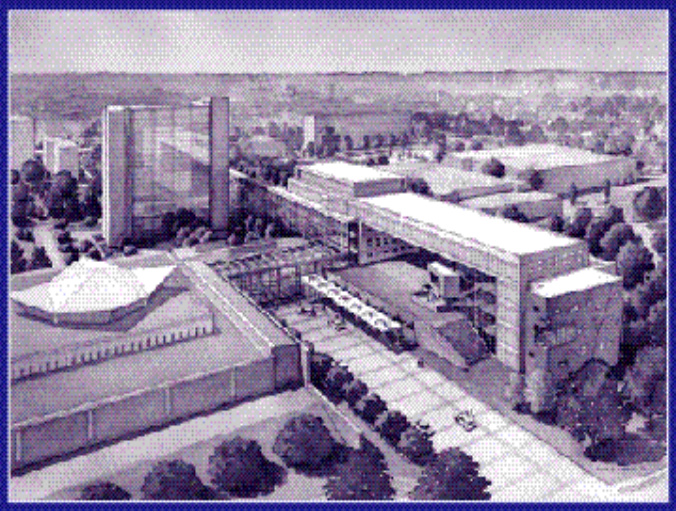
NATIONAL LIBRARY OF MEDICINE, PROGRAMS & SERVICES, FISCAL YEAR 2002

NATIONAL INSTITUTES OF HEALTH
NATIONAL LIBRARY OF MEDICINE
PROGRAMS & SERVICES FY 2002



U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES

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

Office of Communications and Public Liaison
National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894
301-496-6308
E-Mail: publicinfo@nlm.nih.gov
Web: www.nlm.nih.gov

Cover: Artist's drawing of proposed new NLM facility

NATIONAL INSTITUTES OF HEALTH

NATIONAL LIBRARY OF MEDICINE

PROGRAMS AND SERVICES

FISCAL YEAR 2002

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

This edition of our annual report has a special section, "Medicine's Library of the 21st Century," beginning on page 1. As the responsibilities of the National Library of Medicine continue to expand, and as we implement new and even more powerful information systems to serve the health professions and the public, the NLM will require an expanded facility. Fiscal Year 2002, however, saw advances on many fronts. To select a few from the many accomplishments you will see chronicled in this report:

- MEDLINE reached 12 million records this year. Through the PubMed retrieval system, MEDLINE and the other databases were searched more than half a billion times in 2002.

- The information services of the National Center for Biotechnology Information continue to expand. For example, GenBank has 15 million sequences and more than 14 billion base pairs from over 100,000 species; it is accessed daily by some 50,000 researchers.

- MEDLINEplus has emerged as a comprehensive and trusted source of consumer health information and, in 2002, was made available en español.

There are similar important advances in all NLM divisions: from the expanding *Profiles in Science* and *ClinicalTrials.gov* of the Lister Hill Center, to the quick response to 9/11 of the Specialized Information Services staff in mounting special information services on the Web, to the new Network Operations and Security Center installed by Office of Computer and Communications Systems, to new and expanded grant programs for bioinformatics training and IAIMS from the Extramural Programs.

These accomplishments, and the many more described in this report, are the result of the dedicated work not only of our knowledgeable and experienced staff, but of assistance and advice from a cadre of talented consultants and advisors. All of you, working together, have made this institution known throughout the world as an efficient and reliable source of biomedical information in all its forms.

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

MEDICINE'S "LIBRARY OF THE 21ST CENTURY"

FY2002 was a pivotal year for NLM's facilities expansion program, with the delivery of the drawings by the architectural firm Perry Dean Rogers of Boston for the 35% design contract and the awarding of the contract for the completion of the final design. The combination of NLM's historical role as the world's largest collection of books, journals and other information materials pertaining to biomedicine with its major new responsibilities in the fields of biotechnology, health services research, consumer health, and preservation are driving the need for an expansion of space, both for the collections and for people.

Congress recognized "the burgeoning programs of the National Center for Biotechnology Information (NCBI) and the growing limits of available space for the world's greatest medical collection" when it urged NLM in FY2000 to conduct a feasibility study for a new building structure. In FY2001 Congress provided funds for the architectural and engineering design work. Most recently, Congress endorsed the expeditious transition from the design phase to construction in the Senate Report for the FY2003 Appropriations for Labor, Health and Human Services, and Education by stating that:

> Many of the most serious diseases have a molecular basis. The NLM's National Center for Biotechnology Information is an integral player in this research process for it organizes and analyzes the vast volume of genomic information uncovered in the last decade. The Congress believes that if this Center is to make its maximum contribution to our fight against disease, it must very soon have expanded facilities to meet the growing demands being placed on it. The Committee provided funds necessary for the design of such facilities, and it desires that such design, when completed, be rapidly moved into the construction phase. The Committee, therefore, requests a report from the NIH by April 1, 2003, that delineates the features of this new facility, its size and its expected cost, based upon a fast-track schedule.

The new physical facility (see drawing on the cover of this report) is a unique structure designed to accommodate the growing scope and activities of medicine's "Library of the 21st Century." NLM's evolving role is reflected in the expanding variety of information handled, from books or articles, through multimedia, to source databases such as GenBank. This explosion of information to be managed, combined with the heightened demand for access from scientists, health care providers, patients, and the public, has fueled the urgent need for new and innovative facilities.

NLM's ability to perform its expanded mission in an effective and efficient manner is key to capitalizing on Congress's investment in biomedical research in recent years. Molecular biology—arguably the primary driver of medical advances in the 21st century—requires information-handling capabilities that keep pace with the tons of data being generated from the Human Genome Project and related genetic research on many organisms. The Library's NCBI brings together powerful computers, sophisticated software, and highly trained specialists in a collaborative environment that makes the results of genetic research quickly accessible to thousands of scientists on a global basis. NCBI closes the loop between the data generated from basic research and future discoveries by collecting, managing, and annotating sequence data submitted from around the world. Once organized, curated, made searchable in multiple ways, and linked to other information resources, these critical databases enable researchers to identify disease genes, decipher biological mechanisms underlying disease, and design and develop therapeutic strategies for treating and preventing disease. Without such resources, the ability to advance biomedical research at a rapid pace would be vastly diminished.

By also broadening the audience for its information resources NLM plays a significant role in translating the knowledge acquired through NIH research into information that can make an immediate difference in the lives of individuals. Our modern Web-based communications environment has turned the general public into avid consumers of health information. NLM has responded by creating a number of extremely successful services used by professional and public alike. MEDLINE/PubMed, the online database of some 12 million references and abstracts to medical journal articles is now searched close to 500 million times a year. MEDLINEplus, created specifically for the consumer, makes reliable information from the NIH and other authoritative sources easily and freely available, both in English and in Spanish. Its widespread popularity, broad recognition as a reputable source of health information, and growing capabilities reflect the benefits to be accrued from effectively collecting, organizing, and distributing multimedia health information to the public.

NLM's research and information programs have a role to play in medical and public health preparedness for disaster management and terrorist attack. The Library supports the development of basic research tools, including: genomics research databases for targeted development of drugs, vaccines, and other forms of treatment for such diseases as smallpox, anthrax, plague, Ebola, and cholera; informatics R & D related to terrorism and disaster management; training for health professionals

in the use of pertinent information resources; developing experimental information resources targeted at first responders and others involved in disaster management; and improving the information infrastructure so that data can be transmitted and shared during a crisis. One prescient NLM-funded investigator even created a prototype early-warning system for public health emergencies; it was recently demonstrated to President Bush who considered it a model for anthrax-like situations.

As to information programs for disaster management, after September 11, 2001, NLM placed new pages in MEDLINEplus for such health topics as post-traumatic stress disorder, biological and chemical weapons, smallpox, anthrax, and created new special pages on lingering airborne hazards, and biological and chemical warfare agents. Also, recent additions of books and other technical reports, along with the inclusion of older materials to the databases about smallpox and other pertinent topics, have helped to improve access to information useful to the research and public health community in the fight against bioterrorism.

For NLM to continue to leverage the results of biomedical research in ways that foster greater research productivity and contribute to improvements in health care and disease prevention requires not only additional space, but also a facility that will foster collaboration and the sharing of knowledge. The design for the new facility provides a unique resource to promote interaction among scientists, clinicians, and information professionals. It will include space for a "Collaboratory" that will bring together staff from NCBI and other NLM research components, NIH scientists, and medical librarians in a shared space where face-to-face collaboration can be combined with people-to-computer interactions. It is designed to house small working groups, individual study carrels for visiting scholars, and a large briefing facility that also can be used for making public announcements of important projects and new services. The space will be used for a range of NLM activities, such as the production of a variety of online resources, annotation of genome databases, expansion of the Unified Medical Language System, and cooperative activities of the National Network of Libraries of Medicine. The availability of installed network and computational resources will enable more participation by outside individuals in the research aspects of NLM and the NIH itself. The Collaboratory will be unique on the NIH campus and will be a magnet for the best scientific minds.

Another special feature of the new facility will be an innovation demonstration center that will highlight exciting new possibilities for the application of advanced computing and telecommunications technologies to medical research and communication. Examples include new uses of the NLM-developed Visible Human datasets, telemedicine projects to remote areas, and medical applications for the Next Generation Internet.

A two-level underground expansion for the collections will provide both critical space for the growing collections and preserve the historic appearance of the original library building. The rapidly expanding universe of biomedical information in a growing array of formats requires not only additional room for collections, but also for staff with the expertise to organize, manage, and preserve this invaluable resource.

In summary, continued progress in our understanding of the relation between genes and disease requires that NLM's information-handling capabilities keep pace with the voluminous data being generated by scientists. The NCBI in particular is developing the public genome databases and data mining tools that are making advances possible. Its responsibilities for collecting, managing, and analyzing the growing body of genomic data generated from the sequencing and mapping initiatives of the Human Genome Project are essential to scientific progress.

Unfortunately the NLM's present facilities, which predate 1987 when Congress established the NCBI, were made to hold fewer than 650 employees; they now house more than 1,000, including many contract staff. The single biggest reason for this is the National Center for Biotechnology Information and the central place it occupies in 21st century medical science. The promise of 21st century medicine, which is nothing less than preventing—and curing—disease, requires that we invest now in a facility of some 350,000 gross square feet of office, laboratory, and collaborative space to house the collections of the Library and to ensure continued progress in the activities of the National Center for Biotechnology Information.

OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT

Elliot R. Siegel, Ph.D.
Associate Director

The Office of Health Information Programs Development (OHIPD) 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

The NLM Long Range Plan 2000–2005, published in 2000, remains at the heart of NLM's planning and budget activities. Its goals form the basis for NLM operating budgets each year. All of the NLM Long Range Plan documents are available on the NLM Web site.

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 (OCHD). This Committee plans, develops, and coordinates NLM outreach and consumer health activities.

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 is the implementation of a database of NLM outreach, consumer health, and health disparities projects. OCCS is developing, hosting, and supporting this database with assistance from OHIPD staff and the committee. This database will be a major source of data for the National Outreach Mapping Center, which is seeking to use mapping as an aid to ensuring the effective distribution of outreach services by the NLM and the National Network of Libraries of Medicine.

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 in press. 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 are being 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.

Also, in FY2002 NLM/OHIPD partnered with NIH EEO and NLM EEO to participate in the NIH Acting Deputy Director's American Indian Pow-Wow Initiative. This included exhibiting at 7 pow-wows in the Mid-Atlantic area, including the Inaugural Smithsonian National Museum of the American Indian Pow-wow. An estimated 7,000 persons visited the NLM booth over the course of these pow-wows. 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. Several hundred seniors were trained in basic Internet skills during the first year, with the assistance of several dozen teens. The year two emphasis will be on Cyber Health for Seniors with a strong evaluation component, intended to help measure the extent to which the health information seeking behavior and

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 at San Antonio Health Sciences Center to conduct a needs assessment and various health information outreach projects with Hispanic-serving community, health, 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. And the Web environment in which NLM operates is rapidly changing and intensely competitive. These two factors combined 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 operated for about 18 months until it was subsumed by the OCHD coordinating committee. The OCHD is chaired by the NLM Associate Director for Health Information Programs Development, and staffed by the OHIPD. The Web evaluation priorities of the OCHD include: a) quantitative and qualitative metrics of Web usage; and b) measures of customer perception and use of NLM Web sites. During FY2002, the Work Group's evaluation activities included: online surveys of users of select NLM Web sites; several online focus groups; access to a syndicated telephone survey of the US 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 OCHD Committee and OHIPD continue to explore and test a range of internal and external Web evaluation methods and applications.

International Programs

MIMCom.Net: A Malaria Research Network for Africa

NIH has led an international effort to provide malaria researchers in Africa with full access to the Internet and the resources of the Web. This project began with NIH's leadership in the Multilateral Initiative on Malaria in which African scientists identified electronic communication and access to scientific information as critical in the fight against the devastating and economically debilitating effects of malaria in developing countries.

The NLM, working in partnership with organizations in Africa, the United States, the United Kingdom and Europe, has created MIMCom.Net, the first electronic malaria research network in the world. The network provides full access to the Internet and the resources of the Web, as well as access to current medical literature, for scientists working in Africa. The African research sites are of recognized high quality, require improved communications to accomplish ongoing research, and have the necessary resources to purchase equipment and sustain the system.

MIMCom.Net is the result of discussions held at the 1997 Multilateral Initiative on Malaria (MIM) Conference in Dakar, Senegal, where African scientists identified lack of communication as a major barrier to carrying out their work: combating the morbidity and mortality of malaria. As a follow up to the Conference, a working group was formed to discuss ways in which research scientists in Africa might have the same level of Internet access as their colleagues elsewhere in the world and NLM was charged with leading the effort to create the necessary communication network.

MIMCom.Net comprises telecommunications; information access; and new tools for research, training, and evaluation. In collaboration with partners around the world, NLM designs and operates the network and covers all necessary costs. These include: determination of requirements; site surveys; negotiations with African telecommunications regulatory authorities; assistance with equipment purchase and installation; system monitoring; ongoing technical assistance, training and support; handling of monies and agreements; establishment of document delivery systems and information portals; and promotion of malaria research agendas. Individual sites and their funding partners are responsible for equipment costs and the shared cost of using satellite bandwidth.

The network technical hub is located at Redwing Satellite Solutions, Ltd. in the United Kingdom, where a large satellite dish, focused on a geo-stationary satellite 37,000 km above the Atlantic Ocean, is connected directly to the high-speed Internet backbone on the ground. At research sites with no local telecommunications service, a smaller ground station in the form of a Very Small Aperture Terminal (VSAT) is installed. The VSAT dish antenna connects, through a radio unit, to an existing local area network (LAN) used by the researchers. Some sites on the network operate a wireless connection to a local Internet Service Provider (ISP) or to another MIMCom site nearby.

The system provides an open link that allows researchers to send and receive email, search the literature and databases, or share files and images 24 hours a day 7 days a week. Permanent access to information is moving researchers in Africa toward a more efficient way of working with colleagues around the world.

Satellite systems are not subject to the problems and limitations of telephone wires or other more traditional means of obtaining an Internet connection and are, therefore, highly reliable. Although associated costs are high, MIMCom.Net is designed to allow hundreds of researchers in Africa to share satellite bandwidth,

maximizing the usage of satellite capacity and minimizing cost per site.

NLM and the International Centre of Insect Physiology and Ecology (ICIPE) Mbita research site in Kenya are currently leading the first evaluation of MIMCom.Net. The evaluation covers network performance and efficient use of bandwidth, as well as information use and site growth, proposals funded, papers published, and numbers of collaborations carried out. Information is being gathered using baselines that were created before the network was installed.

NLM is using MIMCom.Net to launch two experimental programs that promote increased access to medical literature for malaria researchers in Africa. The Medical Library at the University of Zimbabwe, the Medical Research Council (MRC) in South Africa and NLM have established a pilot document delivery system for malaria researchers in Africa. NLM provides ongoing technical support and training to IT personnel at each site. Additional training opportunities, including individual course work and regular conference calls, are planned to build capacity among African IT specialists at the research sites. NLM is making follow-up visits to each site for updating and trouble shooting.

Special training in the use of IT as it relates to specific research agendas will be offered to malaria researchers in Africa. This may include training in the use of various wireless communication devices as well as personal software agents.

The Web site (http://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 is expanding the site to include special news releases and articles of interest to researchers.

The Network, as of September 30, 2002:

Kenya: Kenya Medical Research Institute(KEMRI)/Centers for Disease Control and Prevention (CDC) in Kisian; KEMRI/Wellcome Trust in Kilifi; KEMRI/CDC/Walter Reed Army Institute for Research (WRAIR) in Nairobi, with links to the US Library of Congress and Wellcome Trust sites; International Centre of Insect Physiology and Ecology (ICIPE) in Mbita, with support from NIH

Ghana: Noguchi Memorial Institute in Accra, with support from NIH, US Agency for International Development (USAID), US Naval Institute of Medical Research (US NIMR) Navrongo Health Research Center with support from NIH, USAID, and US NIMR

Tanzania: (all with support from NIH): National Institute of Medical Research (NIMR) headquarters in Dar es Salaam; Amani Center; Ifakara Center; Malaria researchers at Kilimanjaro Christian Medical Center (KCMC) in Moshi

Uganda: Uganda Viral Research Institute in Entebbe with support from CDC.

NLM is assisting researchers in the following projects to upgrade their access to the Internet: (1) Antimalarial drug resistance project at University of Ibadan, Nigeria, and at Mulago Hospital/Makerere University and Biomedical Lab, Kampala, Uganda; (2) Pediatric malaria project at the College of Medicine and Wellcome Trust in Blantyre, Malawi; and (3) Biotechnology Center, Faculty of Medicine and Biomedical Sciences, University of Yaounde I, Cameroon.

International Network Partnerships

OHIPD is pursuing strategies to develop international network partnerships. 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. NLM is a leader in this field, and several other research and technical organizations now recognize the importance of end-to-end Internet performance. During 2002, NLM built on the earlier phases of end-to-end connectivity testing by conducting outreach to other researchers and organizations actively pursuing this topic. 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 has initiated a collaborative project with the University Corporation for Advanced Internet Development/Internet 2 to conduct research on “critical incidents” where higher bandwidth Internet Connectivity has made or could make a significant difference for biomedical and health applications. Additionally, NLM is developing its own Internet connectivity performance monitoring network, starting

with select US sites but envisioned to extend to international sites in the medium term.

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. 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: 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
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 Information Center, Taipei, Taiwan

International Visitors

In FY2002 the Office of Communications and Public Liaison arranged for 249 tours—119 regular daily (1:30 pm) tours and 130 specially arranged tours. There were 4323 visitors in total. They came from the following 34 countries:

Argentina, Australia, Belarus, Bolivia, Brazil, Canada, China, Colombia, Czech Republic, England, Finland, France, Germany, Iceland, India, Italy, Japan, Kenya, Korea, Mexico, The Netherlands, Pakistan, Palestine, Paraguay, Peru, Poland, Scotland, Senegal, Spain, Switzerland, Taiwan, Uganda, Ukraine, and the United States.

LIBRARY OPERATIONS

Betsy L. Humphreys
Associate Director

The Library Operations (LO) Division selects, acquires, preserves, and organizes NLM's comprehensive collection of scholarly biomedical literature; maintains a subject thesaurus and a library classification used by institutions around the world; produces authoritative indexing and cataloging records; builds and distributes 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 services; and coordinates the National Network of Libraries of Medicine, which improves access to health information services throughout the U.S. These basic services provide an essential foundation for NLM's outreach programs to health professionals and the general public and also support the Library's focused programs in AIDS, health services research, molecular biology, and toxicology and environmental health. The National Information Center on Health Services Research and Health Care Technology (NICHSR) is located within LO.

The largest of NLM's Divisions, LO employs a multidisciplinary staff of librarians, technical information specialists, subject experts, health professionals, historians, museum professionals, and technical and administrative support personnel and relies on the services of a wide variety of contractors. In addition to its basic services, LO develops and mounts major historical exhibitions; carries out an active program of research in the history of medicine; works with other NLM program areas to develop new and enhanced products and services; conducts research and development related to current services as well as advanced information storage and retrieval; directs and sponsors educational programs for health sciences librarians; and contributes to the development of standards for health data and knowledge-based information. LO staff members are active participants in efforts to improve the quality of worklife at NLM, including the Diversity Council and the NLM Intranet.

Program Planning and Management

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. There are four basic goals in the NLM Long Range Plan:

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

Most LO activities directly address the first two of these goals, including special efforts to provide services for the general public and to enhance awareness of these new services. LO contributes to the third goal through training and education for health sciences librarians and activities related to health data standards and information policy. In the informatics research arena, LO collaborates with the Lister Hill Center on the Unified Medical Language System project and automated indexing, with the National Center for Biotechnology Information on gene indexing, and with several NLM program areas on issues related to digital libraries and permanent access to digital information.

In response to the September 11, 2001 attacks and the anthrax incidents, LO participated in NLM-wide efforts to enhance security in the Library's buildings while continuing to provide excellent service to the public. Special efforts to increase NLM's coverage of information related to bioterrorism and disaster response/management are described throughout this report.

On June 19, 2002, LO Division Chiefs and Section Heads held a brainstorming session to identify some of the developments that were likely to affect NLM's operations and services over the next 10 years. The discussion was designed to feed into operational planning and budgeting for FY2003. Among many possible developments, LO staff identified the following as particularly likely to have an impact on NLM in the near future: (1) continuing rapid growth in the biomedical and life sciences literature as a result of the doubling of the NIH budget, (2) emergence of PubMedCentral as a pivotal NLM system affecting many basic LO operations, (3) conversion to electronic-only publishing for the majority of journals indexed for MEDLINE, (4) increasing importance of NLM's role as a selector and organizer of high quality information; (5) heightened attention to physical and network security; and (6) widespread participation of LO staff in work at home arrangements. In line with these expectations, LO is working closely with NCBI to expand the content available in PubMed Central and has also initiated some trial office-sharing arrangements involving a small number of the more than 40 LO staff members who already work some days at home.

During FY2002, LO participated actively in planning for a third NLM building and for the renovation of the existing buildings. In addition to providing input on staff workspaces, LO has lead responsibility for providing functional requirements for the reading rooms, exhibition spaces, and collection storage areas. To assist in designing the collection and exhibition spaces, LO secured the services of an expert in environmental conditions for libraries and museums and also engaged a consultant to advise NLM on replacing the fire suppression system in the rare book storage areas. NLM's interest in the design of library buildings for the 21st century is shared by many other health sciences libraries. A joint NLM/Association of

Academic Health Sciences Libraries symposium on library buildings is planned for 2003.

In the meantime, to make more effective use of current space, LO renovated the Selection and Acquisitions Section work areas, modified the entrance to the History of Medicine Reading Room, divided the Billings Auditorium space into offices and a conference room for the NLM Associate Fellowship program, and renovated the work areas in the B1 stacks used by contract staff working on the First Level Search and Document Delivery contracts.

Collection Development and Management

Many basic NLM services depend on the Library's comprehensive collection of biomedical literature. LO ensures that NLM's collection meets the needs of current and future users by developing and updating NLM's literature selection policy; acquiring and processing literature that meets these selection guidelines in all languages and formats; organizing and maintaining the collection for efficient current use; and preserving materials for future generations. At the end of FY2002, the NLM collection contained 2.4 million volumes and 7.2 million other physical items, including manuscripts, microforms, pictures, audiovisuals, and electronic media.

Selection

LO staff and agents select literature for the NLM collection in accordance with the guidelines in the Collection Development Manual of the National Library of Medicine, which is currently undergoing a major review and revision as it typically does every 5 to 10 years. The Oversight Committee for the revision is chaired by Alison Bunting, also current Chair of the NLM Board of Regents, and includes researchers, practicing health professionals, a consumer health information expert, and senior members of the NLM staff. Many NLM employees are contributing their subject expertise to the revision effort.

In FY2002, TSD and NICHSR took special steps to improve NLM's coverage of works on bioterrorism, biological/chemical warfare, disaster response, and related subjects. The contract arrangement with the New York Academy of Medicine to identify and catalog gray literature on topics related to health policy and public health progressed slowly in the aftermath of the 9/11 attacks on New York City, but is back on track for completion in FY2003.

TSD completed an assessment of NLM's coverage of pre-1960 gerontology literature and found it to be excellent. Dr. Nikolai Krementsov, visiting historical scholar from the St. Petersburg Branch of the Institute of the History of Science and Technology of the Russian Academy of Sciences, assessed NLM's holdings of pre-1918 Russian materials and found them to be outstanding, although as yet largely unknown to Russian scholars. He recommended that the Library take action to improve the bibliographic control and preservation of this important collection. Dr. Anne-Emanuelle Birn, also a visiting historical scholar from the New School University, New York City, reviewed NLM's early holdings of South American materials, which are not extensive, and recommended additional acquisitions that would strengthen NLM's holdings.

Acquisitions

TSD received and processed 156,182 physical books, serial issues, audiovisuals, and electronic media. Net totals of 40,919 physical volumes and 1,009,845 other items (e.g., manuscript pages, pictures, microfilms, audiovisuals, electronic media) were added to the NLM collection. LO employs a variety of agents and vendors to acquire literature published around the world. In FY2002, TSD awarded a new 5-year serials subscription contract covering publications from North America, the United Kingdom, and Western Europe and expanded vendor coverage of materials published in South and Southeast Asia, the Middle East, and the Pacific Islands. Licensing of electronic resources occupies an increasing amount of staff effort and time. Standard licenses generally do not include the ability to provide interlibrary loan, so TSD must negotiate special provisions for practically every license for electronic journals. More than 2,000 electronic journals are now available to onsite users in the NLM Reading Rooms, many as a result of NIH-wide licenses negotiated by the NIH Library.

HMD continued to add important materials to NLM's outstanding collection of early printed books, manuscripts, pictures, and historical audiovisuals. Notable individual items acquired in FY2002 included: Joannes de Ketham's Fasiculo de Medicina...(Venice 1493/1494), a landmark in the history of anatomical illustration that is an Italian translation of an earlier Latin work; In Aristotelis de Anima Commentum (Venice, 1496/1497), an incunable edition of Aristotle's masterwork on the soul, considered to be the first systematic treatise on psychology; and the Kitab-I Kahhali 'Ayn al-Diwa'I (The Book of the Oculist), an illustrated Persian manuscript on ophthalmology, dated 1755.

NLM acquired the papers of Dr. John Eisenberg, the late director of the Agency for Healthcare Research and Quality and a noted leader in health services research, as part of its larger effort to document the history of this field. Other important contemporary manuscript acquisitions included: the papers of Dr. French Anderson, a founder of gene therapy; materials compiled by Judith Robinson in writing her recent biography of Florence Mahoney; a copy of DeWitt Stettin's memoir, "How I Spent My Light"; a nursing photograph scrapbook; and a collection of patent medicine pamphlets, donated by William H. Helfand. HMD also received additions to the papers of Nobel prize winners, Dr. Joshua Lederberg and Dr. Marshall Nirenberg, and former FDA commissioner, Dr. Herbert Ley.

NLM's picture collection was enriched by the addition of drawings by Pietro Berrettini da Cortona, and photographs by Katherine Du Teil and Rosamund Purcell, additions to the collection used in the new exhibition, "Dream Anatomy"; a framed book cover, "Women Will Be Doctors"; an etching commissioned for the Nobel Prize in Medicine in 1984; black and white photographs of the Medical Committee for Civil Rights march in the 1960s; and 280 pieces of medical ephemera donated by William H. Helfand.

Among the historical audiovisuals acquired were approximately 800 films on international health education efforts from the Johns Hopkins University School of Public Health Communications Program; films from Dr. Martine Jozan Work created by her husband, Dr. Telford Work, on epidemiology in the 1940s and 1950s, as part of his epidemiological work for the Rockefeller Institute; and videos from the NIH Office of Women's Health on their celebratory activities, "A Century of Women's Health: 1900-2000."

Sheldon Cohen, M.D., a long-time benefactor of the NLM, commissioned bronze busts of Maimonide (1135-1204), Edward Jenner (1749-1823), and Louis Pasteur (1822-1895) that were installed in the lobby of the NLM building.

Preservation and Collection Management

To preserve NLM's archival collection and keep it readily accessible for current use, LO carries out a range of preservation and collection management activities: binding, microfilming, conservation of rare and unique materials, repair of general collections, maintenance of appropriate storage facilities and conditions; and disaster prevention and response. LO distributes data about what NLM has preserved to avoid duplicate effort by other libraries and provides preservation information helpful 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 and archival-friendly formats in new biomedical publications.

In FY2002, LO bound 26,363 volumes, microfilmed 5,255 volumes, repaired 1,542 items in the onsite book repair and conservation laboratory, made 283 preservation copies of motion pictures, and conserved 66 early books and manuscripts. NLM awarded its own purchase order for library binding on February 1, 2002 and will no longer use the umbrella contract managed by the Government Printing Office for many federal libraries. The new arrangement is a more cost-effective way for NLM to obtain the specialized binding services needed for its national archival collection. Surveys of the condition of the Prints and Photographs collection and of early Japanese manuscripts and books were completed.

A total of 828,448 items were shelved or re-shelved and 89,000 duplicate journal issues were removed from the NLM collection during the year. Consultants were engaged to advise NLM on an appropriate replacement for the carbon dioxide fire suppression system currently installed in the incunabula room and rare book stacks. Staff from throughout NLM responded quickly and effectively to two major floods caused by broken water pipes and several minor water problems in collection areas. As a result, only a few volumes were damaged beyond repair.

The Preservation and Collection Management Section determined that only 6.7% of all currently received serials, 11.5% of U.S. serial titles, and 1.5% of Index Medicus titles are still published on acidic paper. The results of the study were published in a brief article in the Association of American Publishers newsletter.

The initial test run for an overlap study of OCLC records for pre-1950 monographs held by NLM, the New York Academy of Medicine, the Countway Library of Medicine at Harvard, and the College of Physicians of Philadelphia revealed some problems with the algorithms used that will be corrected for the full study. The purpose of this effort is to identify important brittle items held by other libraries, but not by NLM so that preservation strategies can be developed.

Permanent Access to Electronic Information

The long-term preservation of electronic information presents unique challenges that are not yet clearly understood. As outlined in the NLM Long Range Plan for 2000–2005, NLM's general approach to addressing these challenges is to use NLM's own electronic services and publications as test-beds and to work with its sister national libraries, the National Archives and Records Administration, and other organizations to develop, test, and implement strategies and standards for ensuring permanent access to electronic information. LO is working closely with several other NLM program areas on activities related to preservation of digital materials.

PubMed Central, a digital archive of life sciences literature developed by NCBI, is NLM's primary test-bed for the development of procedures and methods for ensuring permanent access to electronic journals. In FY2002, LO contacted publishers of all electronic-only journals indexed for MEDLINE and of major journals in the fields of health policy, health services research, and public health to encourage deposit of these journals in PubMedCentral. LO worked with NCBI and LHC to develop specifications and to award a contract to scan and add to PubMed Central the complete backfiles of many journals that already deposit their current content in this digital archive. LO will manage aspects of the quality review of both scanned retrospective and currently deposited PubMed Central content.

NLM is using its own publications and Web servers as a test-bed for procedures and mechanisms for ensuring permanent access to electronic information published by government and private non-profit

institutions. Following NLM's implementation of the Teamsite Web management software in the summer of 2002, LO is working with OCCS and other NLM program areas to test the use of the NLM standard minimum set of metadata for electronic publications, which was reviewed and revised earlier in the year. This set is based on the Dublin Core, but is more prescriptive. It includes the previously developed NLM permanence ratings which indicate NLM's level of commitment to ensuring that a work remains available. Work is also proceeding on developing a test system for ensuring permanent access to information of historical value (e.g., previous NLM policies that are no longer in effect) without confusing or inconveniencing users who are interested only in currently applicable information.

The NLM Director served on the Library of Congress National Digital Advisory Strategy Advisory Board. One outgrowth of NLM participation was a meeting of institutions funded by the Mellon Foundation to explore various aspects of digital preservation with NCBI's PubMedCentral project team. This led to further interaction on the parameters for new "standard" PubMed Central XML submission and storage DTDs, which will be published in 2003. NLM also worked with the American Medical Publishers Association (AMPA) and a joint AMPA/Association of American Publishers committee to develop plans for a one-day meeting on digital archiving to be held in conjunction with the 2003 AMPA annual meeting.

Bibliographic Control

To improve access to biomedical literature, LO creates authoritative indexing and cataloging records for journal articles, books, serial titles, films pictures, manuscripts, and electronic media. 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 the 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 worldwide.

Thesaurus Development

The 2003 edition of MeSH contains 21,973 main headings, 83 subheadings or qualifiers, 129 publication types, and more than 132,400 supplementary records for chemicals and other substances. One descriptor, nanotechnology, was added after the publication of MeSH for 2002, and 1250 new descriptors were added for the 2003 MeSH, bringing to 1251 the number of descriptors added since the 2002 MeSH was issued. Ninety-three descriptors were replaced with more up-to-date terminology, 20 descriptors were deleted, and 1727 see references (entry terms) were added.

In FY2002, the MeSH Section completed the first phase of a major reorganization and update of terminology related to genetics, enzymes, proteins, and receptors. There was also significant expansion and enhancement of terminology related to bioethics (as part of a joint effort with the Kennedy Institute of Ethics), plants and phytotherapy; microorganisms associated with notifiable diseases (in cooperation with the CDC Epidemiology Program), crustacea, and anatomy. The hierarchical arrangement of terms for Age Groups was revised to provide more logical search groupings.

The MeSH Section manages content editing of the UMLS Metathesaurus, which is now released four times per year. The more frequent update schedule allows addition of MeSH supplementary concepts throughout the year. Other source vocabularies updated in 2002 included: ICD-9-CM, the Medical Dictionary for Regulatory Activities (MedDRA), CRISP, the Alcohol and Other Drug Thesaurus, LOINC, PDQ, FDA's Standard Product Nomenclature (SPN), and the Universal Medical Device Nomenclature.

Two new vocabularies were added to the Metathesaurus in FY2002: the NCBI Taxonomy of organisms and genetic material that are the sources of new sequence data entered into GenBank and RxNorm, a standardized NLM nomenclature for "clinical drugs" (i.e., what clinicians prescribe) developed under the direction of the Head, MeSH in consultation with the HL7 Standards Development Organization, the Department of Veterans Affairs, and the Food and Drug Administration.

RxNorm was created to assist NLM in mapping synonymous terms from various drug vocabularies present in the Metathesaurus and to fill a gap in available drug terminology identified several years ago by HL7 and highlighted in National Committee on Vital and Health Statistics hearings on the administrative code sets to be adopted under the Health Insurance Portability and Accountability Act of 1996. On the one hand, drug names alone lack the strength, dose, and route of administration information that must be specified in a prescription. On the other hand, National Drug Codes (NDC) are too specific (e.g., the 50 and 100 count bottles of the same drug produced by the same manufacturer have different NDCs) for use in medication ordering and decision support systems because clinicians cannot know the specific products that will be used to fill the prescriptions they write. RxNorm combines standard forms for each active ingredient, strength, unit of measurement, and dosage form into structured names for clinical drug preparations. During FY2002, RxNorm work was presented, discussed, and very positively received at several HL7 meetings, the National Committee on Vital and Health Statistics, the e-Gov Consolidated Health Informatics Task Force, the AMIA Spring Congress and the AMIA Symposium, and several special briefings for HHS, VA, and FDA senior staff.

Cataloging

LO catalogs the biomedical literature acquired or selected by NLM both to document what is available from the Library's collection or available on the Web and to provide cataloging and name authority records that can be used by other libraries to reduce their own cataloging effort. LO provides an increasing number of links from its online catalog, LOCATORplus, to the electronic full-text of the items described. In FY2002, TSD developed policy guidelines for establishing links to contemporary monographic Internet resources and print monographic works also available online.

During FY2002, the Cataloging Section cataloged 21,419 contemporary books, serial titles, non-print items, and cataloging-in-publications galleys, using a combination of in-house staff and contractors. Cataloging contractors were provided with high-speed remote access to the Voyager Integrated Library System via DSL or cable, improving the speed, security, and reliability of their access to the system. TSD expanded its name authority file capabilities to accommodate authority control requirements of MEDLINEplus and integrated the majority of MEDLINEplus authority data into the Voyager name authority file.

In the final stage of the transfer of records from the former specialized databases, data from 30,933 monograph and chapter records from HISTLINE, SPACELINE, and BIOETHICSLINE were merged into existing LOCATORplus records, completing a major multi-year effort. Collaborative partners, including the Kennedy Institute of Ethics, ECRI, and NASA, added a total of 2,330 new records for monographs and chapters in specialized fields to LOCATORplus in FY2002. TSD added 434 more journal title records to LOCATORplus to support PubMed retrieval and serial holdings projects.

LO established an NLM-wide Shared Serials Data committee to optimize storage, maintenance, and access to serials data that are used by many different NLM systems, e.g., LOCATORplus, PubMed, DOCLINE, publication programs. Early results of the committee's work included the establishment of a direct link from the PubMed journal browser to LOCATORplus so that PubMed users can more easily obtain full bibliographic records for journal titles of interest to them. The implementation of a new release of the Voyager Integrated Library system brought several search enhancements that helped with this effort.

A full production Web version of the NLM Classification was released in September, replacing the prototype version made available last year. New features include hyperlinks between class numbers in the index and schedules, between terms within the index, and from index terms to the MeSH browser. The new online data creation and maintenance system for the Classification gives NLM the ability to update it annually in tandem with MeSH. The NLM Cataloging Manual was also converted to a Web-based interactive document.

HMD cataloged 363 rare books and early manuscripts. The special project to enhance access to NLM's fine collection of about 1,400 classical Japanese books and manuscripts is proceeding. A shelf-list of the collection was completed, and abstracts for about half of the items were prepared by Dr. Shizu Sakai, a noted scholar of classical Japanese medicine. HMD awarded a contract to upgrade existing records in Images from the History of Medicine and to catalog additional pictures. Another contract was awarded for the initial phase of a project to improve access to NLM's large historical pamphlet collection. The first step is to determine how much of the collection is covered by the Index-Catalogue of the Library of the Surgeon General's Office, which has been recently converted to machine-readable form. HMD continued to make significant progress in cataloging contemporary manuscript collections. In FY2002, 562 linear feet of manuscripts were cataloged, more than five times last year's total; new finding guides were created for several large collections, including a major segment of the archives of the Medical Library Association; and records for 108 finding guides were added to LOCATORplus.

A new Profiles in Science site for Linus Pauling debuted in June 2002. The Pauling papers are held by Oregon State University, which collaborated with NLM in the development of the Pauling site. The Profiles site for Donald S. Fredrickson, renowned lipid researcher and former NIH Director was substantially revised and expanded. The original Fredrickson site released in 1997 served as the prototype for the Profiles project. The new Fredrickson site will be officially released on October 19, 2002 in conjunction with the NIH memorial service for Dr. Fredrickson.

LO worked with LHC and the Office of the Surgeon General to identify and scan all retrospective Surgeon General's reports, to ensure that all of them had appropriate NLM catalog records, and to make them available on a Web site that was officially launched in February 2002. New Surgeon General's reports are added to the HSTAT (Health Services/Technology Assessment Text) database as they are published.

Indexing

LO indexes articles from 4,538 biomedical journals so that users of MEDLINE®/PubMed® database and the products generated from it can locate articles on specific biomedical topics. Existing MEDLINE records are annotated and linked to subsequently published notices or commentaries when the articles to which they refer have been retracted, corrected, or challenged. Under the supervision of the Index Section in the Bibliographic Services Division (BSD), a combination of inhouse staff, contractors, and cooperating U.S. and international institutions indexed 502,056 articles in FY2002, an 8% increase from the previous year. All organizations that index for NLM are now using the new online indexing data creation and maintenance system (DCMS). Indexed

citations were updated to reflect 30 retractions, 4,971 errata, and 25,091 comments. In FY2002, this practice was extended to link evidence-based medicine summaries and patient summaries to the citations for the articles to which they refer. The DCMS was modified to support a more efficient method for handling all of these links, which are now created during initial data entry rather than as a separate task at the end of the indexing workflow. Web-formatted portions of the Indexing Manual are also available to indexers from within the DCMS system. FY2002 was the first year in which annual MeSH updates were applied to citations within the DCMS. In the reinvented system, a process that previously took several months was accomplished in one day.

In March 2002, indexers began creating annotated links between newly indexed MEDLINE citations for articles describing gene functions in six organisms (human, mouse, rat, fruit fly, zebrafish, HIV-1) and corresponding gene records in the NCBI LocusLink database, as a routine by-product of MEDLINE indexing. The full-scale implementation of gene indexing followed a successful test last year and the completion of requisite modifications to the DCMS, LocusLink, and the indexing contracts, More than 18,000 annotated links were created in FY2002, providing an important new service to researchers.

In August 2002, the "Medical Text Indexer"(MTI), developed by the NLM-wide Indexing Initiative project led by LHC, was added as a new feature of the DCMS, where it is available to provide assistance to indexers who wish to use it. MTI uses the text words in a journal article title and abstract to generate a ranked list of potentially appropriate MeSH terms for indexing which indexers may select without rekeying, as they index from the complete text of the journal article. MTI is also used for completely automated MeSH indexing of the Meeting Abstracts database available through the NLM Gateway.

Indexers perform their work after the initial data entry of citations and abstracts is completed by one of three methods: electronic submissions from publishers, scanning and optical character recognition (OCR), or double-keyboarding. Of the citations added in FY2002, 56% were submitted electronically by publishers, 22% were scanned and OCRed, and 22% were keyboarded. The number received electronically, the fastest and most economical method, increased 33% from FY2001. A total of 376 publishers are now supplying XML-tagged electronic citations and abstracts for 2,045 journals. In FY2002, BSD initiated a project to contact publishers who submit electronic data for some, but not all, of their journal titles that NLM indexes to encourage electronic submission for additional titles. NLM is also attempting to find ways to help small publishing operations in developing countries to publish and submit data electronically. In FY2002, Nancy Kamau, international participant in the NLM Associate Fellowship program from Kenya, developed procedures for creating and submitting XML-tagged citation and abstract data for the African Journal of Medical Sciences which is published by her home institution in Kenya. Ms. Kamau briefed a World Health Organization-sponsored meeting of medical journal editors from Sub-Saharan Africa on these procedures.

NLM selects journals for inclusion in MEDLINE and Index Medicus based on the advice of the Literature Selection Technical Review Committee (LSTRC) (Appendix 6), an NIH-chartered committee of outside experts. In FY2002, the Committee reviewed 418 journal titles and rated 115 highly enough for immediate inclusion in MEDLINE; another 89 titles were accepted provisionally, pending receipt of acceptable electronic citation and abstract data from their publishers. A special review of health information journals written for the general public was conducted with assistance from the Consumer and Patient Health Information Section of the Medical Library Association. It led to the addition of another 5 titles. During FY2002, a special effort was made to review all in-scope titles sponsored by the SPARC and BioOne initiatives. All nursing titles formerly indexed for the International Nursing Index were reviewed to determine which should continue to be indexed for MEDLINE.

Information Products

NLM produces databases, publications, and other resources that incorporate its authoritative indexing, cataloging, and thesaurus data and link to other sources of biomedical information. LO collaborates with other NLM program areas to produce some of the world's most heavily used biomedical and health information resources.

Databases and Web Resources

Users conducted about 382 million searches of MEDLINE/PubMed in FY2002, about 1.6 million via the NLM Gateway and the rest directly in PubMed. Groups of journal citations from the former BIOETHICSLINE and HISTLINE databases were transferred into PubMed, after extensive work to identify the unique citations that should receive this treatment. BSD assisted NCBI in designing, developing, and testing many enhancements to PubMed functionality in FY2002, including a new systematic review search filter; links from comments and corrections to the associated citations; the PubMed Text version (which is also useful for PDA implementations), MEDLINE updates to PubMed 5 times per week, and a new PubMed journals database. PubMed's LinkOut for Libraries feature was expanded to allow libraries to created customized displays of print, as well as electronic, holdings based on the holdings data they have submitted to SERHOLD, the serial holdings database that is used by DOCLINE to route document requests. At the end of FY2002, about 450 libraries were participating in LinkOut for Libraries and about 150 were displaying print as well as electronic holdings. NLM itself uses LinkOut to display

electronic journal holdings to patrons in its Reading Rooms.

BSD staff also worked with LHC to design, develop, and test enhancements to the NLM Gateway including improved document ordering options, email verification when sending results, and the addition of the MEDLINEplus encyclopedia to the consumer health category. The Gateway provides the interface to OLDMEDLINE, the database of pre-1966 indexing data. LO continues to make progress on converting its retrospective indexing data to electronic form. Data for the 1957 Current List of Medical Literature was added to OLDMEDLINE in FY2002. Data from 1953–1956 was converted and will be made available for searching in FY2003. Quality review of converted data from all series of the Index-Catalogue of the Library of the Surgeon General's Office was completed, the Encompass software was selected as the initial vehicle for making these data publicly available, and work began on building a searchable Index-Catalogue database.

Use of MEDLINEplus, NLM's Web information service for the general public, increased 87% to 116 million page views in FY2002. Nearly 3.5 million unique users visited the site in the 4th quarter. MEDLINEplus continues to receive substantial publicity and recognition as an unbiased source of high quality health information. Under the direction of PSD's Web Management Team, the number of health topic pages increased 15% to 569 and the number of interactive health tutorials grew more than 230% to 151. The tutorial on anthrax, quickly added to the site in the wake of the anthrax contamination incidents, received substantial favorable publicity and extremely heavy use. A second, more consumer-friendly drug information resource, the MedMaster database produced by the American Society of Health-System Pharmacists, was added in April 2002. The PSD Web Management team and OCCS collaborated with the National Institute on Aging (NIA) to produce NIHSeniorHealth.gov, a new Web resource for seniors that is included in MEDLINEplus. Released in March 2002 with a small number of topics, NIHSeniorHealth employs large fonts, large navigation buttons, short quizzes, and a simplified automatic approach to downloading video clips. Its design reflects the results of NIA-funded research on how seniors learn, as well as extensive usability and focus group testing.

MEDLINEplus en español (medlineplus.gov/esp), a comprehensive Spanish-language consumer health site, debuted on September 9, 2002. Designed with focus group input and usability testing with Spanish speakers, the Spanish site contains more than 480 health topic pages, the ADAM illustrated encyclopedia, 60 interactive tutorials, and "toggle" links on virtually every page between the Spanish and English. The development of MEDLINEplus en español required OCCS to add substantial new functionality to the extensive MEDLINEplus development and management system. The Spanish version is maintained in a parallel workflow by existing MEDLINEplus staff and contractors in addition to two new Spanish-speaking contract staff. PSD and OCCS worked with the University of North Carolina to complete the prototype procedures and technical mechanisms for linking complementary local health services to and from MEDLINEplus. NC Health Info, the first site that will make use of these capabilities, will become publicly available in December 2002. OCCS awarded a small purchase order to test PDA access to MEDLINEplus content. PSD is providing feedback to the developer.

Under the direction of NICHSR, NLM continued to enhance its services for public health professionals. In November 2001, NICHSR released a Healthy People 2010 Information Access site. Developed by NLM and the Public Health Foundation as a project of the Partners in Information Access for Public Health Professionals, this site provides access to information that can assist in developing strategies to meet public health goals. The new site includes evidence-based PubMed strategies that retrieve citations to articles relevant to selected Healthy People 2010 objectives; provides access to the complete text of Healthy People 2010, which has been enhanced with links from its references to corresponding PubMed citations; and points to other relevant resources for the objectives, including MEDLINEplus topics and toxicological and environmental health resources produced by NLM's Specialized Information Services Division.

In July 2002, NLM awarded the New York Academy of Medicine a contract to develop a Web-based Resource Guide for Public Health Preparedness. The Guide will provide a single point of access to essential information resources in public health and disaster preparedness, selected and reviewed by subject specialists from a wide range of disciplines. In addition to linked full-text resources, the Guide will include topical bibliographies linked to pre-scripted searches of PubMed and other literature databases. An advisory board of experts in public health and emergency preparedness will provide oversight for the project.

NICHSR also coordinated NLM's response to a request from the Centers for Disease Control and Prevention for help in making important retrospective documents about smallpox available on the Web. The documents were selected by CDC staff from among the references in two key works: Chapter 6, Smallpox and Vaccinia by D.A. Henderson and B. Moss from Vaccines, 3rd ed. (1999) and Smallpox and its Eradication, published by the World Health Organization in 1988. NCBI mounted the Vaccines chapter on the Bookshelf within Entrez and WHO scanned its 1988 volume and made it available on the Web. NLM staff in BSD, TSD, and NICHSR contacted the publishers of many cited documents for permission to scan them. In some cases, publishers were already working on converting the materials to electronic format; NLM scanned seven books and reports for this project. The electronic full-text is accessible from related cataloging records in LOCATORplus, as well as from locations on CDC's Web site.

FY2002 saw significant improvements in NICHSR's suite of databases for health services researchers. A new Health Services/Sciences Research Resources (HSRR) database contains basic descriptive information about research datasets, instruments, indices, and software tools employed in health services, behavioral, and social sciences research, with links to related information on the Web and pre-formulated PubMed searches that retrieve citations reporting on studies that used the resource described. Examples of resources described in this database include Asthma Quality of Life Questionnaire, the Duke Activity Status Index, the Medical Expenditure Panel Survey, and the AMA Physician Masterfile. A streamlined version of HSTAT (Health Services/Technology Assessment Text) with enhanced search features was released in January 2002 and in July 2002 responsibility for supporting this application was transferred from LHC to OCCS, following a phased transition project involving LHC, OCCS, and LO staff. HSTAT content continues to expand steadily, with more than 1,000 documents in the current collection. FY2002 additions included many AHRQ Evidence Reports and many new sections of both the Guide to Clinical Preventive Services and the Guide to Community Preventive Services. Thanks to expanded efforts by the AcademyHealth (formerly Academy for Health Services Research and Health Policy) and the Sheps Center at the University of North Carolina, HSRProj continues to expand its coverage of health services research-in-progress funded by states, private foundations, and federal agencies. The number of funding agencies with projects represented in HSRProj now exceeds 100. During FY2002, OCCS completed work on a new data input and maintenance system for the HSRProj database, which makes use of some parts of the DCMS.

Machine-Readable Data

NLM leases many of its electronic databases to other organizations in order to promote the broadest possible use of its authoritative bibliographic and thesaurus data. There is no charge for any NLM database, but recipients must abide by use conditions which vary depending on the database involved. The commercial companies, international MEDLARS® centers, universities, and other interested organizations that license NLM data incorporate them into a variety of database and software products and use them in a range of research and development projects.

Demand for MEDLINE data in XML format continues to increase, with the majority of new users interested in using the data for data-mining and research. There are now 170 licensees of MEDLINE data. Fifty-six organizations licensed MEDLINE for the first time in FY2002. Thirty-eight of them receive the data under a new research-only license for international users, developed by BSD in FY2002. BSD also coordinated distribution of the many Specialized Information Services Division files that became available for ftp distribution in XML format in FY2002: ChemIDplus, CCRIS, DIRLINE, Gene-Tox, HSDB, and TOXLINE Special.

To improve access to new additions to the MeSH supplementary concepts, the MeSH Section initiated weekly updates for ftp distribution and also to the MeSH browser, which is heavily used by NLM indexers and catalogers as well as by external users. LO developed XML DTDs for NLM cataloging data and for meeting abstracts data.

The UMLS Knowledge Sources are available via ftp, on CD-ROM, and via the applications programming interface or interactive use of the UMLS Knowledge Source Server. There are 1,932 UMLS licensees. In FY2002, BSD streamlined CD-ROM distribution procedures and instituted new procedures for soliciting the brief annual reports required from licensees. BSD and NICHSR staff members did extensive testing of the interactive features of a new version of the UMLS Knowledge Source Server that was released in February 2002.

Web and Print Publications

NLM's electronic databases and Web site are its primary publication media, although the Library continues to publish some of its authoritative bibliographic and thesaurus data in print publications. In FY2002, LO and OCCS completed the work required to move the production of the monthly Index Medicus, the List of Journals Indexed in Index Medicus, and the List of Serials Indexed for Online Users to NLM's reinvented systems, ending a period of significant publication delays.

Overall use of NLM's main Web site increased 11% to 41 million page views by 5.3 million unique users. Use of the publication pages remained heavy at 2.4 million page views. Web publications include recurring newsletters and bulletins, fact sheets, technical reports, and multimedia catalogs. Issues of the Current Bibliographies in Medicine continue to be popular. Each issue in 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. LO staff members collaborate with outside experts to produce each bibliography. FY2002 additions to the series included: Symptom Management in Cancer: Pain, Depression, and Fatigue; Management of Hepatitis C; Trimethylaminuria and the Flavin Monooxygenases; and Management of the Clinically Inapparent Adrenal Mass (Incidentaloma). The Web-based NLM Technical Bulletin, edited by the MEDLARS Management Section, provides timely, detailed information about changes and additions to a broad range of NLM services and is particularly valuable for librarians and other health information professionals.

PSD's Web Management Team serves as the Web master for NLM's main Web site. There were substantial improvements to the systems used to manage NLM Web sites in FY2002. In November 2001, a revised Web "new files" input system was implemented. In April 2002, PSD and OCCS implemented Teamsite, a comprehensive Web file management application that allows improved quality control and validation of the static (non-database driven) files on the main Web site, MEDLINEplus, and NIHSeniorHealth. Staff throughout NLM were trained in the use of the new systems. In September 2002, the search engine for the main site was replaced by RecomMind, a concept-based retrieval system that groups items retrieved by their source within the Web site. NLM conducted an online survey of users of its main Web site in July: 80% of respondents always or frequently found what they desire, 88% were satisfied with the Web site, and 94% indicated that they were likely to return. These results compare favorably with similar surveys of other Web sites.

Direct User Services

In addition to NLM's information products, 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 members of the National Network of Libraries of Medicine and to international libraries to fill requests for materials that are not readily available from other sources. LO also retrieves documents from NLM's closed stacks for use by onsite patrons.

In FY2002, PSD's Collection Access Section processed a total of 705,069 document requests, a 3% increase from last year. Onsite users requested 331,777 contemporary documents from NLM's closed stacks, a 4% decrease, and 6,870 items from the historical and special collections, a 42% increase attributable to better bibliographic access to these collections via the Web and expanded outreach. The total number of new onsite patrons registered (4,182) decreased by 24%, no doubt due in part to additional building security measures implemented after the September 11, 2001 attacks. The reduction in the number of contemporary items requested from NLM's closed stacks may also have been affected by the increased availability of electronic journals in the main NLM Reading Room.

Other libraries requested 373,292 contemporary documents from NLM, a 10% increase from FY2001. There was an artificial dip in interlibrary loan (ILL) requests to NLM last year due to changes in DOCLINE routing features. Also in FY2002, NLM experienced increased demand for copies of articles in electronic journals and the most current issues of research journals, due to a combination of license agreements that prohibit ILL, budget cuts, and serial price increases that affect the ILL service provided by academic health sciences libraries. NLM handled 87% of the requests within 12 hours and delivered 64% of filled requests electronically. A Relais-Express workstation was installed to replace stand-alone Ariel, eliminating the need for separate equipment for each delivery method. The Collection Access Section created a Web-based form for reporting problems with ILL requests submitted to NLM. The form feeds into NLM's customer service request tracking system.

A total of 3,257 libraries now use DOCLINE: 2,904 in the U.S., 304 in Canada, and 49 in other countries. Work is under way to add several libraries from Mexico to DOCLINE. DOCLINE users entered 3.04 million requests into the system in FY2002, up 4% from FY2001; 91% of the requests were filled. The DOCLINE coordinators at the Regional Medical Libraries in the NN/LM continue to provide invaluable input to LO on system features and document delivery operations. In FY2002, NLM awarded a subcontract to the University of Connecticut to expand and upgrade its Electronic Funds Transfer System (EFTS) to handle payment for document delivery transactions for interested libraries in all NN/LM regions. The system was already in heavy use in four regions. EFTS streamlines and reduces the administrative costs of billing and paying for interlibrary loan transactions.

The requests are routed automatically based on automated serials holdings data in the SERHOLD® database. At the end of FY2002, SERHOLD contained 1.38 million holdings statements for 51,587 serial titles held by 3,028 libraries. In FY2002, NLM worked with OCLC to test procedures for automated exchange of holdings data between SERHOLD and the OCLC database for libraries that use both systems. As previously described, SERHOLD data is now also employed to display libraries' print holdings in the LinkOut for Libraries feature in PubMed.

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. Individuals submitted 924,538 Loansome Doc requests to DOCLINE libraries, 8% more than in FY2001. Twenty libraries outside the U.S. provide Loansome Doc service; these libraries delivered more than 70,000 documents to users around the world. DOCLINE version 1.4, released in September 2002, included a number of enhanced features for using and managing Loansome Doc service.

Reference and Customer Service

LO provides reference and research assistance to onsite and remote users as a backup to services available from other health sciences libraries. LO also has primary responsibility for responding to inquiries from those

seeking information about NLM's products or services or assistance in using these services. PSD's Reference and Customer Section handles all initial inquiries and many of those requiring second-level attention. Staff from throughout LO and NLM assist with second-level service when their expertise is required.

In FY2002, Reference and Customer Service staff handled 97,548 inquiries from onsite users, email messages, and telephone calls. An additional 40,694 "junk" messages were received by the customer service email address, which added to the workload despite steps taken to improve automated handling of such messages. Questions from remote patrons decreased 17% and from onsite patrons 5%. Correspondence and telephone calls may have declined due to expanded MEDLINEplus content, more and better FAQs and information screens, and online tutorials on the NLM Web sites. At least part of the decline in onsite traffic is probably due to increased NIH campus security in the aftermath of the 9/11 attacks which complicates access to NLM. Reference and Customer service staff conducted several studies of onsite use of the collections and the Learning Resource Center as background for discussions about likely future onsite use and space requirements.

During FY2002, LO worked with OCCS to improve automated support for the customer service operation. The Siebel Customer Service Relations software was selected as a replacement for CustQ. The software was configured and tested for initial implementation in October 2002. Staff also experimented with Native Minds, a Virtual customer service representative (Vrep), converting existing Web FAQs into a format that may allow the Vrep to answer some categories of customer service inquiries automatically.

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 which works to equalize access to health information services and technology for librarians, health professionals, and the general public throughout the U.S.; participates in NLM-wide efforts to develop and evaluate outreach programs designed to improve health information access for underserved minorities 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 and other health professionals. Many LO staff members give presentations, demonstrations, and classes at professional meetings and publish 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 for U.S. health professionals, researchers, and the general public, irrespective of their geographic location. The NN/LM program is the core component of NLM's outreach program and its efforts to reduce health disparities. The network includes nearly 5,100 regular and affiliate members. The regular members are libraries with health sciences collections, primarily in hospitals and academic health sciences centers; the affiliate members, including some small hospitals, public libraries, and community organizations, provide health information service, but have little or no physical collection of health-related literature. 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.)

In addition to the basic NN/LM contracts, NLM funds subcontracts for four centers that serve the entire network. The activities of two of these, the National Training Center and Clearinghouse at the New York Academy of Medicine and the Electronic Funds Transfer System at the University of Connecticut, are described elsewhere in this chapter. The Outreach Evaluation Resource Center at the University of Washington provides training and consulting services throughout the NN/LM and assists NLM, the RMLs, and other network members in designing methods for measuring the effectiveness of overall network programs and individual outreach projects. In FY2002, NLM and the RMLs decided to define uniform national measures of the Network's success in reaching two outreach objectives: (1) improving health information access via public libraries and (2) connecting local public health departments to health information services. Individual outreach projects will continue to have project-specific evaluation measures.

The new National Outreach Mapping Center (NOMC), established at Indiana University in Indianapolis in FY2002, will help NLM and the RMLs to display the geographic distribution and impact of NN/LM programs and services and to identify gaps that should be addressed. An NLM-wide group worked with the RMLs and the NOMC to define and collect uniform data for an outreach database that will provide the raw data for mapping NN/LM activities, NLM grant awards, and a range of outreach activities undertaken by other NLM components. In addition to measuring and mapping NN/LM outreach activities, NLM and the RMLs identified other priorities for cross-regional action. These include: outreach to Native American tribal organizations, building on ongoing tribal connections work in the Northwest Region; identification and scanning of consumer health materials in additional languages; and developing strategies for preserving resource sharing options in the era of electronic publication.

The RMLs and other network members develop and conduct many special projects to reach underserved health care professionals and to improve the public's

access to high quality health information. Most of these projects involve partnerships between health sciences libraries and other organizations, including public health departments, public libraries, schools, and community-based organizations. Each RML routinely identifies outreach opportunities in its region by soliciting proposals and through its involvement with regional institutions. Periodically, a national call for outreach proposals is issued simultaneously in all NN/LM regions. One such solicitation was issued on August 1, 2002; awards will be made in FY2003.

In FY2002, the eight RMLs issued a total of 53 outreach subcontracts involving projects in 27 states and the District of Columbia. The projects involve network members in hospitals, academic health science centers, Area Health Education Centers, and public libraries; include a wide range of organizations within communities, such as churches, rural public health departments, non-profit clinics, schools, and community-centers; and target both health professionals and members of the public. LHC assisted in funding several projects that highlighted access to ClinicalTrials.gov.

NLM and the NN/LM collaborate with the CDC, other Federal agencies, and an increasing number of public health associations to improve access to information technology and information services for the public health workforce and to enhance the ability of health sciences librarians to serve this diverse population. Three new partners joined in FY2002: the American Public Health Association, the Association of Schools of Public Health, and the Society for Public Health Education. NICHSR coordinates the Partnership for NLM, and the National Network Office and NLM's Specialized Information Services Division are also heavily involved in Partnership activities. In addition to new public health information resources mentioned elsewhere in this report, the Partnership sponsors satellite training programs and presentations, exhibits, and short courses at public health meetings; publicizes and contributes information to the Public Health Training Finder database maintained by the Public Health Foundation; promotes use of NLM and NN/LM services and funding opportunities to the public health community; and fosters outreach initiatives to the public health workforce. In FY2002, the Partnership organized a number of special training opportunities for public health professionals and health sciences librarians, obtained contact and geographic information for local health departments for use in NN/LM outreach planning and mapping activities, and made connections between the RMLs and the public health distance learning coordinators in each state.

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 organizes the exhibits at the Medical Library Association annual meeting, the American Library Association annual meeting, some of the health professional and library meetings held in the Washington, DC area, and some distant meetings focused on health services research, public health, and history. In FY2002, NLM and NN/LM services were displayed at 207 exhibits at national, regional, and state association meetings across the U.S. These exhibits highlight not just the databases and services to which LO contributes, but also NLM products relevant to attendees at each meeting. New table top exhibit structures were developed for the RMLs to use at smaller meetings. BSD and the NLM Office of Communications and Public Liaison produce a variety of bookmarks, CDs, and other promotional items for distribution at professional meetings. In response to requests from the field, BSD arranged for a wide range of items with the NLM logo, e.g., shirts, mugs, clocks, pens, to be available for purchase online through the NIH Recreation and Welfare Association.

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 disparities and participates actively in the Library's Committee on Outreach, Consumer Health, and Health Disparities. In FY2002, the Office of the Associate Director, LO worked with other NLM components, the American College of Physicians, and the NN/LM to plan a project an FY2003 test of the use of "information prescriptions" for MEDLINEplus in physicians' offices in Georgia and Iowa.

The Office of the Associate Director, LO, the NN/LM Office, and BSD continued to collaborate with the Public Library Association (PLA), a division of the American Library Association (ALA), to improve public library awareness of MEDLINEplus, ClinicalTrials.gov, and other NLM services. After a successful test in eight states last year, NLM and PLA sent a joint mailing, signed by Dr. Lindberg and the PLA President, to all U.S. public libraries on ALA's mailing list describing MEDLINEplus and other NLM services for the general public and including bookmarks and other promotional material. A similar mailing, signed by Dr. Lindberg, was sent to all NN/LM members. A reply card was included in both mailings, inviting recipients to request additional bookmarks if desired. NLM received requests for additional materials from about 20% of both the public libraries and the NN/LM members, an extremely high rate of return for mailings of this sort. LO also organized a well-attended session on consumer health outreach projects involving public libraries at the 2002 annual ALA annual meeting.

LO staff members are actively involved with NLM's partnership with Wilson High School in the District of Columbia. In FY2002, LO provided special training sessions for students, teachers, and school librarians during the school year and provided summer employment and training opportunities for several students and teachers.

Historical Exhibitions and Programs

LO's History of Medicine Division mounts major exhibitions in the NLM rotunda, with assistance from other NLM components. Designed for the interested public as well as the specialist, these exhibitions are an important part of NLM's outreach program. The Once and Future Web: Worlds Woven by the Telegraph and Internet, was installed in May 2001 and remained on exhibit at NLM until July 2002. The electronic version of the exhibition is permanently available on the NLM Web site. The Once and Future Web was the first NLM exhibition to feature an online "Learning Station" designed for teachers and students in grades 6–12. In FY2002, HMD provided tours of the exhibition to groups totaling more than 3,100 people, many of them students and international visitors. HMD also presented a free series of films featuring the telegraph or the Internet, with introductory remarks and a question and answer period provided by scholars and historians.

During FY2002, HMD developed and installed Dream Anatomy, an exhibition focusing on anatomy, medicine, and the artistic imagination. The exhibition, which opens officially on October 9, 2002 and will run through July 2003, features rare anatomical books and illustrations from the NLM collection, as well 20th and 21st century art, holograms, and interactive displays that draw upon the Visible Human datasets. The online version of this exhibition also features a "Learning Station". A number of public programs related to the exhibition are planned for the next fiscal year. Dream Anatomy was a late insertion into NLM's exhibition schedule when it became clear that a planned exhibition on American women physicians would benefit from a longer planning phase. Tenley Albright, M.D., distinguished surgeon and former chair of the NLM Board of Regents, chairs the committee of eminent physicians (both women and men) who are advising NLM on the content of Changing the Face of Medicine: the Rise of America's Women Physicians, now scheduled to open in the fall of 2003.

While HMD worked on new shows, previous NLM exhibitions continued to find new audiences in new formats. On May 7, 2002, the "virtual" DVD version of Breath of Life, a previous NLM exhibition on asthma, was featured at the CDC's observance of the 4th World Asthma Day in Atlanta. ALA, in conjunction with HMD, developed a traveling version of Frankenstein: Penetrating the Secrets of Nature, originally displayed at NLM in 1997/98, with funding from the National Endowment for the Humanities and NLM. ALA solicited proposals from public, academic, and medical libraries interested in displaying the exhibit and was surprised by the number of applications. Four copies of the traveling exhibit will be shown at more than 80 libraries over a 2-year period, beginning in October 2002. Hosting libraries will present a variety of public programs related to science, medicine, and the humanities in conjunction with the exhibit. Rutgers University Press published the catalogue of the exhibition, Frankenstein: Penetrating the Secrets of Nature, An Exhibition by the National Library of Medicine, by Susan Lederer in conjunction with the launching of the traveling version.

Staff from HMD and LHC worked with the British Library to add Vesalius's Humani Corporis Fabrica as the second medically significant book to "Turning the Pages," a remarkable program developed by the British Library that uses computer-animation, high-quality digitized images, and touch-screen technology to simulate that action of turning the pages of rare books. Both the original work by Vesalius and the "Turning the Pages" version, augmented by LHC with interactive links to additional related resources such as the Visible Humans, are featured in the Dream Anatomy exhibition. "Turning the Pages," with Elizabeth Blackwell's A Curious Herbal and De Humani Corporis Fabrica, is now on display in the NLM Visitor Center and also just inside the entrance to the HMD Reading Room. It was removed from the main lobby of the NLM building where glare from overhead lights made viewing difficult.

In addition to the major exhibitions in the rotunda, HMD installs "mini-exhibits" in cases near the entrance to the HMD Reading Room. 'I Swear by Apollo': Greek Medicine from the Gods to Galen was on view from December 2001 to May 2002. Smallpox: A Great and Terrible Scourge, developed by the Office of the Public Health Service Historian, LHC, appeared from June to November 2002. Online versions of these exhibits are available on the HMD Web site. NLM displayed an exhibit on the history of the Pan-American Health Organization, A 100-year Quest for Health in the Americas, 1902-2002" in the Lister Hill Center lobby from March 7–April 5, 2002.

In FY2002, NICHSR made the 2000 NLM video Health Services Research: A Historical Perspective available for viewing on the Web, with and without closed captioning. This video continues to see heavy use in introductory courses on health services research in a wide variety of institutions.

HMD organizes a range of public programs in the history of medicine. On October 29, 2001, NLM held an all-day symposium on New Frontiers of Biomedical Research, 1945–1980, which featured scientists Joshua Lederberg, Julius Axelrod, and Donald Fredrickson and historians of science Nathaniel Comfort, David Hart, David Healy, Ellen Herman, Susan Lindee, Bill Leslie, and Jan Sapp. On May 14, 2002, the Library held a special evening celebration, in conjunction with a meeting of the NLM Board of Regents, to honor donors to NLM's historical collections. This event had originally been scheduled for the night of September 11, 2001. HMD arranges a regular series of seminars by historical scholars as well as special historical lectures in conjunction with the NLM Diversity Council. In FY2002, lecturers included: David Keltz, on "The Illness & Death of Edgar Allan Poe"; Eric Bailey on "Tracing the Roots of Black Folk Medicine: A Cultural Anthropological Approach"; and

Barron H. Lerner on "No Shrinking Violet: Rose Kushner and the Rise of Breast Cancer Activism."

HMD staff members presented historical papers and lectures at professional meetings throughout the year and also published the results of their scholarship in books, chapters, articles, and reviews. HMD continued to play a lead role in preparing the monthly feature "Voices from the Past" and the "Images of Health" feature, which uses items from the NLM collection, for the American Journal of Public Health. As previously mentioned, two distinguished historians spent several months at NLM as visiting historical scholars and evaluated segments of the Library's collection while in residence.

Training and Recruitment of Health Sciences Librarians

LO develops online services training programs for health sciences librarians and other search intermediaries; oversees the activities of the NN/LM 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, the UMLS resources and other topics. LO also collaborates with the Medical Library Association, the Association of Academic Health Sciences Libraries, and the Association of Research Libraries to increase the diversity of those entering the profession, to provide leadership development opportunities, to promote multi-institution evaluation of library services, and to explore specialist roles for health sciences librarians.

In FY2002, the MEDLARS Management Section and the NN/LM National Training Center and Clearinghouse (NTCC) at the New York Academy of Medicine taught PubMed, NLM Gateway, and/or ToxNet searching to 886 students in 61 traditional face-to-face classes, a 31% decrease from the previous year, mostly due to a sharp reduction in attendance at ToxNet classes. The Web-based PubMed tutorial was updated several times to reflect new PubMed features, using viewlet technology. The NTCC released its new Educational Clearinghouse database, which collects information about training courses and Web-based training materials produced by NLM, the RMLs, and other NN/LM members. Information about additional courses and materials can be submitted via the Web. Many NN/LM members provide training to health professionals, researchers, students, and the general public and produce syllabi, handouts, search examples, etc. for different audiences. The goal of the Clearinghouse is to reduce unnecessary duplicate effort by making these materials available for re-use or adaptation in other settings.

There were five first-year and two second-year participants in the Associate Fellowship program in FY2002. Of the five who finished the first year at NLM in August, three elected to continue on the optional second year at The Johns Hopkins University, the University of Maryland, and the University of Tennessee. One accepted a job at the University of British Columbia, and the international Associate returned to the Kenya Medical Research Institute. Of the two second-year participants, one accepted a job in the Reference and Customer Service Section at NLM. Six new first-year Associates entered the program in September 2002. There is no international Associate in the new group.

The NLM Long Range Plan, 2001–2005 recommends that NLM examine the need to expand the supply of specialist librarians in clinical informatics, bioinformatics, and health policy. NLM provided funding to the MLA for an April 2002 conference, held at NLM, which explored the concept of the "informationist," who can be viewed as a librarian or other information specialist who provides very tailored information and educational services as a member of a clinical care or research team. The conference looked at existing models; addressed roles for "informationists" or specialist librarians in clinical care, clinical and basic science research, and education; and discussed how additional demonstrations of the use of informationists could be funded and evaluated. Materials from this conference and from follow-up MLA activities are available on MLA's Web site (http://mlanet.org/research/informationist/). As one follow-up to the conference, NLM will assist the NIH Library in evaluating the use of informationists in clinical research teams at NIH Institutes. In FY2002, the Library introduced several new courses that may assist health sciences librarians in preparing for specialist roles. With assistance from NICHSR, BSD developed an introductory course on the UMLS Knowledge Sources and programs that debuted at the 2002 MLA annual meeting. The course was then expanded to include hands-on exercises and offered twice at NLM in the fall. LO helped NCBI to publicize another new course, NCBI Advanced Workshop for Bioinformatics Information Specialists, which is directed primarily at library-based bioinformatics specialists who assist faculty and students in using genetic databases and software. Several specialists assisted NCBI in developing this course and serve on its faculty.

NICHSR continued to enhance and add to its suite of courses and resources on various aspects of health services research, health policy, and public health. It sponsored a course on "Health Economics Information: The Quest for Efficiency in Health Care" which is currently being reformatted as a distance learning module to be mounted on the NICHSR Web site. Two new e-publications were also promoted to the site: a new case study on bioterrorism (part of the Introduction to Health Services Research Class Manual) and an e-book Guide to Information Resources for Health Technology Assessment. NICHSR also contracted for development of a core library list in health economics.

NLM collaborates with a variety of organizations on librarian recruitment and leadership training initiatives. Individuals from minority groups continue to be under-represented in the profession at a time when outreach to

underserved groups is a high priority. A large percentage of health sciences library directors (and librarians in general) will retire over the next 5–10 years. Over the past few years, LO has contributed to minority scholarship opportunities available through the Medical Library Association, the American Library Association, and the Association of Research Libraries (ARL). In FY2002, LO provided support to the Association of Academic Health Sciences Libraries for a 3-year trial of a new leadership development fellowship program, modeled after an existing successful ARL program. The program will provide leadership training and mentorship for 5 health sciences librarians annually. Directors of AAHSL libraries will serve as mentors. AAHSL has contracted with ARL to provide some of the leadership training.

Health Informatics Activities

In addition to providing the Library's basic services, LO represents NLM in several activities designed to promote more effective health applications of advanced computing and communications technologies. On behalf of several Federal agencies, LO initiated and now manages a contract with the Regenstrief Institute that supports 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 laboratory test results and in the output of devices that perform certain laboratory tests. During FY2002, LO continued to negotiate with the College of American Pathologists for a U.S.-wide arrangement for the use of the SNOMED clinical terminology.

LO continued to serve on the Department of Health and Human Services Data Standards Committee that is overseeing the implementation of the administrative simplification provisions of the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and on the staff to the National Committee on Vital and Health Statistics (NCVHS) Standards and Security Subcommittee. In this latter capacity, LO had lead responsibility for organizing several hearings for the NCVHS Subcommittee on HIPAA code set issues. LO staff also briefed the Subcommittee on the UMLS and on the development of the RxNorm clinical drug vocabulary (described previously in this chapter). In FY2002, LO also began to represent NLM on the Public Health Data Standards Coordinating Committee, which includes federal agencies and representation from state and local public health departments.

Table 1

Growth of Collections

Collection	*Previous Total (9/30/01)*	*Added FY 2002*	*New Total (9/30/02)*
Book Materials			
Monographs:			
Before 1500	583	1	584
1501-1600	5,874	48	5,922
1601-1700	10,172	39	10,211
1701-1800	24,560	53	24,613
1801-1870	41,286	93	41,379
Americana	2,341	0	2,341
1870-Present	696,080	16,105	712,185
Theses (historical)	281,794	0	281,794
Pamphlets	172,021	0	172,021
Bound serial volumes	1,225,098	27,580	1,252,480
Volumes withdrawn	(75,127)	(3,000)	(78,127)
Total volumes	2,384,682	40,919	2,425,403
Nonbook Materials			
Microforms:			
Reels of microfilm	116,431	16,113	132,544
Number of microfiche	438,684	7,110	445,794
Total microforms	555,115	23,223	578,338
Audiovisuals	68,290	2,449	70,739
Computer software	1,992	146	2,138
Pictures	56,960	2	56,962
Manuscripts	3,136,857*	984,025	4,120,882
Total nonbook	3,819,214	1,009,845	4,829,059
Total book and nonbook	**6,203,896**	**1,050,764**	**7,254,660**

*Equivalent to 1,792 linear feet.

Table 2

Acquisition Statistics

Acquisitions	*FY 2000*	*FY 2001*	*FY 2002*
Serial titles received	23,141	20,314	20,350
Publications processed:			
Serial pieces	143,636	142,642	133,908
Other	22,384	21,338	22,274
Total	166,020	163,980	156,182
Obligations for:			
Publications	$4,895,999	$5,155,054	$5,802,023
(For rare books)	($267,300)	($279,710)	($446,039)

Table 3

Cataloging Statistics

	FY 2000	*FY 2001*	*FY 2002*
Completed Cataloging	20,067	19,024	21,419

Table 4

Bibliographic Services

Services	*FY 2000*	*FY 2001*	*FY 2002*
Citations published in MEDLINE	442,168	463,014	502,056
For Index Medicus	434,813	445,041	459,558
Journals indexed for MEDLINE/PubMed			4,538
Journals indexed for Index Medicus	3,472	3,707	3,834
Abstracts entered	341,682	345,624	398,885

Table 5

Web Services

Services	*FY 2000*	*FY 2001*	*FY 2002*
NLM Web Home Page			
Page Views	25,936,000	36,248,000	40,607,752
Unique Visitors	3,572,000	4,490,000	5,300,363
MEDLINEplus			
Page Views	18,437,000	62,069,000	116,335,454
Unique Visitors	2,098,000	4,409,000	9,594,429

Table 6

Circulation Statistics

Activity	*FY 2000*	*FY 2001*	*FY 2002*
Requests Received	749,869	682,777	705,069
Interlibrary Loan	390,574	338,627	373,292
Onsite	359,295	344,150	331,777
Requests Filled:	589,516	535,594	539,274
Interlibrary Loan*	299,182	251,525	268,816
Onsite	292,664	284,069	270,458

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

Table 7

Online Searches—All Databases

	FY 2000	*FY 2001*	*FY 2002*
Total online searches	244,000,000	313,000,000	382,000,000

Table 8

Reference and Customer Services

Activity	*FY 2000*	*FY 2001*	*FY 2002*
Offsite requests	62,971	59,634	49,153
Onsite requests	51,456	51,287	48,395
Total	114,427	110,921	97,548

Table 9

Preservation Activities

Activity	*FY 2000*	*FY 2001*	*FY 2002*
Volumes bound	31,874	31,625	25,609
Volumes microfilmed	4,513	5,131	5,255
Volumes repaired onsite	2,000	1,403	1,542
Audiovisuals preserved	46	225	283
Historical volumes conserved	385	128	66

Table 10

History of Medicine Activities

Activity	*FY 2000*	*FY 2001*	*FY 2002*
Acquisitions:			
Books	226	314	424
Modern manuscripts	1,915,550	1,340,150	840,000*
Prints and photographs	1,391	3,324	3,176
Historical audiovisuals	37	1593	1,361
Processing:			
Books cataloged	49	510	368
Modern manuscripts cataloged	87,150	190,750	984,025**
Pictures cataloged	256	20	0
Citations indexed	1,066	285	846
Public Services:			
Reference questions answered	15,143	15,718	14,898
Onsite requests filled	4,485	4,844	6,870

*Equivalent to 480 linear feet
**Equivalent to 562 linear feet

SPECIALIZED INFORMATION SERVICES

Jack W. Snyder, M.D., Ph.D.
Associate Director

The Toxicology and Environmental Health Information Program (TEHIP), known originally as the Toxicology Information Program, was established 35 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 and effective 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. Resources include chemical and environmental health databases and Web-based information resource collections. 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.

The SIS Web server provides a central point of access for the varied programs, activities, and services of the Division. Through this server (http://sis.nlm.nih.gov) 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. 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 FY2002 SIS focused on several projects for significant re-engineering and initiated several new opportunities to enhance SIS information resources and provide new services in emerging areas. Prototypes have been developed which utilize graphical display of data from our information resources, innovative access and interfaces for consumers, and geographical information systems. Highlights for 2002 include a new user interface and multi-database search capability for TOXNET, our premier collection of databases on toxicology, hazardous chemicals, and related areas; Haz-Map, an occupational toxicology database designed to link jobs to hazardous job tasks that may be associated with occupational diseases and their symptoms; ALTBIB, an improved search of the multiple bibliographies of Alternatives to the Use of Live Vertebrates in Biomedical Research and Testing; several new Toxicology & Environmental Health Special Topic Web resource pages, including Chemical Warfare Agents and Arsenic; TOXMAP, a prototype system that uses maps of the U.S. to help users visually view data about chemicals released into the environment and easily connect to related environmental health information; ToxTown, a graphical portal to chemicals you might encounter in everyday life, in everyday places; successful installation of four PAHO/NLM Disaster Preparedness Information Centers in Honduras and Nicaragua, with the addition of El Salvador this year; expanded Native American outreach initiatives; and new minority outreach activities with the Historic Black Colleges and Universities, United Negro College Fund Special Projects, and the National Medical Association.

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.

Haz-Map database was released in 2002 on the Internet (http://hazmap.nlm.nih.gov). It is an occupational toxicology database designed to link jobs to hazardous job tasks that are linked to occupational diseases and their symptoms. It is a relational database of chemicals, jobs, and diseases. The Haz-Map jobs table is based on the 1997 Standard Occupational Classification system. The industries table is based on the Standard Industrial Classification system. The diseases table is based on the International Classification of Diseases (ICD-9). Information from textbooks, journal articles, and electronic databases (HSDB, ACGIH Documentation of TLVs, ATSDR Toxicological Profiles, NIOSHTIC, and others) was classified and summarized to create the database. A user may search this occupational database by chemical agent, occupational disease and by job type.

ChemIDplus (Chemical Identification File) is an NLM online chemical dictionary, which contains over 360,000 records, primarily describing chemicals of biomedical and regulatory importance, and available on the Internet (http://chem.sis.nlm.nih.gov/chemidplus). ChemIDplus features include chemical structure search and display for over 143,000 chemicals, and hyperlinked locator fields that retrieve data for a given chemical from other resources such as TOXLINE, MEDLINE or HSDB as well as EPA and ATSDR. Over 15,000 records of regulatory interest collectively known as SUPERLIST are also available and hyperlinked in ChemIDplus. During FY2002 over 54,000 queries per month were made of this database. The database was enhanced by the addition of a variety of new locators pointing to international resources, including coverage of agents found in ClinicalTrials.gov. All drug data from the publication USPDDN from the United States Pharmacopeial Convention, Inc. was also reloaded. A new Web-based maintenance system named DBMaint2 was developed and tested, and will be available in 2002. It integrates text and structure input and

modification, which will increase efficiency by staff and contractors. In addition, a new chemical spell checker was developed and tuned for ChemIDplus data. This will be released in 2002, and will help users to retrieve substances by chemical name, a type of data that is highly susceptible to spelling errors by users. In FY2002, a new test version of ChemIDplus was developed for release in 2003.

The **Hazardous Substances Data Bank (HSDB**) continues to be a highly used resource, averaging over 48,000 searches each month (a 20% increase over FY2001). 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. In 2002, there has also been an increased emphasis on adding to HSDB new chemicals with the potential for high toxicity and high human exposure. Over 125 new chemicals were added in 2002, including new pesticides and environmental pollutants. Newer sources of relevant data are being examined for incorporation into new and existing data fields within the current 4,671 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. This will replace the current Remote Data Entry System (RDES) next year. As part of this effort, a test version of a relational HSDB database using the MySQL database application was created, and a new client-server interface was programmed to allow easier updates.

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 FY2002, SIS introduced a new search interface to allow integrated access to the SIS toxicology and environmental health databases. This new search interface (http://toxnet.nlm.nih.gov) allows users to simultaneously search HSDB, TOXLINE, CCRIS, Gene-Tox, DART/ETIC, IRIS, TRI and ChemIDplus from one input screen. Based on recommendations from the Institute of Medicine, 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 was carried out in the fall of 2001, and feedback from this survey is being used for current and future planning. New search screen designs were implemented in 2002, and research and development projects such as a chemical spellchecker, automatic indexing, and a toxicology gateway system were carried out.

Alternatives to Animal Testing—SIS continued to compile and publish references from the MEDLARS files that were identified as relevant to methods or procedures that could be used to reduce, refine, or replace animals in biomedical research and toxicological testing. Staff members search, edit, and categorize citations to create a true value added resource in this field. The 22 bibliographies issued during the past 10 years are available on the Internet through the SIS Web server, and the primary distribution mechanism for this project is now the Internet. In FY2002, a new online resource named ALTBIB was made public, allowing search access to all of the 7,595 citations organized from previous bibliographies. This uses the TOXNET search engine, and is available at (http://toxnet.nlm.nih.gov/altbib.html). A user may search by keyword, author, or one of the 16 subdivisions such as "Quantitative Structure Activity Studies."

TOXLINE (Toxicology Information Online) is a large NLM bibliographic database traditionally produced by merging "toxicology" subsets from secondary sources. By the end of FY2002, the database included over 3 million citations to toxicology literature going back to 1965. In FY2002, we completed 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 journals in the area of toxicology and environmental health to MEDLINE to cover some of the literature formerly provided by outside sources. For the non-standard 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 new TOXNET user interface.

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. The quality and utility of the database continue 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 that also lists Spanish speaking customer service representatives and Spanish language publications.

The **Toxics Release Inventory (TRI)** series of files now includes five online files, TRI95 through TRI2000. 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 environmental release data for air, water, and soil for over 600 EPA-specified chemicals. These files are used in the new SIS R&D project using a geographical information system, TOXMAP.

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 we will continue to 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 on-line 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 240,000 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. In FY2002, we completed the transition to a next generation DART and created two subsets: DART Core on PubMed, containing over 170,000 citations to the journal literature and DART Special containing nearly 70,000 citations to specialized resources (including meeting abstracts, books, technical reports). Easy access to DART Special and to DART Core, is available from the new TOXNET interface. DART is funded by NLM, the EPA, the National Institute of Environmental Health Sciences (NIEHS), and the FDA's National Center for Toxicological Research, and is managed by NLM.

The **Environmental Mutagen Information Center** (EMIC) database contains over 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 EPA, NIEHS, and NLM, collaborating partners in this effort, stopped compiling this special collection as of December 1999, but SIS will keep the collections as part of the TOXLINE Special database on TOXNET.

On March 21, 2002, SIS sponsored a Children's Environmental Health Information Resources Satellite Broadcast via the CDC Public Health Training Network. The program demonstrated selected online resources in the context of important children's environmental health issues. Topics included exposure of children to pesticides, environmental triggers of childhood asthma, methylmercury and fish contamination, the use of Geographic Information Systems for environmental health data, Healthy People 2010 resources, and lead poisoning prevention funding resources. The program was designed for physicians, nurses, physician assistants, nurse practitioners, epidemiologists, public health educators, librarians, counselors, administrators, or anyone else providing environmental health-related services. The Web cast of the broadcast is available at: (http://www.phppo.cdc.gov/PHTN//Webcast/child-env/archivewc.asp). In addition, a children's environmental health resource sampler was developed (http://nnlm.gov/partners/children/sampler.html).

AIDS Information Services

NLM has continued its successful AIDS Community Information Outreach Program with 15 new awards in FY2002, bringing the total number of awards made to 157. In addition to these awards, NLM has been working with other organizations to raise awareness of HIV/AIDS information resources among small community organizations at a grassroots level.

NLM remains as the project manager for the multi-agency AIDS Clinical Trials Information Service (ACTIS) and the HIV/AIDS Treatment Information Service (ATIS). These are in the process of being merged into a new service currently titled "AIDSinfo." This new service will continue to provide access to AIDS-related clinical trials information (through ClinicalTrials.gov) and federally approved treatment guidelines. The contract for this service also provides support services for ClinicalTrials.gov.

Outreach / User Support

Special Population Web Sites: The Arctic Health Web site (http://arctichealth.nlm.nih.gov), initially developed by SIS staff, has been turned over to the University of Alaska, Anchorage for continued development. This will remain a collaborative project between SIS, the Consortium Library and the Institute for Circumpolar Health Studies at the University. A users council has been established, which includes

representatives from many of the stakeholder groups. Work is continuing on developing an Asian American Web site and an American Indian Web site. These Web sites include relevant policy, legislative, and organizational information as well as organized links to health and environmental issues of that particular population.

SIS collaborated in a training project with the DHHS Office of Minority Health. As part of their AIDS initiative, the Office 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 provided training in identifying and using HIV/AIDS information resources for representatives of community-based organizations in six cities across the country.

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 the panel held its annual meeting at the University of Puerto Rico Medical Science Campus. The University of Puerto Rico is one of the two new members of the panel. The assessment of the program was presented to the panel at their annual meeting. The panel has recommended expanding the scope of their activities beyond toxicology and environmental health, to encompass the health disparities that have been identified as disproportionately affecting minority communities.

SIS developed a new program of outreach to HBCUs in conjunction with the United Negro College Fund Special Programs Corporation (UNCFSP). This program gives NLM the opportunity of working with additional HBCUs that may not have graduate health programs. Technical assistance in the form of training in using health information resources will be provided as will other forms of support and assistance. UNCFSP serves as the intermediary in recruiting HBCUs to participate and selecting projects for funding. Several TIOP representatives are serving on the advisory board for UNCFSP as well as serving as reviewers for funding awards.

SIS initiated a new Information Internship program this year. This internship was jointly funded by NLM and the National Center for Minority Health and Health Disparities. Two representatives from the Mandan, Hidatsa, and Arikara People (Three Affiliated Tribes) started a one-year program of working with NLM and learning about health information resources and access. The internship culminates in the development of a local project on their reservation intended to improve access to health information for the tribe. This first project will be the development of a mobile computer training facility that will be moved to different locations on the reservation and for classes and opportunities to access health information.

SIS undertook a major training program with the National Medical Association. In addition to providing training at the NMA Annual Meeting and participating in the sessions of the Community Medicine Section, NLM provided day-long training sessions at six regional NMA meetings. These courses covered all of NLM's online resources including TOXNET, PubMed, ClinicalTrials.gov, and MEDLINEplus.

SIS participated in a health pilot project of the Department of Housing and Urban Affairs Neighborhood Network Centers. Neighborhood Network Centers with computers and Internet connectivity are available in approximately 1,000 HUD assisted multi-unit dwellings. During the pilot project health programs were held at 12 sites. NLM supplied a staff person to demonstrate retrieval of relevant information as part of each health program. In addition, a connection was made between the Center manager and local public and health sciences librarians.

A 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. The main objective of this project is to contribute to disaster reduction by capacity building activities in the area of disaster-related information management. Selected libraries and information centers have been provided with the knowledge, training and technology resources in order to act as reliable information providers to health professionals and others in their countries. Through this initiative, the participating libraries and information centers have been strengthened in several areas:

- Technological infrastructure (Internet connectivity and computer equipment)
- Information management (health science librarian training)
- Information product development (Digital Library, Web sites)

This project is also assisting SIS in developing models for collecting and exchanging health information in geographically isolated and disaster-prone environments and for handling non-traditional or unpublished literature, in this case on the health aspects of 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 HRSA Conference on American Indians/Alaska Natives—HIV/AIDS and Substance Abuse, the DHHS Office of Minority Health's National Leadership Summit on Eliminating Racial and Ethnic Disparities in Health, and 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 the Technology Partnerships Conference held at the Georgia Centers for Advanced Telecommunications and Technology in Atlanta.

Research and Development Initiatives

To meet the mission of providing information on toxicology, environmental health, and targeted biomedical topics to the world, SIS has been developing new ways of presenting the world of hazardous chemicals in our environment to a wider audience. Projects include the following projects:

Household Products Ingredients Database: a Web resource for consumers that links brand name household products with their ingredient chemicals and potential adverse health effects. This pilot database will be ready for beta testing in early FY2003.

ToxTown: (http://toxtown.nlm.nih.gov/): a pilot project that explores how best to provide environmental health information to a general audience. ToxTown offers a graphic view of a typical town and points out environmental hazards that may be in that town. Users can click on a town location, like the school, and see a cutaway view of that building. Toxic chemicals that might be found in the school are listed, along with links to selected Internet resources about school environments. ToxTown will become available to the public from the SIS Web site in October 2002. Plans for an urban view and a rural setting are under way.

TOXMAP: a prototype system that uses maps of the United States to help users visually view data about chemicals released into the environment. It integrates data from the EPA's Toxic Release Inventory (TRI) with information about health effects, research citations, etc found in TOXNET databases. Users can create nationwide or local area maps that show where chemicals are released into the air, water, and ground. TOXMAP also integrates data from other sources, such as demographic data from Census Bureau. TOXMAP provides region-specific links to chemical and bibliographic information. A beta test version is scheduled for release in the first quarter of FY2003

HSDB-in-the-Palm: The objective of this new initiative is to provide critical chemical information quickly and conveniently on a Personal Digital Assistant (PDA) for use by emergency responders (first 24 hours in hot-zone). The application is being developed in partnership with the Agency for Toxic Substances and Disease Registry's.

ToxPortal: a virtual meta-search tool for simultaneous searching of target information systems, displaying search results from targeted systems, and harvesting related concepts. The tool can be configured to define a set of target information/search tools, including SIS databases and searchable resources on the Web. Testing of the prototype is under way and a beta version will be ready for public release in FY2003.

Chemical Spellchecker: Spelling errors in user queries, especially in chemical name searches, and the lack of semantic query assistance are known shortcomings of our present retrieval system. A prototype spellchecker was developed in FY2002, and it incorporates a general dictionary, medical dictionary, and chemical dictionary, as well as a UMLS enhanced lexicon and English language grammar parser. The new spellchecker will be integrated into the TOXNET search engine in FY2003.

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 is a research and development division of the National Library of Medicine. The Center conducts and supports research, develops research tools and systems, and provides training opportunities to individuals at various stages of their careers. The Center has been in existence since 1968, when it was established by a joint resolution of the United States Congress, with the mandate to conduct research supporting the mission of the NLM. The Center's research programs are reviewed biannually by a Board of Scientific Counselors, an external advisory group of researchers from the informatics community (see Appendix 3 for roster). The most current information about Lister Hill Center research activities can be found at (http://lhncbc.nlm.nih.gov/).

The Center's research staff are drawn from a variety of disciplines, including medicine, computer science, library and information science, linguistics, engineering, and education. Research projects are generally conducted by teams of individuals of varying backgrounds and often involve collaboration with other divisions of the NLM, other institutes at NIH, and other academic partners. Center staff publish in the medical informatics, computer and information science, and engineering communities. The Center is often visited by researchers from academic centers around the world. Our ongoing lecture series features presentations from many invited outside speakers.

Lister Hill Center research activities fall into several broad categories. Our training program has grown significantly in the last few years and has brought many talented individuals to the Center to learn from and collaborate with our research staff. Language and knowledge processing research involves basic research in medical language processing and medical knowledge representation. Image processing research involves the development of algorithms and methods to effectively process biomedical images of all types. We have developed and continue to support a number of information systems, all of which are informed by our basic research activities. In addition, Lister Hill Center staff are involved in a number of activities that define and support the research infrastructure for next generation information systems.

The Lister Hill Center is organized into five components, though many research activities involve collaboration across organizational units. The work of each component is described below. An organization chart, with the names of Branch and Office Chiefs, is on the inside back cover of this report.

Organization

Communications Engineering Branch

The Communications Engineering Branch is engaged in applied research and development in image engineering and communications engineering motivated by NLM's mission-critical tasks such as document delivery, archiving, automated production 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. In addition to applied research, the Branch also develops and maintains operational systems for production of bibliographic records for NLM's flagship database, MEDLINE. Research areas include: content-based image indexing and retrieval of biomedical images, document image analysis and understanding, image compression, image enhancement, image feature identification and extraction, image segmentation, image retrieval by "query by image content," 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 archives of large numbers of digitized spine x-rays and bit-mapped document images that are used for intramural and outside research purposes. The most current information about the Communications Engineering Branch can be found at (http://lhncbc.nlm.nih.gov/ceb/).

Cognitive Science Branch

The Cognitive Science Branch conducts research and development in computer and information technologies. Important research areas involve the investigation of a variety of techniques, including linguistic, statistical, and knowledge-based methods for improving access to biomedical information. Branch members actively participate in the Unified Medical Language System project and collaborate with other NLM research staff in the Indexing Initiative project the goal of which is to develop automated and semi-automated techniques for indexing the biomedical literature. The Branch also 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. ClinicalTrials.gov, developed by the Branch, is an excellent testbed for conducting consumer health informatics research. The Branch is currently developing a system designed to provide information about genes and diseases to the public. The

most current information about the Cognitive Science Branch can be found at (http://lhncbc.nlm.nih.gov/cgsb/).

Computer Science Branch

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. Computer Science Branch research staff include the teams that developed NLM's Gateway, Internet Grateful Med and HSTAT programs and the team that annually produces the Unified Medical Language System Metathesaurus. Branch staff coordinate the NIH Clinical Elective in Medical Informatics for third and fourth year medical students. The most current information about the Computer Science Branch can be found at (http://lhncbc.nlm.nih.gov/csb/).

Audiovisual Program Development Branch

The Audiovisual Program Development Branch supports the Lister Hill Center's research, development, and demonstration projects with high quality video, audio, imaging, and graphics materials. From initial project concept through project implementation and final evaluation, a variety of forms and formats of visuals are developed, and staff activities include content creation, editing, enhancement, transfer and display. Included in this effort is the production of a series of video modules documenting the progress of Lister Hill Center research projects. These informational and educational video reviews have been released in a variety of media, including Web delivery. Consultation and materials development are also provided by the branch for the NLM's 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 Web design, the requirement for graphics, video, and audio materials has increased in quantity and diversified in format. Included within the Branch is the Office of the Public Health Service Historian. This Office 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. The most current information about the Audiovisual Program Development Branch can be found at (http://lhncbc.nlm.nih.gov/apdb/).

Office of High Performance Computing and Communications

The Office of High Performance Computing and Communications serves as the focal point for the NLM's High Performance Computing and Communications (HPCC) activities. It coordinates NLM's HPCC planning, research and development activities with Federal, industrial, academic, and commercial organizations, and it collaborates with Lister Hill Center research branches and NLM Divisions in the development, operation, evaluation and demonstration of HPCC research programs and projects. In addition, it plans, coordinates, and administers the interagency HPCC research and development program. Office staff serve as NLM's liaison to scientific organizations at all levels of national, state and international government on planning and implementing research in HPCC. The major research activities of the office center on the Visible Human Project, NLM's Next Generation Internet Program, including telemedicine, the HPCC Collaboratory, and the 3D informatics research program. The most current information about the Office of High Performance Computing and Communications can be found at (http://lhncbc.nlm.nih.gov/ohpcc/).

Training Opportunities at the Lister Hill Center

The Lister Hill Center provides training and mentorship for individuals at various stages in their careers. Fellowship programs may be as short as eight weeks or as long as one year, with possible renewal for a second year. Each fellow is matched with a mentor from the research staff who works closely with the fellow throughout the fellowship program. In all cases, fellows define a research project early in their stay and then give a progress report. At the end of the fellowship period, fellows prepare a final, often publishable, paper and make a formal presentation, which is open to all interested members of the NLM and NIH community.

This past year, we provided training to 46 participants from 16 states and 9 countries. The participants included 9 undergraduate students, 15 graduate or medical students, 15 postdoctoral or post-MD fellows, and 7 visiting faculty scholars. Participants worked on projects in the areas of biomedical knowledge discovery, consumer health systems, history of medicine, image database research, information retrieval research, just-in-time systems, knowledge based research, natural language processing, ontology research, palm technology, semantic Web research, text mining, distance education, and visualization.

We again offered the Clinical Elective in Medical Informatics for third and fourth year medical students in

March and April, and we 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.

In the summer of 2001 we initiated the NLM Rotation Program. This program provides an opportunity for trainees in NLM supported medical informatics training programs to spend eight weeks at the Lister Hill Center learning about our programs and collaborating with our research scientists. Trainees from any NLM sponsored training are eligible for the program. The rotation includes a series of lectures and the opportunity for trainees to work closely with established scientists conducting research at the Center. Trainees who participated in the summers of 2001 and 2002 were members of informatics programs at 7 universities.

Additional information about our training and visiting faculty programs is available at our Web site (http://lhncbc.nlm.nih.gov/) under "Training Opportunities." Interested individuals will find descriptions of each of the training programs including specific application procedures.

Language and Knowledge Processing

Natural Language Processing Research: The Natural Language Systems research team investigates the contributions that natural language processing techniques can make to the task of mediating between the language of users and the language of online biomedical information resources. The successful integration of these techniques with other information retrieval strategies has the potential of contributing to the resolution of some of the most difficult problems underlying biomedical information management.

The focus of our natural language processing work is the development of **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. All components of the SPECIALIST system rely heavily on the linguistic and domain knowledge in the Unified Medical Language System knowledge sources.

Lexical Systems Project: The Lexical Systems project builds and maintains the SPECIALIST lexicon, a large syntactic lexicon of medical and general English that is released annually with the UMLS Knowledge Sources. New lexical items are continually added using a lexicon-building tool maintained by the lexical systems research team. The lexicon currently contains over 180,000 lexical items. Lexical access tools, including LVG, wordind, and norm, are also distributed with the UMLS. Since its initial release with the 2002 Metathesaurus release, the new Java version of the lexical tools has been improved and is significantly faster.

The SPECIALIST 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. The most effective of these have been incorporated into the lexical access tools and are being used by the ClinicalTrials.gov project.

Several modules developed by the Lexical Systems group have been completed and are now available independently as tools for a variety of natural language processing projects. These include a tokenizer, a lexical look-up utility, and a noun phrase extractor.

Semantic Knowledge Representation: Innovative methods for providing more effective access to biomedical information depend 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 NLM resources, including the UMLS knowledge sources and the natural language processing tools provided by the SPECIALIST system. Two programs in particular, MetaMap and SemRep, are being evaluated, 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 UMLS Semantic Network to determine the relationship asserted between those concepts.

The **MetaMap Technology Transfer** program (MMTx) is an exportable, Java-based version of MetaMap that runs under Windows or Unix/Linux and is provided as a resource to the bioinformatics community. A recent release of the MMTx package allows users to exploit the UMLS MetamorphoSys program to exclude or reorder the Metathesaurus vocabularies that MMTx refers to; users can also create MMTx data files independently of the UMLS. MMTx source code is included in the release, and an error reporting and tracking system ensures that problems reported by users are effectively addressed.

Word-sense ambiguity in language constitutes a major impediment to accurate management of biomedical text with automatic methods. The semantic knowledge representation project recently implemented a general framework for research in word-sense disambiguation. The framework depends on UMLS Metathesaurus concepts provided by MetaMap. The implementation is written in Java and includes modules that accommodate multiple disambiguation methods as well as an "arbitrator" for managing output from these methods.

Project resources are being applied in a variety of research initiatives aimed at identifying specific biomedical information in MEDLINE citations, including semantic predications asserting a treatment relationship

between drugs and diseases. Several projects focus on molecular biology. One seeks to identify genes, gene products, and gene functions in abstracts and compares this information to that found in the Gene Ontology. Another supports comparison of protein function by identifying protein-protein interactions in text. Finally, a recent project uses semantic information to support text-based knowledge discovery systems in molecular biology, while another uses such information to help manage the online research literature on the genetic basis of disease.

Indexing Initiative: The Indexing Initiative project is pursuing concept-based indexing methods that go beyond 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 index terms.

Substantial progress has been made in applying the MTI system to both semi-automated and fully automated indexing environments at the NLM. We conducted experiments to evaluate the effectiveness of MTI terms for NLM indexers, and the MTI recommendations are now available to all indexers. In addition, results of the MTI system have recently been assigned as keywords for three collections of meetings abstracts that will not be indexed by humans: AIDS/HIV, health sciences research, and space life sciences. These collections have been made accessible via the NLM Gateway.

Research into the system's indexing methods continues. In particular, a 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.

The Journal Descriptor (JD) project is investigating a novel approach to fully automated indexing based on NLM's practice of maintaining a subject index to journal titles using a set of 127 MeSH terms corresponding to biomedical specialties. The system associates JDs with words in titles and abstracts in a training set of about 435,000 MEDLINE records. Each record "inherits" the JDs from the journal title in the record. Each word in the training set can then be described by a list of JDs ranked according to the number of co-occurrences between the word and the JDs. We index a document based on averaging the word-JD co-occurrences for each word in the document that is also in the training set, ranking the JDs in decreasing order of these averages. The project has extended JD indexing to Semantic Type (ST) Indexing. The set of UMLS Metathesaurus concepts assigned to an ST can be regarded as a document, and therefore undergo JD indexing. We then measure the similarity between the JD indexing of a document to be indexed and the JD indexing of each of the ST documents, ranking the STs in decreasing order of similarity, resulting in a ranked list of ST indexing terms for this document. Preliminary experiments indicate that ST indexing shows promise when applied to the word sense disambiguation problem in automated indexing using MetaMap. If the senses of a word are expressed by STs, we can provide ST indexing of the context surrounding the word (phrase, sentence, abstract) in the expectation that in the ST indexing of the context, the correct ST for the word will rank higher than the other candidate STs for the word. We also are experimenting with using JD/ST indexing in identifying gene symbols (short words used for naming genes but which may have other meanings in the literature).

Unified Medical Language System: We regularly distribute a set of Unified Medical Language System (UMLS) knowledge sources to the research community. These include the Metathesaurus, Semantic Network and SPECIALIST lexicon. 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 UMLS Metathesaurus group continued its two main tasks, producing increasingly comprehensive releases of the Metathesaurus with new and updated vocabulary sources, and developing and deploying new software systems for work on unified concept-oriented terminologies.

For a decade, the **Metathesaurus** had been released annually early in the year. Beginning this year the Metathesaurus is being released quarterly. The size of the Metathesaurus has increased again and the quality has improved with the identification of additional missed synonymy and the addition of more diverse vocabularies. Vocabularies proposed as standards by the Department of HHS in the rulemaking accompanying the Health Insurance Portability and Accountability Act (HIPAA) continue to be added and maintained in each release. The most recent release of the Metathesaurus contains 2.1 million names for 871,000 biomedical concepts in approximately 60 families of vocabularies or thesauri.

Progress has also been made in research and development. Some has resulted in proposals for changes in 2003 that have been circulated to the UMLS user community. One such proposal suggests revised source naming, designed to minimize release file change when content is unchanged and to allow simple update models. Other research in progress will require longer term development; examples are alternative user views of the Metathesaurus and more sophisticated release data formats

that allow complete and transparent representation of full source information.

A collaboration with researchers from the University of Amsterdam is resulting in the development of an interactive editing and collaboration interface for the International Classification of Primary Care (ICPC) medical vocabulary, which has been incorporated into the UMLS Metathesaurus. The ICPC contains concepts in 20 different languages including Hebrew, Japanese, Russian and Greek, their character sets represented in Unicode. The project has continued its related work on a platform-independent Web-based system using the open source tools Apache/PHP/MySQL. The Web-based ICPC system allows multilingual display and editing on clients that have no Unicode capability by means of a Java applet and server-side Unicode manipulation.

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. Finally, the data in the SPECIALIST lexicon are also made available, providing the user with the syntactic and morphologic information about each lexical item it contains.

The most recent release of the Knowledge Source Server incorporates several features designed to enhance performance by allowing faster access to UMLS data, providing flexibility through a rich API set, and facilitating scalability in handling ever-increasing user loads and constituent vocabularies. The redesigned architecture includes a Web server implemented as a collection of Java servlets that provide quick and easy access to UMLS data. The server software connects through the Internet to a backend Remote Method Invocation (RMI) server, which processes all requests for data by first accessing a relational database to obtain relevant information and then forwarding the data through the Internet to the requestor. Open source software from Apache was used for the development of all aspects of the system.

In addition to enhancements to the user interface, XML has been incorporated into the design of the Knowledge Source Server to provide flexibility in delivering data to users. There is an object model for Metathesaurus data that allows users to access XML documents produced by the Knowledge Source Server and to manipulate the data in them in an object-oriented fashion within their programs, thereby providing a mechanism for representing concepts and related data consistently among developers.

A new Application Programming Interface (API) has been written in Java to provide platform independence. With the addition of an XML-based API, both Java and non-Java programs can now access the UMLS through a standard TCP socket. Some 40 API methods have been defined, allowing access to all details of the Metathesaurus. Files associated with the API include documentation for all interface and object model classes, a set of example Java programs for issuing API calls, some sample XML documents that may be used as input to the KSS socket interface, and sample XML output files for each of the API methods.

The **Terminology Server** project provides tools to manage diverse medical vocabularies for diverse purposes. Over the past year, the project moved from the conceptual phase to implementation, and significant progress has been made in several areas. An important function of the Terminology Server is to support the customization of terminologies from the UMLS and other sources to satisfy individual user needs. We are developing filters to help users select subsets of medical terms. The first filter identifies UMLS term variants suitable for natural language processing. Another task is to develop ways to handle UMLS data retrieval and maintenance. A related area of ongoing research is developing models for handling periodic UMLS updates while lessening disruptive effects on Terminology Server clients. In addition, effort is being focused on developing tools to create and edit "local," non-UMLS terminologies. Such tools will allow users to define terminologies relevant to their own domain and to tailor the UMLS data further for specialized needs. Software for specifying and maintaining local terminology using XML technology is under development.

Medical Ontology Research: 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 is used as the primary source of medical knowledge, OpenGALEN, CYC, and WordNet are being explored as well.

During the past year, research focused on two subdomains of biomedicine: anatomy and molecular biology. In one project, the representation of anatomical concepts in two ontologies, the Foundational Model of Anatomy and GALEN, were compared. In another project, we contributed to the integration of the Gene Ontology in the UMLS by studying the properties of this ontology. Biological knowledge is evolving rapidly and ontologies developed for molecular biology must be integrated with those developed for clinical medicine. In these two projects, methods were developed for aligning several knowledge sources. Experience gained from developing

techniques for knowledge visualization and navigation in the UMLS was reapplied to the Gene Ontology. Finally, we also developed methods for assessing the consistency of biomedical terminologies and extending their coverage.

Current research is focused in several areas. Ontologies for molecular biology will be used as background knowledge for information extraction and knowledge discovery from the biomedical literature (e.g., gene indexing). In a subsequent phase, the anatomy project will study how alignment strategies can take advantage of phenomena such as reification.

Image Processing

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 test bed and model for the construction of image libraries that can be accessed through networks. The Visible Human data sets are available through a free license agreement with the NLM. They are 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, legal and industrial uses by over 1700 licensees in 45 countries. We continue to maintain two databases to record information about Visible Human Project use. The first, to log information about the Visible Human Project license holders and record their plans for using the images, and the second, to record information about the products the licensees are developing.

With research support from the NLM, the University of Colorado Health Science Center, Center for Human Simulation has developed a first release Web site version of a head and neck atlas titled "Functional Anatomy of the Visible Human: Version 1.0 The Head and Neck." The atlas is designed in educational modules covering the topics of mastication, deglutition, phonation, facial expression, extraocular motion, and hearing. QuickTime movies have been produced using live human subjects portraying the function of the regional anatomy described from a surface anatomy perspective. Tools include basic anatomic structure identification, a model builder, orthogonal plane browser, and links to the PubMed Web site for automatic key word searches of the literature.

With research support from the NLM, two groups are investigating advanced anatomical methods using the Visible Human data. Brigham and Women's Hospital is investigating the problem of soft tissue expansion due to the use of frozen tissue required for the cryosectioning process used by the University of Colorado to create the original Visible Human datasets. The problem appears to be solved through the development of a completely different tissue preparation method. First the teeth are demineralized in order to achieve improved sectioning. In the original datasets these small objects became brittle and broke off. The hardware used was modified to allow for MRI registration with fiducial screws manufactured from an MRI compatible aluminum alloy. Artery filling (red) and vein filling (blue) was demonstrated to sub-millimeter level. Preliminary results indicate that a new, complete data set at a resolution of 0.1 mm (100 microns) in each of the three dimensions with all artifact problems successfully eliminated can be created. The investigators have already demonstrated an increased voxel resolution in the head from the Visible Human female's 0.33 mm to a resolution of 0.15 mm. Each new transection was cut at a section thickness of 142 microns on an ultracryomicrotome. This allows the collection of a slice of a complete transection, in contrast to the milling method used in the original Visible Human technique. Single intact transections have been histologically stained to differentiate neurovascular structures from adjacent connective tissues.

The Colorado investigators are examining techniques to improve their original milling based method. Tissue differentiation through multi-spectral imaging to enhance automated segmentation is being attempted. Ultraviolet illumination and visible wavelength fluorescence appears to be very promising. Spectral imagery recordings were taken following excitation in the ultraviolet region of the spectrum. What are interpreted as distal peripheral nerves in muscle tissue were seen for the first time in 100 micron sections identified by their intrinsic fluorescence. Recordings were made of the reflectance patterns as a narrow aperture ultraviolet scanner captured the absorbance characteristics of the anatomic structures. Spectral profiles of the basic tissue types were obtained. Surface freezing between slices has been successfully accomplished and automated with manual intervention every hour. Continuous cutting has been achieved for a time of 36 hours. This supports the concept of continuous cutting 24 hours per day from head to foot of an entire human. This process will reduce banding present in the Visible Human Male data and stabilize tissues with a continuous freeze.

Another Visible Human Project inspired initiative, the **Insight Toolkit**, began beta testing last year. The toolkit makes available a variety of open source image processing algorithms for computing segmentation and registration on a variety of hardware platforms. This work is being conducted by a consortium of university and commercial entities.

Building on the earlier **AnatLine** system, the object-oriented database of Visible Human images indexed for the male thorax region, Lister Hill Center researchers created AnatQuest with the goal of providing widespread access to the Visible Human images. AnatQuest offers users thumbnails of the cross-section, sagittal and coronal images of the Visible Male, from which detailed (full-resolution) views are accessed. Low bandwidth connections are accommodated by a combination of adjustable viewing areas and image compression done on the fly as images are requested. Users may zoom and

navigate through the images. Since its release in June 2002, AnatQuest has averaged about 60,000 hits per month, about five times the number of hits for AnatLine.

In addition to its main purpose, AnatQuest serves as an access point for AnatLine which allows access through anatomic terms to high resolution cross-sectional images and segment masks (useful for rendering anatomic objects). The tools needed to use AnatLine are also available: VHParser and VHDisplay. The first is for unpacking the data files into their individual components. The second is for displaying both cross-sectional and rendered images. Also, VHDisplay is augmented to voice names of anatomic structures as the images are displayed. Other resources accessible through AnatQuest are: 195 surface-rendered objects created at the Lister Hill Center as well as from outside sources (e.g., VoxelMan); and the FTP server for bulk transfer of high resolution image files. In addition to the Web-mediated version of AnatQuest, we developed a 'kiosk' version for the Dream Anatomies exhibit at NLM as a Java application suitable for onsite patrons with operation through a touch screen monitor.

Phase II of the high resolution scanning project, which involves scanning all Visible Female 70mm images, continued during this past year. Some technical issues involving resolution and color balance have been resolved, and a thorough quality review process of the digital scans and their derivative files is ongoing.

WebMIRS: The Web-based Medical Information Retrieval System (WebMIRS) is a Java application that allows remote users to access data from two surveys conducted by the National Center for Health Statistics. These are the National Heath and Nutrition Examination Surveys II and III (NHANES II and III), carried out in 1976–1980 and 1988–1994, respectively. WebMIRS has been reported in past years.

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, the 17,000 x-ray images collected in NHANES II may also be accessed with WebMIRS and displayed in low-resolution form. WebMIRS allows a user to control a graphical user interface to construct a query for the NHANES II or NHANES III data. A sample query might be equivalent to the English statements: "Find records for all individuals who reported chronic back pain. Return their age, sex, race, age when the pain began, and longest duration of pain. Also, return the record data required for statistical analysis and display their x-ray images."

WebMIRS enhancements done this year include packaging and deployment of WebMIRS as a Java application using Java WebStart technology, streamlining the capability to request image display, and redesigning Web pages for WebMIRS. In addition to the program enhancements, a new MySQL database was created to maintain information about WebMIRS usage, and an expert in Graphical User Interface usability was consulted to analyze the WebMIRS interface. Also, collaborative work was initiated with Texas Tech University to develop a Wavelet compression capability to allow delivery of the WebMIRS images in compressed form. A software interface to support decompression and display of these images was designed.

The Digital Atlas of the Spine is a dataset of cervical spine 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 images collected in the NHANES II survey.

The Atlas may be accessed either as a Java applet or downloaded as a Java application. In addition, we provide a version of the Java application on CD. The Java application version allows the user to add images (either grayscale or color) in a special "My Images" section, and to annotate and title those images for later use. Version 2.1 of the Atlas is now being distributed. Though started and led by research engineers at the Lister Hill Center, this project has the following collaborators: the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), the NIH Clinical Center, and rheumatology experts from the University of Miami, the University of California at San Francisco, and Johns Hopkins University.

Content-Based Image Retrieval: The Content-Based Image Retrieval project develops methods for effective extraction of biomedical information from digital images of the spine. This work has implications both for indexing of image data and for search of those data. For example, for the NHANES II images, the only indexing data available is the collateral (alphanumeric) data collected in the questionnaires and examinations; no indexing information derived directly from the images is available, and the high cost of employing radiological experts to compile such data by physical viewing and interpreting each image makes it unlikely that such information will ever be acquired by purely manual means. These circumstances could be reversed if reliable, biomedically validated software could produce image interpretations automatically. Even in the more likely case that only semi-automated methods should prove feasible, the reduction in labor costs could be sufficient to allow the creation of databases of significant biomedical information where this is not currently economically feasible. This is the implication of research into computer-assisted image indexing.

Computer-assisted image searching is a potential enabler of enhanced information extraction from a database that has already been indexed. The most popularized form of this type of search is query by example or a variant, query by sketch. In query by example, the user inputs an image, perhaps by selecting from a set of choices provided by the system, or by providing a completely new image, and queries the

database by asking, in effect, "Find records with images like this one," usually with respect to one or more characteristics of the example image, such as shape, histogram, or texture. In query by sketch, the input image is replaced by a sketch by the user, using drawing tools provided by the system. In either case, the system analyzes the input into component features, then searches the images in the database for those with similar features. Results are usually returned as a similarity ranking.

We implemented an initial prototype Content-Based Image Retrieval (CBIR) system for the retrieval of images based on simple vertebral shape models. This program allows the user to specify up to nine control points and the geometric configuration of these points to define an approximate vertebral shape to search for. The prototype database contains 100 cervical and lumbar images, and will rotate and scale each vertebra in each image to identify the best match to the input shape. Alternatively, the user may specify an example vertebra, and the program will search for the best shape match to the example.

The second CBIR prototype was created with significantly enhanced capabilities, including an indexing function with capability to do active contour segmentation, to create detailed representations of vertebrae boundaries, and to convert these boundaries into multiple shape representations (global shape descriptors, invariant moments, polygon turn functions, and Fourier descriptors). In addition, we implemented a retrieval function that supports retrieval of shapes by any of the above shape representations. The database created includes NHANES text data, and supports query by sketch, image example, and text, in addition to hybrid text and image-based queries. Current project work is directed toward continued completion of segmentation functions for indexing, analysis of effectiveness of the various shape methods implemented for spine x-rays with significant osteoarthritis features, implementation of spatial data trees for feature vector organization, and creation of a database of segmented vertebrae of significant size and segmentation accuracy, to serve as testbed data for ongoing work.

Engineering Laboratories: The Document Imaging Laboratory supports 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, optical character recognition (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 a variety of research.

The **Image Processing Laboratory** is equipped with a variety of high end servers, workstations and storage devices connected by 100 Mb/s Ethernet. The 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, the laboratory also has a variety of image content. Most of the machines housed in the laboratory 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. ATM switches connect the Ethernet and FDDI networks to other local area networks throughout the Lister Hill Center building, to the Internet, and to experimental ATM, in addition to Abilene, the infrastructure for the Next Generation Internet and Internet-2 initiatives.

The **Document Image Analysis Test Facility**, designed, developed and maintained by Lister Hill Center staff, is an off-campus facility housing high-end Pentium workstations and servers that constitute the MARS production system. 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.

Multimedia Research and Development: Our multimedia R&D efforts concentrate on 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 by the staff of new methods and processes, the facilities and hardware infrastructure must reflect state-of-the-art standards in a very rapidly changing field. High definition video is a technology area being developed that represents the future for improved electronic image quality. Multimedia systems, scientific visualization and networked media are being pursued for the performance, educational, and economic advantages that they offer. Three dimensional computer graphics, animation techniques, and photorealistic rendering methods have changed the tools and products of the graphic artists in the Center. Digital video and image compression techniques are central to projects requiring storage of large images and rapid transmission.

The "Breath of Life Virtual Tour" DVD was presented to the Joint Commission on Sports Medicine and Science in March 2002 in St. Louis in a session titled "DVD Delivery of Medical and Health Information." The NLM distributed 1500 copies of the DVD through the National Collegiate Athletic Association to athletic trainers

at their member institutions. The DVD was included in the 2002–2003 NCAA Sports Medicine Handbook, which was delivered to colleges throughout the country. The DVD and traveling exhibit was also on location at the Centers for Disease Control and Prevention in Atlanta in May and June. The "Breath of Life Virtual Tour" was the centerpiece of the opening activities on World Asthma Day, and continued as part of program events on the CDC campus. The traveling exhibit returned to NLM in June and was reinstalled in the NLM Visitors Center.

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 clips of neurologically based movement disorders. The video database of patients with a variety of clinically diagnosed movement disorders underwent updated editing and compression to capitalize on advanced digitization and compression technology. The patients featured in the series display varying degrees of Parkinson's and Huntington's disease, and cervical dystonia, or Torticollis. The video sequences represent standard diagnostic routines used clinically to measure degrees of illness, including gait, speech and hand and foot coordination. Forty-five video sequences have been edited and converted in format and resolution. This is the first step in a larger, ongoing effort to investigate the preparation of high quality, compressed video for distribution on the Web. The methodology and production of the next phase will be part of a collaboration with the Institute of Neurodegenerative Disorders and the Center for Advanced Instructional Media.

"Expanding the Medical Universe," the new video shown daily in the Visitors Center, was completed in January 2002. This is the first NLM video to be produced in the High Definition format. The new video also is presented in "surround sound" and is delivered on a DVD which has both non-captioned and captioned versions. The video is available with Chinese, Japanese, Korean, Russian, and Spanish soundtracks as well as English.

Information Systems

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, reported previously, uses innovative digital technology to make available the manuscript collections of prominent biomedical scientists. 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 nine prominent American biomolecular researchers. Added this year were Linus Pauling and Donald Fredrickson. We also recently made available the Reports of the Surgeon General (1964–2000).

Several research projects this year continued to enhance the effectiveness of the *Profiles in Science* site, including study of system designs such as the Open Archival Information System reference model. Architecture planned for the next generation Profiles in Science system served as the basis for information collection and dissemination for the NLM's upcoming History of American Women Physicians exhibit. Transition to a new underlying database for the metadata entry program commenced. We have conducted some studies of current approaches to digital preservation, including the development of an image migration framework for testing a variety of image conversion methods. We have also carried out some experiments in extracting embedded metadata and creating preservation metadata, especially for complex digital objects.

Document image analysis and understanding: Research into document image analysis and understanding combined with database design, GUI design for workstations, image processing, string pattern matching, lexical analysis, 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. From bitmapped images of the first page of the articles, this system is designed to automatically extract the article title, author names, affiliations and the abstract. Our current research centers on the identification of rules for algorithms for page segmentation, zone labeling, OCR error correction, affiliation ranking and other steps. Operators enter fields (other than the ones automatically extracted), as well as perform text verification before the records are made available to indexers.

While other fields are detected with a high degree of accuracy, correct detection of the affiliations field remains problematic. The reasons for this include: words in italics and very small font size, as well as the fact that indexing conventions require the inclusion of only the first author's affiliation and the removal of all others. In 2002, we pursued an approach that involved exploiting author

names and zip codes (which are usually recognized correctly) as cues to detecting the affiliations. To support this investigation, two tools were developed: a test data generator (TDG) and a test program. The TDG generates input data for the test program consisting of the OCR output and correct text from the MARS production database. The TDG was used to extract author and affiliation data from 1,228 articles processed by MARS during December 2001. Of the 1,228 author names tested, almost 50% were found in the existing author/affiliation database. The test program allows our research assistants to view, for a given author name or zip code, three items: the OCR affiliation, an affiliation already in the database and the correct (verified) affiliation. For each such triplet, the research assistant selects from the OCR affiliation or the database affiliation the one more like the correct affiliation, and this choice and the score are recorded. Initial testing shows an improvement in matching that increased the true positive rate (106 to 229) and decreased the false positive rate (45 to 35).

While the 5-engine OCR system used in MARS has shown a high degree of accuracy and reliability, it does not recognize Greek letters and biomedical symbols. To address this problem, we conducted research toward a prototype recognition system based on features calculated from the output of multiple OCR systems, string pattern matching, and a set of rules derived from an analysis of document content, journal specifications, and medical dictionaries. Our technique uses two passes of a document image page through OCR systems designed for different languages. Document content information and journal specifications are derived from an analysis of the page contents of each journal. The low-confidence words containing Greek characters from previous documents are analyzed and their features (contents, attributes, and frequencies of occurrence) are recorded for use in recognizing characters in subsequent documents. Preliminary evaluation conducted on a sample of medical journal page images shows that the system is capable of improving the recognition of Greek characters embedded within predominantly English language text: 89% of the Greek characters were correctly identified.

The **MARS** system relies on image analysis and lexical analysis algorithms to correctly extract bibliographic data from images. These algorithms are based on rules constructed from features extracted from the layout geometry and OCR output. To date, 3,058 journal titles have been tested, and 2,376 titles can be processed by MARS, i.e., there are suitable rules for these. Of these, bibliographic data is supplied directly by publishers for 853, leaving 1,523 titles as MARS-compliant. Since there are 4500+ titles indexed at NLM, this still leaves 1,500 titles for which no rules have been extracted. Our goal is to develop a method to rapidly extract features from which rules may be constructed for the zoning and labeling algorithms.

We believe that the publication of ground truth data from the large set of images and extracted data collected in MARS would be an important contribution to the field by facilitating the development of new document analysis methodologies. Accordingly, the effort to develop the mechanism to disseminate this ground truth data for research by the computer science and informatics communities is under way. The tasks accomplished are as follows: First, a program was developed to export the MARS production data into XML format, using Visual C++ and Xerces, part of the Apache XML Project (http://xml.apache.org). This export program allows an end user to use a wizard style approach to pick destination directories for the XML/TIFF files, the journal titles and the specific page images desired. The application then converts the MARS data to XML data. In addition, the export routines create line and word-level information that, in some cases, might not have been part of the MARS production data, but would be useful for future research. The second task was to select the initial ground truth data set. We have identified approximately 1,000 page images from different journals to accommodate a broad range of page layout types for this purpose. The third task was to design and develop a Web site to serve as a repository for the ground truth data and tools to analyze the data.

DocView: DocView is a system that facilitates the delivery of library documents directly to the patron via the Internet in multiple ways, but it is most commonly used by library patrons to receive scanned journal articles from libraries that use Ariel software for interlibrary loan services. While Ariel, a product of the Research Libraries Group, is used by libraries and document suppliers, there are few options for end users to directly receive them. Once documents in bitmapped image form are received, the user may use DocView to retain them in electronic form, view the images, organize them into "folder" 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.

DocMorph provides additional functionality for DocView users. DocMorph enables online users to convert files from one format to another for easier exchange or delivery. The system allows the conversion of more than 50 different file formats to PDF, for instance, to enable multi-platform delivery of documents. Also, by combining OCR with speech synthesis, DocMorph enables the visually impaired to use library information. It has been used by librarians for the blind and physically handicapped to convert documents to synthetic speech recorded onto audio tapes for blind patrons. DocMorph handles all files types from the Ariel system and many others.

Turning The Pages Information System: In 2001–2002, NLM and the British Library collaborated in the production of two virtual books, Blackwell's Herbal and Vesalius's Anatomy in photorealistic "Turning The Pages" form. The pages of these were scanned, and these high quality color images were manually processed by Adobe Photoshop, animated by Macromedia Director software, and displayed on a touch screen monitor. The library patron may "touch and flip through" each of these books. Using the TTP books as a starting point, we began an investigation of different ways to extend them beyond their application as beautiful museum pieces to information systems (TTP+). We followed two different approaches, the "discovery" and the "storyline" models, to implement the extensions. The TTP+ version of Blackwell's Herbal (discovery model) retains the photorealism of the original TTP, while allowing a patron to "travel" to live sites on the Internet. For example, from highlighted text on the St. John's Wort page, one can go to a PubMed search and get citations, or go to ClinicalTrials.gov and get information on clinical trials of this drug. Also, links are available for plant descriptions and photographs on sites of the USDA, Forest Service and U.S. Herbaria, among others.

The TTP+ version of Vesalius followed the storyline model. Here, while elements of the design strategy for the Blackwell TTP+ were employed, the page images and images from other sources (e.g., rendered Visible Human images, pictures of Italian cities, etc.) were interlinked to present the patron with several multimedia "stories," including "Man of Padua," and "Modes of portraying anatomy."

NLM Gateway: NLM offers an increasing number of Internet-based information resources, each with its own user interface. We have created the NLM Gateway to let users initiate searches in multiple retrieval systems from a single interface at one Web site. 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, released in October 2000, provides the capability, with one query, of simultaneously searching 11 document collections using a variety of retrieval methods on different systems.

Several document collections, notably OLDMEDLINE and the collections of abstracts have been expanded this year. Access to a new collection, the MEDLINEplus Medical Encyclopedia, was added in the Consumer Health Information category. Numerous interface enhancements have been added in response to user comments and informal usability surveys. As NLM moves increasingly toward XML-based systems, the ability of the Gateway infrastructure to accept new content in XML and to manipulate XML data has been substantially improved.

One of the most significant enhancements to the NLM Gateway is the result of a collaborative effort with the Indexing Initiative project. To enhance retrieval, all of the meeting abstracts in the Gateway's collections have been automatically indexed by the Indexing Initiative systems. Other enhancements in progress include the addition of search filters that will allow user-specified views of the NLM information from several data collections with an effect similar to the earlier searching of AIDSLINE, TOXLINE and SPACELINE. Initial access to additional data resources from NLM's Specialized Information Services division is complete and is in testing.

HSTAT System: The HSTAT system this year has served as a model for technology transfer of a system developed through the Lister Hill Center R&D process to production status in the Office of Computer and Communications Systems (OCCS), NLM's operations division. A transfer plan was developed and successfully executed by joint teams from LHNCBC and OCCS. Staff from NLM's National Information Center for Health Services Research were trained to manage the HSTAT data. OCCS acquired appropriate hardware and software to support the system. HSTAT was ported to the latest versions of compilers and software tools on the OCCS hardware by the development team, and software licensing and security issues were resolved.

Consumer Health Informatics Research: The Consumer Health Informatics research projects explore the needs, information seeking behavior, and cognitive strategies of health care consumers. The goal is to use medical informatics and information technologies to study ways to develop, organize, integrate, and deliver accessible health information to the members of the public at all levels of health literacy.

ClinicalTrials.gov provides members of the public with comprehensive information about clinical trials. The site not only simplifies access to research protocols but also directs visitors to appropriate background information, such as health topics at MEDLINEplus and the biomedical literature at PubMed. Currently, ClinicalTrials.gov has nearly 6,600 protocol records sponsored by the Federal government, the pharmaceutical industry, and nonprofit organizations in over 70,000 locations, mainly in the United States and Canada. The site hosts over 8,000 visitors daily.

This year, several new search features were introduced in ClinicalTrials.gov to help users find relevant studies more easily. For example, "Search Within Results" enables users to narrow their search results with additional criteria and "TryIt" automatically suggests alternative queries when no studies are found. A revised syntax for encoding complex queries as Web links allows users to launch searches with a single mouse click. In addition, ClinicalTrials.gov received over 400 protocols from pharmaceutical industry sponsors within six months following the release of the Food and Drug Administration's "Guidance for Industry: Information Program on Clinical Trials for Serious or Life-Threatening Diseases and Conditions." The Resources and What's New pages were redesigned on the site to better explain the role of clinical research in medical practice.

The Genetic Disease Home Reference: The Genetic Disease Home Reference is a new project that seeks to provide information about genes and diseases to members of the public. This resource will focus on diseases that are caused by single genes and, in turn, on the genes that cause diseases. As knowledge of genetics expands, the interrelationships between genes and diseases will continue to unfold. Our goal is to provide a bridge between the clinical questions of the public and health professionals and the richness of data emanating from the Human Genome Project. Other resources will delve more deeply into the clinical aspects of the diseases and the details of the genes. This system is meant to serve as a guide into those other resources. We have developed a research prototype, and we are currently further developing and testing the system.

Research Infrastructure and Support

Next Generation Internet: We are working to define and support 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 have collaborated with other NLM staff on the Multilateral Initiative on Malaria in Africa. We developed a statement of work for performance evaluation of the network linking malaria research sites in Africa, and the procurement was awarded in March 2002 to Infinite Global Infrastructures (IGI). Together with IGI personnel, we conducted a review of Redwing Satellite Solutions, the space segment and Internet access provider located near London. Following this visit, we gave a presentation on NLM's communications work and the performance measurement task at an NLM-sponsored in Kenya.

We 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. This involves using existing state owned fiber where available and using current right-of-ways the state possesses to add additional fiber in underserved regions such as the Eastern Shore, and Western and Southern Maryland. The plan provides equity of access to all regions of the state, and supports multiple segments of society. In the near future the state will conduct a 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 to help develop pilot projects in health care involving remote oncology treatment planning and remote intensive care support.

We continue to explore applications of smart card technology. As noted in previous reports, a smart card is a credit card sized plastic card with an embedded circuit chip. The card can be used both for authentication and for data storage. Recent applications often 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 cosponsored the Western Governors' Association Health Passport Project, one of the largest health-oriented smart card pilot programs in this country. Kiosks in public places allow clients to check and print information from the card. Kiosk functionality is now being expanded to include Internet access to consumer health information, including that in NLM's MEDLINEplus system. Phase II of the Health Passport Project is getting under way in the San Diego area. This project will incorporate biometric authentication on the card, digital certificates, and trusted third party systems to facilitate the safe transfer of private medical and demographic information over the Internet in encrypted form.

During 2001 NLM made the decision to eliminate phase three of the originally designed three phase NGI initiative. The third phase was to be a test of scalability of applications to a national scope on public networks featuring quality of service. Such a network is not available because of the rapid increase in available bandwidth and a decrease in the price of that bandwidth. In place of phase three, during this past year we began a new initiative in Scalable Information Infrastructure. The purpose of this initiative is to encourage development of health related applications of scalable, network aware, wireless, geographic information systems, and identification technologies in a networked environment. The initiative focuses on situations that require or greatly benefit from the application of these technologies in health care, medical decision-making, public health, large-scale health emergencies, health education, and biomedical, clinical and health services research. Projects must involve the use of testbed networks linking one or more of the following: hospitals, clinics, health practitioners' offices, patients' homes, health professional schools, medical libraries, universities, medical research centers and laboratories, or public health authorities.

Lister Hill Center staff continue to participate in the monthly meetings of the multi-agency Joint Telemedicine Working Group. We participated in several demonstrations for Congress and the Administration on state-of-the-art telemedicine and e-health projects and solutions. The congressional Steering Committee on Telehealth and Healthcare Informatics sponsored a demonstration and roundtable discussion as part of its 2002 ten-session educational series for Members of Congress and staff, federal agency officials, healthcare and technology organization representatives, and the public. The NLM/OHPCC display featured wireless PDA access to PubMed.

Collaboratory for High Performance Computing and Communication: The Collaboratory for High Performance Computing and Communication investigates innovative means for assisting health science institutions in their use of online distance learning technologies and explores advanced computer and network technologies for distance interactivity including wireless technology and virtual reality research. During this past year we conducted a four-session workshop in the Collaboratory, "Using the Internet and Online Resources in Health Care Practice." The NIH's Foundation for Advanced Education in the Sciences designated this educational activity for 12 hours of category-1 credit towards the AMA Physician's Award. The Collaboratory streaming video and multipoint videoconferencing servers had major equipment and software upgrades this past year. A new wavelet-based codec for videoconferencing was acquired and simple telephony technology was put in place to demonstrate voice over IP. Hardware and software systems were installed in the Collaboratory enabling Lister Hill Center staff to demonstrate applications developed under the NGI initiative. Examples include an imaging, annotation, and collaborative video workstation in embryology developed by George Mason University, Johns Hopkins and other collaborators; an imaging and haptic collaborative environment developed by Stanford University and its collaborators; and a televideo application and video enhanced patient record system developed by Indiana University enabling NLM to be part of the university's virtual private network. In addition, MPEG2 videoconferencing, wireless network technologies involving use of PDAs, and other internal applications were inaugurated.

We conducted a series of experiments combining wireless and streaming video technology offsite to do live webcasts from NLM that originated at remote sites. Video encoding software on a laptop roaming the scientific poster session at the annual meeting of the American Medical Informatics Association (AMIA) annual meeting sent the video stream through a wireless network installed onsite back to NLM for broadcast via the Collaboratory streaming server. We established a 802.11b wireless network for attendees at the AMIA NLM booth for the duration of the exhibits.

We used wavelet codecs to demonstrate wireless videoconferencing and application sharing between the Collaboratory and the Slice of Life Conference at the University of Toronto. The Collaboratory's videoconferencing and streaming technology enabled it to become a local viewing site for the Internet2 Virtual Member Meeting when the physical meeting was canceled as a result of September 11 last year. We also used videoconferencing in distance learning tutorials conducted at the University of Washington and the Radiological Society of North American annual meeting. Live webcasts of selected meetings of the Washington Area Computed Assisted Surgery special interest group meetings continued and have been archived for on demand viewing.

Collaboratory staff participated in the creation of the NIH Handheld Users Group. The Group's mission is to provide NIH users with information on handheld resources, services, and technologies, to offer a forum for the exchange of information and ideas, and to promote awareness of the benefits of handheld technologies in fulfilling the NIH mission. Because of the convergence of technologies and devices towards IP (Internet Protocol), the group is interested in all wireless devices that have voice and data capabilities.

System Security and Advanced Network Planning: This group's work during the year focused on computer security, the NLM network, the Next Generation Internet (Internet2 and NGI), and the upgrading of Lister Hill Center systems. Computer security concentrated on the refinement of access controls and the development of a security classification organization. A Secure Subnets working group last year developed a classification of NLM systems used to categorize different levels of required network access between each system and the Internet. The first phase of the Secure Subnets initiative has been implemented, with most of the desk-based systems placed on subnets that are not accessible from outside NLM. These systems can themselves access sites outside NLM but transmissions originating outside of NLM cannot access them. The effect of this grouping demonstrably has been to make these systems less vulnerable to external security attacks.

In response to the changing threat conditions, network access policies to subnets accessible from outside NLM were modified. The previous policy had been to permit all types of traffic except those specifically denied. The new policy is the opposite: to deny all types of traffic except those specifically permitted. The addition of redundant gigabit firewalls will help to secure LHC systems against both security intrusions and denial of service attacks.

Work on the network has continued with the development of a Gigabit backbone. New Extreme switches have been installed that can handle Gigabit connections to the desktop. These switches have been connected to two core Gigabit switches (Extreme Black Diamond) that provide a redundant connection between the local switches, the NGI networks, and the Internet. This provides fully redundant paths from NLM to the Internet and increases the Internet2 and NGI connection links from OC-12 to gigabit speed.

NLM's Next Generation Internet project continues to evolve. The NGI networks are being used for multimedia applications involving voice and video. The Abilene network supports full IP multicast. In that mode, NLM can receive and transmit multicast voice and video sessions.

A new network performance testing system can generate live traffic for analyzing and tuning the performance of the NLM network. A new very high capacity tape library system based on LTO tape

technology is now used for backups of the LHC computer systems. Most backup media can now be online at all times, available for restoration of older data and programs. Backup volumes are created in duplicate, with one set for offsite storage.

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. The Office was involved in the development of two exhibits during this past year: "Smallpox: A Great and Terrible Scourge" and an exhibit to commemorate the centennial of the Pan American Health Organization. The Office completed the reorganization of the historical reference files and the revision of the Office's Web site.

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 14 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 FY2002, approximately 3.3 million sequences were added to GenBank, and the base count rose from 13.5 billion in August 2001 to 22.6 billion in August 2002. The 15 million sequences in GenBank represent data from over 130,000 organisms. During the first quarter of FY2002, GenBank release 127.0 held the largest increase in the number of sequences and basepairs in the database's history.

Important sources of data for GenBank are direct sequence submissions from individual scientists and genome sequencing centers. NCBI produces GenBank from thousands of sequence records submitted directly from researchers and institutions 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.

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 nine years of processing direct submissions, NCBI has issued over 823,000 accession numbers, with approximately 26% of these assigned in FY2002. There are now over 709,000 direct submission accession numbers that are publicly available and approximately 71,000 accession numbers pending release. Sequence data submitted in advance of publication is maintained as confidential, if requested.

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 available 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, both nucleotide and protein, 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 useful external resources such as biological databases and sequencing centers. In addition to literature information, GenBank provides links to related information in other Entrez databases. 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.

In FY2002, NCBI began publicly releasing Third Party Annotation (TPA) database sequences. This database, created in conjunction with our international counterparts EMBL and DDBJ, supports third party annotation of sequence data already available in the public domain. Sequences in the TPA database are predicted or assembled from such sources as ESTs, genome data, and other unannotated sequences. Publication of the analysis in a peer-reviewed scientific journal is a requirement of this database. Another development this year is the acceptance of submissions from Whole Genome Shotgun (WGS) sequencing projects. Annotations are allowed in these assemblies and will be updated as sequencing progresses and new assemblies are computed.

Improvement of NCBI's sequence submission software continues to be a high priority. A new version of Sequin, NCBI's stand-alone submission tool, was released in FY2002. In this new version, improvements have been incorporated to facilitate sequence annotation by the GenBank indexing group including an improved graphical view, sequence update/alignment functions, and global tools for editing multiple entries simultaneously. Other changes and enhancements include added support for the Third Party Annotation (TPA) database; the ability to accept discontinuous alignments of phylogenetic data for display in Entrez; and improved annotation support for transgenic sequence submissions, environmental sequence sets, unusual molecule types, and complete genomes. BankIt, another sequence submission software tool, is now in its eighth year of use. Some of the improvements made to BankIt this year include the ability to identify sequences appropriate for the TPA database, options for including strain name for mouse, rat and influenza virus, and a more explicit example of features that can be added to a record.

A new GenBank submission tool, Sequin MacroSend, was released in July of this year. This tool is designed to upload large Sequin files to avoid problems that occur when large submission files are sent via email, such as message corruption or truncation of large files. This tool allows submitters to upload a Sequin file from their computer directly to the GenBank indexing staff where their submission is immediately given a temporary identification number.

GenBank has evolved to contain several types of sequence information, 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). As of August 2002 there were 12,455,889 public EST entries stored in dbEST.

Another rapidly increasing segment of GenBank is the Genome Survey Sequences (GSS) 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. As of August 2002 there were 3,654,997 public records stored in dbGSS.

The Sequence Tagged Site (STS) division of GenBank consists of short sequences that are operationally unique in the genome and used to generate mapping reagents. This division continues to experience growth and as of August 2002 there were 124,064 entries in dbSTS. The 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. UniSTS was recently added to the Entrez search system, allowing searching with Boolean operators and providing cross-referencing to other Entrez databases. Each marker report contains primer information, mapping data, and cross-references to other NCBI resources, such as Map Viewer and LocusLink.

Entrez Genomes contains records representing over 1,000 species including bacteria, archaea, and eukaryotes, over 700 viruses, complete microbial genomes, and a number of viroids, mitochondria, and broad host range plasmids. The genomes represent both completely sequenced organisms and those for which sequencing is in progress. Some of the 25 new complete genomes added to the database in FY2002 include: Anopheles gambiae, Salmonella typhimuruim LT2, Brucella melitensis, Streptomyces coelicolor strain A3(2), Xantomonas axonopodis, Streptococcus pyogenes MGAS8232, Schizosaccharomyces pombe. The Genomes group also installed map data into the Entrez Genomes Map Viewer for the following organisms: Danio rerio (zebrafish),

Saccharomyces cerevisiae (baker's yeast), Schizosaccharomyces pombe (fission yeast), Plasmodium falciparum (malaria), Rattus norvegicus (rat), Arabidopsis thaliana (thale cress), Avena sativa (oat), Hordeum vulgare (barley), Oryza sativa (rice), Triticum aestivum (wheat), Glycine max (soybean).

A new Web page devoted to Anopheles gambiae (malaria mosquito) was made available in FY2002. This resource provides access to both the original sequence data deposited in GenBank as well as the first version of the genome assembly. From this page, users can access information from the Map Viewer, BLAST and the taxonomy database. This pest is the primary vector responsible for an estimated 200 million clinical cases of human malaria each year. Further sequencing information will undoubtedly impact public policy and health issues related to the malaria parasite.

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. 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 released four builds of the assembled human genomic sequence in FY2002. With each new build, 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 characterizing, 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 (Reference Sequence) 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 annotated with many 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 one to view an organism's complete genome, as well as integrated maps for each chromosome.

Various computational approaches are used by NCBI investigators to accomplish the task of predicting novel genes. Alignment with short segments of expressed genes called Expressed Sequence Tags 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

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 researchers' use of this data.

NCBI's Web resource, "The Human Genome, A Guide to Online Information Resources," serves as a nexus for the collection and storage of diverse human 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 provides 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 a graphical display 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), the location of genes, gene expression, exons, STSs, FISH mapped clones, ESTs, GenomeScan models, and sequence variation. Maps from other sequencing centers are also available. Genes or markers of interest can be found by submitting a query against the whole genome, or by querying one chromosome at a time. Results are available in both graphical and tabular format.

In FY2002, NCBI continued to improve its Map Viewer. In addition to providing more organisms, the improvements 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.

Two new features were added to the Map Viewer in FY2002. The Evidence Viewer provides the user with a graphical display of the biological evidence supporting a particular gene model. It displays all RefSeq models, GenBank mRNAs, annotated known or potential transcripts, and ESTs that align to the genomic sequence region of interest. Another feature is the Model Maker. This tool allows users to view the evidence used to build a gene model on assembled genomic sequence, and to create a version of the model by selecting exons of interest. Model Maker is accessible from sequence maps that are analyzed at NCBI and displayed in the Map Viewer.

NCBI's Human-Mouse Homology Map is designed to allow navigation 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 Radiation Hybrid (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 importantly, how an understanding of the human genome will contribute to improving diagnosis and treatment of disease. This site was expanded in FY2002 to include a number of additional diseases and now contains descriptions for over 150 genetic diseases and provides links to databases and organizations that can supply additional information. This site was recently incorporated into the NCBI Books site, making it available for searching within the Entrez system. For each disease-causing gene there is 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 "Online Mendelian Inheritance in Man" catalog of human genes and genetic disorders. The database, produced at The Johns Hopkins School of Medicine, contains over 14,000 records and usage exceeds 8,000 users per day. OMIM is part of the Entrez retrieval system and provides links to related OMIM records as well as links to several other databases. This provides greater flexibility in field searching and increased relevance of retrieved information.

LocusLink 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 records, UniGene clusters, map information, and relevant Web resources. LocusLink has rapidly expanded over the past year from 88,560 records last year to 276,719 records this year. Other LocusLink features include annotation for human genes; gene ontology terms for human and other genomes; domain names from CDD-based analysis of RefSeq proteins; and links to other NCBI resources such as UniGene, HomoloGene and Human-Mouse Homology Maps. LocusLink also provides one of the windows into NCBI's annotation of the human genome, with direct links to the Map Viewer, graphical sequence viewer, evidence viewer and model maker. In FY2002, LocusLink's organism list was expanded to include HIV-1, rat, and zebrafish.

The Reference Sequence (RefSeq) database 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. Curated genomic annotations can be retrieved via LocusLink and the Map Viewer. In addition to man, mouse, and rat RefSeq was expanded in FY2002 to include Drosophila melanogaster (fruit fly), Danio rerio (zebrafish), and Saccharomyces cerevisiae (budding yeast). The database holds over 150,000 reference sequence records for these six organisms and over 120,000 corresponding RefSeq protein records. In FY2002, a new group of accession numbers beginning with NR_ was added to RefSeq and represents non-protein coding RNAs.

The most common forms of sequence variations are single nucleotide polymorphisms, or SNPs. SNP detection and discovery is expected to facilitate large-scale association genetic studies. To accommodate this high volume 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.

In FY2002, dbSNP was made available for search and retrieval in the Entrez retrieval system. This allows searching using the Entrez features as well as an extensive array of limits. Each SNP record also contains links to other Entrez databases, LocusLink, genomic sequence data, and external SNP resources such as the SNP Consortium. Several new builds of dbSNP were released throughout the year, improving accuracy of annotation on the human genome. In order to improve dissemination of information, a "dbsnp-announce" mailing list was created to report the release of new builds, announce new features, and report corrections or problems with past or present builds. In April, a data dictionary for the dbSNP schema was made available online to provide a better understanding of the database for those maintaining the database on their own system. dbSNP contains information

from 15 organisms. The SNP database consists of 5.4 million submissions and a non-redundant set of 3.6 million refSNP clusters.

The dbSNP sister database, dbHLA, is being developed to define molecular haplotypes for the common human tissue-typing alleles. 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. Haplotype sets and individual reports are available at this time via dbSNP.

From Human to Mouse: Model Organisms for Research

The public mouse sequencing effort made significant progress over the last year. The ultimate goals of the project include the construction of a robust physical map and a high quality, finished sequence of the mouse, since 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 deposited into GenBank.

NCBI expanded its mouse resources with the release of the whole genome shotgun assembly of the genome of Mus musculus strain C57BL/6J by the Mouse Genome Sequencing Consortium. The mouse genome resources page was extensively revised, making it easier to navigate the mouse genome BLAST pages, mouse Map Viewer, trace repository, and the Mouse/Human Homology viewer. Links are also available to information on sequencing progress, sequencing centers, strain resources, and a monthly newsletter designed for the mouse research community. Mouse data is being accumulated in both the RefSeq and LocusLink databases and investigators have assembled the data set in order to generate larger contigs. The mouse sequence reads are of immediate use for both human and mouse genetics and there are many examples of mouse genes that have been cloned using the available public information.

Literature Databases

PubMed is an innovative, Web-based literature retrieval system developed by NCBI to provide access to the MEDLINE database of citations and abstracts for journal articles in the biomedical sciences. It is the bibliographic component of the NCBI's Entrez retrieval system and provides links to full-text journal articles at Web sites of participating publishers, as well as to other related Web resources.

PubMed services have expanded in all aspects over the last year. Full-text journals that link to PubMed have increased from 2,285 in October 2001 to over 3,000 in October 2002. Approximately 40% of all PubMed citations from 1990–2002 now have links to full-text. Usage of PubMed by the scientific and lay communities has also grown considerably since its introduction in 1997, with up to 1.02 million searches and approximately 200,000 users per day.

LinkOut is a feature of Entrez designed to provide users with links from PubMed and other Entrez databases to a wide variety of relevant Web-accessible online resources, including full-text publications, biological databases, consumer health information, research tools, and more. As of October 2002, over 700 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. The LinkOut resources page has grown by providing information on help and tutorials, utilities, lists of providers and journals, DTD specifications, and contact information.

The LinkOut for Libraries program continues to provide biomedical libraries the ability to link patrons from a PubMed citation directly to the full-text of an article. LinkOut for Libraries has been expanded to include print holdings. Libraries that store their print holdings information in the NLM SERHOLD database can now display this information in LinkOut. As of October 2002 over 420 libraries were participating in this program.

System enhancements were made to PubMed throughout the year, including the release of a PubMed Text Version. This version was developed to assist physically challenged users who require special adaptive equipment to access the Web. It provides basic PubMed search and retrieval functionality. Its text-based page design and easy-to-use navigational controls help simplify the process of searching and viewing results and can be used on PDAs.

Additional system enhancements made to PubMed include the addition of history of medicine and bioethics-related journal citations unique to the former NLM HISTLINE and BIOETHICSLINE databases. Also, History of Medicine, Bioethics, and Space Life Sciences selections were added to the Limits Subset pull-down menu. A new search filter, Systematic Reviews, was added to the Clinical Queries screen. This feature retrieves citations for systematic reviews, meta-analyses, reviews of clinical trials, evidence-based medicine, consensus development conferences, guidelines, and citations to articles from journals specializing in clinical review studies.

The NCBI Bookshelf is a growing dataset of books available online via the Entrez Retrieval system. A search feature allows searching of all or individual books. At this time, there are more than a dozen reference books available, with an "NCBI Handbook" coming soon which

will provide an in-depth description of NCBI's major information resources. Some titles added to the Books database this year include, Molecular Biology of the Cell, Introduction to Genetic Analysis, Modern Genetic Analysis, Molecular Cell Biology and the chapter "Smallpox and Vaccinia" from the text entitled Vaccines.

In September NCBI released the new "Journals" Entrez database. This new database replaces the Journal Browser and provides additional search and display features. The Journals database is available from the Search pull-down menu and from the PubMed sidebar and provides links to the NLM catalog, Locatorplus.

PubMedCentral (PMC) was established as a digital archive of life sciences journal literature providing barrier-free access to the public. This repository is based on a natural integration with the existing PubMed biomedical literature database of abstracts. PMC currently provides free and unrestricted access to the full text of over 55 life sciences journals, with more forthcoming. Some of the new journals added to PMC this year include, Cell Biology Education, Eukaryotic Cell, and Journal of Biology.

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 an unknown sequence against the database of all known sequences to determine likely matches. Hundreds of major sequencing centers and research institutions 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. BLAST also provides users the option of retrieving results with a request ID anytime within 24 hours of searching.

Many enhancements were made to the BLAST suite of programs throughout FY2002. At this time, version 4.0 of the BLAST database is fully supported. A BLAST Program Selection Guide was created to help users identify which BLAST program is best suited for their needs. In March, Linkouts were made available from BLAST search results to the LocusLink and UniGene databases with links to additional databases under development. A new Discontiguous MegaBLAST program was also released. This version of MegaBLAST is designed specifically for comparison of diverged sequences, which have alignments with a low degree of identity. It can be used to search the Trace Archive to compare nucleotide sequence data against the raw data underlying all of the sequence generated by the various genome projects.

The genome-specific BLAST pages have been enhanced extensively. MegaBLAST options were added for all BLAST organism pages. BAC ends, HTGS (all phases), and ESTs were added to the Human organism page. Early in FY2002, the Mouse BLAST pages were revised to conform more closely to the format of the Human BLAST page with a single interface and pull-down menu for database selection. New BLAST pages were added for specific searching of the rat and zebrafish genomes.

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 200,000 BLAST searches are performed, with users submitting their requests through server/client programs and the World Wide Web. BLAST e-mail submissions were discontinued, including the e-mail retrieval of BLAST results from Web searches in June of this year. Most BLAST searches are now conducted through the BLAST Web page. The popularity of BLAST has resulted in regular expansion of computing capacity to accommodate the growing volume of users. Standalone BLAST software is distributed to allow users to run BLAST searches within their own institution.

A new resource called BLink (BLAST Links) displays the graphical output of pre-computed BLASTP results of all Entrez proteins against the protein non-redundant database. BLink integrates heterogeneous NCBI resources, offering a variety of display options that include the distribution of hits by taxonomic grouping, sorting by taxonomic proximity, the best hit to each organism, protein domains in the query sequence, and similar sequences that have known 3-D structure.

Other Specialized Databases and Tools

A new NCBI resource called ProtEST was introduced in FY2002. With this tool, the alignment of UniGene sequences with their translational products can be displayed. This tool uses BLASTX to compare UniGene mRNA and EST sequences with protein sequences from eight organisms, recording the best match for each case. Therefore, UniGene nucleotide sequences can have up to eight matches. The ProtEST Web site is accessible from the UniGene homepage and any UniGene cluster or sequence page.

Plant Genomes Central provides access to plant data from large-scale genomic and EST sequencing projects. Organism names are linked to the corresponding taxonomic information in NCBI's Taxonomy database. Organisms listed under "large scale sequencing projects" and "genetic maps" are represented in the Map Viewer. Organisms listed under "large-scale EST sequencing projects" are also linked to their EST sequences in Entrez. In October 2002, there were over 20 organisms included the Plant Genomes database.

The Viral Genomes Web site provides a convenient way to retrieve, view and analyze complete genomes of viruses and phages. This site now contains

1000 viral genomes and over 1000 viral genomic reference sequences.

The Spectral Karyotyping and Comparative Genomic Hybridization Database (SKY/CGH) database provides a repository of publicly submitted data which are complementary fluorescent molecular cytogenetic techniques. The SKY/CGH database provides a means for detecting and mapping chromosomal breakpoints; detecting previously unknown chromosomal translocations; characterizing complex chromosomal rearrangements; and identifying marker chromosomes for genome mapping.

In FY2002, NCBI publicly released the Karyotype Converter in order to speed up the entry of cytogenetic data into the SKY database. The Karyotype Converter is a computer programs that automatically reads short-form karyotypes, extracts the information therein, converts it into a readable SKY format, and inserts it into the database. As a result of the database entry, the cytogenetic information can be displayed, and integrated into other genomic analyses, whereas the short-form karyotype would otherwise be intractable.

Microarray technology—a method for generating gene expression data—is another 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. The Gene Expression Omnibus, or GEO is the NCBI tool designed to support the public use and dissemination of gene expression data. GEO represents NCBI's effort to build an expression data repository and online resource for the storage and retrieval of gene expression data. In FY2002, the original platform types of arrays were retired and new platform types put into place. GEO data can be retrieved by GEO accession number, through the GEO current holdings page, or through the Entrez ProbeSet search interface. ProbeSet is deeply indexed and is reciprocally linked to Entrez Nucleotide, PubMed, and Taxonomy. At this time, there are 163 platforms, 2602 samples, and 93 series of microarrays.

Serial Analysis of Gene Expression, or SAGE, is an experimental technique designed to quantitatively measure gene expression. The SAGEmap tool compares computed gene expression profiles between SAGE libraries generated by the Cancer Genome Anatomy Project (CGAP) and submitted by others through GEO. SAGEmap also includes a comprehensive analysis of SAGE tags in human GenBank records. Data can be retrieved by tag, sequence, UniGene cluster ID and library name. This year, the Web site was re-launched with additional features, including new mapping methods, links to genomic sequence via the Map Viewer, and new libraries from the GEO database.

The NCBI is participating in the NIH-sponsored Mammalian Gene Collection (MGC). The goal of the MGC is to provide a complete set of full-length (open reading frame) sequences and cDNA clones for each human and mouse gene. As of October 2002, there were approximately 13,900 distinct human clones and 10,000 distinct human genes. There were also about 8,100 distinct mouse clones and 6,800 distinct mouse genes. All MGC resources generated are fully accessible by the biomedical research community.

NCBI's Molecular Modeling DataBase (MMDB), is Entrez's "Structure" database. The goal of this 3D-structure database is to make structure information easily accessible to molecular biologists. MMDB 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) and the European Bioinformatics Institute (EBI).

In FY2002, revisions were made to the Structure Web servers to improve presentation of biological annotation. The MMDB sequences, retrieved via Entrez, provide links to related information such as: MEDLINE citations, Entrez's organism Taxonomy database, sequence neighbors, the Conserved Domain Database (CDD), Structure neighbors for protein chains, and Entrez's integrated viewer, Cn3D. Entrez's "Structure summary" provides a concise description of the contents of an MMDB entry and available annotation. Improvements were made to the "structure summary" page to show locations of protein domains and/or aligned regions using a new, and easier to understand graphical style. At this time, MMDB serves about 50,000 queries a day and contains about 20,000 structures, up from approximately 15,000 last year.

NCBI's three-dimensional structure viewer, Cn3D, provides easy interactive visualization of molecular protein structures from Entrez. Cn3D also serves as a visualization tool for sequences and sequence alignments. What distinguishes Cn3D 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, coloring by alignment conservation, and a variety of file export formats that together make Cn3D a powerful tool for structural analysis. In FY2002, NCBI released two new versions of Cn3D. The latest version, 4.0, includes a new user interface, a complete alignment editing system including capacity to construct new alignments, improved annotation features, and a built-in help system.

The Conserved Domain Database is a collection of sequence alignments and profiles defining protein domains as recurrent evolutionary modules. It includes

domains from Smart and Pfam—two popular Web-based tools for studying sequence domains—as well as domains contributed by NCBI researchers. This year, CDD was also made part of the Entrez retrieval system. Conserved Domains are indexed for retrieval by keywords, and links between Conserved Domains and Proteins, PubMed, and Taxonomy have been added. Conserved Domains are also linked to other Conserved Domains by two neighboring mechanisms. "Similar" domains are defined as those giving overlapping annotations on sets of protein sequences; "Co-occurring" domains are defined as those giving non-overlapping annotations on sets of protein sequences. Identification of conserved domains within a protein sequence is also available via the CD-search service, which is now run by default for each protein BLAST search.

VAST, or the Vector Alignment Search Tool, is a structure-structure similarity search service that compares three-dimensional coordinates of newly determined protein structures to those in the MMDB. VAST uses an algorithm developed at NCBI for identifying similar three-dimensional protein structures and creates a list of structure neighbors, or related structures, that a user can then browse interactively. The "VAST summary" provides a series of controls for selecting and sorting the structure neighbors. Structure-structure alignments can be easily viewed in Cn3D by selecting neighbors and clicking on a button. Other controls sort structure neighbors by measures of similarity and select subsets that include only one representative of sequence-similar subgroups. These summary pages were updated this year to provide a better and easier to use graphical style. There are currently about 50 million structure-structure alignments recorded in VAST.

Protein Reviews on the Web (PROW) an online resource that features PROW Guides, was included in MEDLINE indexing this year. PROW is an authoritative short, structured reviews on proteins and protein families. It provides approximately 20 standardized categories of information (biochemical function, ligands, etc.) for about 200 human CD antigens.

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, contains the names and lineages of greater than 130,000 organisms, both living and extinct, represented by at least one nucleotide or protein sequence in the NCBI genetic databases. New organisms are added to the database as sequence data are deposited for them. The database is recognized as the standard reference by the international sequence database collaboration.

The Taxonomy browser is an NCBI search tool, and part of the Entrez system, that allows an individual to search the taxonomy database. Using the browser, information may be retrieved on an organism or taxon's lineage with links available to nucleotide, protein, structure, and genome records. Users also have the ability to browse the taxonomic tree. Searches of the NCBI Taxonomy database may be made on the basis of whole, partial, or phonetically spelled organism names, with direct links to organisms commonly used in biological research also provided. The Taxonomy system also provides a "Common Tree" function that allows one to build a tree for a selection of organisms or taxa.

Taxonomy LinkOut has been expanded to include more than 50 linkout providers. The Taxonomy group has improved this service by providing comprehensive links at the species level to several taxonomic databases, including FishBase, Tree of Life and Amphibian Species of the World. Additional links include image databases such as CalPhotos and NPPI. An important new application of linkout involves GenBank entries derived from the University of Alaska Museum, which in turn have links to the Museum's specimen pages. Similar links are planned for the future with other museums and herbaria.

The Taxonomy group developed a new Taxonomy Name/Id Status Report page designed to assist outside groups in maintaining Taxonomy LinkOut links. This page reads files of names or taxids and reports their current status in the taxonomy database. This page is also useful for anyone keeping track of a set of taxnames or taxids in some context to stay current with the taxonomy database.

TaxPlot 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. This tool can show similarity at both the genome and gene level.

UniGene (Unique Human Gene Sequence Collection) is NCBI's system for automatically partitioning transcribed sequences into a non-redundant set of gene-oriented clusters. Each UniGene cluster contains sequences that represent a unique known or putative gene, as well as related information such as the tissue types in which the gene has been expressed and map location. During FY2002, Drosophila melanogaster (fruit fly) and Anopheles gambiae (mosquito) were added to the UniGene database, bringing the total number organisms to thirteen. As of October 2002, 8,351,504 sequences were included in UniGene, with the number of clusters (sets) totaling 321,059.

HomoloGene is a database of curated and calculated homologs for genes represented by UniGene, LocusLink and more recently, by genomic sequences, and assembled contigs using whole genome shotgun (WGS) reads. HomoloGene allows users to explore possible homology relationships among the less well studied genes. 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. In FY2002 HomoloGene was expanded to include zebrafish, clawed frog, thale cress, barley, wheat, maize, and rice, bringing the total number of organisms represented to 13. As of October 2002, there were 96,842 homologous groups included in HomoloGene.

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, Taxonomy, OMIM, ProbeSet, Books, and BLink. This year additional databases added to the Entrez retrieval system include UniGene, SNP, 3D Domains, UniSTS, and Journals.

A new version of the Entrez program was made public in FY2002. Included are XML options for displaying GenBank records. Another new feature allows search results to be displayed in the format of links to other databases. For example, the "OMIM Links" display option will provide the OMIM links for all results of a GenBank search rather than retrieving the links one at a time from individual records.

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. Approximately 100,000 Entrez nucleotide and protein queries are handled per weekday and the number continues to rise.

The Query e-mail server, which provided e-mail access to a subset of Entrez databases, was discontinued in April. NCBI resources are primarily Web-based and most applications, if not all, can be accommodated by Web Entrez, which provides access to more databases and features than were possible through the e-mail interface.

Other Network Services

Usage of NCBI's Web services, first introduced in December 1993, continues to expand as more information and 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. At the end of FY2002, NCBI's site was averaging over 25,000,000 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 ISP monitoring sites across the U.S. and overseas, the average time to load the entire NCBI home page is under 1.5 seconds, an average PubMed search takes less than 3 seconds and availability has been better than 99.5 percent.

During FY2002, NCBI began a transition to Linux-based hardware for its public services. PubMed has been ported to a set of eight 8-way servers and BLAST is in the process of being re-implemented on Linux platforms. Currently, BLAST searches are run on a network of approximately 25 computers, containing a total of about 200 CPUs. Internal research computing is carried out on a "compute farm" of some 40 machines with a total count of 190 CPUs. The compute farm is heavily used for the internal assembly of the human genome as well as genomes of other species.

Due to the high volume of data associated with genome projects, much of NCBI's data is centralized in high-availability network storage devices, now totaling over ten terabytes. Connectivity among internal servers is via a gigabit-switched network; outside connections to the Internet and Internet-2 are via two 155 megabit per second lines. The computing and network architecture has been specifically designed to accommodate the dramatic increase in data resulting from genome projects as well as the increased public demand for NCBI services. Relatively low-cost hardware can be added in increments to the various public services in order to maintain stable levels of performance. Throughout FY2002, the NCBI computer environment has kept pace with a greater than 30 percent increase in demand.

Research

Research is at the core of NCBI's mission. The Computational Biology and the Information Engineering Branches are the main research branches of NCBI, with the latter branch concentrating on applied Research and Development. Each Branch comprises a multidisciplinary team of scientists that carries out research on a broad range of fundamental problems in molecular biology by developing and applying mathematical, statistical, and other computational methods to the life sciences. The research approach taken relies on theoretical, analytical and applied approaches, as, in the field of bioinformatics, these lines of research prove mutually reinforcing and complementary. Research conducted by NCBI investigators has led to the 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.

NCBI's basic research group is within the Computational Biology Branch and consists of 71 senior scientists, staff scientists, research fellows, and

postdoctoral fellows. Research projects focus on computer methods for the analysis of genome sequences and for analyzing and predicting macromolecular structure and function. Other projects underway include techniques in analysis of particular genomes of several pathogenic bacteria, viruses, and other parasitic organisms. Topics of current research include: database searching algorithms, low-complexity sequences, sequence signals, mathematical models of evolution, statistical methods in virology, dynamical behavior of chemical reaction systems, statistical text-retrieval algorithms, protein structure and function prediction, comparative genomics, taxonomic trees, and population genetics.

Currently, the intramural group is engaged in over 20 major 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 110 publications this year with more 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 HHS. NCBI also hosted numerous outside speakers throughout the year focusing on a wide variety of topics.

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 is designed to provide for concentrated efforts on improving and strengthening GenBank and genetic resources. 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 FY2002, NCBI expanded its outreach and education programs to increase awareness of its myriad public databases and specialized tools and services. NCBI staff implemented a general Web site on NCBI resources; 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.

General information Web site: About NCBI

Early this year, a new Web site was launched titled, "About NCBI." This Web site was designed to introduce researchers, educators, students, and the lay public to NCBI's role in organizing, analyzing, and disseminating information in the fields of molecular biology and genetics. "About NCBI" topics include information on NCBI services and resources; science primers, or general explanations, of bioinformatics, ongoing projects, and the various types of sequence information contained on the NCBI site; general information on databases and tools; outreach and education information; and NCBI news.

Education: Mini-Courses and Lecture Presentations

A new mini-course, "BLAST Quick Start" was developed to provide a practical introduction to the BLAST family of programs. Other mini-courses, "Unmasking Genes in Human DNA," "Making Sense of DNA and Protein Sequences," and "GenBank and PubMed Searching," were presented both within and outside the NIH community.

Education: Bioinformatics Training

The intramural Core Bioinformatics Facility reached many NIH scientists throughout the year and continues to gain interest within the NIH community. The purpose of this course is to help NIH researchers make optimal use of computer science and technology to address problems in biology and medicine. 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 CoreBio listserv allows the expertise of the entire group to focus on the diverse array of problems encountered by researchers at the NIH. The training program lasts nine weeks, with each week dedicated to exploring a major topic over a period of four days. The daily sessions consist of an hour of lecture followed by an hour of hands-on work.

Education: Extramural Educational Collaborations

A new five-day course, "NCBI Advanced Workshop for Bioinformatics Information Specialists" was offered late this year. The course was designed for individuals based in medical libraries who provide user training and support for NCBI information resources. The course was developed and taught as a collaborative project

among NCBI and six university-based biologists and librarians who currently provide support at their institutions. Course participants included molecular biologists and librarians with advanced degrees in science who currently provide, or are developing, bioinformatics support programs for their institutions. This course will be offered yearly in August. Course developers and participants have formed the foundation of a bioinformatics support network (BSN) that will facilitate continued learning and communication among the group. Future course participants will be added to the BSN.

A three-day introductory course, "Introduction to Molecular Biology Information Resources" is also in development and will be offered at NLM each fall and spring starting in November 2002, with eventual plans for offerings at regional medical libraries.

Outreach: User Guides for NCBI Resources

NCBI has continued to develop a more comprehensive list of fact sheets that outline the services and databases offered by NCBI. 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 such as Entrez, PubMed and BLAST. A Structure tutorial was added to the Education page during the past year. The "About NCBI" site has an Education page that serves as a comprehensive resource for all tutorials.

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 and available genomes; a section on frequently asked questions; NCBI investigator profiles; and a bibliography of recent staff publications. In FY2002, over 19,000 printed copies of the NCBI News were distributed quarterly. Access to the newsletter via the NCBI Web site has increased dramatically as more people have become aware of its availability online.

"Coffee Break" 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 FY2002, NCBI was referenced in over 150 articles.

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 whole genome, and gene sequence and mapping information. This is due in large part to the 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, plays an active and collaborative role in deciphering the human, as well as other genomes 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) of NLM continues to receive its budget under two different authorizing acts: the Medical Library Assistance Act (unique to NLM), and Public Health Law 301 (covers all of NIH). The funds are expended mainly as grants-in-aid, and in some instances as contracts, to the extramural community in support of the goals of the NLM. Review and award procedures conform to NIH policies. The Web site at (www.nlm.nih.gov/ep/extramural.html) lists grants awarded in FY2002.

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 research in informatics and bioinformatics
- Publication grants to support preparation of scholarly manuscripts
- SBIR/STTR
- Special Projects

Two categories of NLM grant were discontinued in FY2002: Information Access Grants and Internet Connection Grants.

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. During FY2002 the Connection Grants program and the Information Access program were discontinued. Outreach goals will now be addressed by a new program, Internet Access to Digital Libraries (IADL), that combines many features of the discontinued programs plus additional refinements. The change was made in response to comments from applicants and from members of the BLRC that the existing outreach programs were now overlapping, and did not fully reflect recent changes in potentials of information technology.

Accordingly, there are now three types of Resource Grant that range in complexity as well as in dollar amounts and duration. They are considered "seed" grants designed to initiate a service that is expected to become self-sustaining. All three Resource Grants are open to public and private, nonprofit health institutions engaged in health education, research, patient care, and administration all strongly encourage some health science library involvement in the project.

Internet Access to Digital Libraries (IADL) Grants

The IADL grant program replaces the Access and Connection programs, and recognizes that multi-media, digital libraries, and new technologies are now features of a desirable outreach program. The theme of these grants is to facilitate decision-making by timely, efficient provision of reliable information.

Many health-related organizations, particularly smaller ones and those in rural or urban health-underserved areas, lack resources to take full advantage of the Internet's ability to facilitate informed decision making by health professionals and consumers. IADL grants enable organizations to offer access to health-related information provided by NLM and others, to transfer files and images, and to interact by e-mail and videoconferencing with colleagues throughout the world.

IADL grants provide up to $45,000 for a single institution and up to $8,000 each for up to 15 additional performance sites. The applicant may propose two years as the project period, but a longer project period does not increase the total size of the award. The initial RFA resulted in 112 applications, 52 of which were considered meritorious. NLM was able to fund most in FY 2002 and will fund the others in FY2003. The successful applicants had wide geographic diversity, and a gratifying mix of rural and inner city organizations. The enthusiastic response is a useful indicator of the continuing need for outreach programs, despite the increased availability of information technology in the U.S. On February 1, 2003, IADL becomes a regular resource grant program, accepting new applications three times per year.

Information Systems Grants

Information Systems Grants, which average $150,000 per year for up to three years, are suitable for a broad variety of information management projects. They emphasize the use of information technology to bring usable, useful health-related information to end users. This flexible grant mechanism is often used to apply a new technology in a way that improves management of health information or to create unique digital information resources and services.

Integrated Advanced Information Management Systems (IAIMS) Grants

The NLM provides IAIMS grants to health-related institutions and organizations that seek assistance for projects to plan, design, test and deploy systems and techniques for integrating data, information and knowledge resources into a comprehensive networked information

management system. In March 2002, NLM introduced the updated IAIMS grant program, shifting emphasis from infrastructure to content, with emphasis on projects that bridge disparate information systems.

There are now five types of IAIMS grant (two resource grants, two investigational grants and a fellowship). IAIMS Planning Grants provide up to $150,000 per year for one or two years, with an optional infrastructure supplement of $100,000 in the second year; IAIMS Operations Grants provide up to $400,000 per year for up to four years. The IAIMS Pilot Study Grant provides up to $75,000 per year for one or two years; the IAIMS Testing & Evaluation Grant provides up to $100,000 for one or two years. Because of their research aspects, the latter two sub-programs provide support for facilities and administration costs (not provided for resource grants). The IAIMS fellowship gives $50,000 for one or two years and is no longer restricted to funded IAIMS grantees. Applicants for IAIMS grants must base their work in one of three fundamental areas of IAIMS activity (context-appropriate information, standards-based information management, and digital libraries).

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 biomedical informatics as applied to clinical medicine and to basic research. NLM provides both institutional and individual training support.

NLM-Supported Training Programs

Five-year institutional training grants support approximately 150 trainees at predoctoral and postdoctoral levels. Twelve institutions were receiving support in the 5-year funding period that expired June 30, 2002. In accordance with the new competition held in FY2001, 18 training programs were funded for a new 5-year period beginning July 1, 2002. Eleven of the previous 12 were again funded, and seven new programs were added to the set. NLM is expanding its support for such programs in response to the marked recent interest in biomedical computing and the consequent need for trained informaticians. Among our programs, training for bioinformatics is now receiving significantly more attention and opportunity than in previous years, and, for the first time, a program dedicated to imaging informatics is included. For the latter, NLM will receive some co-funding from NIBIB, the new NIH Institute for bioengineering and imaging. NIDCR will continue to contribute funds to NLM to help support slots at these training sites for applicants interested in dental informatics.

The 18 programs currently funded are at the following universities: California (Irvine), California (Los Angeles), Columbia, Harvard, Indiana, Johns Hopkins, Minnesota, Missouri, Oregon Health Science, Pittsburgh, South Carolina, Stanford, Rice, Utah, Vanderbilt, Washington, Wisconsin, and Yale.

Individual Fellowships

As a step in the revision of its individual informatics training fellowships, a new announcement was published that combined opportunities for both basic and applied training, and offered a new stipend schedule created to facilitate recruitment of computer scientists, engineers, librarians, and nurses into informatics. Planned for FY2003 is a new informatics training opportunity intended for those who have had ten or more years of professional experience in some appropriate field.

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, increasingly, projects in electronic publishing, video, and. and other media are being supported.

Minority Support From EP Authorized Grant Funds

NLM continues its support of the NIH program "Research Supplements for Underrepresented Minorities." In FY2002 computer science doctorate candidate, Jonathan Allen, received a minority supplemental award as part of the program. This continuation of support award to the Institute for Genomic Research, with mentored support by Dr. Steven Salzberg as doctoral advisor, is to develop new algorithms to 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

The new Internet Access to Digital Libraries grants included a much higher percentage of awards to minority-serving institutions, community-based organizations, state and local governments, health clinics and other organizations serving rural and inner-city populations than do other NLM grant programs. The

description of this program emphasizes outreach to disadvantaged and geographically remote 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. These 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

In the early years of the grant program, the majority of NLM's research support in informatics 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). In recent years there has been marked expansion in research support for informatics issues related to biological and medical research. However, NLM plans to continue support for clinically relevant informatics. There were a number of meritorious informatics research applications that could not be funded in FY2002 and will be held over for awards in FY2003.

Bioinformatics and BISTI

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 (the original name, "biotechnology" was subsequently changed to "bioinformatics").

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 continuing 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.

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. Vanderbilt and the University of Washington were added in FY2002. How the implementation grants for these centers will be handled, and when the requests for applications will be issued remains to be determined.

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 NIH Projects

NLM participates in a number of other multi-institute projects including bioengineering, pharmacogenetics, imaging, and nanotechnology. In FY 2001 Congress created a new Institute, National Institute of Biomedical Imaging and Bioengineering (NIBIB). Several overlap areas between the interests of NLM and those of NIBIB have been identified.

Other Support

- Conference Grants: Support for conference 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 the indexing and cataloging of bioethics-related materials for inclusion in 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, Extramural Programs Division 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. 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. Target for total project budget from all sources is $50 million over 5 years. The last installment of NLM commitment to this program was in FY2002.

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

This program, begun in FY1998, received 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. 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. Although some small supplements were added to several of these projects in FY2002, funding for the NHAAP informatics program is essentially complete. Work is still on going. A contractors' conference is planned for FY2003.

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 FY2002 include the National Human Genome Research Institute (NHGRI), National Center for Research Resources, and National Science Foundation.

NLM received co-funding for grants from other organizations, including the National Center on Minority Health and Minority Health Disparities, NCI, NIDCR, NHGRI, NIMH, NIBIB, NIA, and the Department of the Army.

SBIR/STTR (PHS 301)

All NIH research grant programs mandate that a fixed percentage of available funds every year be allocated 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 four employees, issued a total of 222 awards for FY2002 (Table 11), as well as administrative supplemental awards for underrepresented minorities and extensions of support for some specifically identified research and resource grants.

Board of Regents

The Board of Regents met three times in FY2002 on February 12-13, May 14-15, and September 10-11. The Extramural Programs Subcommittee and the Subcommittee on Outreach and Public Information were held during each of these meetings. A special Planning Subcommittee meeting was held on May 13, to develop a report to present to the full Board on the design schedule of the new NLM Facility. The report, "Medicine's Library of the 21st Century" was approved by the Board on May 14. The Board approved 523 grant applications, including any special reviews made by the EP Subcommittee. Three new grant concepts were approved for Bioterrorism, a New IAIMS program, and a K-22 Award to facilitate the transition of investigators to independent careers. A new Chair was elected to the Board of Regents, Ms. Alison Bunting, from the University of California, Los Angeles.

The Electronic Council Book was successfully implemented at the May 14-15 meeting in order to streamline the review of grant applications by the Board and its EP Subcommittee. Applications were previously distributed in paper form. The members of the Board have shown an overall positive response to the use of this Web-accessible database.

On May 14, a special awards ceremony, "Lasting Legacies: Gifts to Medicine and the American People" was hosted by the NLM and the Board of Regents to honor 12 prestigious individuals and organizations who have given generous donations to the historical collection of the Library.

Three new Regents were appointed in FY 2002: Ernest L. Carter, M.D., Ph.D., Howard University; A. Wallace Conerly Sr., M.D., University of Mississippi School of Medicine; and Thomas Detre, M.D., University of Pittsburgh. Two special subcommittees were created by the Board at its September 2002 meeting: Biomedical Imaging and Bioengineering Review Subcommittee; and Bioethics Subcommittee.

Grant Review Activities

NLM's initial review group, the Biomedical Library Review Committee (BLRC), evaluates grant applications for scientific merit. BLRC met three times in FY2002 and reviewed 102 applications. The Committee (see Appendix 5 for roster of members) operates as a "flexible" review group. It is composed of three 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.

The Amended Charter of the Biomedical Library and Informatics Review Committee was approved, reflecting the broader scope of research applications in the areas of clinical informatics, bioinformatics, biomedical computing, management of health science information, as well as library science. In addition, a subcommittee name change was approved, from the Biomedical Library Review Committee, to the Networked Information Access Subcommittee. This subcommittee is concerned with resource grant programs that focus on the application of networked computers to improving access to high-quality health information, with emphasis on improving access for rural and urban underserved health professionals, librarians, and consumers.

Special Emphasis Panels:

- Grants: 14 Special Emphasis Panels were held during FY2002. These panels are convened on a one-time basis to review applications for which the regularly constituted review group lacks appropriate expertise, or when a conflict of interest exists between the applicant and a member of the BLRC. The panels reviewed a total of 178 applications during FY2002. One site visit to evaluate an IAIMS application was also carried out by an ad hoc panel.
- Contracts: 4 contract Special Emphasis Panels were held during FY2002. Three reviewed contracts from the University of Pennsylvania School of Dental Medicine, and one contract was for Network Infrastructure Health and Disaster Research.
- Second Review: A second peer review of applications is performed by the Board of Regents as described above. 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.

Summary

- Continued emphasis on BISTI and biomedical computing.
- New Library Resource program—Internet Access to Digital Libraries
- Closure of Connections and Access programs
- Program announcement expressing interest in research in disaster informatics
- Increase in Informatics Research Training programs from 12 to 18 with marked increase in training slots for bioinformatics.
- Significant revision of NLM's individual informatics training fellowships.

Table 11

Extramural Grants
(Dollars in thousands)

	FY 2000		*FY 2001*		*FY 2002*	
	No.	$	No.	$	No.	$
MLAA	92	25,508	174	29,551	152	33,896
PHS	68	19,325	82	22,848	70	24,087
Total	160	44,833	256	52,399	222	57,983

Table 12

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

		FY 2000		*FY 2001*		*FY 2002*	
Category	Program	No.	$	No.	$	No.	$
IAIMS	IAIMS Ph. I	8	1,175	5	745	2	302
	IAIMS Ph. II	3	1,650	1	550	3	1,599
	Total IAIMS	11	2,825	6	1,295	5	1,901
Training	T15	12	7,919	12	6,250	18	10,769
	BISTI Supp.	12	2,000	12	1,948	-----	-----
	Fellowship	13	799	12	705	9	550
	Total Training	37	10,718	36	8,903	27	11,319
Publications		6	268	39	2,406	36	2,141
Resource	Inf. Sys. G08	13	1,879	20	2,119	19	2,383
	Access G07	9	696	13	760	45	3,287
	Connect. G0	5	207	47	1,572	9	418
	Total Resource	27	2,782	80	4,451	73	6,088
Bioethics		1	530	1	697	1	724
Other	Distance Ed.	2	199	-----	------	-----	-----
	AMI Alert*	-----	------	3	1,758	1	182
	AIDS	-----	------	1	74	1	74
NN/LM	Contracts	8	8,186	11	9,967	8	11,467
Total MLAA		**92**	**25,508**	**174**	**29,551**	**152**	**33,896**

*Contracts (includes $858 from NHLBI)

Table 13

Grants Awarded with PHS 301 Funds
(Dollars in Thousands)

		FY 2000		*FY 2001*		*FY 2002*	
Program		No.	$	No.	$	No.	$
Med. Informatics	R01	34	8,590	46	8,770	42	11,032
	DL2	1	1,000	1	1,000	1	1,000
	Total Med. Info.	35	9,590	47	9,770	43	12,032
Bioinformatics	R01	19	4,757	16	3,988	20	4,315
	BISTI	1	300	4	2,852	11	4,360
	Resource P41	7	2,974	9	2,765	6	2,480
	PDB	1	150	1	150	1	150
	Total Bioinfor.	28	8,181	30	9,755	38	11,305
DL2		-----	------	-----	------	-----	-----
SBIR/STTR		4	424	4	502	4	600
Bioethics		-----	------	-----	------	-----	-----
NIH Taps		0	1,030	0	2,671	0	0
Chairman's Grant		1	100	1	150	1	150
Total PHS		**68**	**19,325**	**82**	**22,848**	**86**	**24,087**

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 to support NLM's research and management programs.

OCCS develops and provides the NLM backbone computer networking facilities, 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; develops software; and provides extensive customer support, training courses, and documentation for computer and network users.

OCCS helps to coordinate, integrate, and standardize the vast array of computer services available throughout all of the organizations comprising NLM. 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.

Executive Summary

Enhanced MEDLINEplus: Implemented the Spanish version of MEDLINEplus that included over 400 health topics. Improving functionality contributed to an increase of 70% in pages viewed in the past year. Deployed an advanced search engine to support the Spanish MEDLINEplus. Also:

- Improved software architecture and deployment scripts.
- Improved listserv capability allowing users to post multipart e-mail messages with text and HTML components.
- Added American Society of Health-System Pharmacists (ASHP) drug information and integrated it with existing United States Pharmacopeia Drug Information (USPDI) drug data.
- Added new quarterly update from ADAM, the health and medical content provider.
- Added 30 new Patient Education Institute (PEI) modules.

High Speed Communication Network: Upgraded the NLM backbone communication network capacity from 0.2 Gigabits to 2 Gigabits. Increased the NLM fiber backbone from 1 Gigabit to 10 Gigabits. Enhanced the fault-tolerance level of NLM networks which resulted in elevating the availability of NLM network services to 99.9%. Also:

- Implemented a new network architecture for the NLM backbone network. This included the upgrade of critical components to Gigabit speeds, provided equipment/connection redundancy and higher levels of security.
- Upgraded the network fiber backbone with a new fiber cable plant. This will enable future data transmission of up to 10 Gigabits per second to the LAN closets.
- Expanded broadband access (DSL and cable modem) to NLM contractors and employees. The Citrix remote access system was upgraded and became the primary method of access to NLM services for NLM contractors and staff.

Multi-faceted IT Security Program: Established a multi-faceted and multi-layered IT security program that successfully prevented over 6000 virus attacks, detected more than 150,000 probes, scans, denial of service attacks and other security events on a monthly basis. Also:

- Established the NLM Security Committee, composed of representatives from each NLM program area.
- Reengineered and implemented the NLM network security architecture, which included multiple levels of security protection mechanisms (defense in depth).
- Conducted independent vulnerability assessments from inside and outside of NLM IT security perimeter to identify system vulnerabilities and formulate corrective action plans.
- Deployed intrusion-detection systems to detect external and internal IT security attacks.
- Deployed an NLM-wide virus scanning and spam control mechanism.

New Senior Health Web Site: Deployed the Senior Health Web site, with more than 90 videos and three senior health topics. The site provided enlarged images and large fonts to serve low-vision users. This was a joint effort between NLM and the National Institute of Aging.

Automated Desktop Production: Automating desktop systems led to an increase of 38% in OCCS desktop production capacity. The automation also resulted in standardized configurations and effective help desk support. This year, 687 desktop systems were built or rebuilt, compared to an average under 500 for each of the three previous calendar years.

New Customer Service Support System: Consolidated CustQ and HelpQ, the NLM customer service and help

desk support systems. Deployed Siebel, the new NLM customer service and help desk support system. This new system will strengthen customer relationship management and improve the internal technical support services.

Improved MEDLINE Indexing and Year-End-Processing: Expanded the functionality of DCMS, the MEDLINE indexing system, to facilitate automatic indexing and gene indexing. Enhanced the functionality, reliability, and robustness of MEDLINE applications which led to the improvement of MEDLINE data quality and the decrease of MEDLINE year-end-processing time by 30%.

Enhanced DOCLINE: Expanded the functionality and improved the usability of DOCLINE, the NLM interlibrary loan system, to support 3,200 domestic and international libraries and to process over 3 million interlibrary loan transactions in the past year. Version 1.4 was released in September 2002 and included 47 enhancements in response to user and Library Operations requests. An online form was added to collect user and platform data when DOCLINE users submit questions or suggestions to NLM Customer Service.

Fault-tolerant Computing Platforms: Expanded the capacity and fault-tolerance of computing platforms and storage systems. The expansion reduced application failures by 80%, increased the online backup and recovery capacity, and reduced the backup cycle time by 50%.

Multi-lingual MESH Support: Developed the MeSH Translation Maintenance System (MTMS) to create an inter-lingual database of translations and to support a concept-centered vocabulary maintenance system. The MTMS will make the MeSH vocabulary useful for non-English speakers.

Upgraded Voyager: Upgraded Voyager, the NLM integrated library system, to improve the acquisition, organization, and management of NLM serials, journals, and other health information.

New Search Engine: Deployed an advanced concept-based search engine for the NLM home page and the Spanish MEDLINEplus. The search engine automatically extracts topics and concepts from the data.

MEEC Licensing Savings: OCCS renewed the licensing agreement that provides a bundle of Microsoft products at the lowest cost available in the U.S. Renewing seats, priced this year at $12.88, provide for the current desktop operating system, Microsoft Office Professional, Visual Studio .Net, and BackOffice clients and updates to each of these products. GSA prices for these same products total $1,712.00, by contrast.

Computer Facility Enhancements: Established a Network Operations and Security Center (NOSC) to consolidate system administration, system monitoring, IT security, and after hour help desk in a centralized location. This created additional space and efficiently utilized raised floor space for IT equipment. Other enhancements include installing five 20-ton A/C units to cool the facility and allow for future additions of computer equipment. Installed biometrics security devices to securely control the access of NLM computer room. The devices also serve as a backup to the current proximity card system.

The following describes in more detail OCCS accomplishments in FY 2002.

Customer Services

The IT Services Center (ITSC) Help Desk continues to deliver on its goal to effectively provide a single point of contact for OCCS systems support. This year LAN Support call center services were consolidated into the OCCS Help Desk. As a consequence, first-level ITSC Help Desk services now also include:

- New account creation services for administrative systems;
- Administrative desktop and network password resets (Novell and NT environments);
- Tracking and dissemination of routine system checkout reports; and
- Clearance and dissemination of NLM-wide e-mail broadcasts.

The Help Desk entered and tracked over 7,100 requests for IT support from its NLM customers this year, up from 6,000 last year. First call resolution rose from 367 last year (6.12% of total calls) to 1,002 calls this year (14.11% of total calls) resolved by first-level Help Desk staff. This is an improvement our customers noticed and appreciated. The ITSC staff also considerably improved its accuracy with second tier support assignments after developing a support-staff competencies and responsibilities guide.

Desktop Support

OCCS realigned its desktop support staff assignments and duties with those of the new technical support contractors. As an outcome of this process, DSS staff members took on the role of providing training to NLM staff on OCCS-provided commercial off-the-shelf software. Training this year has included many offerings for SSH Secure Shell, HelpQ, GroupWise, PowerPoint and Windows 2000 Updates, scheduled conveniently and free of cost for NLM staff.

OCCS has implemented a new problem reporting system based on the Siebel Call Center product. This product was acquired to provide problem reporting and resolution tracking for both the Customer Services (external NLM customers) and ITSC HelpDesk

environments. The Siebel product family will replace the products currently used by these two groups respectively, CustomerQ and HelpQ.

Although Windows 2000 was studied and found to offer several advantages over Windows NT two years ago, several core NLM applications could not be supported on a Windows 2000 (W2K) operating system until they were upgraded. This obstacle was overcome and OCCS has advanced moving to W2K wherever possible.

This year, 687 desktop systems were built or rebuilt, compared to an average under 500 for each of the three previous calendar years, resulting in an increase of approximately 38% in our production capacity. A contributing factor to this production increase was the adoption of a new technical approach for automating operating system and software distribution, the On Technologies product called On Command CCM. This product has led to automating many steps previously requiring manual intervention, and consequently has reduced the staff labor time required for each PC's software installation and configuration

Network Support

During FY 2002, OCCS continued in its mission to provide reliable LAN and Internet communications services, meet the communication needs for new IT systems, provide security services, end user assistance and training, implement new network based applications and operating systems, explore new technologies and plan for systems to meet NLM's continued growth in networking, services and communications.

Public Internet connectivity services continued to be provided through an OC3 (155Mbps) circuit to the Genuity network node in Washington DC. The Digicon/Genuity 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.

During FY 2002, the LAN perimeter connections to external networks were upgraded to an aggregate of 2 gigabits per second (Gbps) from 200 Mbps. These upgrades will increase performance and reliability between key network resources and between connections to LHNCBC, NCBI, and NIH/CIT. The interconnection between NLM and the NIH campus backbone has been upgraded from 100 Mbps to 1 Gbps.

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 staff monitors the health of the NLM networks 24/7. These activities take place within the new Network Operations and Security Center (NOSC). The NOSC allows centralized monitoring and management activities. Public status displays were added to the hallway windows of the NOSC. These four, flat panel displays show near real time and trend information about MEDLINEplus, PubMed, and the Internet connections.

Many of the systems in the computer room now rely on the new NLM System Console, which provides KVM (keyboard, video and monitor) connections for access to system consoles. This system frees up valuable space on the computer room floor for various computing systems. These consoles are centrally located in the NOSC and are shared by multiple systems administrators.

Network support continues to provide cable modem, DSL, ISDN and 56K dial-in access for a wide range of NLM users. OCCS recommended small footprint computers 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 are implemented, where possible, for contractors and staff. OCCS recommends the NIH/CIT Parachute dial-in service for users who do not qualify for cable or DSL. Cable and DSL do require the use of additional security software.

In addition to supporting the indexing system, a new Citrix terminal server solution has been implemented as a good fit for flexi-place workers. The terminal server system provides authentication into the NLM network, access to office and NLM business applications, network-based files, and the Internet.

OCCS has implemented a redundant set of CiscoSecure servers that provide secure authentication for users and systems administrators of the various OCCS-managed equipment. These services provide authentication for remote access users dialing in, for systems administrators who need access to the NLM System Console in the NOSC facility, for wireless devices that will be connected to LAN, and for systems administrator access to network devices such as routers and switches.

Systems Support

It was thought that FY 2001 was the second and final year of major support transition for the OCCS systems support team. However, after transitioning from a staff that was responsible for maintaining both legacy/mainframe systems and Unix-based systems in FY 01, this year the team had to transform itself again to a team that supports both Unix-based systems and PC-based systems. Approximately 130 Unix and Windows systems are already built or under construction. The main system support activities included:

- Upgrading and isolation of the Web and Oracle architectures
- Installation, maintenance, and support (IMS) of NIS, NFS, DNS, and Web services
- Unix O/S IMS for approximately 110 systems
- Windows O/S IMS for approximately 20 systems
- Hardware IMS for approximately 130 systems

- Monitoring, Performance, Analysis and Tuning for approximately 130 systems
- Oracle/Versant database IMS for 24 applications
- Security and Account administration for approximately 130 systems
- Reading Room support for several dozen workstations
- Integration of 20+ Windows systems into support architecture

One of the most exciting projects this year was the deployment of the new HP OpenView/Reporter software. This Reporter software is being used to graphically display and report on a database of archived performance parameters for all critical systems. These parameters include cpu busy, disk busy, file utilization, memory busy, network busy, and various other resource values. The availability of these reports and displays will make it possible to see when systems are over-taxed, what happens to a system when new applications are deployed, what happens to a system when it is upgraded itself, and various other key life-cycle behavior.

Upgrades to improve both the Web and Oracle environments within our systems were made this fiscal year. In addition to the deployment of an additional pair of load-balancing front-end systems, thus permitting the additional isolation of the Test and Production subsystems, numerous server and storage upgrades and new installations were performed. Inter-system Gigabit connectivity was also deployed.

Host hardware and software platforms were upgraded for numerous existing OCCS-supported applications including TeamSite, HP OpenView, Relais, Mesh 2000, CustQ/HelpQ, Cold Fusion Development, Production Web Front-End, MEDLINEplus, DOCLINE, Oracle Development, and LOCATORplus. Hardware was procured and deployed to support new applications such as Hstat, Oracle QA, Cold Fusion QA, Encompass, NativeMind NeuroServer, Siebel, the RecomMind Search systems, and the NOSC displays. Additionally, 7 terabytes of storage were procured and deployed to support the new host hardware for both the existing and new applications.

NLM's IBM mainframe was de-installed this fiscal year. This event was made possible by a multi-year LO and OCCS-wide effort to migrate to a distributed client-server Unix and PC environment. Staff provided software support for the ceremonial shutdown ceremony performed by Dr. Lindberg during the Getting to Know OCCS session on October 16, 2001.

IT Security

Throughout the year, NLM demonstrated the benefits of its fundamentally strong security program orientation. NLM's tradition of proactive security initiatives had put in place a technology basis and organizational culture that required only minor augmentation to strengthen measures in the wake of increased security attacks and threats in 2002.

There were a number of security improvements made this year. The NLM Security Committee, composed of representatives from each NLM program area, was established. The committee meets twice a month to discuss general directional issues of organizational security strategy, security practices, and preparedness including needs to augment current security awareness by general NLM staff members. Also a Vulnerability Assessment Management Program (VAMP) was created and NLM now subscribes to the leading security awareness message sites and reports on alerts issued by these services. An Incident Response Program, backed by full onsite forensic analysis capabilities, was also created. Finally, upgraded network security architecture resulted in multiple levels of security protection (defense in depth).

In September, the security team and the network team enhanced NLM's security posture regarding virus detection/deletion by implementing a new "virus wall." The virus wall filters potential threats at the infrastructure boundary and filters e-mail attachments. Also this year, the SpamAssassin was implemented to help NLM users identify spam e-mail messages. SpamAssassin performs header analysis and content analysis, tagging the spam messages and moving them to a folder. Recipients still have full control of the e-mail message.

In order to strengthen the NLM network and Internet security posture, Riptech was awarded a contract to perform independent vulnerability assessments from inside and outside of NLM. Riptech submitted a final draft of the "Network Security Assessment Report." The assessment included external and internal scanning to identify vulnerabilities on workstations, servers, and network devices. The assessment assured the Library that NLM administrators have the necessary talent and tools to provide significant security of the NLM network. Increased cooperation and knowledge sharing initiated during the assessment process will strengthen each division subnet's security and thereby improve the NLM security posture.

Successful implementation and utilization of intrusion-detection systems is a viable defense mechanism for protection of NLM networks; Internet attacks are detected and blocked daily. After the NLM beta testing, two Gig IntruVert appliances were purchased and were successfully implemented into our new network infrastructure. One is running at our new perimeter architecture and the second is monitoring all Internet-2 traffic pending the move of I-2 to outside of the primary firewall.

Computer Facilities

NLM systems continue to be supported in a safe environment in NLM's computer facility, which is available 24/7. The Network Operations and Security Center was established in 2002 to centralize several

functions, including public information display systems, IT system and service monitoring, IT system administration, IT security event monitoring, and after-hours Help Desk support.

The NOSC display system consists of four 32-inch wide-screen plasma displays that are visible on the outside of the computer room. The intended audience of this display system is the general public and NLM staff. The system consists of information "panels" with descriptive text, statistical charts and near-real time activity monitors. Each panel focuses on a particular NLM service or IT infrastructure component. They include near-real-time utilization counters for MEDLINEplus and for PubMed/MEDLINE, NLM services as seen by remote users at 35 locations around the world, and near-real-time utilization data of NLM's Internet-1 and Internet-2 data communication links.

Major computer facility accomplishments this year included:

- Successfully terminating mainframe platform life cycle and removing all outdated and unnecessary computer systems, hardware, storage and furniture. Centralization and consolidation enabled OCCS to create additional space and efficiently utilize raised computer floor space for computers. Increased space also allowed NLM to densely populate the computer facility with an array of current and new computer systems.
- Installing additional cooling to accommodate the influx of equipment into the facility. Purchasing and installing five 20-ton A/C units sufficiently cooled the facility and allowed for future additions of computer equipment. These systems also allowed NLM to create a redundant cooling system within its largest facility.
- Biometrics security devices—iris scanners and hand-geometry units—were installed at the computer facility entrance, the NOSC and the main computer facility floor.

Consumer Health

MEDLINEplus: Throughout 2002 MEDLINEplus continued to be one of NLM's most publicly visible Web sites. On September 9, the Spanish language MEDLINEplus site was officially opened to the public, a major milestone. Throughout the year, the traditional MEDLINEplus site received at least one major upgrade each quarter. In addition to the major changes required by the introduction of MEDLINEplus in Spanish, there were improvements in the listserv capability, the ADAM encyclopedia, and the addition of drug information from the American Society of Health-System Pharmacists. In addition, 30 new Patient Education Institute modules were added to MEDLINEplus. The MEDLINEplus development environment was ported to the TeamSite, a fully featured environment that supports more efficient code/version control and more tightly managed workflow.

Senior Health Project: This project seeks to provide health information on the Web using content and modes of delivery that are appropriate for older Americans with a range of access limitations (notably, low vision and low hearing). The objectives overlap with general efforts to comply with section 508 regulations. Work began as a joint effort of the National Institute on Aging and the NLM, and participation has expanded to other NIH institutes.

In the first quarter OCCS developed a set of Web page templates to meet the needs of low-vision and blind users. For low-vision users, the Web pages include enlarged photographs and large fonts. These templates were used to create pages on Alzheimer's Disease and Exercise for Older Adults. The templates also include important Section 508 compliance features for support of screen readers ("talking pages"). Pages produced with these templates were tested with the JAWS screen reader.

Another early goal was to provide video streaming with maximum transparency for users. To this end, auto detection technology was integrated to detect user connection speed, video player software, and operating system; this enables the Web server to automatically apply the appropriate streaming method. Testing was performed with PC and Mac platforms operating with connections from 56K to high speed DSL and cable modem. The site currently supports video streaming with the Microsoft Windows Media Player and QuickTime.

On March 19, 2002 the project beta version was deployed for in-house use and testing. A large number of videos were incorporated and enhancements added, including video rewind. With new organizations joining the project, all software was ported to the TeamSite environment in the fourth quarter, to improve configuration management and better control the development workflow. As testing and template refinement continue, the current focus is on increasing the automation of Web page development.

Virtual Customer Service: OCCS has established a joint project team with Library Operations to implement virtual customer service on the NLM Web site. Virtual customer service refers to the use of intelligent software to respond to customer questions. For example, if a user types "when is the library open?" the system hyperlinks the user to a Web page with library hours of operation. The benefits to the user are 24/7 service and the ability to ask questions in a conversational mode. The benefits to NLM are reduced customer service costs and increased capacity to respond to inquiries. The project team has selected NativeMinds vReps technology. A development and testing environment has been set up and programs are being written to import health topics from MEDLINEplus. Content for responses to typical Library Operations requests is being developed.

Professional Health Information

NLM Classification System: The NLM Classification System includes two functions: a viewer application that allows public access over the Internet for searching and browsing the NLM classification schedules, and an editing application that is used by the NLM Cataloging Section staff for system maintenance. Both applications were released for public use this year. The Cataloging Section has used the editor to rework the classification production database.

Mesh Browser: The MeSH Browser is an online terminology lookup aid accessible to many NLM applications whose users frequently consult MeSH terms. It locates descriptors of possible interest and shows the hierarchy in which those descriptors appear. Virtually complete MeSH records are available, including the scope notes, annotations, entry vocabulary, history notes, allowable qualifiers, etc. In 2002 the MeSH Browser build process was upgraded with significantly better performance. The generation time for a new MeSH Browser release has been reduced from 24 hours to 2.5 hours. As a result, the MeSH Browser is now updated weekly, rather than quarterly as in previous years.

DOCLINE: DOCLINE is NLM's online interlibrary loan request routing and referral system. It processes more than 3 million interlibrary loan requests annually. DOCLINE allows users to make document requests that are routed automatically to libraries that report owning the specific year or volume requested. SERHOLD provides journal holdings information. Through Loansome Doc Patron Administration the system allows libraries to maintain administrative information on their Loansome Doc users. Loansome Doc is a stand-alone module that allows individuals to submit DOCLINE requests under pre-established agreements to a participating NN\LM library.

OCCS and the National Center for Biomedical Information established a data sharing procedure to support the new LinkOut feature in PubMed. LinkOut allows a library that is an NLM SERHOLD participant to display its print holdings information to users searching in PubMed. As a user reviews the match list of citations for a given PubMed query, this feature provides a quick way to check whether the corresponding article is in the holdings of the library. NCBI obtains the SERHOLD information by triggering a data extraction and transfer process from the DOCLINE server each night.

A new version of DOCUSER was planned and specified in 2002. Target changes include improved ease of use, increased user efficiency, and response to various user requests for enhancements. Development began early in the fourth quarter with completion now anticipated for the second quarter of 2003.

HSTAT: On September 30, the Health Service Technology Assessment Text (HSTAT) system was completely transferred to OCCS from the Lister Hill Center's Information Technology Branch. HSTAT is a free, Web-based resource of full-text documents that provide health information and support health care decision-making. HSTAT's audience includes health care providers, health service researchers, policy makers, payers, consumers and the information professionals who serve these groups. HSTAT also provides links to external databases, including PubMed, the CDC's Prevention Guidelines Database, and the National Guideline Clearinghouse.

HSTAT is built upon technology that was new to OCCS. One of the most difficult aspects of preparing HSTAT for transfer was the need to upgrade all software to the latest available version for future maintenance and stability. This was accomplished along with updating and completing all the documentation. OCCS also investigated and analyzed HSTAT system hardware structure and indicated possible system failure points as well as performance issues. The OCCS project team set up a new hardware infrastructure and worked to configure the new hardware, solve the system issues, and set up the proper system monitor/clean up tool. The new hardware infrastructure resulted in increased system performance.

Relais: Relais, an application from Relais International, manages the fulfillment of requests for copies and interlibrary loans. Requests arrive from remote customers via DOCLINE and from reading room customers. An important function of Relais is its delivery capability. Modes of electronic delivery include fax, e-mail attachment, post-to-Web e-mail, and various file transfer methods including Ariel. A new delivery FTP module from the vendor was implemented in the first quarter of 2002 and eliminated the need for three intermediary delivery servers in the OCCS computer room. Relais 3.8, a new release scheduled for implementation in the fall of 2002, adds electronic delivery support (e-mail attachment and Post-to-Web e-mail) for customers submitting requests in the NLM reading room. These customers have had only hardcopy delivery as an option until now. In conjunction with Relais 3.8, OCCS is porting the application to Windows2000, which is the standard operating system platform for both the vendor and the NLM LAN support group.

Literature Selection Technical Review Committee: The Literature Selection Technical Review Committee, which meets periodically to recommend medical journals for inclusion in *Index Medicus*/MEDLINE, uses an application to track the progress of its business. In the fourth quarter, the journal history table underwent extensive testing. The table contains historical data on each medical journal, including information such as the journal's date of creation, date of inclusion in Index Medicus, date of exclusion, etc. A batch process has been developed to automatically update the table on a nightly basis.

OLDMEDLINE: OLDMEDLINE is the set of citations dating from 1966 and earlier that were not included in MEDLINE. Some of these citations are converted from the mainframe, others are scanned from hardcopy publications. The mainframe citations have been imported into the OLDMEDLINE application database and work is ongoing to capture citations from hardcopy sources. As of the fourth quarter of 2002, the years 1965 through 1957 were completed and citations for 1954 to 1956 were in process. Preliminary conversion work is in progress for 1953.

History of Medicine: OCCS supported NLM's History of Medicine Division on two projects in 2002—Images from the History of Medicine and Reports of the Surgeon General. Images from the History of Medicine is a database of images in NLM's historical prints and photographs collection. The images are Web accessible and indexed for retrieval by keyword or topic. OCCS worked with HMD on various maintenance tasks, including index maintenance, COTS software upgrades, and conversion of the index catalog from SGML to XML. In the second quarter the entire Images application was moved to the OCCS server cluster to simplify production support.

The Reports of the Surgeon General project entails Web hosting publications issued by the U.S. Surgeon General's office throughout its history. The documents were captured in SGML and are being converted to XML for insertion into the project database. OCCS is also assisting with the generation of the index catalogs.

Medical Subject Headings (MeSH): MeSH is NLM's controlled vocabulary thesaurus. It is used for cataloging, indexing, and searching citations in MEDLINE and as an online reference tool for users of other NLM custom applications. Portions of the MeSH database are published annually in hardcopy. OCCS supported the MeSH Section in distribution of various MeSH releases, including ASCII MeSH (which is integrated into the DCMS application allowing quick lookups by citation indexers), XML MeSH (used by NCBI for the MeSH Browser), and PDF MeSH (sent to GPO for printing).

OCCS supported maintenance of the application used by the MeSH staff to develop the thesaurus. Tasks included installing software upgrades, adding search capabilities, and developing software to distribute Pharmacological Actions and Pharmacological Drugs in XML. Important upgrades of quality assurance functionality included new methods to monitor the integrity of MeSH databases and a rework of the application's audit module.

OCCS is working with the MeSH Section to implement the MeSH Translation Maintenance System which will allow an interlingual database of translations. With the translations integrated into a single database, modifications of the terminology tree is automatically tracked for all languages. The system, which also will allow continual updating of the translations, will be released in mid-2003.

Data Creation and Maintenance System (DCMS): The Gene LocusLinks function was put into production in the first quarter. Indexers create gene links during the normal indexing and the link is flagged for review by a gene specialist. Integrating this function with the general citation indexing workflow increases overall productivity at minimal additional cost.

Lister Hill's Medical Text Indexer (MTI) was integrated with the DCMS RelatedRecord function. It was tested by several indexers through the summer and is now available for use by all indexers. In this context, the MTI application provides "decision support"; that is, it presents a picklist of candidate terms from which the indexer chooses. A fully automated MTI deployment for meeting abstracts is currently in progress.

Significant progress was made in 2002 further consolidating NLM's legacy medical citation databases on the NLM Web site and distributing them to licensees in open (XML) formats. Most notably, the transfer of HISTLINE and BIOETHICS files to DCMS was completed. POPLINE has also been imported into DCMS but is in the process of being exported to PubMed and licensees.

An initiative seeking faster propagation of MeSH changes into the MEDLINE citation database began in 2002. Synchronizing the full citation set with MeSH updates has traditionally waited for year end processing. In April a daily update function was put into production whereby changes in Chemical Concept terms are propagated on a nightly basis. The chemical term change is detected by the software and triggers a process that revises all associated citations in DCMS. OCCS also included over 60 enhancements to DCMS in maintenance releases during the fiscal year in response to user requests. These included posting the Indexer rules online. They are now accessible in a searchable form to indexers through a single mouse click.

NLM Web Page

Advanced Search Engine: The new RecomMind search engine is a concept-based system that performs a search on the concepts associated with words in a query, not just on the exact words. It was first integrated with the NLM public Web site earlier this year and has supported the MEDLINEplus Spanish site since its debut in September. Work on the MEDLINEplus English implementation is in progress. This new technology was a high priority item for the MEDLINEplus Spanish site because it is capable of handling diacritics. The RecomMind integration has involved developing XSLT style sheets, cascading style sheets, XHTML, HTML, and XML. The project has also required Java servlet development and PL/SQL programming. The RecomMind engine automates URL

verification and concept learning, and these features have required extensive configuration work. Corrective steps have included software modifications by the vendor as well as configuration changes.

Web Content Management: NLM's new content management software, TeamSite, a commercial product from Interwoven, Inc., was deployed in April. OCCS and Library Operations provided extensions to TeamSite to ensure optimum use at NLM. TeamSite is now a major asset in the quality assurance program. It also enhances productivity by simplifying the addition of Web content. It manages creation, maintenance and promotion of Web documents to the NLM public Web site and Intranet and is also used in conjunction with development of a large number of NLM applications.

Web Content Technical Support: Throughout the year, the OCCS Web Support Team provided a range of technical consulting services to NLM contributors on projects such as the Once and Future Web (exhibit), Greek Medicine (exhibit), MEDLINEplus, and Senior Health. The provided services included debugging platform- and browser-related presentation problems (Mac vs. PC, Internet Explorer vs. Netscape, Version x vs. Version y, etc.), style sheet methods, site engineering, and linking problems when relocating content to new servers.

Section 508 and Section 508 Plus: OCCS continued to support various NLM branches in Section 508 accessibility compliance efforts. Typical support tasks included running compliance tests on existing Web pages and providing technical assistance to fix discrepancies. Commercial evaluation software is used. Sample projects include the NLM Administrative Office site and the Senior Health Project site.

The goal of the Section 508 Plus R&D initiative is to explore technologies and implementation methods that will ease section 508 compliance and accelerate progress across NLM's many Web applications in making the Library's vast Web assets accessible to all users. In 2002, the project's first year, the Section 508 Plus team has collected and analyzed requirements, investigated new technologies, identified implementation architecture options and approaches, and built a prototype Web site for empirical testing with target user populations. Goals of the project include cross platform support, eliminating the need for additional software or hardware on the part of the user, and addressing the needs of special populations with limited physical abilities. Testing so far has indicated that new technology can be used with various populations other than the visually impaired, including non-native English speakers and children and adults with reading/cognitive disabilities whose sight is normal.

Outreach

Consumer Health and Outreach System: This system is being developed by OCCS in support of NLM's Outreach Consumer Health Project. Initial requirements analysis and design began in the first quarter of 2002. Implementing the diverse requirements of three different user populations led to significant design challenges in the following months. After extensive revisions based on user input, a prototype was demonstrated to the Project Committee in September. The response was positive and led to a last round of feature additions for the production system. The extensions now contemplated include the use of map images. The production system will also include a more fully developed user interface and additional database fields.

Health Services/Sciences Research Resources (HSRR): OCCS staff continued maintenance support on this application, which it developed for NLM's National Information Center on Health Services Research and Health Care Technology. The HSRR databases contain data surveys and other information such as clinical practice guidelines and health care technology. To improve performance, OCCS implemented the Oracle InterMedia search engine in the second quarter. Search time improved dramatically from 10–15 seconds to 2–4 seconds. Testing continued with the search coverage extended to more fields of the data. NICHSR is reviewing requirements for inclusion of other fields in the search scope.

The Maintenance module has new interface mechanisms that ease user access to data elements when administrators need to update a dataset or instrument record. A link checker program was developed that tests the URLs in the database and generates a list of stale links. In response to user requests, a Java-based spellchecker program (JSpell) has been selected for integration with the maintenance module; the procurement process is under way. Finally, a user manual was drafted and is being put online.

Administrative Support Systems

New Customer Service System: NLM operates two customer support applications: the Reference Call Center for external users (known as CustQ), and the OCCS Call Center for internal NLM users (known as HelpQ). The Reference Call Center assists users of NLM's public applications such as MEDLINEplus and DOCLINE. The OCCS Call Center deals with a range of information technology issues, from computer room maintenance to desktop PC setups. Both help systems currently run on CustomerQ, a discontinued commercial product. As a replacement, NLM acquired Siebel Enterprise in early 2002. This will provide a common base for all customer support applications anticipated by NLM for the present and the future.

Health Resource Information System for Personnel: This year, the OCCS project team made several enhancements to the Health Resource Information System. The team added security data protection by using a scramble-unscramble function. The system automatically scrambles

the data when it is saved into the database. The data is then automatically unscrambled when personnel staff initiates a retrieval. In addition, the OCCS project team successfully implemented an award module in the system which allows personnel to enter the Award data and track the status.

ADMINISTRATION

Jon G. Retzlaff
Executive Officer

Table 14

FINANCIAL RESOURCES AND ALLOCATIONS, FY 2002
(Dollars in Thousands)

Budget Allocation:	
Extramural Programs	$60,580
Intramural Programs	203,599
Library Operations	(78,484)
Lister Hill National Center for Biomedical Communications	(54,884)
National Center for Biotechnology Information	(57,939)
Toxicology Information	(12,292)
Research Management and Support	11,613
Total Appropriation	275,792
Plus: Reimbursements	8,000
Total Resources	**$283,792**

Personnel

In October 2001, **Maricel G. Kann, Ph.D**., joined the staff of the Computational Biology Branch of NCBI as a Fogarty Visiting Fellow. Dr. Kann received her Ph.D. in chemistry from the University of Michigan. Her thesis research involved optimization of a score function for detection of distant homologies using protein sequence alignment, applications of this optimization procedure to hidden Markov models and protein structure prediction, and other applications of statistics and probability theory to computational molecular biology. At NCBI, Dr. Kann is performing research in protein classification algorithms.

In November 2001, **Kira S. Makarova, Ph.D**., joined the staff of the Computational Biology Branch of NCBI as a Staff Scientist. Dr. Makarova, received her Ph.D. in molecular evolution from the Institute of Cytology and Genetics of the Russia Academy of Science. Dr. Makarova began as a Post-doctoral Fellow with the Department of Pathology of the Uniformed Services University of the Health Sciences (USUHS) in 1998. At the NCBI, she will work on comparative analysis of eukaryotic genomes, which is central to the development of a new generation of protein databases.

In December 2001, **Olga D. Ermolaeva, Ph.D.,** joined the staff of the Information Engineering Branch of the NCBI as a Staff Scientist. Dr. Ermolaeva, received her Ph.D. in molecular biology from the Shemyakin-Ovchinnikov Institute, Russian Academy of Sciences. At the Shemyakin-Ovchinnikov Institute, Dr. Ermolaeva worked on the theoretical analysis of the methods of subtractive hybridization. Dr. Ermolaeva brings a unique mix of skills to NCBI—expertise in genomics, database management and programming. Since joining the NCBI, Dr. Ermolaeva has been instrumental in launching the MapViewer system for the human genome.

In December 2001, **Mr. Scott D. McGinnis**, joined the staff of the Information Resources Branch of NCBI as a Staff Scientist. Mr. McGinnis received a M.A. degree in biochemistry from the University of Buffalo. Mr. McGinnis started at NCBI in 1998 as a member of the help desk staff. He developed a support team that helped to free up programmer staff from having to respond to user questions and requests. He will be the lead person responsible for monitoring abuse of NCBI Web service and taking corrective action and will assist in NCBI's outreach efforts with users and in suggesting potential directions for future BLAST developments.

In December 2001, **Ms. Joyce Backus** was selected to lead the Division of Library Operation's team for library systems, databases, and network services. Ms. Backus received her Masters in library science from Catholic University. She has been with NLM since 1985 when she joined the staff as a Library Associate. In her new position, Ms. Backus will focus on leading a Web group of six librarians as well as support and contract staff who manage the NLM's main Web site, the Intranet, and MEDLINEplus. The group will also provide technical consultation in selecting and implementing software solutions.

In December 2001, **Mr. Kenneth Niles** was appointed to the position of Head, Collection Access Section (CAS), Public Services Division, Division of Library Operations. Mr. Niles received his Masters in library science from State University of New York. He joined NLM in 1991 and has represented the Library extremely well with scholars and researchers worldwide, ensuring that their visits to use the collection were successful. Prior to his selection, he served as Head, Onsite Unit, CAS. In his new position, Mr. Niles will focus on matters pertaining to new computer technology, integrated library systems, and coordination of these products with other NLM Divisions and the Regional Medical Libraries.

In January 2002, **Mary Moore, Ph.D**., was appointed to the position of Head, Reference Section, Public Services Division, Division of Library Operations. She received her Ph.D. in library and information science from the University of Texas at Austin, where her research focused on diffusion of information innovations. Prior to joining NLM, Dr. Moore was Dean of Library and Information Resources and Associate Professor, College of Communications, Arkansas State University. Dr. Moore has extensive experience in health sciences libraries, has taught courses on information networks, Internet issues, and World Wide Web tools, and has published and lectured widely on telemedicine and on the role of the library and the use of technology in continuing education

and outreach. As Head, Reference Section, Dr. Moore will manage NLM's centralized customer services.

In January 2002, **Sergey V. Bazhin**, joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Mr. Bazhin received a M.S. degree in physics with special training in biophysics from the Moscow Institute of Physics and Technology. He continued his studies in the Ph.D. program at the Moscow Institute of Molecular Genetics. At NCBI, Mr. Bazhin currently supports the input of data from High Throughput Genomic (HTG) Sequences projects. He is also responsible for SMART UPDATE, a system which allows NCBI staff to interactively update data, which is already in the NCBI Repository.

In January 2002, **Alexandre Souvorov, Ph.D.,** joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Souvorov received a Ph.D. in physics and mathematics from Latvian State University, Riga, Latvia. Dr. Souvorov served as a Researcher/Mathematician at the Carolinas Medical Center, Charlotte, NC 1996–2000. At NCBI, he works on a new tool for the comparison of complete genomes that provides a genome-wide approach to the study of gene and protein functions and a graphical interface for BLAST with complete and unfinished genomes. In addition, he works on an alignment viewer: a graphical interface and program for global alignment.

In March 2002, **Yuenyin Kathy Kwan** joined the staff of the Information Engineering Branch of NCBI as a Staff Scientist. Ms. Kwan received a MLS degree from the State University of New York at Buffalo and a M.A. in computer science from the Queens College of the City University of New York. Prior to joining NCBI, Ms. Kwan held a variety of positions in information technology development and management in several libraries, including Harvard University. At NCBI, Ms. Kwan will serve as project coordinator for LinkOut ensuring the quality of the information links and working with staff and users to support service expansion.

In April 2002, **Barbara Rapp, Ph.D**., was selected for the position of Coordinator of the Associate Fellowship Program, NLM's post-masters degree training and internship program for health sciences librarians. For the past 12 years, Dr. Rapp worked at the NCBI where she directed the user services, outreach, and training programs. Dr. Rapp received her M.S. degree in library and information science from the University of Illinois in 1978, and her Ph.D. in information science from Drexel University. Upon completion of her doctorate, she served on the faculty of the School of Library and Information Science at the Catholic University of America before joining NLM in 1988.

In April 2002, **Sergey Krasnov, Ph.D**. joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Krasnov received his Ph.D. in computational mathematics from the Moscow Institute of Physics and Technology. Dr. Krasnov has worked at NCBI as a contractor with InforMax Inc. with several key roles in the development and implementation of the PubMed software systems since 1997. Dr. Krasnov's primary project is with the PubMed Central (PMC) team at NCBI as the software development project leader. Under his leadership the entire PMC system was redesigned and the new system was launched in July 2001.

In June 2002, **Farideh Chitsaz, Ph.D**., joined the staff of the Computational Biology Branch, NCBI as a Research Fellow. Dr. Chitsaz, a native of Iran, received her Ph.D. in molecular and cell biology from the University of Maryland. Prior to joining NCBI, Dr. Chitsaz was employed in the Malaria Vaccine Development Unit at NIAID. Dr. Chitsaz has completed an analysis of the large, highly diverse and unusual gene family encoding the major antigenic protein, PfEMP1, expressed on the surface of infected red blood cells. Her study will help to predict structural characterization of PfEMP1 and its role in malaria infection. At NCBI, she will apply successful computational technologies to the antigenic complement encoded by the entire P. falciparum genome.

In June 2002, **Zhan Zhang, Ph.D**., joined the staff of the Office of High Performance Computing and Communications, Lister Hill National Center for Biomedical Communications (LHNCBC) as a Staff Scientist (VP). Dr. Zhang, a native of China, received his Ph.D. in Image Processing from Wuhan University. His doctoral research work focused on noise reduction and image contrast enhancement based on wavelet analysis and the perceptual characteristics of the human visual system. Before coming to Lister Hill, Dr. Zhang held positions at the State University of New York at Buffalo and the University of Iowa. At LHNCBC, Dr. Zhang will work in the area of image processing, including the design and development of prototype systems for the efficient segmentation, alignment, storage and retrieval of biomedical images.

In July 2002, **Mr. Anton Golikov**, joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Mr. Golikov, a native of Russia, received his Master's of Science in mechanical engineering from the State University of Aircraft Technology, Moscow. In 1999 Mr. Golikov worked for Informax, Inc. as a Systems Analyst assigned to NCBI. He quickly became involved in the NCBI PubMed/Entrez development group due to his strong background in computer science and software development. He developed search software for the PubMed database. His current position with NCBI entails work on the project "Cubby" which provides PubMed users with a personalized search facility where each user may set their own parameters to interact with the system.

In August 2002, **Donald W. King, M.D**. was appointed NLM Deputy Director for Research and Education. Dr. King received his M.D. from Syracuse University and he completed his internship and residency in pathology from the Presbyterian Hospital, New York. He has held various executive and academic appointments including Director, Given Institute of Pathobiology, University of Colorado Medical Center; Dean, Division of

Biological Sciences, University of Chicago; and Dean and Vice President of the Pritzker School of Medicine. Most recently, he served as the Executive Director of the American Registry of Pathology at the Armed Forces Institute of Pathology in Washington, D.C. Dr. King is very familiar with the Library, having served as a consultant and member on numerous long range planning panels.

In August 2002, **Jack W. Snyder, M.D., Ph.D**., was appointed Associate Director for Specialized Information Services, NLM. Dr. Snyder received his B.S. and M.D. from Northwestern University, a J.D. from Georgetown University, a Master of Public Health from Johns Hopkins University, a Master of Forensic Science from George Washington, and a Ph.D. in Pharmacology and Toxicology from the Medical College of Virginia. Prior to joining NLM, he held various executive positions in private industry. In addition, he has held various academic appointments at the Jefferson Medical College/Thomas Jefferson University, Philadelphia, Pennsylvania. Most recently, he served as Physician, Immediate Health Care, Medical Toxicology-Occupational Medicine at St. Vincent's Medical Center in Bridgeport, Connecticut. As Associate Director for SIS, Dr. Snyder will be responsible for providing direction and leadership in planning, developing, and administering a national toxicological and environmental health information program.

In August 2002, **Jon G. Retzlaff**, was appointed NLM Executive Officer. Mr. Retzlaff received his MPA from the School of Public and Environmental Affairs, Indiana University, Bloomington, and an MBA from the Sloan School of Management, Massachusetts Institute of Technology. Mr. Retzlaff came to NIH in 1993 as a participant in the Presidential Management Intern Program. His rotational assignments included: administrative management, budget formulation and execution, Congressional, legislative and public affairs. From 1995 to 1998, he worked in the NIH Office of Legislative Policy & Analysis. In 1998, he was detailed to the House of Representatives Committee on Appropriations. After that assignment, Mr. Retzlaff joined the Office of the Assistant Secretary for Legislation at the Department and later became a senior legislative advisor at the National Institute of Neurological Disorders and Stroke. From NINDS, he was detailed to the United States Senate Committee on Appropriations. There he helped prepare the FY2001 appropriations bill and report. At NLM, he provides advice to the Director and other senior staff on administrative management matters and directs the administrative programs including budget, acquisitions, human resources, space management, travel as well as other administrative services.

In August 2002, **Yi-Kuo Yu, Ph.D.,** joined the staff of the Computational Biology Branch, NCBI as a Staff Scientist. Dr. Yu, a native of Taiwan, received his Ph.D. in physics from Columbia University, New York. His dissertation was on "Directed Polymers in Random Media and Kinetic Roughening in Interfact Growth". From 1994 to 1997, he conducted postdoctoral research at Case Western Reserve University. In 1997, he became an assistant professor of physics at Florida Atlantic University, Boca Raton. Dr. Yu's recent research has had a major bearing on problems in biological sequence analysis, which is one of the most important research and service components of the NCBI. It is the statistical theory on which the BLAST algorithm is based.

In August 2002, **Medha Bhagwat, Ph.D**., joined the Information Resources Branch, NCBI as a Staff Scientist. Dr. Bhagwat received her Ph.D. in biochemistry from the University of Maryland at College Park. Her thesis research elucidated the mechanism of action of a DNA repair enzyme, FPG protein, from E. coli. In 1998, Dr. Bhagwat joined NCBI as a contractor with ComputerCraft corporation working as a scientific data analyst for the GenBank database. Dr. Bhagwat has been working for the User Services section and is involved in bioinformatics training that NCBI offers.

In August 2002, **Chunlei "Charlie" Liu, Ph.D**., joined the staff of the Computational Biology Branch, NCBI as a Staff Scientist. Dr. Liu received his Ph.D. in physiology from Dartmouth College. He completed his postdoctoral research on tumor vaccines based on melanoma associated antigens at NIH in 1997. In addition, Dr. Liu completed a M.S. in computer science from the Johns Hopkins University in 2001. At the NCBI, Dr. Liu will work on the design of a sequence-display graphics library from C++ Web servers, and modification of existing servers for Molecular Modeling Database and Conserved Domain Database to make use of this library. Dr. Liu will also design a new server and associated database for display of Related 3D Structure links from protein sequences in PubMed/Entrez.

In August 2002, **Mehmet M. Kayaalp, M.D**., Ph.D., joined the staff of the Cognitive Science Branch, LHNCBC as a Staff Scientist. Dr. Kayaalp, a native of Turkey, received his M.D. from the Istanbul School of Medicine and his Ph.D. in intelligent systems (medical informatics track) from the University of Pittsburgh. Dr. Kayaalp served as the chief medical officer for the Turkish military before coming to Southern Methodist University in 1990 where he conducted research and pursued a Master's Degree in computer science. Since 1998, he served as a researcher in an IAIMS project. Dr. Kayaalp will bring his skills of knowledge representation, modeling, inference and connecting specific queries of clinical questions to be part of the project team being assembled to focus on biomedical knowledge discovery.

In August 2002, **Stephen E. Wilhite, Ph.D**. joined the staff of Information Engineering Branch, NCBI as a Staff Scientist. He received his Ph.D. in plant biology from the University of Maryland at College Park. Dr. Wilhite has a solid background in molecular biology and genetics with a concentration on plant and fungal biology. While at NCBI, Dr. Wilhite will participate in all phases of processing submissions to GenBank. Dr. Wilhite is

presently assisting in the development and maintenance of Taxonomy LinkOut. In addition, he will assist in other special projects related to GenBank processing.

In August 2002, **Mr. Kenneth E. Thompson** joined the staff of the Computational Biology Branch, NCBI under the Pre-Doctoral, Intramural Research Training Award (IRTA) program. Mr. Thompson is currently a pursuing a Ph.D. in biology at Johns Hopkins University. He received his B.S. degree in biology from the University of South Carolina, Columbia. At NCBI, he will be conducting research on clustering algorithms based on 3D structure similarity, to link protein domain families identified previously by sequence similarity.

In August 2002, **Mr. Gelio Alves** joined the staff of the Computational Biology Branch, NCBI under the Pre-Doctoral, Intramural Research Training Award (IRTA) program. Mr. Alves is presently enrolled in the doctoral program in physics at Florida Atlantic University, Boca Raton. He also received his B.S. degree in biochemistry and physics from Florida Atlantic University. Mr. Alves will assist in researching problems relating to protein interactions. He will investigate to what extent sequence information can help to predict the protein-protein interactions and if it is possible to predict the interaction between two proteins.

In August 2002, **Timothy P. Doerr, Ph.D**. joined the staff of the Computational Biology Branch, NCBI under the Post-Doctoral, Intramural Research Training Award (IRTA) program. Dr. Doerr received his Ph.D. in physics from Case Western Reserve University, Cleveland, Ohio. Dr. Doerr's experience as a researcher has included studies of polymer, liquid crystal and polymer-liquid crystal systems by analytical methods and by computer simulations. At NCBI, Dr. Doerr is assisting in researching problems relating to biological sequence analysis. He will investigate the behavior of large polymer and liquid crystal systems and will utilize quantitative results obtained through molecular dynamic.

In September 2002, **Anjanette Johnston, Ph.D**. joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. She received her Ph.D. in Biology from the University of Virginia. Following graduate school, Dr. Johnston came to the NIDDK as a postdoctoral fellow. At NCBI, she is part of a rotation that is responsible for acting upon e-mail correspondence sent to GenBank. Dr. Johnston has gained extensive experience in the annotation of GenBank submissions and will continue to participate in all phases of processing submissions to GenBank.

In September 2002, **Mr. Aleksey Y. Ogurtsov** joined the staff of the Computational Biology Branch, NCBI as a Staff Scientist. Mr. Ogurtsov received his M.S. degree in mathematics with emphasis on computer science from Lomonosov Moscow State University. Mr. Ogurtsov has extensive experience in creating and applying software for biological research. He was responsible for creating software for producing and analyzing genome alignments. A software tool for interactive hierarchical alignment, OWEN, developed by Mr. Ogurtsov, is used in several laboratories throughout the USA. His skills and experience will be valuable for comparative analysis of multiple eukaryotic genomes.

In September 2002, **Ms. Carolyn Tilley** assumed the position of advisor on activities of the Unified Medical Language System (UMLS) in the office of the Chief, Bibliographic Services Division, Library Operations. Ms. Tilley will provide coordination of administrative functions, data distribution, training, user support and documentation, and expansion of access to the Unified Medical Language System. Prior to her reassignment, Ms. Tilley was Head of the MEDLARS Management Section, BSD, LO.

Retirements and Separations

In October 2001, **Ms. Elizabeth A. Pope** resigned from her position of Staff Scientist with the Information Engineering Branch, NCBI. While at NCBI, she assumed a lead role in the PubMed Central project, NIH's repository for primary research reports in the life sciences. She was the primary contact for publishers, scientific societies and other organizations submitting data and research reports to PubMed Central. Ms. Pope accepted a position as Director of Electronic Publishing with WebMD in New York.

In January 2002, **Ms. Dianne McCutcheon**, Head of the Serial Records Section, Technical Services Division, Division of Library Operations, left the NLM after more than 21 years to join the Library of Congress. As Section Head, Ms. McCutcheon was responsible for the acquisition, initial bibliographic control and processing of over 22,000 print, audiovisual and electronic biomedical and related serials; the maintenance of SERHOLD, the serial holdings database; the control of serials data about publications which the NLM indexes; and the development and maintenance of automated systems to support these activities.

In January 2002, **Ms. Carol Unger**, Assistant Head, Preservation and Collection Management Section, left to serve as a Digital Projects Coordinator for the Library of Congress. Ms. Unger worked at NLM for 21 years and served as Assistant Head, Preservation and Collection Section since 1986. Ms. Unger played a crucial role in developing and managing all aspects of the preservation program, most notably microfilming, binding, and automation of records and workflow. During the past year, she completed studies of the amount of acid paper in the NLM collection and the percentage of materials currently being received on alkaline paper.

In April 2002, **Donald C. Poppke**, NLM Associate Director for Administrative Management, left to serve as the NIH Budget Officer. Over the past 6 years, Mr. Poppke superbly managed all facets of the Office of Administration. The administrative requirements of an organization as diverse as the NLM have been considerable and he handled them admirably. Mr. Poppke

demonstrably raised the quality of the Library's administrative services which include the office of human resource management, budget, and acquisitions. The success of NLM's diverse administrative duties owes much to his innovative leadership and perseverance.

In April 2002, **Andrew S. Trotman** resigned from his position as Staff Scientist at the Information Engineering Branch, NCBI. Mr. Trotman came to NCBI in 1999 bringing with him commercial experience in electronic publishing. Mr. Trotman and his development team created tools that made it possible to make available online through the Web (and for free) the full text of biomedical journal articles. Mr. Trotman will be returning to New Zealand to continue his research.

In July 2002, **Dennis E. Black, Ph.D**. retired from his position as Contracts Officer and the Federal government after 31 years of service. Dr. Black received his Ph.D. in Public Administration from American University. He joined NIH in 1977 as a Contract Specialist and he came to NLM in 1982 as a Supervisory Contract Specialist with the Office of Contracts Management now the Office of Acquisitions Management. Dr. Black's responsibilities included overseeing the special acquisition authorities designed to streamline the procurement process within NLM as a result of OAM's designation as a Reinvention Laboratory and as a "Competitive Service Center" with added responsibilities for providing acquisitions support to organizations outside of NLM.

In August 2002, **Joseph W. Hutchins** retired from his position as Supervisory Computer Specialist and the Federal government after 32 years of service. Mr. Hutchins served as Chief, Applications Branch, OCCS. Mr. Hutchins was a key designer and developer of many of NLM's principal legacy systems. He had Project Manager responsibility for the development and implementation of NLM's reinvention efforts to replace all of NLM's legacy systems. This included the Integrated Library System, Interlibrary Loan System, Publications System, and Database Creation and Maintenance System. He provided NLM management with critical advice on a wide variety of technical issues.

Awards

The NLM Board of Regents Award for Scholarship or Technical Achievement was awarded to Mr. Rodney Long for the design and development of the Web-based Medical Information Retrieval System (WebMIRS).

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 2002 award was Ms. Karen A. Kraly (OCCS) in recognition of continuing innovative and substantial contributions to NLM's DOCLINE System.

The NIH Director's Award was presented to Ms. Kathleen G. Cravedi, Office of the Director (OD) for her contributions to NLM's outreach program that have enhanced the Library's reputation for service to the public.

The NLM Director's Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to three employees: Joseph W. Hutchins (OCCS) in recognition of exceptional vision, planning, and implementation of the NLM's System Reinvention Initiative; Mr. Donald C. Poppke (Office of Administration) in recognition of exemplary management and leadership of the NLM's administrative programs and services; and Ms. Naomi Miller (LO) for outstanding management of the selection and organization of MEDLINEplus content, including the definition and maintenance of its high standards for quality, authority, and accuracy.

The NIH Merit Award was presented to five employees: Ms. Roma P. Samuel (LO) for dedicated support in locating materials and providing reference assistance for researchers and users of the NLM Staff Library; Ms. Catherine R. Selden (LO) for unique and significant contributions to the NLM's Health Services Research Information Program; Mr. German E. Tello (LO) for outstanding productivity and accuracy as an indexer, reviser, and trainer of new indexers for MEDLINE; Ms. Janice H. Willis (LO) for superior, long-term contributions to the accuracy and quality of the MEDLINE database through extensive examination and intensive testing of data; Dr. Terry S. Yoo (LHNCBC) for continuing support and leadership of the Visible Human Insight Toolkit Project; and, as a group, Mr. David L. Nash (OD), Dr. Elliot R. Siegel (OD), and Dr. Frederick B. Wood (OD) in appreciation for their efforts in addressing health disparities and recruitment concerns through the successful implementation of the NIH Native American Powwow Outreach Initiative.

The Philip C. Coleman Award recognizes significant contributions to the NLM by individuals who demonstrate outstanding ability to motivate colleagues. The recipient of the 2002 award was Ms. Julia C. Royall (OD) for outstanding support to improve accessibility for disabled individuals.

The NLM EEO Special Achievement Award was presented to Dr. Colette Hochstein (SIS) for her outstanding role in conducting Computer Support Coordinators meetings in a manner that supports the diverse needs of its members.

The Sidney M. Edelstein Award was presented by the History of Medicine Division, American Chemical Society, to Dr. John L. Parascandola (LHNCBC) for his outstanding achievement in the history of chemistry, as evidenced by scholarly publications on biological and medical chemistry.

The prestigious 2002–2003 Kelly Lectureship prize, given to outstanding scientists and engineers by the Department of Chemistry and the School of Chemical Engineering, Purdue University, was awarded to Dr. Michael Y. Galperin (NCBI).

The National Medical Association presented a plaque to Ms. Cassandra R. Allen (SIS) for her outstanding service and support.

TABLE 15

FY 2002 Full-Time Equivalents (Actual)

Office of the Director	13
Office of Health Information Programs Development	7
Office of Communication and Public Liaison	8
Office of Administration	58
Office of Computer and Communications Systems	55
Extramural Programs	18
Lister Hill National Center for Biomedical Communications	79
National Center for Biotechnology	117
Specialized Information Services	31
Library Operations	298
Total FTEs	**684**

NLM Diversity Council

The NLM Diversity Council began the year by welcoming four new members: Kathleen Cravedi, Felicia Derricott, Ihsia Hu, and Michael Simpson. Each will serve a two-year term from January 2002 through December 2003. Continuing on the Council are Vivian Auld, Carole Brown, Tamar Clarke, Kimberlee Ford, James Knoben, Dawn Lipshultz, and Marta Melendez. The Council continues to receive support from its ex-officio members—Donald Poppke followed by Jane Griffith and subsequently Jon Retzlaff from the Executive Office, David Nash from the Equal Employment Opportunity Office, and Nadgy Roey from the Office of Human Resources, as well as its distinguished alumni. Vivian Auld accepted the responsibilities of Council Chair and Tamar Clarke became the Council Vice-Chair.

FY2002 Accomplishments:

NLM Director's Employee Education Fund: Continued coordination of the NLM Director's Employee Education Fund. In FY2002, the Fund enabled 57 staff to take 88 classes from 14 area schools. This is up from 49 staff taking 59 classes in FY2001. Staff who have taken advantage of the Fund represent 60% from the Division of Library Operations, 28% from the Office of the Director, 7% from the Office of Computer and Communications Systems, and 5% from the National Center for Biotechnology Information. Undergraduate classes made up the majority of classes supported. The school with the largest number of NLM enrollees is the University of Maryland (44%) with Montgomery College coming in second (29%). Other institutions being attended are the American University, University of the District of Columbia, Johns Hopkins University, George Mason University, Shepard College, Bowie State University, Prince George's Community College, Virginia Tech, Capital College, Coppin State College, Morgan State University, and Strayer University. Courses enrolled in included computer science, communications, business, English, environmental studies, law, history, Spanish, math, science, psychology, art, and library science. In addition to traditional classroom instruction some courses were taken on the Internet. The Diversity Council continues its effort to publicize the availability of the fund.

Facility Accessibility and Reasonable Accommodation: The Council continued efforts to upgrade access at NLM for people with disabilities. $33,000 of mid-year funds was added to the Diversity Council budget to address access issues. Council members met with the Chief of the Office of Administrative Management Services and the Chief of the Audiovisual Program Development Branch of the Lister Hill Center to discuss alternatives for spending these funds. These funds were allocated as follows:

- Two portable encoders for use in the Lister Hill Auditorium and Conference Room B. The encoders translate text produced by CART software for projection on television or LED displays. Portable encoders were selected so they can be used in other meeting rooms as needed.
- LED Caption Display for the Lister Hill Auditorium. This device provides scrolling LED display of CART and real-time captioning to be seen by everyone in a large meeting room.
- Wireless, portable FM listening system for the NLM Visitor's Center. This device provides hearing assistance for use during tours and meetings. The device will be available for use by other areas of the Library.
- Accessible water fountains in Building 38A. Four water fountains in Building 38A will be modified or replaced so they are accessible to those in wheel chairs. Two water fountains are on the B1 level, one is on the 1st floor by the elevators, and one is on the 4th floor.

Communication of NLM Diversity: The Diversity Council again 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.

English Language Courses: The Council began work on a new program to enable NLM employees to improve their linguistic proficiency speaking and writing English. Following the model used by local literacy programs, the NLM program would entail one-on-one

tutoring with NLM staff serving as tutors whenever possible. Having received approval from Dr. Lindberg the Council will poll staff for interest by January 2003 and begin the program soon thereafter.

National Minority Health Month: In celebration of National Minority Health Month, the Diversity Council and the NLM Office of Equal Opportunity sponsored a presentation by Beverly Coleman-Miller, M.D. entitled "National Minority Health Month: A Historical Perspective and Current Efforts to Alleviate Health Disparities."

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
(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:

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

CARTER, Ernest L., M.D.
Director, Telehealth Sciences
Howard University
Washington, D.C.

CONERLY SR., A. Wallace, M.D.
Dean, University of Mississippi
School of Medicine
Jackson, MS

DEAN, Richard H., M.D.
President, Wake Forest University
Health Sciences
Winston-Salem, NC

DETRE, Thomas, M.D.
Distinguished Service Prof. of Health Sciences
University of Pittsburgh
Pittsburgh, PA

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

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

STEAD, William W., M.D.
Professor of Biomedical Informatics
Vanderbilt University
Nashville, TN

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:

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

CHEN, Hsinchun, Ph.D.
Professor of Management Information Systems
University of Arizona
Tucson, AZ

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

FULLER, Sherrilynne S., Ph.D.
Professor of Biomedical and Health Informatics
University of Washington School of Medicine

GIUSE, Nunzia B., M.D.
Associate Professor of Biomedical Informatics
Vanderbilt University
Nashville, TN

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 NCBI Board of Scientific Counselors meets periodically to review and make recommendations on the NLM's biotechnology-related programs.

Members:

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

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

DELISI, Charles, Ph.D.
Director, Graduate Program in Bioinformatics
College of Engineering
Boston University
Boston, MA

FIRE, Andrew Z., Ph.D.
Staff Scientist
Department of Embryology
Carnegie Institution
Baltimore, MD

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

MAYO, Stephen L., Ph.D.
Associate Prof. of Biology & Chemistry
California Institute of Technology
Pasadena, CA

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:

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

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

BALAS, Andrew, M.D., Ph.D.
Dean, School of Public Health
Saint Louis University
Sr. Louis, MO

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

CAMPBELL, James R., M.D.
Professor of Internal Medicine
University of Nebraska Medical Center
Omaha, NE

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

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.
Associate Professor
Department of Medical Informatics
Columbia University
New York, NY

JENKINS, Carol G., M.L.S.
Director, Health Sciences Library
University of North Carolina
Chapel Hill, NC

KAZIC, Toni, Ph.D.
Associate Professor of Computer Engineering
University of Missouri-Columbia
Columbia, MO

KOHANE, Isaac S., M.D., Ph.D.
Associate Professor
Department of Medicine
Children's Hospital
Boston, MA

MCGOWAN, Julie J., Ph.D.
Associate Dean, Information Resources
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

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

SAHNI, Sartaj K., Ph.D.
Distinguished Professor and Chair
Computer Science and Engineering
University of Florida
Gainesville, FL

SHAVLIK, Jude W., Ph.D.
Professor of Computer Science
University of Wisconsin-Madison
Madison, WI

Silverstein, Jonathan C., M.D.
Assistant Professor of Surgery
University of Chicago
Chicago, IL

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

WONG, Stephen T.C., Ph.D.
Assistant Professor
Department of Radiology and Neurology
University of California, Scan Francisco
San Francisco, CA

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:

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

BOROVETZ, Harvey S., Ph.D.
Professor, Dept. of Bioengineering and Surgery
Center for Biotechnology and Bioengineering
University of Pittsburgh
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.
Professor of General Dentistry
Director, Oral Biology Division
Tufts University School of Dental Medicine
Boston, MA

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

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

MCCLURE, Lucretia W., M.A.
Special Assistant to the Director
Countway Library of Medicine
Harvard University
Boston, MA

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

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

SIEGEL, Vivian, Ph.D.
Editor, Cell
Cell Press
Cambridge, MA

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

TOM-ORME, Lillian, Ph.D.
Research Assistant Professor
Dept. of Family and Preventive Medicine
University of Utah
Salt Lake City, UT

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

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

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

GINSPARG, Paul, Ph.D.
Professor of Physics and Computer Science
Cornell University
Ithaca, NY

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

JOSEPH, Heather D., M.A.
President and CEO
BIOONE
Washington, D.C.

KAPLAN, Samuel, Ph.D.
Professor and Chair
Microbiology and Molecular Genetics
University of Texas Health Science Ctr.
Houston Medical School
Houston, TX

KAUFMAN, Paula T., M.B.A.
University Librarian
University of Illinois at Urbana-Champaign
Urbana, IL

KHOSLA, Chaitan S., Ph.D.
Prof. of Chemistry & Chemical Engineering
Stanford University
Stanford, CA

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

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

WATSON, Linda A., M.L.S.
Dir., Claude Moore Health Science Lib.
University of Virginia
Charlottesville, VA

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

APPENDIX 8: ACRONYMS AND INITIALISMS USED IN THIS REPORT

AAHSL	**Association of Academic Health Science Librarians**
ACTIS	AIDS Clinical Trials Information Service
ALA	American Library Association
ALTBIB	Alternatives to the Use of Live Vertebrates in Biomedical Research and Testing
AMIA	American Medical Informatics Association
AMPA	American Medical Publishers Association
API	Application Programming Interface
ARL	Association of Research Libraries
ASHP	American Society of Health-System Pharmacists
ATIS	HIV/AIDS Treatment Information Ser
BISTI	**Biomedical Information Science and Technology Initiative**
BLAST	Basic Local Alignment Search Tool
BLRC	Biomedical Library Review Committee
BSD	Bibliographic Services Division
BSN	Bioinformatics Support Network
CAS	**Collection Access Section**
CBIR	Content-Based Image Retrieval
CCRIS	Chemical Carcinogenesis Research Information System
CDC	Centers for Disease Control and Prevention
CDD	Conserved Domain Database
CENIDS	Centro Nacional de Informacion y Documentacion sobre Salud
ChemID*plus*	Chemical Identification File
CISTI	Canada Institute for Scientific and Technical Information
CIT	Center for Information Technology
CRID	Regional Disaster Information Center for Latin America and the Caribbean
DART	**Developmental and Reproductive Toxicology**
DbEST	EST database
DCMS	Data Creation and Maintenance System
DDBJ	DNA Data Bank of Japan
DIMDI	German Institute for Medical Documentation and Information
DIRLINE	Directory of Information Resources Online
DSL	Digital Subscriber Line
EBI	**European Bioinformatics Institute**
ECRI	Emergency Care Research Institute
EEO	Equal Employment Opportunity
EFTS	Electronic Funds Transfer System
EMBL	European Molecular Biology Laboratory
EMIC	Environmental Mutagen Information Center
EP	Extramural Programs Division
EPA	Environmental Protection Agency
ESTs	Expressed Sequence Tags
ETICback	Environmental Teratology Information Center backfile
FAQs	**Frequently Asked Questions**
FDA	Food and Drug Administration
Gbps	**Gigabits per Second**
GEO	Gene Expression Omnibus
GPRA	Government Performance and Results Act
GSS	Genome Survey Sequences
HCBU	**Historically Black Colleges and Universities**
HHS	Health and Human Services
HIPAA	Health Insurance Portability and Accountability Act
HLA	Human Leukocyte Antigen
HMD	History of Medicine Division
HPCC	High Performance Computing and Communications
HSDB	Hazardous Substances Data Bank
HSRR	Health Services/Sciences Research Resources
HSTAT	Health Services/Technology Assessment Text
HUD	Department of Housing and Urban Development
IADL	**Internet Access to Digital Libraries**
IAIMS	Integrated Advanced Information Management Systems
ICD-9	International Classification of Diseases
ICIPE	International Centre of Insect Physiology and Ecology
ICPC	International Classification of Primary Care
IGI	Infinite Global Infrastructures
ILL	Interlibrary Loan
IP	Internet Protocol
IRIS	Integrated Risk Information System
ISP	Internet Service Provider
IT	Information Technology
ITSC	IT Services Center
JD	**Journal Descriptor**
JST	Japan Science and Technology Corporation
KCMC	**Kilimanjaro Christian Medical Center**
KEMRI	Kenya Medical Research Institute
LAN	**Local Area Network**
LHC	Lister Hill Center
LHNCBC	Lister Hill National Center for Biomedical Communications
LO	Library Operations
LOINC	Logical Observations, Identifiers, Names, Codes
LSTRC	Literature Selection Technical Review Committee
MARS	**Medical Article Records System**
Mbps	Megabits per Second
MedDRA	Medical Dictionary for Regulatory Activities

MEDLARS	MEDical Literature Analysis and Retrieval System
MeSH	Medical Subject Headings
MGC	Mammalian Gene Collection
MIM	Multilateral Initiative on Malaria
MIME	Multipurpose Internet Mail Extensions
MLA	Medical Library Association
MLAA	Medical Library Assistance Act
MMDB	Molecular Modeling DataBase
MMTx	MetaMap Technology Transfer program
mRNA	messenger RNAs
MTI	Medical Text Indexer
MTI	Medical Text Indexer
MTMS	MESH Translation Maintenance System
NASA	**National Aeronautics and Space Administration**
NCAA	National Collegiate Athletic Association
NCBI	National Center for Biotechnology Information
NCI	National Cancer Institute
NHGRI	National Human Genome Research Institute
NCVHS	National Committee on Vital and Health Statistics
NDC	National Drug Codes
NGI	Next Generation Internet
NHAAP	National Heart Attack Alert Program
NHANES	National Heath and Nutrition Examination Surveys
NHLBI	National Heart, Lung, and Blood Institute
NIA	National Institute on Aging
NIAMS	National Institute of Arthritis and Musculoskeletal and Skin Diseases
NIBIB	National Institute of Biomedical Imaging and Bioengineering
NICHSR	National Information Center on Health Services Research and Health Care Technology
NIDCR	National Institute of Dental and Craniofacial Research
NIDDK	National Institute of Diabetes, Digestive, and Kidney Diseases
NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
NIMH	National Institute of Mental Health
NIMR	National Institute of Medical Research
NIOSH	National Institute for Occupational Safety and Health
NLM	National Library of Medicine
NMA	National Medical Association
NN\LM	National Network of Libraries of Medicine
NOMC	National Outreach Mapping Center
NOSC	Network Operations and Security Center
NSF	National Science Foundation
OAM	**Office of Acquisitions Management**
OCCS	Office of Computer and Communications Systems
OCHD	Coordinating Committee on Outreach, Consumer Health and Health Disparities
OCLC	Online Computer Library Center
OCR	Optical Character Recognition
OD	Office of the Director
OHIPD	Office of Health Information Programs Development
OMIM	Online Mendelian Inheritance in Man
PAHO	**Pan American Health Organization**
PDA	Personal Digital Assistant
PDB	Protein Data Bank
PEI	Patient Education Institute
PHS	Public Health Service
PLA	Public Library Association
PMC	PubMedCentral
PROW	Protein Reviews on the Web
PSD	Public Services Division
RCSB	**Research Collaboratory for Structural Bioinformatics**
RDES	Remote Data Entry System
RefSeq	Reference Sequence
RFA	Request for Applications
RMI	Remote Method Invocation
RML	Regional Medical Library
RTECS	Registry of Toxic Effects of Chemical Substances
SAGE	**Serial Analysis of Gene Expression**
SBIR	Small Business Innovation Research
SIS	Specialized Information Services
SKY/CGH	Spectral Karyotyping and Comparative Genomic Hybridization Database
SNPs	Single Nucleotide Polymorphisms
ST	Semantic Type
STS	Sequence Tagged Site
TDG	**Test Data Generator**
TEHIP	Toxicology and Environmental Health Information Program
TIOP	Toxicology Information Outreach Project
TOXLINE	Toxicology Information Online
TOXNET	Toxicology Data Network
TPA	Third Party Annotation
TRI	Toxics Release Inventory
TSD	Technical Services Division
TTP	Turning the Pages
UMLS	**Unified Medical Language System**
UNCFSP	United Negro College Fund Special Programs Corporation
UniGene	Unique Human Gene Sequence Collection
USAID	US Agency for International Development
USNIMR	US Naval Institute of Medical Research
USPDI	United States Pharmacopeia Drug Information
VA	**Department of Veterans Affairs**
VAMP	Vulnerability Assessment Management Program
VAST	Vector Alignment Search Tool
Vrep	Virtual customer service representative
VSAT	Very Small Aperture Terminal

WebMIRS	**Web-based Medical Information Retrieval System**
WGS	Whole Genome Shotgun
WRAIR	Walter Reed Army Institute for Research

NIH Publication No. 03-256