical libraries.

The reinvented LSTRC was completed this year, allowing access via a web browser utilizing ORACLE for specific LSTRC data with interfaces to the Serial Extract Database for serial information.

MEDLINEplus Enhancements

MEDLINE*plus* generated 62 million page views in FY2001, and the consumer health information service now contains 500 health topics, up from 22 at the system's debut. Six versions of MEDLINE*plus* were released this year.

Document Delivery Enhancements

A new Relais delivery method, Post-to-Web, was introduced, which delivers documents to a web server and emails the requester a URL for remote downloading and printing. The Interim Binding Module was deployed which allows users to keep track of instructions for binding serials as well as track titles ready for binding. A web-based Overnight Photocopy Service (OPS) was deployed. OPS will allow reading room patrons to request NLM staff to photocopy articles.

Administrative Support Systems Enhancements

The online customer Ordering/Inventory Control System for the NLM Office of Administrative Management Services was enhanced by implementing approximately 600 inventory photos into the system. Staff are now able to view a picture of the supplies they are ordering before their order is actually placed. The Personnel Administrative Control system was successfully released in September 2001. This system tracks personnel information including employee information, recruitment actions, personnel actions, and award information. Due to the sensitivity of the personnel data, a secure access to personnel information was also implemented.

Infrastructure Improvements

Among the infrastructure improvements made in FY2001: NLM's public Internet connectivity was upgraded to an OC3 (155Mbps) circuit to the Genuity network node in Washington, D.C.; OCCS provided broadband access (DSL and cable modem) to contractors and employees of the BSD Indexing Section in support of DCMS.; extensive planning was coordinated by OCCS to lay the groundwork for a new core network design that will permit Gigabit speeds and higher levels of security; plans have been completed to upgrade the OCCS network fiber backbone that will enable future data transmission of up to 10 Gigabytes; and a storage area network was implemented to increase storage central network storage capability to 800 Gigabytes.

IT Security

Among the IT security improvements: vulnerability scans of network-based systems took place on a regular basis during 2001; over 200,000 Code Red Worm, W32/SirCam, Love Letter, EICAR, VBS/PeachyPDF, SnowWhite, BadTrans and Magistr viruses were removed from inbound NLM emails; Lucent firewalls were implemented at our Internet point of connection and firewall policies were implemented.

Customer Support Enhancements

OCCS joined a software licensing agreement already in place between Microsoft and the University of Maryland System's Maryland Educational Enterprise Consortium (MEEC) for software acquisitions. As a result of this agreement, costs per PC will be $29, resulting in a 5,970% savings over GSA pricing.

OCCS negotiated an agreement with NIH's Center for Information Technology for acquisition of Novell products and Network Associates' McAfee antivirus suite at a savings of 30% over last year's costs.

Computer Facility Enhancements

After major equipment shifts and re-alignments, the NLM computer facility was physically reorganized to provide more efficient use of floor space and designated cable paths. This will allow for future equipment growth as well as a more secure facility.

Construction of the Network Operations & Security Center (NOSC) started at the end of this fiscal year. The NOSC will enable NLM support personnel to perform system monitoring, intervention, administrative maintenance and security activities from one central location within the NLM computer facility. In addition, staff and visitors will have the opportunity to view NOSC system activity through the computer facility windows.

Electrical power in the computer room has been upgraded and simplified. Power panels have been consolidated allowing for expeditious isolation of problems or potential problems. As more UNIX servers are brought in, power lines supporting lower amperage are being installed. The computer facility continues to be supported by Uninterrupted Power Supply (UPS).

Section 508

OCCS led the implementation of Section 508 (pertaining to accessibility of web sites) at NLM. A team of Section 508 coordinators worked diligently for eight months reviewing NLM web sites for 508 compliance. The NLM Section 508 Web Implementation Plan was in place for the June 21st enforcement date.

New Technical Support Contract

A technical support contract task was awarded to the team of CSC and AAC on September 1, 2001. The task was awarded under the NIH CIO-SP2 contract. The period of performance is for one base year, with nine option years. The immediate scope of support is LAN, desktop, and security. Systems administration, facilities management, and various ad-hoc tasks can be added as required.

Retirement Celebrations

Several long-time OCCS employees retired in FY2001. After 37 years with the NLM, Mr. Philip Neilson retired on December 28, 2000. Phil was the NLM Y2K coordinator for which he received the NIH Director's Award. After 36 years with the Federal Government and 24 of those years at the NLM, Mr. Richard Wiles retired on July 27, 2001. Rich worked in the Systems Technology Branch, first as a Mainframe Computer Operator then as a Mainframe Systems Programmer specializing in Data Communications.

The following sections describe, in detail, accomplishments by OCCS in each major functional area for FY2001.

Customer Services

The IT Services Center (ITSC), formed in 2000, is a single point of contact for OCCS systems support. The ITSC input and tracked over 6,000 requests for IT support this year. To assist management in tracking operations, a daily status report is emailed to users and managers. It provides a status summary of all OCCS systems and statistics on requests for services. In concert with the Library Operations Customer Service Center, NLM's external customer help desk, the IT Services Center has been exploring the use of a more powerful problem reporting and tracking tool. Once implemented, the new tool will permit a more rapid response to customer interaction.

Desktop Support

Workstation Operating System Upgrades

Extensive testing of the Windows 2000 (W2K) desktop operating system was conducted in FY2000. W2K was found to be a generally improved desktop operating system over Windows NT, and was adopted as the NLM

desktop standard operating system. Windows 2000 has been deployed on 30% of the PCs in NLM. Upon the release of Windows XP, OCCS will undergo compatibility testing with NLM applications and hardware.

Consolidated Software Acquisition and Cost Savings

OCCS has identified and initiated a low-cost "academic model" contract vehicle for software licenses based on the NLM's standing as a public library. NLM has acquired Microsoft and other software at one third the price of GSA Schedule offerings. In this year's review of alternatives to this expiring contract, OCCS identified an even cheaper contracting model and on behalf of all NLM organizations, joined a software licensing agreement already in place between Microsoft and the University of Maryland System's Maryland Educational Enterprise Consortium (MEEC). The MEEC agreement provides for a basic bundle of software licensing and maintenance, and is augmented by a pre-negotiated academic pricing model for Microsoft products that fall outside of the bundle. Operating system upgrades, Microsoft Office and its upgrades, Microsoft development products and the licenses needed to connect to Microsoft servers are all included in the bundle. Contrasted to the same product set acquired at GSA schedule prices, $1,712, this year's bundle price of $29 per PC permitted cost savings of over 5,970 percent (5,970%). The MEEC program may be extended for two option years, during which time our bundle seat price will drop further, to $13 per PC for the year.

NLM PC Acquisition Consolidation

In a similar licensing review effort, the OCCS evaluated various acquisition vehicles for Novell and other software products. As a consequence of this review, OCCS developed and negotiated an agreement with NIH's Center for Information Technology that includes Novell products under academic licensing terms, as well as Network Associates' McAfee antivirus suite. While not as dramatic as the MEEC savings, this new arrangement saves NLM an estimated 30% over last years' costs.

Further direct and administrative costs were also saved this year by using consolidated acquisitions to purchase PCs based on standard specifications produced by the Desktop Services Section and the NLM Personal Computer Advisory (PCA) committee. Approximately 200 PCs were acquired this year following the PCA methodology.

Each year, OCCS provides hardware technical specifications for standard computer hardware for review. Then, the agreed-upon specifications are used to make hardware product selections that lead to consolidated acquisitions. This process demonstrated reduced acquisition and desktop support costs of $85,000 this year.

Future Services

The Desktop Services Section (DSS), and the IT Services Center in particular, intends to realign itself more closely to match customer and organizational needs and opportunities. In short, OCCS plans to strengthen customer service and provide more effective ways to meet IT support needs at NLM. Proposed improvements include: ITSC staff will "own" problem calls from when they arrive until the ticket is closed, and will be accountable for improving follow-up activities with our customers; DSS will offer to provide oversight of printer maintenance on NLM printers; the IT Services Center will improve first-call problem resolution of trouble calls; OCCS will reassess our customer's IT training needs; and OCCS will expand and more frequently update the DSS web pages to provide more current and extensive NLM access to ITSC and Desktop Services resources.

Network Support

During FY2001, OCCS continued in its mission to provide reliable LAN and Internet communications services, meet the communication needs for new systems, provide security services, provide end user assistance and training, implement new network-based applications and operating systems, and explore new technologies and plan for systems to meet NLM's continued growth in networking,

services and communications. Looking forward, the Network Engineering Section is taking further steps to increase capabilities of networks and of storage, by providing for: enhanced monitoring and management, Increased security, increased performance and through-put for networks, additional redundancy, enhanced back up, and expanded, centralized and efficient storage. Activities accomplished toward meeting these goals included:

Internet Connectivity Upgrade

Public Internet connectivity services continued to be provided through a contract with Genuity. The T3 (45Mbps) circuit was upgraded to an OC3 (155Mbps) circuit to the Genuity network node in Washington DC in July 2001. The contract also provides an OC3 link for CIT/NIH to the Genuity node in New York. NLM and NIH collaborate in using these links to back up each other's Internet connectivity. PSC and NCI links to the Internet are also provided through this contract.

Network Infrastructure Upgrade

During FY2001 plans were made for upgrading the LAN core connections among OCCS routers and switches. These upgrades will also take place between key network resources and between connections to LHNCBC, NCBI, and NIH/CIT. These upgrades will increase the connection speed from 100 Mbps to multi-Gbp and possibly higher speeds.

Network Management

HP OpenView Network Node Manager remains the primary system used within OCCS to monitor a wide range of hardware and software, such as routers, switches, high-speed connections, Unix systems and Oracle databases. OCCS/FMS staff monitor the health of the NLM networks on a 24-hour basis, seven days per week. These activities will soon take place within the new NOSC (Network Operations and Security Center). The NOSC will allow centralized monitoring and management activities.

Additionally, Cisco Secure was installed for controlling access to the Cisco routers and Cisco Works 2000 was installed for configuration of the Cisco Switches and routers. Additional monitoring packages for specific systems are also used. These include MailCheck and MailCentral to monitor the GroupWise email system, Compaq Insight Manager and Dell OpenManage for Dell and Compaq servers, and DS Expert and DS Analyze for Novell NDS.

Reconfiguration of Network Systems and Cabling

Due to the computer facility reconfiguration and in order to prepare for the NOSC construction, many of the network systems were reconfigured and relocated. Extensive rewiring of cable connections within the computer facility was accomplished to accommodate the equipment moves. In addition, a new distribution fiber backbone is planned for Buildings 38 and 38A to use dedicated cable trays in the ceilings of main hallways. The NOSC will require running new KVM (keyboard, video and monitor) connections for many, if not all, systems connections within the computer room. The equipment to begin a KVM setup has been purchased and additional KVM equipment will be needed.

Network Operating Systems

In order to increase NLM/LAN central storage capability and redundancy, a SAN (Storage Area Network) was implemented. The system started with 300 Gigabytes and an additional 500 GB has been ordered. Many Novell Netware shared department and project network files have been migrated from Novell servers with local disk space to the SAN system. The intention is to create an environment where users' electronic files can be reliably, routinely, and transparently backed-up to storage devices that can be shared across multiple platforms and operating systems.

Upgrades were made to many networked-based applications. Within the GroupWise system, HTTPS support was added to the web access system, Blackberry and Palm support was tested and implemented, and critical

problems were regularly resolved regarding email, mailboxes, security and mailing lists. GroupWise 6.0 was tested for possible release at NLM. Numerous projects were completed for end users and departments such as the implementation of a Filemaker Pro server for HMD.

Extramural Programs Support

The IT functions of the NLM Extramural Programs (EP) Office continued to be supported in EP's off-campus location in the Rockledge I building in Bethesda. Onsite technical support is provided for the PC, network, and IMPAC II systems.

Remote Access

Network support continues to provide 56K dial-in access, cable modem, DSL, and ISDN access for a wide range of NLM users. OCCS recommended Compaq Ipaq computer systems and cable modems as the most effective solution for high-speed access. Unfortunately, cable modems are not yet available at many users' locations. DSL, the second choice, is also not universally available. These technologies were implemented, where possible, for Index Contractors. OCCS intends to provide a dial-in service for users who do not qualify for cable or DSL. Cable and DSL do require the use of additional security software called a VPN (Virtual Private Network) client and a local firewall software package, which were tested, purchased and installed.

In addition to supporting the indexing system, a 3-terminal server setup has been tested and found to be a good fit for flexiplace workers. The terminal server system provides authentication into the NLM network, access to office applications, network based files and the Internet.

Systems Support

FY2001 was the second and final year of major support transition for the OCCS system support team. Although staff continued to be responsible for maintaining legacy systems until they were phased out at the very end of the year, the support focus was deploying and maintaining the hardware and software platforms for new client/server applications. Approximately 100 Unix systems are already built or under construction. The main system support activities for FY 2001 included: installation, maintenance, and support (IMS) of NIS, NFS, DNS, and Web services; Unix O/S IMS for approximately 100 systems; hardware IMS for approximately 100 systems; monitoring, performance, analysis and tuning for approximately 100 systems; Oracle database IMS for 23 applications; security and account administration for approximately 100 systems; Reading Room support for several dozen workstations; and final year of legacy O/S, program product, and application support.

Monitoring

One of the most exciting projects this year was the complete deployment of the HP OpenView/ITO software. This HPOV/ITO software is being used to monitor and, under certain conditions, repair the NLM IT network, OCCS Unix systems, applications, and Oracle databases. After the ITO deployment was completed, the Vantage Point Performance component (HPOV/VPP) was also acquired and deployed. This new product gathers real-time system resource utilization on the Unix systems which is placed into historical databases. These databases are then used to create various reports and graphs of system performance and utilization and are also used to predict future system performance and load.

The Web-Based Daily Historical System Configuration Tracker was installed and customized. This software will be used to view a historical (daily) record of critical configurations and parameters on any system. A daily report of changes made to critical Oracle databases was also deployed.

Applications

Six virtual Apache servers were configured and deployed to support three versions of both the test and production versions of Locatorplus. It enables different types of Locatorplus users, inside and outside NLM, to

use different views of the system. The Voyager system was upgraded in February from Version 99.1 to 2000.1.2 with extensive support provided by OCCS.

Scripts and procedures to create both test and production distributions of MEDLINE XML data on Unix DLT tapes were created, tested, and deployed. OCCS and MMS staff ensured that the data tapes were distributed to and successfully read by the licensees of the data.

OCCS worked with Library Operations to disassemble the List Servers supported by the Unix Systems team. These old List Servers used a limited Shareware product. The operations were converted to a full-featured COTS product maintained by NIH/CIT.

Systems Security

Network security was increased during FY2001. Early in the year, two Lucent firewalls were implemented at the point of connection to the Internet. The firewall policies are configured to deny all traffic, except that which is specifically allowed to pass. This creates the opportunity to better identify unwanted traffic and intrusions.

Security vulnerability scans of network-based systems took place on a regular basis using scanning software. These scans identify weaknesses within Unix, Microsoft, Linux systems and applications. Virus scans occur at the server, email, and desktop levels on an ongoing basis. During 2001, thousands of NIMDA, Red Code, SirCam and other viruses were removed from inbound NLM emails. Many systems were updated with patches, operating system and application upgrades to combat worms, viruses and hacker attacks.

One of the security improvements made early in the year was the activation of a system that requires users of Unix systems to change their passwords at least every six months. This was implemented to conform to NIH policy. Unix systems are still receiving scans from all over the world looking for system vulnerabilities to exploit. During the year, many systems administrators outside the NLM were notified that their systems were being misused. They in turn reported their systems were compromised but were unaware until they received our notifications. Software was installed and configured to scan all incoming/outgoing emails on our Unix email servers for viruses. As a result, every day viruses are detected and quarantined.

Computer Facilities

NLM systems continue to be supported in a safe, secure environment in NLM's Computer Facility, which is available 24-hours-a-day, 7 day-a-week, 365 days a year. OCCS staff provides system monitoring, immediate response and system support services to users within OCCS and other NLM organizations.

A major accomplishment this year was the completion of the NLM computer facility reallocation project. It now provides for better monitoring, more floor space, and improved electrical power support. New Unix and client/server systems were then moved to the computer facility from LHNCBC, NCBI and SIS. These computers previously had been located throughout the building. Key aspects of the upgrade to the Computer Facility included:

- Tiles for the 72 x 90 square foot, 18-inch raised floor that was covered by the mainframe and other decommissioned computer hardware was replaced.
- Dormant mainframe power was removed and new electrical power was added to the computer facility to meet the specifications of various UNIX and client server hardware. Two additional UPS-protected power panels were installed and populated to allow for NCBI projected growth.
- Surplus mainframe computer hardware, furniture and supplies within the computer facility were removed to create useable floor space.
- Environmental surveys on both the air-cooling system and overall electrical usage were conducted. As a result of these surveys, additional UPS protection and air-cooling will be introduced to the facility in the next fiscal year.
- Electrical circuits within the computer facility were moved to provide each

NLM organization with isolated power sources. This allows each organization to perform installations, removal and maintenance of equipment without affecting systems managed by other NLM organizations. Electrical circuits within the electrical power panels were moved in order to provide every NLM organization with potential electrical growth.

- OCCS identified and updated documentation on each computer system's power source and relabeled each circuit in all electrical power panels.

There were a number of key computer facility accomplishments this year. Standard processing each weekend by facilities staff includes complete system pack and database off-site backups. The Tuesday immediately following each weekend, the backup tapes were shipped to a secured class-A volt for storage. There was no unscheduled downtime for mainframe systems. The processing and year-end shipping for MEDLINE licensees was successful.

Facilities Management staff have undertaken certification programs that will provide the Section with a more up-to-date, required skill set. There continue to be daily turnover meetings and discussions are held on the previous 8 hours of production processing as well as on upcoming production scheduling, new programs, procedural changes and scheduled maintenance and/or shutdowns. NLM staff and vendors are welcome to attend.

System Reinvention Initiative

Data Conversion

This year data conversion particularly deserves recognition. The transition from the legacy systems to NLM's reinvented system was complex requiring several working groups. Data conversion was a huge portion of FY 2001 workload.

In the legacy system there were specialized, non-MEDLINE databases, including HISTLINE, SPACELINE, AIDSLINE, BIOETHICSLINE, and POPLINE. Records in these databases contained material from serial/journal titles, monographs, audio visuals, meeting abstracts and journal citations. Conversion from the legacy format to a format suitable for loading in the appropriate target system was a monumental task. It was not a straightforward process and required thousands of hours of analysis and review of legacy data. This activity was completed using the spiral model with seemingly countless cycles of review and modification to specifications and programs.

Voyager Integrated Library System (ILS)

NLM uses the Voyager ILS for acquisitions, serials control, cataloging, collection management, circulation, and preservation. NLM's Online Public Access Catalog, known as Locatorplus, is also a feature of Voyager. Most activities related to Voyager involved the loading of legacy data produced by NLM's data conversion activities, distribution of data to licensees in various formats, and upgrading to the latest release of the Voyager software.

The major ILS accomplishments this year were:

- Accommodating the updated MARC 21 standard, which included modified character set conversion mappings.
- Establishing workflow procedures to automatically send data daily to the Library of Congress and distribute status reports to NLM staff.
- Data created and maintained in Voyager is extracted in XML and distributed to NCBI for use in the journal browser.
- POPLINE, BIOETHICSLINE and monographic chapter records formerly in MEDLINE were added.
- CATFILE*plus* was developed to support US MARC distribution system.

Major ILS Upgrades

OCCS performed a major upgrade of the Voyager system this year. Voyager Gold 2000 was released in the first quarter of 2001 as

projected. Due to the new testing environment, the development team conducted a comprehensive and efficient testing methodology on all upgrades associated with this software system, meeting all projected goals for the Voyager Gold 2000 system.

Interim Binding Module

OCCS developers worked with Library of Operation staff and created a binding module that interfaces with the Voyager system. The system contains Oracle tables for existing binding data elements, user interface/screens for displaying existing binding data and links to the Voyager system. It also performs record editing and can be used to generate Impromptu reports. The Interim Binding Module was deployed on March 21st, 2001 and will be used until the Voyager binding module is available. The purpose of the Interim Binding Module is to allow users to keep track of instructions for binding serials, such as color, binding frequency, and special instructions.

Overnight Photocopy Service (OPS)

A Web-based Overnight Photocopy Service (OPS) system was deployed on February 20th, 2001. The OPS system is unique to NLM. While Voyager is designed to support reading room activities, it does not fully support OPS. When the new release of Voyager was deployed this fiscal year, it no longer supported client software in the reading rooms. Therefore, OCCS developed a web-based replacement for OPS support. Overnight Photocopy Service allows reading room patrons to request NLM staff to photocopy articles and to pick them up the next day. There is a fee for this service.

Data Creation and Maintenance System (DCMS)

The new web-based Data Creation and Maintenance System (DCMS) was initially deployed in June of 2000 to replace the legacy Automated Indexing Management System (AIMS) which had been used to support the creation of MEDLINE for more than two decades. The DCMS supports all journal articles that are indexed with MeSH headings. All data (keyboarded, scanned/OCR, electronic) is received in XML format. The activities of FY2001 focused on completing the transition from the legacy AIMS system to the DCMS. The DCMS was a dramatic change from the legacy system with improvements ranging from software changes to increased throughput via high speed Cable and DSL lines whenever possible.

A component of the DCMS is a relational database that contains more the 12 million journal citations. A primary service of the NLM is to provide this information to licensees. Early in the NLM System Reinvention Initiative it was determined that all journal citation data would be available in the eXtensible Markup Language (XML). A Data Type Definition (DTD) developed by OCCS/LO and NCBI is used to define the structure of this XML. This XML standard is the only distribution format for MEDLINE indexing data created in 2001. NLM leased data for 2001 was available by either FTP or DLT tape for large quantities of data such as the entire retrospective class-maintained MEDLINE load for 2001.

During FY2001 enhancements to the DTD were made to support the non-MEDLINE journal citations. Distribution in FY2002 will contain data extracted from all legacy citation databases.

The DCMS provides all legacy mainframe-based functionality as well as various enhancement especially in support of electronic journals. It provides citation creation, online journal assignment, journal tracking function, SGML article review, SGML issue verification, and citation maintenance functions. The DCMS has interfaces with several other systems, including Voyager, MeSH2000, PubMed and all licensees of MEDLINE data.

The initial deployment supported the creation of MEDLINE citations only. Phase two of the implementation supported maintenance of completed citations. Journal citations in the DCMS are also in PubMed, they may be Pre-MEDLINE records, they may be completed records, and they may or may not be distributed to data licensees.

The final phase of implementation of the DCMS supported creation and maintenance of

the non-MEDLINE citations, which are now in a single relational database.

Publications

The *List of Journals Indexed* (LJI) and *List of Serials Indexed* (LSI) publications were produced from extracts of the Voyager systems. In addition to hardcopy, the publications are now available online on the NLM public FTP service in PDF and DOS text format.

Index Medicus had been produced in the legacy system for more than two decades. Transition to the new environment was extremely complex, tedious and time consuming. A workflow was developed that describes the process from data extraction to publisher delivery. *Index Medicus* will now be produced in a PDF format that will significantly streamline the process.

DOCLINE

DOCLINE version 1.2 was released on May 6, 2001. DOCLINE is NLM's online interlibrary loan request routing and referral system that processes more than 3 million interlibrary loan requests annually for 3,000 U.S. and Canadian medical libraries. OCCS and LO staff continue to improve the production of the more than 30,000 reports that are distributed to DOCLINE libraries. These reports are available in electronic form via FTP. The four main functions of DOCLINE are:

- DOCUSER, which provides directory and interlibrary loan information on participating libraries;
- REQUESTS, which allows users to make document requests that are routed automatically to libraries that report owning the specific year or volume requested;
- SERHOLD, which provides journal holdings information; and
- Loansome Doc Patron Administration, which allows libraries to maintain administrative information on their Loansome Doc users.

The web-based DOCLINE system is a key step in NLM's Systems Reinvention. Developed at NLM, the system interfaces seamlessly with other NLM products and services including PubMed and the online catalog Locator*plus*. DOCLINE 1.1 was released on October 30, 2000.

Software was developed and implemented to make weekly distribution of DOCUSER data available to the RML's. A subsystem of DOCLINE was developed to manage NLM's quarterly invoices for ILL activity and produce files for NTIS to send out invoices.

Relais

Relais 3.2 was implemented on September 3, 2001. It introduces the new delivery method, Post-to-Web, which delivers documents to a web server and emails the requestor a URL for remote downloading and printing. The new Web delivery method allows NLM to proceed with differential pricing for electronic delivery. Relais 3.2 also introduced a document resend feature, improved workflow monitoring, administrator tools, and improved usability.

The NIH Library purchased and implemented the Relais system this year. NLM and NIH systems staff worked together with Relais International to implement the transition of NIH to its own system. NIH continues to use NLM-supported files to download and update DOCLINE requests.

Classification

Every five years, the *NLM Classification* is published. The fifth edition was published in 1994. The *NLM Classification* is a scheme for the shelf arrangement of medical literature in libraries. Significant progress on building a new web-based classification system to replace the existing client/server version was made this year. The legacy database was converted to an Oracle database. The OCCS project team developed a prototype to demonstrate the search results to the Cataloging Section in order to finalize the implementation

approach. The search methodology was then implemented. OCCS and the Cataloging Section made some necessary changes to the classification website to demonstrate it to the NLM Web Editorial Committee at the MLA annual meeting (where it received a standing ovation). Full implementation of all phases of the classification is scheduled for FY 2002.

MeSH2000

MeSH is NLM's controlled vocabulary thesaurus. It is used for cataloging, indexing, and searching MEDLINE and other NLM databases. OCCS completed development of the new custom client/server MeSH2000 in October 1999. As part of the reinvention project, the underlying data structure of MeSH was altered to afford a concept-based representation that is more compatible with the UMLS Metathesaurus. Together with the new DCMS, this system will simplify the annual maintenance of MEDLINE records. The Applications Branch staff handled problem reports and performed the following enhancements or modifications:

- Cut over for chemicals to prepare 2001 Supplementary Concepts.
- Distributed 2001 MeSH to vendors worldwide.
- Upgraded underlying infrastructure (UNIX/ORACLE) to enhance performance and to provide UNICODE support.
- Developed routines to monitor DCMS/MeSH daily activities.
- Developed and implemented software necessary for the implementation of procedures to synchronize the DCMS with the current year of MESH. These include changes in Supplementary Chemical Records, Descriptors and Qualifiers Records.
- Developed Data Type Definition and programs to support creation of XML files for Descriptors, Qualifiers, and Supplementary Chemicals.
- Developed and implemented new subsystem for data entry and editing capabilities to Qualifier Trees.
- Major architectural changes to provide Merge/Promote capabilities in Current SCR's and features to provide Merge/Promote functionality in the New Qualifier Records.
- Distributed MeSH files for the MeSH Browser.

FY2002 will be the initial execution of the annual updating of MeSH terms in journal citations to the 2002 MeSH year with software developed during NLM System Reinvention. Critical tasks for Year-end-processing (YEP) implementation were completed, including production of the M2000 Descriptor, Qualifier, and Chemical transactions for the M2000/YEP processing. All preparation is complete for YEP which is scheduled for FY 2002.

Health Services Research Projects (HSRPROJ)

The legacy HSRPROJ system was re-engineered to utilize methodologies common to NLM’s reinvented systems. A DTD was developed and data were extracted from the mainframe in XML. The new system is Web-enabled utilizing an Oracle database as the depository. The data were converted to an ORACLE database and will be maintained with software tailored to the HSRPROJ application.

OLDMEDLINE

The legacy OLDMEDLINE system was re-engineered to utilize methodologies common to NLM’s reinvented systems. A DTD was developed and extracted from the mainframe in XML. Many data issues and rules were discussed and decided with MMS in order to prepare the data for conversion. There are approximately 1 million records. The new system is Web-enabled utilizing an Oracle database as the depository. The data were converted to an ORACLE database and will be maintained with software tailored to the OLDMEDLINE application.

Literature Selection Technical Review Committee (LSTRC)

Beta versions of the reinvented LSTRC system had been under development for more than two years. This work could not be completed until other key components of the system reinvention were finalized. The reinvented LSTRC system was placed in production in the last quarter of FY2001. The new system is accessed via a Web browser, utilizing ORACLE for specific LSTRC data with interfaces to the Serial Extract Database for serial information.

All journal titles reviewed at any LSTRC meeting can now be searched by the respective review date. Also, system updates can easily be made for categories of data that are used for entries to journal records. This includes adding or deleting optional "types of contents" that are entered into the records when the title is screened prior to a committee meeting and the names of review groups used for in various medical disciplines can be added at any time. New data fields were carefully formatted and added to the database.

MEDLINEplus Consumer Health Information

NLM's consumer health information service, MEDLINE*plus*, contains carefully selected links to Web resources with health information. MEDLINE*plus* generated 62 million page views in FY2001. Six versions of MEDLINE*plus* were released during the past fiscal year. Version 5.01 added a frequently asked questions (FAQs), a tour of the site, a welcoming video by Dr. Lindberg, and a how-to-link-to-us page. Version 5.02 provided Library Operations with the ability to edit and deploy some of the pages on MEDLINE*plus*. Version 5.5 added a news feed from the daily print media. Version 5.7 included interactive Patient Education Institutes Health tutorials. Version 6.0 consisted of the locally hosted ADAM.com medical encyclopedia. Version 6.5 was primarily a maintenance release; the entire application was upgraded to support the latest release of the ColdFusion server and new hardware and file system structure.

Data Sharing

NLM is interested in collaborating with other organizations to organize local, state and regional health information, linking this information to the national and international information organized by MEDLINE*plus*. To facilitate this, a Data Sharing project was initiated. Phase I included making a delimited flat file of national-level, consumer health records available, so institutions could load MEDLINE*plus* data in local systems on a regular basis. The file was released February. OCCS developers worked closely with the MEDLINE*plus* team to build an automated process, so a Data Sharing flat file will be created automatically each time MEDLINE*plus* is refreshed.

Phase II focused on distribution of a flat file with health topic names and associated URLs. OCCS developers worked with a LO functional group on user and system requirements. Teleconference meetings with outside users were conducted and a draft flat file was made available for user feedback.

National Center for Complementary and Alternative Medicine (NCCAM)

This project was initiated to create a database of citations in all areas of CAM so that a searcher can be assured of reasonably accurate and timely results. Organizations involved in this project were LO, OCCS, NCBI, and NCCAM. OCCS provided suggestions regarding project implementation and provided user requirements documentation and data flow to the team. OCCS also provided technical support to NCCAM by creating several web sites and by building a link to PubMed and other links. The current CAM Citation Index (CCI), which is located on the NCCAM web site, was replaced by a link to PubMed. PubMed will utilize a search strategy to flag a subset of citations that are associated with Complementary and Alternative Medicine to link to NCCAM. The project was completed and deployed on February 5, 2001. OCCS received a letter of appreciation from NCCAM representatives for its work.

Senior Health Project

The Senior Health project is a joint effort between the National Institute on Aging, the National Health Council, and the NLM. The prototype web site contains a tutorial, quizzes and health topics information, etc which will help senior Americans better understand and remember health information. OCCS solved problems related to QuickTime, tested accessibility under different versions of browsers, and converted several HTML files into style-sheets in order to ensure that web pages display properly under the PC and Mac. A test environment for RealMedia server, window Media Server, and QuickTime streaming server will available in FY 2002.

Outreach

The Outreach Database Project was established to develop a database of outreach projects and activities in order to provide a centralized depository of NLM's Outreach efforts. OCCS team members met with the project coordinator to discuss the user requirement and data elements.

In another project, OCCS staff worked on the development of the HSRTools database and interface for NLM's National Information Center on Health Services Research and Health Care Technology (NICHSR). NICHSR's mission is to make the results of Health Services Research (HSR) available to the Health Services Research community (researchers, health care practitioners, and public health professionals). The HSR databases contain data surveys and other information such as clinical practice guidelines and health care technology. Based on input and feedback received, revisions were made to enhance usability of the HSRTools database. The database was demonstrated at the Health Research Conference in June 2001.

OCCS also completed a PL/SQL program to load R.O.W. ASCII data to an Oracle database. More effort is required on the DTD for data submission and the search engine.

NLM Web Page

Search Engine

An improved search engine was one enhancement to the NLM Web site this year. NLM uses ht://Dig as a search engine for the Main NLM Web site. OCCS is constantly monitoring and refining the search engine, as well as periodically reviewing the capabilities of engines other than ht://Dig.

A spell checker from Wintertree that has both a regular English dictionary and a medical dictionary is being used to improve search capabilities on the NLM home page. The team successfully deployed Spell Check into MEDLINE*plus* production. It will help users with misspellings and selecting wrong medical words. In addition, OCCS/LO worked with CIT to build one custom dictionary, which contacts all USP drug names/words.

The team successfully worked with the MEDLINE*plus* team to implement a solution where MEDLINE*plus* users can automatically switch to a search engine without any interruption when Spell Check is down. OCCS team members also worked with an LO Associate on a project to compare several search engines to identify their advantages and disadvantages.

Web Content Management Software

A web content management tool called TeamSite has been chosen. It features a thin-client, server-driven design to manage large-scale operations with minimal overhead on client systems. Completed Web content can be served from the TeamSite server or deployed to any number of production Web servers.

NLM Application Server

All production servers and development servers have been upgraded to ColdFusion 5.0. The OCCS Web Support team worked with

Systems Technology Branch to establish a robust production environment. The team developed DB/ColdFusion stress tests to simulate database connection problems. They also developed ColdFusion applications to exercise the tables (via Microsoft's Web Application Stress Tool).

Technical Bulletin

The MEDLARS Management Section requested functionality to allow users to print the entire issue of the *NLM Technical Bulletin* instead of only one article at a time. OCCS created a new template to support a "print all" function.

Statistics Reporting Package

Active Concepts' FunnelWeb Pro was selected to analyze the various log files, including streaming media analyses, cluster analysis, proxy analysis, support for virtual domains, click-stream analysis, incremental log analysis, remote administration via web, and online advertising analysis. All daily, monthly, and quarterly log reports are available on the NLM Intranet.

The OCCS development team worked closely with the LO Tag team to design and implement on-demand reports. Most of the on-demand reports have been implemented and testing has started for annual reports.

Administrative Support Systems

This year, OCCS continued to increase support for internal customers at NLM. OCCS worked on four administrative support systems this year: Inventory Control, Online Personnel Policy; NLM Administrative Manual; and Online Request for Service.

The OAM Inventory Control System is an online customer Ordering/Inventory Control System for the Office of Administrative Management Services. This system allows NLM users to order office supplies online and assists OAMS in inventory management. The team has completed the last enhancement successfully. Approximately 600 photos are in the production and staff can review the online catalog before placing an order. The team also designed a user interface screen to ensure that the system could be accessed from the Intranet.

The Personnel Administrative system tracks personnel information, including employee information, recruitment actions, personnel actions, and award information. The application provides easy entry using drop-down menus. It supports information validation and dynamic report generation. Because of the sensitivity of the personnel data, a secure access to personnel information was implemented. The Personnel Administrative Control system was successfully released to production in September 2001.

The NLM Administrative Manual is divided into four sections: Manual Chapters, Delegations of Authority, Functional Statements, and Organizational Charts. A separate searching feature is available that limits searches to Manual Chapters or Delegations of Authority. The application features include a standard template format for the NLM Manual Chapters and Delegations of Authority, conversion of existing Manual Chapters and Delegations of Authority into the new template, and the capability for the Web user to perform full-text searches of chapters and delegations.

The OAMS Request for Service System project is an automated online system for receiving and tracking requests for service from NLM individuals or program areas. The services covered by the system include maintenance trouble calls, telecommunications work/trouble calls, and transportation and messenger service. All requests for service were being tracked via a paper trail. The process is very time consuming and OAMS has requested an automated online system for receiving and tracking requests for service from NLM individuals/program areas. A new Request for Service System is being developed.

OAMS Phone Directory Project

OAMS produced a paper version of the NLM phone and staff directory. The process was very time consuming and the information was not consistent with other databases in the NLM. Work so far has focused on centralizing the information and working with NIH technical

staff to resolve issues with the NIH staff directory. Three phone directory formats have been created and presented to OAMS management and NLM senior staff for review.

ADMINISTRATION

Donald C. Poppke
Associate Director for Administrative Management

NLM Facilities Expansion

The design for expanding NLM's existing facilities is moving forward. On July 24, 2001, President Bush signed the 2001 Supplemental Appropriations Act (P.L. 107-20). The Conference Report accompanying the bill (H. Rept. 107-148) contained the following language:

> *Of the amount appropriated in the Departments of Labor, Health and Human Services, and Education, and Related Agencies Appropriations Act, 2001 (as enacted into law by Public Law 106-554) for the National Library of Medicine, $7,115,000 is hereby transferred to Buildings and Facilities, National Institutes of Health, for purposes of the design of a National Library of Medicine facility.*

This transfer of funds, along with funds previously transferred to the NIH Buildings and Facilities account, clears the way financially for the completion of the design for the new and expanded facilities.

Working through the Army Corps of Engineers as the contracting office, the NIH, on August 16th, made an award for the completion of the design to the 35 percent level. Clearance was provided by the Small Business Administration (SBA) to award the contract to CETROM, with the major design work subcontracted to Perry Dean Rogers of Boston. These are the two firms NLM has been working with to develop the initial Program of Requirements and will streamline the entire process in terms of time, effort and cost. Final arrangements are in place through the SBA to allow options to complete 65 percent and 100 percent of the design without recompeting the contract.

We estimate that it will take approximately eight months to complete the 35 percent contract and another 7 or 8 months to bring the design to one hundred percent. This level is equivalent to complete working drawings. At that time, to move forward, Congress will need to appropriate funds to the NIH Buildings and Facilities account to begin the actual construction. Assuming work begins in September uninterrupted, funding to begin construction could be needed by December 2002 (FY 2003).

System Reinvention Activities

The NLM System Reinvention was a high-priority initiative conducted by NLM in support of its role as a reinvention laboratory under the National Performance Review. The project was designed 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. This multi-year effort was completed in FY2001. A summary of FY 2001 activities include:

Integrated Library System: NLM acquired and installed Voyager, an integrated library system (ILS) in FY1999. Voyager supports all aspects of traditional library services. A major effort of FY2001 was to supplement the Voyager database with monographic material from specialized NLM legacy databases including SPACELINE, HEALTHSTAR, HISTLINE, BIOETHICSLINE and POPLINE. A collaborative effort among several NLM organizations resulted in Voyager becoming the single database where serial information is maintained. This centralization resulted in more consistent data for all NLM products and services (such as PubMed's Journal Browser) and also reduced the workload for maintaining multiple copies of the data.

NLM has agreements with many organizations to provide electronic copies of material cataloged by the NLM. These agreements vary in content and scope. NLM continues to be responsive to requests of this type. Customized procedures were developed to

ensure data recipients received the data in a form that could be processed in a most efficient manner.

PubMed Retrieval System: PubMed is a World Wide Web retrieval service developed by NLM that provides access, free of charge, to MEDLINE, a database of more than 11 million bibliographic citations and abstracts in biomedicine. As part of the System Reinvention initiative, the MEDLINE database in PubMed was expanded to include the journal citations that have been in the HEALTHSTAR database. PubMed also contains links to the full-text versions of articles at participating publishers' Web sites. In addition, PubMed provides access and links to the integrated molecular biology databases maintained by the NCBI. These databases contain DNA and protein sequences, genome mapping data, and 3-D protein structures. MEDLINE/PubMed has been widely accepted by the biomedical community and consumers as a useful, complete, confidential and authoritative source of health information.

Data Creation and Maintenance: The Data Creation and Maintenance System (DCMS) replaced several legacy systems used for online indexing and editing of bibliographic citations for MEDLINE and derived files. All completed citations available to the world via PubMed and other retrieval services are created using the DCMS application. This Web based system began a phased implementation in FY2000.

Efforts during FY2001 covered a wide range of activities. The DCMS replaced a legacy system that had been in operation for more than two decades. User training was provided with follow-up to ensure optimum use of the new system. Software revisions were made as a result of user feedback. In order to ensure high throughput, high-speed remote access was provided. The initial implementation of the DCMS supported record creation only. Phase 2 made it possible to maintain records previously created.

The final phase of implementation introduced support for what has been referred to as MEDLINE derived files. In the legacy system this included HEALTHSTAR, AIDSLINE, HISTLINE, SPACELINE, BIOETHICSLINE and POPLINE. In the legacy system a separate database and separate software system were required to support each of these systems. The DCMS was designed to fold these specialized citations into a single database and software system. Data conversion from the legacy system was extremely tedious and time consuming. Workflow and record ownership rules were established and implemented to ensure a collaborative effort between NLM and outside data providers.

The distribution of MEDLINE data to licensees is a major service of the NLM. The DCMS made it possible to streamline this data distribution. All completed citations are available via FTP in a single Data Type Definition (DTD) in the eXtensible Markup Language (XML). This industry standard format was widely accepted by all licensees. The initial distribution supported MEDLINE only. Data distribution in FY2000 supported only MEDLINE data. Beginning in FY2002, the DTD and XML will be modified to add support for MEDLINE and data from the MEDLINE derived files.

The DCMS resulted in a database of relational citations. Redundant databases and redundant data have been eliminated. A major factor for this design was the annual updating of journal citations (now more than 12 million) to the current year of Medical Subject Headings (MeSH). In the legacy system, redundant data and databases resulted in a 9-month effort for planning and updating of all the legacy data. This 9-month effort required thousands of person hours from a team of 15–20 members. With the DCMS this effort has been greatly reduced. There is still a significant effort by a small group to ensure the citation update transactions are accurate. Once completed and verified, however, the machine processing of updating all 12 million citations will take only 1 day as compared to the legacy system where it was a very long process indeed.

Document Delivery: DOCLINE is the Library's automated interlibrary loan (ILL) request and routing and referral system. The purpose of this system is to provide improved document delivery service among libraries in the National Network of Libraries of Medicine (NN/LM).

The new DOCLINE was implemented in FY2000. This Web-based system replaced 3 legacy systems that had served the biomedical community for over 15 years. As the transition to the new system evolved in FY 2001, significant coordination and communication with the DOCLINE community was required. LISTSERV and email facilities were established. Conference calls were held frequently. Enhancement requests were prioritized and are being implemented by the project team. DOCLINE currently supports 3 million interlibrary loan requests annually.

The NLM Gateway: The NLM Gateway was introduced in FY2001. It presents a single interface that lets Internet users search simultaneously in multiple NLM retrieval systems. It is intended to provide an overview scan for the user who comes to NLM not knowing exactly what is here or how best to search for it. The Gateway provides "first-stop shopping" for an increasing number of NLM resources—currently searching 11 document collections using 5 retrieval access methods. Journal articles; books, serials and audiovisual materials; consumer health information; meeting abstracts; and other information are available from a single search. The Gateway group has most recently added access to the DIRLINE database on the TOXNET system from NLM's Division of Specialized Information Services. We have also begun planning for access to the important and timely information in the Hazardous Substances Data Bank.

Financial Resources

In FY2001, the Library had a total appropriation of $239,189,000. Table 11 displays the FY2001 authority plus reimbursements from other agencies.

The FY 2001 appropriation language authorized the Library to use personal services contracts and provided for the availability of $4.0 million without fiscal year limitations. These authorities are key elements of NLM's system reinvention initiative.

Table 14

Financial Resources and Allocations, FY 2001
(Dollars in Thousands)

Budget Allocation:

Extramural Programs	$51,445
Intramural Programs	177,516
Library Operations	(71,733)
Lister Hill National Center for Biomedical Communications	(51,966)
National Center for Biotechnology Information	(43,530)
Toxicology Information	(10,287)
Research Management and Support	10,228
Total Appropriation*	239,189
Plus: Reimbursements	11,568
Total Resources	**$250,757**

*Excludes $47,000 for the Secretary's 1% transfer.

Personnel

In October 2000, **Jane Bortnick Griffith** joined the Library as the Assistant Director for Policy and Legislative Development. In this new position, Ms. Griffith will serve as a key advisor in policy development affecting the Library, especially as it relates to science and technology information. Ms. Griffith has worked for more than 25 years in the field of information science and technology policy analysis. For the last two years, she was director of a Task Force under the aegis of the National Academy of Sciences, National Academy of Engineering, and the Institute of Medicine that examined the goals, organization, and operational effectiveness of the National Research Council. Before that, she was a senior specialist in the Library of Congress in many areas of critical interest to NLM, including federal funding for advanced information technology. Ms. Griffith holds a B.A. in American history from the University of Wisconsin and a M.A. in American history from Rutgers University.

In October 2000, **Olivier Bodenreider, M.D., Ph.D.**, joined the staff of the Lister Hill Center as a Staff Scientist. Dr. Bodenreider, a native of France, received his M.D. degree in 1990 from the University of Strasbourg School of Medicine in France. He received his Ph.D. in 1993 in medical informatics from Henri Poincare University, Nancy, France. For six years he held a position as Assistant Professor of Medical Informatics and Biostatistics at the University of Nancy, while also working as an attending physician at the university hospital. For the past several years, Dr. Bodenreider has been a Lister Hill Center research contractor, working on the Unified Medical Language System project, the Indexing Initiative project, and the Clinical Trials Database project. Dr. Bodenreider recently developed sophisticated mapping methods for the browse capability in the ClinicalTrials.gov system. As a Lister Hill Center scientist, Dr. Bodenreider will work on medical knowledge representation research.

In October 2000, **Colleen M. Guay-Broder** was appointed to the staff of NCBI as a Program Analyst. Prior to coming to NCBI, Ms. Guay-Broder was employed as a Program Analysis Officer at NIDDK. Ms. Guay-Broder holds a B.S. in biological sciences from the Florida Institute of Technology, Melbourne, Florida and a M.H.S.A. in health service administration and policy from George Washington University. She first came to the NIDDK in 1991 as a Biologist in the Laboratory of Cell Biology and Genetics. Ms. Guay-Broder serves as a Special Projects Officer for NCBI, responsible for program planning and development in scientific areas that cut across organizational lines of the Center and have significant public policy implications. She's also responsible for developing long-range planning documents and other related studies and will track progress on program objectives.

In November 2000, **Mr. Michael L. Feolo** joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Mr. Feolo received his Bachelor's Degree in biology in 1996 from the University of Utah. Mr. Feolo is currently a candidate for a M.S. degree in the medical informatics program at the University of Utah, with a specialization in genetic epidemiology. His research project has focused on the genetic analysis of Celiac Disease. In his research, Mr. Feolo both developed a strategy and operational protocol for high-throughput HLA-DQ typing, and conducted linkage analysis of candidate genes and whole-genome association scans in large (90+) pedigrees of Celiac Disease. Mr. Feolo's familiarity with the genetic, statistical and clinical aspects of HLA data and the ability to organize and compute on the data make him an ideal addition to the NCBI software group.

In December 2000, **Wolfgang M.C. Helmberg, M.D.**, joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Helmberg, a native of Austria, received his M.D. in 1992 and his specialization degree in transfusion medicine 1999 from the University of Graz, Austria. At NCBI, Dr. Helmberg serves as a project manager whose duties involve a two-pronged approach: that of a curator of genotype and phenotype data from both the human immune system (HLA) and other human genes surveyed for their possible role in autoimmune disorders as well as that of public liaison. As a curator, Dr. Helmberg will design and implement an internal NCBI database to serve as both permanent archive and point of public redistribution of data. He will also work with other staff scientists and software engineers at NCBI to integrate this data with NCBI resources such as GenBank, LocusLink, PubMed and dbSNP.

In January 2001, **Christopher J. Lanczycki, Ph.D.**, joined the staff of the Computational Biology Branch NCBI as a Staff Scientist. Dr. Lanczycki received his Ph.D. in computational physics in 1995 from the University of Maryland. Prior to joining NCBI, Dr. Lanczycki was a Staff Scientist with the Center for Information Technology, NIH. His projects have focused on applying high-performance and parallel computing techniques in the computationally intensive areas of protein structure determination and three-dimensional virus structure reconstruction. At NCBI, he will develop the computational infrastructure for a new generation of protein databases that combine structural and evolutionary information for the purpose of robust classification of proteins and reliable prediction of their activities and functions. In addition, he will oversee and

implement the software that is required for the database of Clusters of Orthologous Groups of Proteins.

In January 2001, **Carol Bean, Ph.D.**, joined NLM's Division of Extramural Programs to help determine the direction of programs in the area of informatics as applied to health care delivery and to medical scientific research. Dr. Bean is herself a graduate of one of the NLM training programs at Columbia University. Dr. Bean has also worked for the Cognitive Science Branch within the Lister Hill Center. Most recently, Dr. Bean was Assistant Professor at the School of Information Sciences, University of Tennessee. She has an M.S. in Medical Informatics from Columbia University, an M.L.S. in Information Science from the University of Maryland, and a Ph.D. in Biopsychology from the University of Georgia. Dr. Bean's background, membership in appropriate professional associations, and wide acquaintanceship with peers within the informatics community will be invaluable.

In January 2001, **Darren A. Natale, Ph.D.,** joined the staff of the Computational Biology Branch of NCBI as a Staff Scientist. Dr. Natale received his Ph.D. in Molecular Biology from the State University of New York at Buffalo in 1993. He performed post-doctoral work at the Roche Institute of Molecular Biology and then at the NICHD. Previously, Dr. Natale was employed by Computercraft Corporation and worked at the NCBI as a contractor. His responsibilities involved the maintenance, curation, and advancement of the Clusters of Orthologous Groups of proteins (COG) database, and during the last year, he has been leading a group of 5-6 expert contract curators of this database. Also, he is contributing to the construction of interfaces between the COG database and other databases and retrieval systems such as GenBank and Entrez, and is helping to create the database of Reference Sequences for complete genomes.

In February 2001, **Valerie Florance, Ph.D.**, was named Grants and Contracts Program Specialist within the Division of Extramural Programs. Dr. Florance received her B.A. in cultural anthropology and an M.A. in medical anthropology from the University of Utah. In addition, she completed an M.L.S. degree from Brigham Young University, and a Ph.D. degree in library and information sciences from the University of Maryland. For the past two years, Dr. Florance worked as Project Director at the Association of American Medical Colleges where she was responsible for the design and execution of a project to create a set of recommendations for the best ways for American academic medicine to utilize information technology during the next 10 years. At NLM, Dr. Florance will be responsible for providing scientific leadership and direction for a program in the field of biomedical information management.

In February 2001, **Richard M. von Sternberg, Ph.D.,** joined the staff of the Computational Biology Branch of NCBI Branch as a post-doctoral Fellow. Dr. von Sternberg received his Ph.D. in systems science (theoretical biology) from Binghamton University, Binghamton, NY, in 1998, and a Ph.D. in biology from Florida International University, Miami, FL in 1995. He was a Research Associate at the National Museum of Natural History, Smithsonian Institution, Washington, D.C., where he studied the relationship between genomic organization and morphological traits. At NCBI, Dr. Sternberg will research taxonomic issues. He will perform systematic analysis and develop novel pattern recognition programs for the information analysis of protein, nucleotide, and morphological databases.

In February 2001, **Eva Czabarka, Ph.D.,** joined the staff of the Computational Biology Branch of NCBI Branch as a Research Fellow. Dr. Czabarka, a native of Hungary, received her Ph.D. in combinatorial mathematics from the University of South Carolina, Columbia in 1998, followed by two years of course work in statistics before coming to NCBI in January 2000 as a Fogarty Visiting Fellow. Since then, she has been working on the statistics of structural alignments. In particular, she has demonstrated the ability to solve difficult mathematical problems and program computer solutions in C++. Her work has resulted in the possibility of introducing gapping into structural statistics, which is very likely to improve the VAST structure matching

computer-program at NCBI. She is presently working to improve the VAST statistic further.

In April 2001, **Ms. Carol Myers** joined the staff of the Information Engineering Branch of NCBI as a Staff Scientist. Ms. Myers received her MLS from Catholic University of America, Washington, DC, in 1992. Ms. Myers has 10 years' experience managing the technical services departments of two Washington, D.C. law firms. Ms. Myers was previously employed for 12 years at the Navy Ships Parts Control Center in Pennsylvania. Ms. Myers has worked at NCBI as a contractor since February 2000. She will be NCBI's first-line point of contact for the representatives of publishers and other data suppliers who submit electronic citation files to PubMed. She will also coordinate NCBI's requirements with the needs of the data suppliers and other NLM departments, and provide routine technical review of sample XML data from providers prior to processing. In addition, Ms. Myers will help write documentation of NCBI and NLM procedures and policies for publication on the Web.

In April 2001, **Clifford O. Clausen, Ph.D.,** joined the Information Engineering Branch of NCBI as a Staff Scientist. Dr. Clausen received his Ph.D. in information technology from George Mason University, Fairfax, VA in 1999. Prior to working for NCBI, he spent 10 years as an officer in the Army in various capacities including an operations research analyst responsible for developing new simulation software. Following military service, Dr. Clausen worked for the Unisys Corporation for 14 years where he developed information and decision support systems for government agencies. Dr. Clausen has previously worked for NCBI as a contractor, where he developed infrastructure applications and supported the NCBI effort to convert its reusable software library from C to C++. As an NCBI staff member, Dr. Clausen will continue to work on infrastructure projects including development and enhancement of internal web applications to support software management functions.

In April 2001, **Olivier Lespinet, Ph.D.,** joined the staff of the Computational Biology Branch of NCBI as a Visiting Fellow from France. Dr. Lespinet received his Ph.D. in molecular and cellular genetics from the Centre de Genetique Moleculaire of the CNRS in Gif-sur-Yvette Cedex, France in 2001. His primary experience is in evolutionary biology and development, but he also received professional training and has teaching experience in bioinformatics. At NCBI, Dr. Lespinet will perform research on evolutionary genomics of animals using methods of computational biology. Genome-wide comparisons of protein sequences are expected to facilitate the solution of fundamental problems of evolutionary biology such as the origin of animal developmental mechanisms and the relationships between the major animal taxa.

In April 2001, **Yoshimi Toda, Ph.D.,** joined the staff of the Computational Biology Branch of NCBI Branch as a Visiting Fellow from Japan. Dr. Toda received her Ph.D. in bioinformatics from Keio University, Tokyo, Japan in 2000. She is an expert in vertebrate repetitive elements; her thesis focused on computational analyses of the Alu elements in primate genomes. At NCBI, she will be responsible for maintaining a repetitive elements database. This database is used for screening and masking genomic sequences for repeats, a crucial step for efficient searching against genome sequence databases. She will also create a comprehensive collection of repetitive elements from fungi, part of a long-term collaboration with the RepBase, an online resource for genomic repetitive elements.

In May 2001, **Ilya V. Dondoshansky, Ph.D.,** joined the staff of the Information Engineering Branch of the NCBI as a Staff Scientist. Dr. Dondoshansky received his Ph.D. in applied mathematics from the University of Maryland, Baltimore in 1996. Following work as a software developer at Bloomberg, L.P., Dr. Dondoshansky began his employment at NCBI as a contractor for Management Systems Designers in the BLAST group in November 1999. Dr. Dondoshansky has worked on BLASTCLUST and a version of TBLASTN. BLASTCLUST is a program for clustering protein and nucleotide sequences, and TBLASTN uses newly developed sum statistics that are specifically designed to handle the case of proteins encoded by multiple exons. By combining experience in BLAST mathematics and sophisticated software development, Dr.

Dondoshansky continues to play an important role in maintaining and developing one of the most heavily used resources at the NCBI.

In May 2001, **Jeffrey D. Beck** joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Mr. Beck received his B.S. degree in English and Mass Communication from Towson University, MD in 1987. Prior to working for NCBI, he spent seven years at Cadmus Corporation, where he served as a production editor for the Journal of Biological Chemistry. Mr. Beck was also an E-Doc project manager, where he led a team responsible for the production of more than 100 online journals. Since March 2000, Mr. Beck has been working as a contractor for the Kevric Company working on the PubMed Central project at NCBI. Mr. Beck's skills and experience will be extremely valuable in testing, validating and refining data in various formats sent by publishers.

In June 2001, **Jian Ye, Ph.D.,** joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Ye received his Ph.D. in microbiology and immunology from the University of North Carolina, Chapel Hill, in 1995. Dr. Ye did work on bioinformatics as a Postdoctoral Fellow at the NCBI from October 1998 to May 2000. During this period, Dr. Ye worked on a wide variety of projects, including IgBLAST that uses the BLAST algorithm to search for immunoglobulin sequences. Dr. Ye then joined Curagen Corporation as a Senior Research Scientist in May 2000 where he developed a system to run, parse, and analyze BLAST results. BLAST is one of the most heavily used resources at the NCBI. In addition to his ability to perform sophisticated software design.

In June 2001, **David I. Hurwitz**, joined the staff of the Computational Biology Branch, NCBI as a Staff Scientist. Mr. Hurwitz received a B.S. degree in electrical engineering at Brown University, Providence, RI in 1981; an M.S. in biomedical engineering at Case Western Reserve University, Cleveland, OH, in 1985; and a second M.S. degree in chemical physics at the Weizmann Institute of Science, Rehovot, Israel, in 1991. Mr. Hurwitz held several positions in software engineering between 1981 and 1996 where he worked on control systems for medical instrumentation and developed image processing algorithms. In 1997, Mr. Hurwitz returned to research in the field of computational chemistry. Since January 2000, Mr. Hurwitz has been under contract with Management Systems Designers, Inc. and has been part of the NCBI group, where he worked on projects related to Cn3D.

In July 2001, **Siqian He, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as a Staff Scientist. Dr. He received his Ph.D. in biophysics from the University of Minnesota, Twin Cities, MN in 1992 and his Sc.D. in applied mathematics from MIT, Cambridge, MA in 1996. Dr. He has worked with NCBI as a contractor since August 2000. Dr. He has applied his background in 3D structure information services to modify the SYBASE database which is used to archive and distribute protein 3D structure data for NCBI's Entrez retrieval service. Dr. He has also developed a new tracking and retrieval database for structure alignments of 3D-domain pairs. Through his experience in scientific programming and algorithmic manipulation of 3D structure data, Dr. He has made important contributions to the 3D-structure information services of the Computational Biology Branch.

In July 2001, **Ron Edgar, Ph.D.,** joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Edgar received his Ph.D. in chemistry in 1998 from the Weizmann Institute of Science, Rehovot, Israel. He joined NCBI as a Visiting Fellow in August 1999. Dr. Edgar has coupled his extensive knowledge of computer programming languages and operating systems to training on the Gene Expression Omnibus project (GEO). He has also assisted in the development of a sophisticated indexing engine for an Entrez GEO database and is experienced in the process used to build a suite of user-friendly Common Gateway Interface (CGI) programs. Dr. Edgar's understanding of the mathematics used by BLAST and his ability to develop sophisticated software will provide the opportunity to play an important role in maintaining and developing one of the most heavily used resources at the NCBI.

In August 2001, **Mr. Joe Thomas** was selected as the new Head of Unit B, Index Section, Bibliographic Services Division. Mr. Thomas received his B.S. from the University of

Maryland. He began his career in 1984 in the Cataloging Section of Library Operations and was reassigned to the Index Section in 1989. For more than 10 years, Mr. Thomas has served as the Index Section liaison to the MeSH Section of Library Operations and as a senior indexer and reviser. Over the years, he gained expertise in the journals indexed for MEDLINE, especially in the area of molecular biology. He has also served as Project Officer for the contract to create commentary linkages between citations in MEDLINE.

In August 2001, **Deanna M. Church, Ph.D.**, joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Church received her Ph.D. in human genetics from the University of California, Irvine in 1997. In her postdoctoral work at NCBI, she has become familiar with relational databases and the SQL, together with some programming experience in Perl and C++ and has worked on the design and implementation of NCBI web pages. Dr. Church has successfully applied her computational skills to the analysis of the human and mouse genomes. Her work on constructing the mouse/human homology map has not only provided a useful web resource, but also formed a major section of the paper describing the initial sequencing and analysis of the human genome that was published earlier this year in *Nature*. She is a member of Mouse Sequencing Network, an international collaboration aimed at completing the DNA sequence of the mouse genome.

In August 2001, **Wonhee Jang, Ph.D.**, was appointed a Staff Scientist with the Information Engineering Branch, NCBI. Dr. Jang, a native of Korea, received her Ph.D. in human genetics at the University of Michigan, Ann Arbor, Michigan in 1998. Dr. Jang came to the NCBI in 1998 and trained for two years as a postdoctoral Visiting Fellow and an additional year as a Research Fellow. Dr. Jang has worked on several informational aspects of the human genome and developed the information resources necessary to link two different views of the genome. Dr. Jang was part of a team that developed an NCBI database to integrate STS data from multiple sources. She then used the method of "electronic PCR" to localize the STSs from this database within the human DNA sequence. In addition, Dr. Jang was a member of the working group that created the basic design of the NCBI Map Viewer.

In August 2001, **Mikhail Domrachev** joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Mr. Domrachev received his B.Sc. in applied mathematics and physics and an M.S. degree in computational physics in 1994 from the Moscow Institute of Physics and Technology. Mr. Domrachev has been working under contract as a programmer with NCBI since April 1999. He supported the internal NCBI project to convert major software systems from C to C++ and developed web-based applications using the C++ toolkit. He redesigned the NCBI Taxonomy server in a more object-oriented approach, and used that interface to implement web-based Taxonomy resources. He also worked on the SourceTrack and RefTrack databases used to support the NCBI RefSeq project. His skills and experience in software development and object-oriented DBMS for data storage will continue to be crucial for ongoing projects at NCBI such as the GEO and Taxonomy projects.

In September 2001, **Joyce Mitchell, Ph.D.,** joined the Lister Hill Center on a detail assignment under the Intergovernmental Personnel Act. Dr. Mitchell is on the faculty of the University of Missouri, Columbia where she is Associate Dean for Integrated Technology Services. She received her Ph.D. in Population Genetics and Statistics from the University of Wisconsin at Madison and is an expert in both informatics and medical genetics. At LHNCBC, Dr. Mitchell will focus on research projects related to the Human Genome Project, bioinformatics, and information designed specifically for the public.

In September 2001, **Michael Kimelman, Ph.D.**, joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Kimelman, a native of Russia, received his Ph.D. degree in computer science from the Moscow Institute for System Programming in 1996. Prior to joining NCBI, he worked as a systems analyst for Informax, Inc, where he became involved in the NCBI dataflow group. Dr. Kimelman has made many important contributions to existing software and created new programs to handle the flow of GenBank

data. At NCBI, he has responsibility for supporting and developing programs for real-time delivery of the data, GenBank releases, daily and cumulative updates, daily creation of BLAST databases as well as internal consistency checking of GenBank.

In September 2001, **Vyacheslav Chetvernin** joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Mr. Chetvernin, a native of Russia, received a Master's degree in mathematics from Novosibirsk State University in 1983. Since then he has acquired expertise in operating system design and implementation, artificial intelligence, application software development, system and network administration and web development. Mr. Chetvernin has expertise as a software developer writing applications for the visualization of large-scale genomic data. He created MapViewer, a tool that provides integrated access to the genome data, and he has participated in the design of the search and retrieval system and the database backend for MapViewer. He is currently maintaining and constantly improving MapViewer for such organisms as Arabidopsis thaliana, Danio rerio (zebrafish), etc.

Retirements and Resignations

In November 2000, **Sharee Pepper, Ph.D.**, resigned her position of Scientific Review Administrator with the Division of Extramural Programs. Dr. Pepper joined the NLM in November 1997, and during her tenure she was responsible for the review of grant applications assigned to the NLM. Dr. Pepper began a new position with the State of Hawaii.

In April 2001, **Barbara (Bonnie) Kaps** retired with 29 years of service in the Federal Government. Ms. Kaps joined the NIH in 1984 and she served as NLM's Committee Management Officer for the past 4 years. As the Committee Management Officer, she served as the focal point for the operation and management of all NLM's chartered advisory and peer review committees. Ms. Kaps' work with the NLM's Board of Regents was especially noteworthy.

In August 2001, **Wojciech Makalowski, Ph.D.,** resigned from his position as Staff Scientist with the Computational Biology Branch of NCBI. Dr. Makalowski, a native of Poland, received his Ph.D. in Molecular Biology at Poznan University in Poland. Dr. Makalowski came to NCBI in 1994 as one of the first "GenBank Fellows." His detailed knowledge of evolutionary sequence variation was invaluable to NCBI's work. Dr. Makalowski accepted an associate professor position of biology at Pennsylvania State University.

In September 2001, **Johnie Sullivan** retired from the Federal government after 18 years of Federal service. Mr. Sullivan served as Chief, Systems Technology Branch, Office of Computer and Communications Systems (OCCS) where he oversaw the development of new operating systems languages and association software and hardware as well as the investigation of new computer technology in support of the NLM. Prior to joining NLM, Mr. Sullivan was employed by the Federal Bureau of Investigations as the Chief Information Systems Security Officer for the National Security Division.

Awards

The 2001 Cosmos Club Award was presented to the Club's 38th recipient, Dr. Donald A.B. Lindberg. The Award honors individuals of national or international standing who have made outstanding contributions in science, literature, the fine arts, the learned professions or the public service. Dr. Lindberg was recognized for his "vision, creativity, and leadership in making the immense and ever-expanding universe of medical information and knowledge available and easily accessible to all who care about sick people anywhere in the world. He has inspired and engineered the creation of simple, clear, and free systems for obtaining from the greatest collection of medical literature on earth immediate answers to inquiries made by anyone with access to a computer or medical library, providing incalculable benefits to patients and their health care workers."

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 2001 award was Mr. John R. Butler (OCCS) for technical achievement in software development that has substantially improved NLM's processing of bibliographic material.

The NIH Director's Award was presented to Mr. Ronald Stewart (OD, OA) for outstanding initiative and persistence in marshaling generic clearance from OMB to conduct customer satisfaction surveys for the NLM.

The NLM Director's Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to four employees: Ms. Becky Lyon (LO) for her sustained contributions to the development and enhancement of NLM's outreach programs for the general public; Mr. Robert Mehnert (OD, OCPL) for his intellectual contributions linking the NLM to the press and public, and his graceful navigation of the Office of Communications and Public Liaison through new territory; Ms. Bonnie Kaps (retired in April 2001) for outstanding service to the NLM Board of Regents and dedication to the mission of the NLM; and Mr. Stanley Jablonski (NLM Scholar) for continuing scholarly achievement in developing the Online Multiple Congenital Anomaly/Mental Retardation Syndromes database and making this important resource available worldwide through the NLM Web site.

The NIH Merit Award was presented to five employees: Mr. Richard Banvard (LHC) for continuing support and leadership of the Visible Human Project; Ms. Jana Brightwell (LO) for her consistent, exemplary performance which significantly contributed to the success of various important projects and products produced by the Public Services Division; Mr. Reginald Frazier (LO) for his diligence in improving the infrastructure that allows the MEDLINE database to expand and be more valuable to health professionals; Ms. Alice Jacobs (LO) for leadership on projects which contributed to a 25% growth in bibliographic records in NLM's online public access catalog; and Dr. Frederick Wood (OD) for extraordinary achievement in developing and evaluating NLM outreach and web metrics initiatives.

The NIH Quality of Work Life Award was presented to Ms. Deborah Katz (LHC) for expert management, communication, and personal skills in creating and fostering an environment that encourages personal growth, creativity, flexibility, and a truly enjoyable workplace.

The Philip C. Coleman Award recognizes significant contributions to the NLM by individuals who demonstrate outstanding ability to motivate colleagues. The recipient of the 2001 award was Mr. Anthony Pirrone, III, for his continued efforts in furthering Equal Employment Opportunity at the NLM and inspiring others to do the same.

The NLM EEO Special Achievement Award was presented to Mr. George Franklin and Mr. Pierre Levermore for their initiative and dedication, as part of the NLM outreach initiative to Native Americans, in promoting NLM's unique information services for the public, especially MEDLINEplus and ClinicalTrials.gov.

The journal, *Federal Computer Week*, presented two awards in July 2001: 1) The Monticello Award was presented to the NLM for the Multilateral Initiative on Malaria Communication Network; and 2) The "Federal 100 Award," which honors executives who had the greatest impact on the government systems community, was presented to Ms. Julia Royall (OD, OHIPD), who led the effort to create a real-time, satellite-based research network for scientists working in Africa to find a better treatment for malaria.

Table 15

FY 2001 Full-Time Equivalents (Actual)

Office of the Director	13
Office of Health Information Programs Development	7
Office of Communication and Public Liaison	9
Office of Administration	54
Office of Computer and Communications Systems	57
Extramural Programs	16
Lister Hill National Center for Biomedical Communications	79
National Center for Biotechnology	101
Specialized Information Services	27
Library Operations	293
TOTAL FTEs	**656**

NLM Diversity Council

The NLM Diversity Council began the year by welcoming five new members: Carole Brown, Tamar Clarke, Kimberlee Ford, Dawn Lipshultz, and Marta Melendez. Each will serve a two-year term from January 2001 through December 2002. Continuing on the Council are Vivian Auld, Nadine Benton, James Dean, Julian Owens, Tony Pirrone, and Julia Royall. After Julian Owens left NLM, James Knoben was appointed to the Council. The Council continues to receive support from its ex-officio members, Donald Poppke, David Nash, and Nadgy Roey as well as its distinguished alumni. Julia Royall accepted the responsibilities of Council Chair and Vivian Auld became the Council Vice-Chair.

FY2001 Accomplishments:

- *NLM Director's Employee Education Fund*: Continued coordination of the NLM Director's Employee Education Fund. In FY2001, the Fund enabled 49 staff to take 59 classes. Staff who have taken advantage of the Fund represent 35% from the Division of Library Operations, 18% from the Office of the Director, 18% from the Office of Computer and Communications Systems, 13% from the Lister Hill Center, 10% from the NCBI, and 6% from Specialized Information Services. Undergraduate classes made up 87% of the classes supported. The school with the largest number of NLM enrollees is Montgomery College (20%). Other institutions being attended are the University of Maryland, University of the District of Columbia, Johns Hopkins University, George Washington University, Shepard College, Bowie State University, and Strayer University. Course disciplines enrolled in included, computer science, business, English, math, religion, foreign language, science, psychology, art, and logic. In addition to traditional classroom instruction, courses were taken on the Internet and Voice Mail formats. The Diversity Council continues its effort to publicize the availability of the fund.
- *Getting to Know NLM*: The Council continued the Getting to Know NLM Series, scheduled to end with a grand finale on December 11, 2001 with a special closing. This series is designed to promote the different operational units at NLM, highlighting the major programs of each area and the skills, education, and expertise needed to succeed in each unit. Each office within NLM is being featured successively for one month. During its month, each office provides a presentation to all NLM staff detailing their mission, goals, and areas of responsibilities. In addition, each office creates a bulletin board showcase that is on display during the entire month. The series, a popular and creative success, has provided an opportunity for individuals to see how their duties and responsibilities contribute to the accomplishments of their office; and ultimately, to the success of NLM. While it serves to enhance employees' knowledge of the library, it fulfills the Director's effort to promote diversity at the Library.

Operational units covered in FY2001 were Library Operations, Extramural Programs, Specialized Information Services, Lister Hill Center, National Center for Biotechnology Information, and the Office of Computer and Communications Systems. Transcripts of the series are produced for each program by the captioning services. The Getting to Know NLM program also requires that the videos incorporated in the programs include captioning. Videos are made of each program and are available in the NLM Staff Library. Later they will be archived in the History of Medicine Division.

- *Communication of NLM Diversity:* The Diversity Council collaborated with the Office of Communications and Public Liaison to promote various activities on the NLM Staff Bulletin Board located outside the cafeteria. This display has provided an excellent setting for celebrating the diversity found at the NLM. The Council purchased two additional bulletin board panels to accommodate this collaboration.
- *Facility Accessibility and Reasonable Accommodation:* The Council continued efforts to upgrade access at NLM for people with disabilities. To facilitate the discussion, the Council met on two occasions with the Chief of the Office of Administrative Management Services. Council members attended NIH events relating to access for people with disabilities, including the NIH Disability Awareness and the NIH Technology Awareness Expo.
- *Installation of Multimedia Equipment:* The Council requested a decoder/encoder system that will make it possible to display captioning on the monitors in Conference Room B. In addition, the Council also requested wireless microphones to be provided in Conference Room B.
- *Shepherd's Table:* The Council planned and carried out a food drive that resulted in NLM food items for the Shepherd's Table, a community center for people in need.

APPENDIX 1: REGIONAL MEDICAL LIBRARIES

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

2. **SOUTHEASTERN/ATLANTIC REGION**
University of Maryland at Baltimore
Health Science and Human Services Library
601 Lombard Street
Baltimore, MD 21201-1583
(410) 706-2855 FAX (410) 706-0099
States served: AL, FL, GA, MD, MS, NC, SC, TN, VA, WV, DC, VI, PR
URL: http://www.nnlm.nih.gov/sar

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

4. **MIDCONTINENTAL REGION**
University of Utah
Spencer S. Eccles Health Sciences Library
10 North 1900 East
Salt Lake City, Utah 84112-5890
Phone: (801) 581-8771
Fax: (801) 581-3632
States Served: CO, KS, MO, NE, UT, WY
URL: http://nnlm.gov/mcr

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

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

7. **PACIFIC SOUTHWEST REGION**
University of California, Los Angeles
Louise M. Darling Biomedical Library
Box 951798
Los Angeles, CA 90025-1798
(310) 825-1200 FAX (310) 825-5389
States served: AZ, CA, HI, NV and U.S. Territories in the Pacific Basin
URL: http://www.nnlm.nih.gov/psr

8. **NEW ENGLAND REGION**
University of Massachusetts Medical School
The Lamar Soutter Library
55 Lake Avenue, North
Worcester, MA 01655
Phone: (508) 856-2399
Fax: (508) 856-5039
States Served: CT, MA, ME, NH, RI, VT
URL: http://nnlm.gov/ner

APPENDIX 2: BOARD OF REGENTS

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

Appointed Members:

FOSTER, Henry, M.D., Ph.D.
Professor Emeritus
Meharry Medical College
Nashville, TN

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

BUNTING, Alison, M.L.S.
Associate University Library for Science
Louise Darling Biomedical Library
University of California, Los Angeles
Los Angeles, CA

KLEIN FEDYSHIN, Michele, MSLS
Manager of Library Services
University of Pittsburgh Medical Center
Pittsburgh, PA

LEDERBERG, Joshua, Ph.D.
Sackler Foundation Scholar
Rockefeller University
New York, NY

LINSKER, Ralph, M.D.
IBM–T.J. Watson Research Center
Yorktown Heights, NY

NEWHOUSE, Joseph, Ph.D., Director
Division of Health Policy Research & Education
Harvard University
Boston, MA

PARDES, Herbert, M.D.
President and CEO
New York Presbyterian Hospital
New York, NY

PRIME, Eugenie, MS, MBA
Manager, Hewlett-Packard Libraries
Palo Alto, CA

WEICKER, Lowell, Governor
Alexandria, VA

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:

SRINIVASAN, Padmini, MSC, Ph.D.
Dir. School of Library & Info. Science
University of Iowa
Iowa City, IA

FERRIN, Thomas E., Ph.D.
Professor in Residence
U. of Cal. Computer Graphs. Lab.
San Francisco, CA

FRIEDMAN, Carol, Ph.D.
Professor, Dept. of Medical Informatics
Columbia University
New York, NY

MARSHALL, Joanne G. Ph.D.
Dean, School of Information & Library Science
University of North Carolina
Chapel Hill, NC

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

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

SIEVERT, MaryEllen C., Ph.D.
Professor of Library and Information Science
University of Missouri
Columbia, MO

SRIHARI, Sargur N., Ph.D.
Distinguished Professor
Computer Science & Engineering
State University of NY
Buffalo, NY

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:

DELISI, Charles, Ph.D. (Chair)
Dean, College of Engineering
Boston University
Boston, MA

KWITEK-BLACK, Anne E., Ph.D.
Asst. Professor, Dept. of Physiology
Human and Molecular Genetic Center
Medical College of Wisconsin
Milwaukee, WI

LEE, Christopher J., Ph.D.
Assistant Professor
Laboratory of Structural Biology
University of California
Los Angeles, CA

MATISE, Tara Cox, Ph.D.
Assistant Professor
Department of Genetics
Rutgers University
Piscataway, NJ

PREUSS, Daphne K. Ph.D.
Assistant Professor
Molecular Genetics and Cell Biology
University of Chicago
Chicago, IL

TRASK, Barbara J., Ph.D.
Head, Human Biology Division
Fred Hutchinson Cancer Research Ctr.
Seattle, WA

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:

NILAND, Joyce, Ph.D., Chair
Chair, Division of Information Sciences
City of Hope National Medical Center
Duarte, CA

ALTMAN, Russ. B., M.D., Ph.D.
Associate Professor
Stanford Medical School
Stanford, CA

BALAS, Andrew, M.D., Ph.D.
Assistant Professor
University of Missouri

BYRD, Gary D., Ph.D.
Director, Health Sciences Library
State University of NY at Buffalo

CHUTE, Christopher G., Dr.P.H., M.D.
Section Head and Professor
Medical Informatics
Mayo Foundation
Rochester, MN

CLARKE, Neil D., Ph.D.
Associate Professor
Dept. of Biophysics and Biophysical Chemistry
Johns Hopkins School of Medicine
Baltimore, MD

DALRYMPLE, Prudence, Ph.D.
Dean and Associate Professor
Graduate School of Library Information Science
Dominican University
River Forest, IL

DIMITROFF, Alexandra, Ph.D.
Associate Professor
School of Library Science
University of Wisconsin
Milwaukee, WI

GUARD, J. Robert, MLS
Chief Information Officer
University of Cincinnati Medical Center
Cincinnati, OH

HRIPCSAK, George, M.D.
Chief Information Officer
University of Cincinnati Medical Center

HUANG, H.K., D.Sc.
Director, Radiological Informatics
University of California at San Francisco
San Francisco, CA

KOHANE, Isaac S., M.D., Ph.D.
Associate Professor
Department of Pediatrics
Harvard Medical School
Boston, MA

MCGOWAN, Julie J., Ph.D.
Director, Ruth Lilly Medical Library
Indiana University School of Medicine
Indianapolis, IN

McKNIGHT, Michelynn, M.S.
Director, Health Sciences Library
Norman Regional Hospital
Norman, OK

MILLER, Perry L., M.D.
Professor of Anesthesiology & Medical Informatics
Yale School of Medicine
New Haven, CT

MILLER, Randolph A., M.D.
Chairman, Department of Biomedical Informatics
Vanderbilt University Medical Center
Nashville, TN

OHNO-MACHADO, Lucila, M.D., Ph.D.
Assistant Professor, Radiology Department
Brigham and Women's Hospital
Harvard Medical School
Boston, MA

PINSKY, Seth, Ph.D.
Senior Director
Merck and Company, Inc.
Rahway, NJ

SAHNI, Sartaj K., Ph.D.
Distinguished Professor
Computer & Information Science
University of Florida
Gainesville, FL

SHAVLIK, Jude W., Ph.D.
Professor of Medical Informatics
University of Wisconsin
Madison, WI

SWEENEY, Latanya K.
Assistant Professor of Computer Science
Carnegie Mellon University
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:

COLLEN, Morris F., M.D.
Consultant and Director Emeritus
Kaiser Permanente Medical Care Program
Oakland, CA

BIRKMEYER, John D., M.D.
Assistant Professor of Surgery
Veterans Affairs Medical Center
White River Junction, VT

BOROVETZ, Harvey S., Ph.D.
Professor of Bioengineering
University of Pittsburgh School of Medicine
Pittsburgh, PA

BRANDT, Cynthia A., M.D., Ph.D.
Assistant Professor
Center for Medical Informatics
Yale University
New Haven, CT

CHEN, Jinkun, DDS, Ph.D.
Associate Professor of Pediatric Dentistry
University of Texas Dental School
San Antonio, TX

COOPER, James N., M.D.
Director, INOVA Institute of Research
Chairman, Department of Medicine
Fairfax Hospital
Falls Church, VA

COPELAND, Robert L., Ph.D.
Associate Professor of Pharmacology
Howard University School of Medicine
Washington, D.C.

DOUGLAS, Janice G., M.D.
Professor of Medicine
Case Western Reserve University
School of Medicine
Cleveland, OH

FUNK, Mark E.
Samuel J. Wood Library
Weill Medical College
Cornell University
New York, NY

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

O'DONNELL, Anne Elizabeth, M.D.
Assistant Professor
Pulmonary and Critical Care Medicine
Georgetown University School of Medicine
Washington, D.C.

PICOT, Sandra J. Fulton, Ph.D.
Associate Professor
School of Nursing
University of Maryland
Baltimore, MD

SHEPRO, David, Ph.D.
Professor, Depts. of Biology and Surgery
Boston University
Boston, MA

TOLEDO-PEREYA, Luis H., M.D.
Director, Surgery Research & Molecular Biology
Borgess Medical Center
Kalamazoo, MI

VALENTINE, Joan S., Ph.D.
Professor of Chemistry and Biochemistry
University of California
Los Angeles, CA

WEISSMAN, Norman, Ph.D.
Professor, Health Services Administration
University of Alabama
Birmingham, AL

WILLIAMS, Benjamin T., M.D.
Cape Haze, FL

APPENDIX 7: PUBMED CENTRAL NATIONAL ADVISORY COMMITTEE

The PubMed Central National Advisory Committee meets twice a year to review and make recommendations about the information resource, PubMed Central.

LEDERBERG, Joshua, Ph.D. (Chair)
Sackler Foundation Scholar
Rockefeller University
New York, NY

BROWN, Patrick O. Ph.D., M.D.
Associate Professor
Department of Biochemistry
Stanford University, School of Medicine
Stanford, CA 94305-5323

COZZARELLI, Nicholas, Ph.D.
Professor of Molecular and Cell Biology
Division of Biochemistry and Molecular Biology
University of California
Berkeley, CA

DAVIDOFF, Frank, M.D.
Editor, *Annals of Internal Medicine*
Philadelphia, PA 19106

FRANCKE, Uta, M.D.
Professor of Genetics
Stanford University Medical Center
Stanford, CA

GINSPARG, Paul, Ph.D.
Theoretical Physicist
Los Alamos National Laboratory
Los Alamos, NM

HOMAN, J. Michael, M.A.
Director of Libraries
Mayo Foundation
Rochester, MN

MARINCOLA, Elizabeth, M.B.A.
Executive Director
American Society of Cell Biology
Bethesda, MD

McINERNEY, Suzanne, M.A.
Health Writer/Patient Advocate
Hummelstown, PA

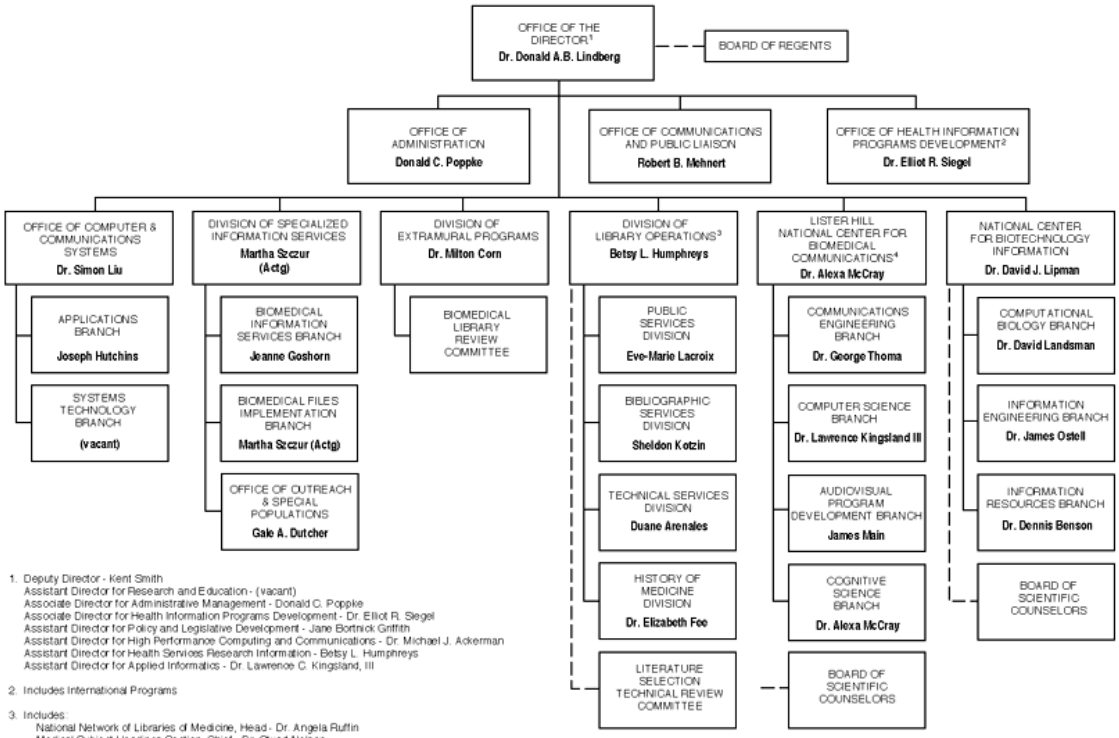
RABB, Maurice F., M.D.
Professor of Ophthalmology
College of Medicine
University of Illinois at Chicago
Chicago, IL

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

VARMUS, Harold, M.D.
Director and CEO
Memorial Sloan-Kettering Cancer Center
New York, NY

WILLIAMS, James F., M.S.L.S.
Dean of Libraries
University of Colorado
Boulder, CO

National Library of Medicine



1. Deputy Director - Kent Smith
   Assistant Director for Research and Education - (vacant)
   Associate Director for Administrative Management - Donald C. Poppke
   Associate Director for Health Information Programs Development - Dr. Elliot R. Siegel
   Assistant Director for Policy and Legislative Development - Jane Bortnick Griffith
   Assistant Director for High Performance Computing and Communications - Dr. Michael J. Ackerman
   Assistant Director for Health Services Research Information - 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, Head - Dr. Angela Ruffin
   Medical Subject Headings Section, Chief - Dr. Stuart Nelson
   National Information Center on Health Services Research and Health Care Technology - Marjorie A. Cahn

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

February 2002