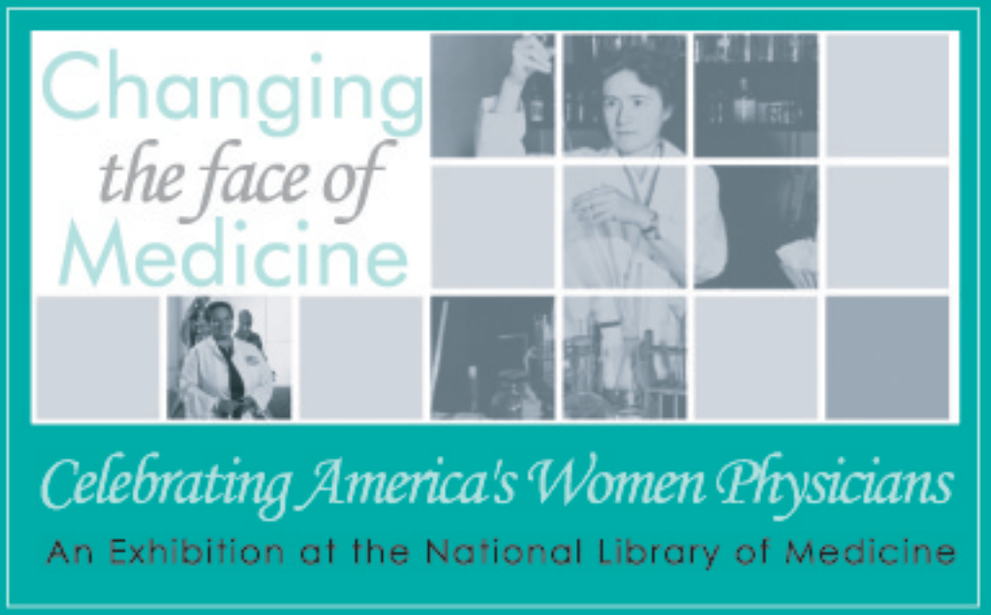
NATIONAL LIBRARY OF MEDICINE, PROGRAMS & SERVICES, FISCAL YEAR 2003

NATIONAL INSTITUTES OF HEALTH
NATIONAL LIBRARY OF MEDICINE
PROGRAMS & SERVICES FY 2003



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: "Changing the Face of Medicine," an exhibition at the NLM, honors the lives and achievements of American women in medicine

NATIONAL INSTITUTES OF HEALTH

NATIONAL LIBRARY OF MEDICINE

PROGRAMS AND SERVICES

FISCAL YEAR 2003

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

Tables

Appendixes

Preface

Fiscal Year 2003 saw important advances on several fronts. The National Library of Medicine is providing more services to a wider audience than ever before. Among the highlights of this year's *NLM Programs and Services*:

- MedlinePlus continues its amazing growth, both in the breadth of its content and the volume of Web usage. There are some 650 health topics. MedlinePlus en español has been expanded and now approaches the English MedlinePlus in richness. More than sixteen million unique visitors sought out MedlinePlus this year and there were some 215 million page hits. This year we launched the first "Go Local" site, tying MedlinePlus users to local health resources in the state of North Carolina. These are all described in the Library Operations chapter.
- NLM's National Center for Biotechnology Information logs 25 million hits per day.
- New Web-based information resources for the public have been introduced, including the Genetics Home Reference (Lister Hill Center chapter); and Tox Town, the Household Products Database, the Arctic and Asian American Health database (Specialized Information Services chapter).
- "Profiles in Science," the Lister Hill Center's popular educational Web site was enriched this year with the papers of Linus Pauling and Donald Fredrickson.
- The NLM exhibition program continues to expand. "Dream Anatomy" was a great success and an ambitious new exhibition, "Changing the Face of Medicine: Celebrating America's Women Physicians," will be launched early in FY2004 (Library Operations chapter).

In a move that may have far-reaching consequences, the NLM in FY 2003 arranged for a license for the use of the SNOMED CT systematized nomenclature by government agencies and the private sector. The vocabulary will be distributed within the NLM's UMLS Metathesaurus. The arrangement was made by NLM with the assistance of other federal agencies, and it was announced on July 1, 2003, by HHS Secretary Tommy G. Thompson. SNOMED CT is discussed in the Library Operations and Lister Hill Center chapters.

I trust this report demonstrates that the National Library of Medicine is accomplishing what it is mandated to do, that is, to be a worthy steward of the world's medical literature, in all its forms, and to make this knowledge available to all for the improvement of human health. To the extent we are succeeding, the credit goes to our fine staff and to our many consultants and advisors.

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

OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT

Elliot R. Siegel, Ph.D.
Associate Director

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

- 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;
- conducting NLM's international programs; and
- establishing, planning, and implementing the NLM Long Range Plan and related planning and analysis activities.

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.

NLM Coordinating Committee on Outreach, Consumer Health and Health Disparities

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. A number of the activities described below are conducted under the auspices of the OCHD.

American College of Physicians/American Society of Internal Medicine
Physician Information Prescription Project

Doctors often prescribe medication after seeing a patient. But what if that doctor also wants to direct the patient to up-to-date, reliable, consumer-friendly information about a health concern? The American College of Physicians Foundation teamed with the National Library of Medicine in FY2003 to create the "Health Information Prescription" program. Now, doctors in several states have been given customized prescription pads that they can use to point patients to first-rate online health information in NLM's MedlinePlus database.

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. The continuing 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 FY2003, the OCHD continued to pursue an integrated approach intended to encourage exchange of information and learning within NLM, and help better inform NLM management decision-making on Web site research, development, and implementation. The year'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 U.S. public's online and offline health information seeking behavior; analysis of NLM Web site log data; and access to Internet audience measurement estimates based on Web usage by user panels organized by private sector companies. The Committee and OHIPD continue to explore and test a range of internal and external Web evaluation methods and applications.

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) and Phase 2 (Pacific Southwest) of tribal connections are complete, with a final project evaluation published in Wood, et al., "Tribal Connections Health Information Outreach: Results, Evaluation, and Challenges," *Journal of the Medical Library Association*, Vol. 91, January 2003, pp. 57-66. 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. Phase 3 was completed in 2003, and an evaluation report is being prepared. NLM has funded a Phase 4, in collaboration with the University of Utah (Midcontinental Regional Medical Library), emphasizing the development of Web-based tribal health information resources in the Four Corners Region (AZ, CO, NM, UT).

Also in 2003, NLM/OHIPD again partnered with NIH and NLM Equal Employment Opportunity offices to participate in the NIH American Indian Pow-Wow Initiative. This included exhibiting at eight pow-wows in the Mid-Atlantic area. An estimated 8,000

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 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. The initial Lower Rio Grande Project is complete. Based on the project results, NLM has funded a follow-on project focusing on outreach to Hispanic populations in the Lower Rio Grande Valley who live in colonias. The follow-on project involves collaboration with Texas A&M University as well as the University of Texas at San Antonio.

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 on basic Internet skills during the first year, with the assistance of several dozen teens. The year two emphasis was on Cyber Health for Seniors. The Cyber Health Web-based curriculum development and initial training are now complete, and evaluation is under way.

International Programs

MIMCom: 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 World Wide 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. Using satellite technology, 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.

Three separate evaluations of the network have been conducted. In the course of these evaluations, African researchers have confirmed the importance of MIMCom. "We're not so far away anymore," said one researcher. "We're finally 'here.'"

A senior researcher from Cameroon, working at the remote ICIPE research site in Kenya, underscores and personalizes the importance of this connectivity tool. "Improved working relationships with our colleagues is perhaps the most important result of the network. Before, we were left out of the south-south network and the south-north network, because we didn't have access to information or a way to contact our partners. MIMCom has dramatically improved our working relationships with colleagues at sites in the south and north.. . . MIMCom assists developing countries to enter and participate fully in the information revolution. For us, it is a real support tool for sustainable development."

Another senior scientist at Kenya Medical Research Institute (KEMRI) in Kisian, Kenya, shares this sentiment: "It is fantastic because it removes those old barriers which were about controlling information, because information is power. Those who control information control the systems, so if you break those barriers, you can access resources. Moreso, you have much better access to information which you can use for policy formulation or designing projects or gathering data, presenting findings, packaging it for policy makers in our country or elsewhere…. We manage projects, some set in Maryland, some set in U.K. We run projects in Africa. We forward mail to each other. We plan and agree and disagree. It is not one man writing a letter, giving instructions. There is a difference here. It's a completely different way of communicating."

The Web site, http://www.nlm.nih.gov/mimcom, has 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.

MIMCom Partnerships

Site	Local Partners	International Partners	Status
Cameroon	The Biotechnology Centre, Faculty of Medicine and Biomedical Sciences, University of Yaounde 1 (Yaounde)	US National Library of Medicine US National Institute of Allergy and Infectious Diseases/ National Institutes of Health	Operational
Gabon	Medical Research Unit, Albert Schweitzer Hospital (Lamebrains)	US National Library of Medicine US National Institute of Allergy and Infectious Diseases/ National Institutes of Health	Operational
Ghana	Noguchi Memorial Institute (Accra) Navrongo Health Research Center (Navrongo)	US National Library of Medicine US National Institute of Allergy and Infectious Diseases/ National Institutes of Health Naval Institute of Medical Research US Agency for International Development	Operational
Kenya	Kenyan Medical Research Institute(KEMRI) (Nairobi) KEMRI/ Wellcome Trust (Kilifi) KEMRI/CDC (Kisian) International Center of Insect Physiology and Ecology (ICIPE) (Mbita)	US National Library of Medicine US Walter Reed Army Institute of Research US Centers for Disease Control US National Library of Medicine Wellcome Trust (U.K.) US National Library of Medicine US Centers for Disease Control US National Library of Medicine US National Institute of Allergy and Infectious Diseases/ National Institutes of Health	Operational
Mali	Malaria Research and Training Center, Faculte de Medecine (Bamako)	US National Library of Medicine US National Institute of Allergy and Infectious Diseases/ National Institutes of Health (Initial support for a microwave connection which led to installation of an independent VSAT system.)	
Malawi	College of Medicine/Pediatric Malaria Project/ Wellcome Trust	US National Library of Medicine UK Wellcome Trust US National Institute of Allergy and Infectious Diseases/ National Institutes of Health	Operational
Tanzania	National Institute for Medical Research (NIMR) Amani Medical Research Center Ifakara Health Research and Development	US National Library of Medicine US National Institutes of Health/Office of the Director	Operational

	Center		
Uganda	Makerere University/ Mulago Hospital Ugandan Viral Research Institute	US National Library of Medicine US National Institute of Allergy and Infectious Diseases/ National Institutes of Health US Centers for Disease Control	Operational

International Network Partnerships

OHIPD is pursuing strategies to develop international network partnerships. One initial area for exploration is international DOCLINE. In FY2003, letters of invitation to join DOCLINE were sent to selected libraries in Mexico, following the 1.5 release of DOCLINE, which added Region 21 for Mexico. The NN/LM South Central Region, housed at the Houston Academy of Medicine–Texas Medical Center Library, is serving as Region 21's Regional Medical Library in its initial phases. A number of Mexican libraries have joined and they now add holdings to SERHOLD, enabling them to share resources among themselves and border libraries in Texas and other U.S. libraries agreeing to reciprocal borrowing with Mexico.

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 2003, NLM completed a collaborative project with the University Corporation for Advanced Internet Development/Internet2 to conduct research on "critical incidents" where higher bandwidth Internet connectivity has made or could make a significant difference for biomedical and health applications. The final project report is in preparation. Additionally, NLM has implemented Phase I of its own Internet connectivity performance monitoring network, starting with select U.S. sites (the eight Regional Medical Libraries) but envisioned to extend to other U.S. sites and some international sites in the medium term.

International MEDLARS Centers

Bilateral agreements between the Library and more than 20 public institutions in foreign countries (below) 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.

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 FY2003 the Office of Communications and Public Liaison (and HMD) arranged for 221 tours—93 regular daily (1:30 p.m.) tours and 128 specially arranged tours. In addition, the History of Medicine Division arranged 43 tours of the "Dream Anatomy." There were 5,703 visitors in all. They came from the following 46 countries:

Armenia, Australia, Austria, Azerbaijan, Bahamas, Barbados, Bolivia, Brazil, Canada, Chile, China, Costa Rica, Cuba, Czech Republic, Ecuador, England, Finland, France, Georgia, Germany, Greece, Guatemala, Hungary, India, Ireland, Israel, Japan, Jordan, Kazakhstan, Korea, Kyrgyzstan, Malaysia, Mexico, The Netherlands, Poland, Russia, Spain, St. Kitts, South Africa, Switzerland, Taiwan, Ukraine, United States, Uzbekistan, Venezuela, Wales.

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.

Based on the Long Range Plan, OHIPD documents NLM's progress in achieving its goals for a variety of purposes, including the Government Performance and Results Act (GPRA) and appropriations hearings, as well as NLM's involvement in a variety of disease and policy-related areas.

The OHIPD also 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.

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 OCHD is the implementation of a database of NLM outreach, consumer health, and health disparities projects. This database, which includes projects from all of the Regional Medical Libraries as well as NLM, is 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.

Library Operations

Betsy L. Humphreys
Associate Director

The Library Operations (LO) Division is responsible for the basic services that ensure access to the published record of biomedical science and the health professions. LO acquires, organizes, and preserves NLM's comprehensive collection of biomedical literature; creates and disseminates controlled vocabularies and a library classification scheme; produces authoritative indexing and cataloging records; builds and distributes bibliographic, directory, and full-text databases; provides back-up document delivery, reference service, and research assistance for the nation; helps varied user groups to make effective use of NLM products and services; and coordinates the National Network of Libraries of Medicine to improve access to health information services across the United States. These services provide an essential foundation for NLM's outreach programs to health professionals and the general public. They also support the Library's focused programs in AIDS, health services research, molecular biology, and toxicology and environmental health.

In addition to its basic services, LO develops and mounts major historical exhibitions; carries out an active research program in the history of medicine; works with other NLM program areas to enhance NLM products and services; conducts research related to current operations and services; directs and sponsors training programs for health sciences librarians; and contributes to the development of national health data standards policy and to the production and dissemination of clinical vocabulary standards.

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 range of contractors. LO is organized into four major Divisions: Bibliographic Services, Public Services, Technical Services, and History of Medicine; three units: the Medical Subject Headings (MeSH) Section, the National Network of Libraries of Medicine Office, and the National Center on Health Services Research and Health Care Technology (NICHSR); and a small administrative office. LO staff members participate actively in efforts to improve the quality of work-life at NLM, including the Diversity Council and the NLM Intranet.

Program Planning and Management

LO sets its priorities in accordance with 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. Many of the program plans outlined in these documents focus on the opportunities and challenges arising from electronic publishing. In FY2003, LO continued to review and revise policies and procedures for processing electronic publications, to explore additional ways to use electronic information to enhance basic operations and services, and to work with other NLM program areas to expand PubMed Central as a permanent archive for electronic journals. Specific projects undertaken to improve access to—and handling of—electronic information are described throughout this chapter.

In the current economic environment, many health sciences libraries across the country are cutting print subscriptions to journals in favor of providing access to the electronic versions that are more useful to their primary clienteles. In FY2003, LO and the Regional Medical Libraries examined the impacts of this trend, of the increasing amount of electronic full-text that is available free, and of the national maximum interlibrary loan charge on resource sharing within the NN/LM and on access to documents by unaffiliated health professionals. These background studies will assist in determining whether changes are needed in NLM and Network resource sharing policies.

The NLM Long Range Plan for 2000–2005 also emphasizes the enhancement of NLM services directed toward patients, their families, and the general public and expanded outreach to the public and to health professionals, with a particular emphasis on minority populations and the under-served. LO's contributions to the development, assessment, and improvement of MedlinePlus and other services directed toward the public are described throughout this chapter. In FY2003, LO and the Regional Medical Libraries put in place specific national plans for expanded NN/LM outreach to public libraries and public health departments. LO contributed to the development of an NLM-wide database for tracking outreach activities and helped to direct the MedlinePlus Information Rx pilot project in collaboration with the American College of Physicians. This project encourages physicians and their office staffs to provide patients with "prescriptions" for information available in MedlinePlus.

Although many of its efforts are directed toward creating and promoting use of electronic information resources, LO also devotes substantial resources and attention to the care and handling of NLM's extensive collections of physical library materials and to the space and environment in which staff and patrons work and collections are stored. In FY2003, LO continued to assist with plans for the new NLM building and for the accompanying major renovations of the existing NLM facilities. LO took the lead for NLM in collaborating with the Association of Academic Health Sciences Libraries (AAHSL) to sponsor a two-day symposium on "The Library as Place" to be held at NLM in early FY2004. Meanwhile, LO continued to make improvements to

conditions in the existing NLM library building and to develop strategies for handling the projected growth of the collections until the new facility becomes available.

FY2003 saw a number of changes in the organizational units and automated systems that handle personnel and administrative functions within the National Institutes of Health and the Department of Health and Human Services. These changes presented significant challenges to LO's Administrative Office, which worked effectively to minimize the negative impact on LO operations and services. LO continues to encourage its staff to take advantage of flexiplace work arrangements as appropriate. More than 60 LO employees work at home at least one day per week.

Collection Development and Management

NLM's comprehensive collection of biomedical literature provides the essential underpinning for many of the Library's services. LO ensures that this collection meets the needs of current and future users by updating NLM's literature selection policy; acquiring and processing literature in all languages and formats that meets the selection guidelines; organizing and maintaining the collection to support current use; and preserving it for future generations. At the close of FY2003, the NLM collection contained 2.46 million volumes and 5 million other physical items, including manuscripts, microforms, pictures, audiovisuals, and electronic media.

Selection

The *Collection Development Manual of the National Library of Medicine* guides LO staff and their agents in selecting literature for the NLM collection. The *Manual* is currently undergoing a major review and revision, as it does every 5 to 10 years. This process is guided by an external Oversight Committee consisting of researchers, librarians, practicing health professionals, a consumer health information expert, and senior members of the NLM staff. The Committee is chaired by Alison Bunting, immediate past Chair of the NLM Board of Regents.

During FY2003, the Board established two Working Groups of external experts to review NLM's coverage—in its collection, the MeSH vocabulary, and its databases—in two specialized subject areas. Dr. Thomas Detre, a current Regent, chaired a group charged with reviewing coverage and services in the area of bioethics. For nearly 30 years, NLM has provided funding to the Kennedy Institute of Ethics to complement NLM's own bioethics coverage by indexing and acquiring bioethics materials published outside the core literature of biomedicine and health. The working group is reviewing whether the combined efforts of NLM and the Kennedy Institute are meeting the needs of users of the literature in this multidisciplinary field. Dr. Morris Collen, immediate past chair of the NLM Literature Selection Technical Review Committee which reviews and recommends journals for inclusion in PubMed/MEDLINE, chaired the second group which reviewed NLM's coverage to assess its ability to support the programs of the newly established National Institute of Biomedical Imaging and Bioengineering. The findings of the working groups will provide input to the overall revision of the *Collection Development Manual.*

Dr. Preston Reynolds, visiting historical scholar, assessed NLM's coverage of materials related to the history of African Americans in medicine, produced a guide to the available resources, and made recommendations for acquiring additional manuscripts and pictures. Dr. Reynolds found that NLM's collection of contemporary works on this subject is quite comprehensive. She located some interesting manuscript and picture collections that had not been previously identified as containing materials related to African Americans. HMD is upgrading the records for these collections as a result.

Acquisitions

TSD received and processed 159,102 contemporary physical items (books, serial issues, audiovisuals, electronic media). This is about 2 percent more than in FY2002, indicating that the increase in electronic biomedical publishing has not yet led to a noticeable decline in the number of new physical items NLM acquires. A net total of 30,096 volumes and 442,168 other items (including nonprint media and manuscripts and pictures acquired by HMD) were added to the NLM collection. LO uses several agents and vendors to acquire current literature published around the world. In FY2003, the blanket purchase order arrangements for contemporary monographs and monographic series were recompeted. TSD arranged for its serials subscription agent to provide consolidated shipping of serials issues from the United States and Canada, improved procedures for claiming missing issues of indexed journals, and studied the timeliness of book receipts from U.S. vendors. Considerable time and effort was expended in negotiating with HHS legal counsel and NIH procurement officials to achieve acceptable procedures for negotiation and review of NLM licenses for electronic publications.

HMD acquired many splendid early printed books, manuscripts, images, and historical films for the NLM collection in FY2003. Important books acquired included: Bartolomeo de' Sacchi's *D'la Hoesta Voluptate & Valitudin* (1487); *De Prandii ac Caenae Libellus* by Matteo Cortii (Rome, 1562), a book about eating for health which draws on the works of Hippocrates and Galen; Marc Antonio Madero's *Apologia pro Sanguinis Circuitione Noviter Sufflaminata ab Homobono Pisone* (Venice, 1698), a defense of William Harvey's theory of circulation; *Memoria Chirurgica sul Labbro Leporino* by Giuseppe Sonsis (Cremona, 1793), a work on reconstructive

surgery that includes a description of an operation for cleft lip; and two works by disciples of Paracelsus, Gerhard Dorn's *Dictionarium Theophrasti Paracelsi* (Frankfurt, 1584) and *Paracelsus His Aurora, Describing the Matter of, and Manner how to attain the Universal Tincture & Treasures of Philosophers* (London, 1689).

Additions to the manuscript collections included the papers of Winifred Sewell, often referred to as the "mother of MeSH"; the papers of James Bosma, a specialist in swallowing disorders; the Adrian Kantrowitz papers from his laboratory in Detroit; the Howard Bishop papers; portions of the Murray Bowen papers; several additions to the C. Everett Koop papers; and the last of the John Eisenberg papers. William Helfand continued his generosity to the Library by donating more than 450 items to the prints and photographs collection: 235 prints (including a Rembrandt), 100 posters, and many images related to women in medicine. Other important additions to the picture collection included a fine print "La Malade Imaginaire," numerous posters from the 1910s to the 1930s, 88 black and white photographs by Martha Tabor of contemporary medical scenes, and 19 posters dealing with pregnancy donated by John Parascandola. HMD acquired films on consumer protection and rat eradication from the Food and Drug Administration and a U.S. Public Health Service film about C. Everett Koop and the Commissioned Corps.

During FY2003, HMD revived the NLM archives program by beginning an effort to schedule and acquire LO records at the Associate Director and Division chief levels.

Preservation and Collection Management

LO undertakes a broad range of activities to preserve NLM's archival collection and keep it readily accessible for use. These activities include: binding, microfilming, conservation of rare and unique materials, book repair, maintenance of appropriate storage and environmental conditions, and disaster prevention and response. LO distributes data about what NLM has preserved to avoid duplicate effort by other libraries. LO works with other NLM program areas to conduct experiments with new preservation techniques as warranted and to promote the use of more permanent media and archival-friendly formats in new biomedical publications.

In FY2003, LO bound 15,646 volumes, microfilmed 2,795 volumes, repaired 1,285 items in NLM's onsite repair and conservation laboratory, made 500 preservation copies of motion pictures, and conserved 111 rare items. A total of 786,856 items were shelved or re-shelved and 20,357 duplicate journal issues were removed from the collection. In FY2003, LO set up an onsite audiovisual inspection laboratory that will begin to inspect the condition of audiovisuals in the NLM collection in FY2004. In the fourth quarter of FY2003, LO reduced the rate of preservation microfilming, pending an early FY2004 review of preservation priorities. Since 1985, NLM has microfilmed close to 100,000 brittle serials volumes and books, with an emphasis on titles covered by *Index Medicus* or the Surgeon General's *Index Catalogue*. The amount of brittle paper in the NLM collection is still substantial, but the Library will be reassessing the allocation of preservation funds among various activities and categories of materials, including historical motion pictures.

After investigating several options, NLM replaced the CO_2 fire suppression system in the incunabula room and the rare book stacks with a water fire suppression system, coupled with advanced fire detection sensors. LO reviewed the new binding module of the Voyager Integrated Library System and found it to be unsuitable for use by NLM. As a result, the Library will continue to use and enhance an internally developed system for managing the binding operation.

Permanent Access to Electronic Information

The preservation of electronic information presents unique challenges that are not yet fully understood. 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 other national libraries, the National Archives and Records Administration, and other interested organizations to develop, test, and implement strategies and standards for ensuring permanent access to electronic information. LO works closely with other NLM program areas on activities related to the preservation of digital materials.

PubMed Central, a digital archive of medical and life sciences literature developed by the National Center for Biotechnology Information (NCBI), is NLM's primary test-bed for the development of procedures and methods for ensuring permanent access to electronic journals. In FY2003, LO assisted NCBI in expanding current deposits to PubMed Central by soliciting the participation of additional journals, primarily in the fields of clinical medicine, health policy, health services research, and public health. LO organized a joint NLM-American Medical Publishers Association (AMPA) seminar on digital archiving, held in conjunction with AMPA's annual meeting in March 2003, in which PubMed Central was featured. This was one of several methods LO used to help to publicize NCBI's new Journal Publishing Document Type Definition (DTD) and the Archiving and Interchange DTD as potential standards for the publishing, library and information science, and informatics communities.

Throughout FY2003, the Public Services Division worked closely with NCBI on the major project to scan and add to PubMed Central the complete backfiles of journals depositing current issues in the digital archive. PSD had principal responsibility for assembling more than 2.5 million pages to be scanned, collating and shipping them to the scanning contractor,

and managing aspects of the quality review of the scanned images, accompanying OCR data, and the XML-tagged citations for articles that predate current MEDLINE/PubMed coverage. Since bindings are cut to make scanning more efficient, NLM is not using volumes from its archival collection in this project, but has solicited copies from other libraries and the publishers. LO is particularly grateful to the Medical Library Center of New York, the Medical Library Association, and the National Agricultural Library for donating issues for the scanning effort. The initial group of scanned back issues will be added to PubMed Central in early FY2004.

NLM is using its own publications and main Web site as a test-bed for procedures and mechanisms for ensuring permanent access to electronic information published by government agencies and private nonprofit institutions. In FY2003, LO and OCCS evaluated *DSpace*, software developed by the Massachusetts Institute of Technology for institutional repositories, and OCLC's Digital Archive but found them unsuitable for providing permanent access to NLM Web publications that are of historical interest, but are no longer currently applicable. At year's end, LO staff had begun to use the TeamSite Web management software to assign NLM standard metadata, including permanence ratings for documents on the NLM main Web site. The permanence ratings will provide a prospective system for labeling NLM Web documents that will become part of the permanent NLM archives.

Vocabulary Development and Standards

LO produces and maintains the Medical Subject Headings (MeSH®), a subject thesaurus used by NLM and many other institutions to describe the subject content of the biomedical literature and other types of biomedical information; develops, supports, or licenses for the U.S. vocabularies designed for use in patient records and clinical decision support systems; and works with the Lister Hill Center to produce Unified Medical Language System® (UMLS®) Metathesaurus®, a large vocabulary database that incorporates many vocabularies, including MeSH and other vocabularies produced or supported by NLM. A multi-purpose knowledge source used in operational systems and informatics research, the Metathesaurus also serves as a common distribution vehicle for classifications, code sets, and vocabularies designated as standards for U.S. health data.

LO represents NLM in federal initiatives to select standard vocabularies and other health data standards for use in patient record information and in administrative transactions governed by the Health Insurance Portability and Accountability Act (HIPAA). In this capacity, LO serves on the Department of Health and Human Services Data Standards Committee, on the staff of the National Committee on Vital and Health Statistics (NCVHS) Standards and Security Subcommittee, and on the Public Health Data Standards Coordinating Committee. In FY2003, LO organized the standards and vocabulary track for an HHS-sponsored symposium on "Developing a National Agenda for the National Health Information Infrastructure" (NHII '03).

Medical Subject Headings (MeSH)

The 2004 edition of MeSH contains 22,568 main headings, 83 subheadings or qualifiers, 130 publication types, and more than 139,000 supplementary records for chemicals and other substances. For the 2004 edition, the MeSH Section added 666 new descriptors, replaced 109 descriptors with more up-to-date terminology, deleted 20 descriptors, and added 484 entry terms or "see" references.

The 2004 vocabulary reflects continuing work to reorganize genetics-related terminology and revise and expand protein terminology; a new policy for adding or modifying pharmacologic actions and a substantially revised pharmacologic action tree; revised hierarchies for physiology, algae, fungi, and bacteria; revised and reorganized terminology for human and animal population groups; and addition of public health terminology.

MeSH is translated into many other languages by organizations around the world, including a number of NLM's international MEDLARS partners. The MeSH Section has been working with OCCS to develop a MeSH translations database with a Web interface that could be used by remote users to improve the currency and accuracy of the translations. The database and interface will be tested by the German Institute for Medical Documentation and Information (DIMDI), the producer of the German translation in early FY2004. During FY2003, additional organizations which translate MeSH were recruited to use the new database once it is operational.

Clinical Vocabularies

The MeSH Section also produces RxNorm, a clinical drug vocabulary originally developed to assist NLM in mapping synonymous terms from various drug vocabularies present in the Metathesaurus. RxNorm was developed in consultation with the Department of Veterans Affairs and the Food and Drug Administration. It fills an important gap in available drug terminologies and in FY2003 was recommended as a provisional U.S. standard by the interagency Consolidated Health Informatics (CHI) eGov initiative. RxNorm is currently updated quarterly and is released within the UMLS Metathesaurus. At the end of FY2003, it contained about 104,000 concepts (terms).

On behalf of many federal agencies, LO's NICHSR supports the continued development and free distribution of LOINC (Logical Observations: Identifiers, Names, Codes) by the Regenstrief Institute. LOINC was officially designated a U.S.-government-wide clinical terminology standard in March 2003. In February 2003, NICHSR issued a new 5-year contract to the Regenstrief

Institute for basic maintenance and distribution of LOINC. In September 2003, NIH provided additional funds to enable the Regenstrief Institute to assist NIH-supported clinical research networks in implementing LOINC.

In FY2003, NLM concluded a three-year negotiation with the College of American Pathologists that resulted in U.S.-wide perpetual license for the use of the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT®), as distributed within NLM's UMLS Metathesaurus, and 5 years of updates. Under the terms of the license, which NLM negotiated on behalf of HHS, the Department of Veterans Affairs, and the Department of Defense, SNOMED CT is available for use throughout the U.S. by government agencies and the private sector, including software vendors. Many federal agencies, including NLM, contributed to the one-time perpetual license fee. NLM will pay for the annual updates. Secretary of HHS Tommy G. Thompson announced the license and contract at the NHII '03 conference on July 1, 2003. SNOMED CT was formed by the merger of a previous edition of SNOMED and the U.K. National Health Service's Clinical Terms. Also recently recommended as the U.S. government-wide standard, it is the most comprehensive clinical terminology available (more than 340,000 concepts) and has an advanced knowledge structure. SNOMED CT will appear in the first 2004 edition of the UMLS Metathesaurus. *[See further discussion of SNOMED CT in the UMLS section of the Lister Hill Center chapter, page 32.]*

UMLS Metathesaurus

The MeSH Section manages the content editing of the UMLS Metathesaurus, which at the close of FY2003 contained 975,354 concepts and 2.4 million concept names from 102 source vocabularies. A number of the vocabularies are present in multiple languages. Two new vocabulary sources—the Gene Ontology (GO) and the FDA National Drug Code Directory—were added to the Metathesaurus in FY2003; dozens of vocabularies were updated; and additional translations were added.

In preparation for the addition of SNOMED CT to the Metathesaurus in 2004, LO assisted the Lister Hill Center in defining a new and expanded Metathesaurus format that allows completely accurate representation of all relationships present in the source vocabularies and supports the representation of purpose-specific mappings between source vocabularies. The Metathesaurus is expected to become the vehicle for distribution of standard mappings between clinical terminologies and other terminologies, including the code sets required in electronic health care claims under the provisions of the Health Insurance Portability and Accountability Act of 1996 and specialized vocabularies, such as those used in adverse event reporting. In September 2003, NIH provided funds to NLM to support mappings of particular interest to the clinical research community.

Bibliographic Control

LO creates authoritative indexing and cataloging records for journal articles, books, serial titles, films, pictures, manuscripts, and electronic media, using MeSH to describe their subject content. LO also maintains the NLM Classification, a scheme for arranging physical library collections by subject that is used by health sciences libraries worldwide. NLM's authoritative bibliographic data improves access to the biomedical literature in the Library's own collection, in thousands of other libraries, and in many electronic full-text repositories.

Cataloging

LO catalogs the biomedical literature acquired or selected by NLM to document what is available from the Library's collection or on the Web and to provide cataloging and name authority records that reduce the level of cataloging effort required in other libraries. Cataloging is performed by a combination of staff in TSD's Cataloging Section, staff in HMD, and contractors.

In general, LO adheres to the *Anglo-American Cataloging Rules, 2nd edition,* when creating cataloging and name authority records. After studying the changes in the 2002 Revision to the rules, however, NLM concluded that one of the changes could have a negative impact on citing and obtaining journal articles covered in MEDLINE/PubMed. As result, NLM did not implement this change. A full explanation appears in the *NLM Technical Bulletin*.

In FY2003, TSD's Cataloging Section cataloged 19,927 contemporary books, serial titles, non-print items, and cataloging-in-publication galleys. The number of monograph catalog records that include table of contents data increased. All NLM cataloging records for Chinese language materials were converted to the pinyin Romanization system. The previous Wade-Giles Romanization entries were retained in the records as cross-references. Changes in workflow enabled by the Voyager Integrated Library System prompted TSD to move the Serials Bibliographic Unit from the Serial Records Section to the Cataloging Section.

The full production Web version of the *NLM Classification* was released in October 2003 for use by NLM staff and contractors and health sciences librarians around the world. Links between the *Classification* and the MeSH browser make it easy for catalogers to move back and forth between them. During FY2003, the system used to edit the *Classification* was upgraded to permit validation against features of the Indexing Data Creation and Maintenance System and MeSH.

HMD cataloged 368 rare monographs and 1,048 pictures and 116 linear feet of manuscript

collections. New finding guides were created for several collections. In other projects to improve access to existing historical collections, cataloging records were created for 495 Japanese items, primarily from the Edo period (1600–1869) and about 11,000 historical pamphlets. New Profiles in Science Web sites were released for Fred Lowe Soper (1893–1977), an international public health scientist, and Florence Rena Sabin (1881–1953), the first woman elected to the National Academy of Sciences. The Sabin site includes materials from the Sophia Smith Collection at Smith College and the American Philosophical Society.

Indexing

LO indexes articles from 4,697 biomedical journals for the MEDLINE/PubMed database to assist people in identifying articles on specific biomedical topics. The indexing workload continues to rise, due in part to the selection of additional journals for MEDLINE/PubMed, but primarily due to increases in the number of articles published in journals already being indexed. The Index Section is training larger numbers of new contract indexers. In FY2003, a combination of in-house staff, contractors, and cooperating U.S. and international institutions indexed 526,338 articles, about 5% more than the previous year. Previously indexed citations were updated to reflect 62 retractions, 5,185 corrections, and 38,560 comments found in subsequently published notices or articles. To improve the currency of MEDLINE indexing, an increasing proportion of journals are indexed from the online electronic version.

In FY2003, indexers created 32,318 annotated links between newly indexed MEDLINE citations for articles describing gene function in selected organisms and corresponding gene records in the NCBI LocusLink database. By the end of the year, links were being created for relevant articles about 10 organisms: human, African clawed frog, cow, fruit fly, HIV-1, mouse, nematode, rat, sea urchin, and zebrafish.

In FY2003, the Index Section completed the data collection phase of a prospective study of indexing consistency. The last study of MEDLINE indexing consistency occurred more than 20 years ago and was based on articles inadvertently indexed twice between 1974 and 1980. Many factors, such as a significant increase in the size of the MeSH vocabulary, automated systems used by indexers, and characteristics of the literature indexed, that could effect indexing consistency have changed significantly since 1980. The methodology for the current study, which was developed with advice from a statistician, assigned each of 39 journal issues (with a combined total of 586 articles) to 4 different indexers. The indexers knew that a study would be conducted some time during the year, but they did not know when, nor did they know which (if any) journal issues assigned to them were part of the study. Analysis of the data is under way. The results should assist NLM in improving indexing training and the indexing process and will provide a baseline for measuring the effect of additional efforts to provide automated assistance to indexers. The current online indexing system already gives indexers the option to make use of a ranked list of potentially relevant MeSH terms generated by the "Medical Text Indexer" developed by the NLM-wide Indexing Initiative Project led by the Lister Hill Center. This system continues to be refined based on indexer feedback.

Indexers perform their work after the initial data entry of citations and abstracts is accomplished by one of three means: electronic submission from publishers (the fastest and most economical method), scanning and optical character recognition (OCR), and double keyboarding. Of the citations added in FY2003, 61% were received electronically from publishers, an increase of 10% from the previous year. A total of 489 publishers are now supplying XML-tagged electronic data for 2,458 journals. BSD continues to encourage publishers to submit electronic data for additional indexed journals. In a change in workflow initiated in FY2003, the citation and abstract data are now reviewed and corrected prior to indexing, which improves the quality of "in process" citations in PubMed. LO now also reviews and corrects citations and abstracts for articles that are present in PubMed, but not indexed for MEDLINE, e.g., out of scope articles in selectively indexed journals.

NLM selects journals for indexing with the advice of the Literature Selection Technical Review Committee (LSTRC) (Appendix 6), an NIH-chartered committee of outside experts. In FY2003, the Committee reviewed 422 journals and rated 85 highly enough for NLM to begin indexing them immediately. Another 91 journals ranked sufficiently highly to be indexed if their publishers are able to supply electronic citation and abstract data. In addition to review of newly published journals, the LSTRC reviewed groups of older titles from Sub-Saharan Africa and countries in other parts of the developing world. NLM is working with NIH's Fogarty International Center and editors of a number of prestigious Western medical and public health journals to assist editors of African journals in improving the quality of their journals. NLM's role is to arrange for communications support for the African editors so they use the Internet to recruit authors and reviewers, communicate with Western editors, etc.

Information Products

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

Databases

LO manages the creation and maintenance of the content of MEDLINE/PubMed, NLM's database of indexed citations; Locatorplus, the Library's online catalog; MedlinePlus and MedlinePlus en español, NLM's primary information resources for patients, their families, and the general public; and a number of specialized databases, including many in the fields of health services research and public health. These databases are richly interlinked with each other and with other important NLM resources, such as ClinicalTrials.gov, PubMed Central, other Entrez databases, and SIS toxicological, environmental health, and AIDS information services. In FY2003, LO worked with other NLM program areas to ensure that all relevant NLM services were alerting users to new NIH Clinical Alerts and Advisories, which are issued when preliminary results of clinical trials warrant broad announcement to health professionals and public.

Use of MEDLINE/PubMed rose to 504 million searches, most directly in PubMed and some via the NLM Gateway. The separate OLDMEDLINE database of pre-1966 *Index Medicus* citations (previously available only via the NLM Gateway) was merged into PubMed, bringing the total number of citations in PubMed to more than 14 million. LO continues to make progress on converting retrospective data from NLM's printed indexes to machine-readable form. An additional 430,000 citations from 1953–56 became available online in 2003. HMD and TSD continued work with OCCS and the Endeavor company on the project to make records from all five series of the *Index-Catalogue of the Library of Surgeon General's Office* available for public searching. By year's end, many of the technical problems arising from size and mixed record formats had been resolved and the data were being loaded into the Encompass software.

BSD assisted NCBI with the design, development, and testing of many enhancements to PubMed, including e-mailing of search results, icons to indicate the availability of free full-text, new journals and MeSH databases, and a new cancer subset. BSD staff also worked with LHC to design, develop, and test enhancements to the NLM Gateway, including search subsets for AIDS, bioethics, history of medicine, and space life sciences; phrase detection; addition of ClinicalTrials.gov in the consumer health category; and document ordering for OLDMEDLINE citations.

In March 2003, NLM made its Voyager Integrated Library System generally available to other external computer systems via the Z39.50 protocol. TSD worked with OCCS and Endeavor to implement a new release of the Voyager and make significant improvements to Locatorplus, the NLM catalog database. NLM's name and series authority records are now publicly available through Locatorplus. Users searching on alternate forms of names and series now retrieve cataloging records that contain the preferred names. Records for older individual serial issues were removed from Locatorplus displays, improving the speed of display to catalog users. Despite these improvements, some types of searching remain difficult and cumbersome within Locatorplus (and therefore also via the NLM Gateway), due to Voyager limitations. In FY2003, NLM began a project to create an NLM catalog database in Entrez, the NCBI retrieval system used for MEDLINE/PubMed, using XML output from Voyager.

Use of MedlinePlus continued to increase dramatically. There was a combined total of 214 million page hits from 16.3 million unique users for MedlinePlus and MedlinePlus en español. Both sites received substantial and highly favorable coverage in the media. Early in the fiscal year, NLM and the University of North Carolina, Chapel Hill, inaugurated the first "Go Local" connection between MedlinePlus and NC Healthinfo, a Web directory of health services available throughout North Carolina. Work is underway to establish additional "Go Local" connections.

PSD and OCCS continued to expand and improve the basic content and features of the English and Spanish MedlinePlus sites. Sixty-one new health topics were added to the English MedlinePlus site to bring the total to 630; 99 were added to the Spanish site for a total of 589. Fifteen new interactive tutorials were added in both languages, raising the total to 165. Other enhancements included: a more flexible three-column site design that allows news and sites of current interest to be highlighted more effectively, an English medical dictionary, a database-driven look-up feature that assists users in finding U.S. libraries that provide consumer health information service, a biweekly Spanish listserv announcing new links from the Spanish site, and the ability to print and email MedlinePlus content pages. During FY2003, PSD also worked with OCCS and other NIH institutes to expand the content and provide a "talking" version of NIHSeniorHealth, to be released in early FY2004.

Under the direction of NICHSR, NLM continues to enhance its services for health services researchers and public health professionals. New methodological search filters that will assist users in focusing their MEDLINE/PubMed on cost and quality of care issues were developed by Brian Haynes and colleagues at McMaster University under contract to NICHSR and are currently available for testing from the NICHSR homepage. The PubMed clinical queries filters are also being reviewed and updated. NICHSR began work with NCBI to move the contents of HSTAT (Health Services and Technology Assessment Text) to the Entrez system, as part of the Bookshelf. This project, to be completed in FY2004, will allow more robust linking between HSTAT documents (including current evidence reports from the Agency for Healthcare Research and Quality) and MEDLINE/PubMed, PubMed Central, and other Entrez databases. It will also reduce the number of database maintenance streams that NLM must maintain.

NICHSR continued to work through AcademyHealth and the Sheps Center at the University of North Carolina, Chapel Hill to expand the content of HSRProj (Health Services Research Projects) to incorporate more private foundation and state-funded projects, in addition to those funded by federal agencies. In FY2003, projects funded by the Centers for Disease Control and Prevention, Public Health Practice Program, Florida Center for Medicaid and the Uninsured, Georgia Health Policy Center, and the National Patient Safety Foundation were added for the first time. The HSRR (Health Services and Sciences Research Resources) database also continued to expand to cover additional datasets, survey and other research instruments, and software packages used with datasets. New resources added included: OASIS GIS Mapping Tool, Georgia Division of Public Health, Virginia Reportable Disease Surveillance Data, Oklahoma Health Status Indicator Profiles, and the American Legacy Longitudinal Tobacco Use Reduction Study.

LO seeks to improve its databases and Web services based on the results of usability testing, focus groups, and user surveys. BSD awarded a contract to conduct usability testing of PubMed in early FY2004. Early in FY2003, LO staff helped to arrange and observed online and in-person focus groups for MedlinePlus en español. In February 2003, NLM conducted Web surveys of users of MedlinePlus and MedlinePlus en español which indicated very high levels of satisfaction with both sites. The majority of users of the then relatively new Spanish site were from outside the U.S. In comparison to results obtained from a comparable survey of the English site conducted in 2001, more current users of the English site are likely to have found out about MedlinePlus from a search engine. A summary of the survey results is available at http://www.nlm.nih.gov/medlineplus/survey2003/index.html.

In FY2003, PSD and SIS collaborated on a successful proposal for NIH evaluation set-aside funds to cover the inclusion of five NLM Web sites, including MedlinePlus, MedlinePlus en español, and the main NLM Homepage, in the American Consumer Satisfaction Index (ACSI). Produced through a partnership of the University of Michigan Business School, the American Society for Quality (ASQ), and the international consulting firm, CFI Group, the ACSI generates comparable user satisfaction ratings for many government and private sector Web sites along a number of different dimensions, permits the use of some site specific survey questions, and permits tracking of changes in user satisfaction over the course of a year as features are added or modified. NLM will begin to receive ACSI results in early FY2004.

Machine-Readable Data

NLM leases many of its electronic databases to other organizations to promote the broadest possible use of its authoritative bibliographic, vocabulary, and factual 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 obtain NLM data incorporate them into many different database and software products and use them in a variety of research and development projects.

Demand for MEDLINE/PubMed data in XML format continues to increase, with the majority of new users interested in the data for data-mining and research. There are currently 219 MEDLINE licensees, 57 under the non-US research only license. In FY2003, for the first time, licensees had the option to use FTP download to obtain the entire MEDLINE baseline database or 12+ million records (in 396 segments). Weekly update files have been available via FTP for several years. Thirty licensees reported back to NLM on their download speeds for the complete baseline—which ranged from 30 minutes for one U.S. site to 25 hours for one German site. Beginning in FY2004, licensees will receive all prospective quality-reviewed citations in the PubMed database, including those not indexed for MEDLINE. A small number of organizations license NLM catalog records in MARC21 format or one or more of SIS toxicological and environmental health data files in XML format. Many users execute the online Memorandum of Understanding that allows them to FTP the MeSH files in XML, ASCII, or MARC format.

In FY2003, 2,103 licensees obtained the UMLS Knowledge Sources via FTP, on CD-ROMs, or via the applications programming interface to the UMLS Knowledge Source Server. BSD, OCCS, and LHC developed plans for a Web-based licensing system for the UMLS to be implemented in FY2004, when all UMLS users must sign a new UMLS license, which incorporates provisions covering the inclusion of SNOMED CT in the UMLS Metathesaurus. The new license applies only to the UMLS Metathesaurus, which contains content produced by many different organizations. The other UMLS Knowledge Sources will be available under Open Access principles. Staff from BSD assisted LHC in quality control and testing of Metathesaurus releases and features of the UMLS Knowledge Source Server.

Web and Print Publications

NLM's databases and its Web sites are its primary publication media. Demand for the Library's few remaining print publications continues to decline due to increasing electronic access to NLM data. After assessing the content and use of the four MeSH publications, LO decided to expand the introductory material in the "Black and White" MeSH that is published as a separately available supplement to *Index Medicus* and to cease publication of the other printed MeSH tools after release of their 2003 editions. Following the 2003 release of the Web version, LO also decided to cease publication of the

infrequent printed editions of the *NLM Classification*.

A total of 4.7 million unique users obtained 37 million page hits on NLM's main Web site, which provides access to many NLM publications. PSD's Web Management Team serves as the Web Master for the main Web site. Informed by results of usability testing conducted in FY2002, PSD awarded a contract to support a major redesign of the NLM homepage and the second level pages to which it refers. During FY2003, several alternative designs were reviewed by NLM staff and tested with potential users. The redesigned site, with a more flexible three-column format that can accommodate news and highlight time-sensitive content, is scheduled for release in mid-2004.

Publications available from the main Web site include recurring newsletters and bulletins, fact sheets, technical reports, and multimedia catalogs. Users downloaded close to 42,000 copies of issues in the *Current Bibliographies in Medicine* series, which is edited by the Reference and Customer Services Section. LO staff members collaborate with outside experts to produce each bibliography, which addresses a topic of current interest to NLM, NIH, or other federal agencies. The MEDLARS Management Section edits the *NLM Technical Bulletin*, which provides timely, detailed information about changes and additions to a broad range of NLM services and policies for librarians, information professionals, and other interested parties.

Direct User Services

In addition to building databases and supplying other information products, LO provides document delivery and reference and customer service to remote users, as a national and international backup to services available from other health sciences libraries and information suppliers. LO also serves a substantial onsite clientele in the NLM reading rooms.

Document Delivery

LO provides interlibrary loan service to other members of the National Network of Libraries of Medicine and to international libraries to fill requests for materials not readily available from other sources. LO also retrieves documents from NLM's closed stacks for use by onsite patrons.

In FY2003, PSD's Collection Access Section processed a total of 653,916 document requests, a 7% decline from the previous year. Onsite users requested 290,564 contemporary documents from NLM's closed stacks, 12% less than in FY 2002. The decline reflects fewer service days (due to weather emergencies), increased security measures that hamper access to the NIH campus, and increased online availability of biomedical journals. In contrast, onsite users requested 16,163 items from the historical and special collections, a 135% increase due to heavy use of prints, photographs, and historical films by invited scholars preparing their presentations for the *Visual Culture and Public Health* symposium to be held at NLM in early FY2004. An NLM patron photo-ID card is now also valid for clearing security when entering the NLM building and must be worn while working in the NLM reading rooms.

Other libraries requested 363,352 contemporary documents from NLM, down 3% from FY2002. The Collection Access Section handled 83% of the requests in 12 hours and delivered 76% of the filled requests electronically. In the 4th quarter NLM's fill rate improved to 76% from 73% in the previous year, due to a new procedure that retrieves some journal issues from the indexing workflow if they are needed to fill interlibrary loan requests.

A total of 3,254 libraries now use DOCLINE, 2,881 in the U.S., 312 in Canada, and 61 in other countries. Eleven Mexican libraries became DOCLINE users in FY2003. DOCLINE users entered 2.86 million requests into the system in FY2003, a 6% decrease from the previous year; 92% of the requests were filled. Although the absolute number of requests received by NLM declined, the Library's share of all DOCLINE requests increased by half of a percent to 12.7%. Individuals submitted 863,001 document requests to DOCLINE users via the Loansome Doc feature in MEDLINE/PubMed and the NLM Gateway, a 7% decline from FY 2002. The trend in document request traffic is down throughout all Regions of the NN/LM, reflecting increased access to electronic full-text journals.

In FY2003, NLM and the NN/LM took several steps to highlight the availability of free electronic full-text for MEDLINE/PubMed users and for DOCLINE and Loansome Doc users, who continue to request some articles that are freely available to them on the Web. In addition to the PubMed icons previously mentioned, NCBI staff made it easier for libraries to include all PubMed Central titles in their implementations of LinkOut for Libraries. Staff at the Regional Medical Libraries continued to promote the use of PubMed's LinkOut for Libraries to NN/LM members, as a means for customizing PubMed to display their electronic and print holdings to their primary clientele. Locatorplus was updated to provide more comprehensive information on journals freely available in PubMed Central or from publishers' Web sites. DOCLINE and Loansome Doc were modified to alert users when they request an article that is free in PubMed Central. Work is under way to extend this latter capability to include journals that participate in LinkOut and are free on the publisher's Web site. These changes help users to obtain needed information more quickly and also reduce unnecessary work in libraries.

DOCLINE requests are routed to libraries automatically based on automated holdings data in the SERHOLD® database. At the end of FY2003, SERHOLD contained 1.39 million holdings statements for 52,731 serial titles held by 3,066 libraries. In FY2003, LO and OCCS implemented the automated

transfer of holdings data from SERHOLD to the OCLC database for NN/LM members who requested the service. This will reduce workload for a number of network libraries and improve the quality and currency of holdings data in OCLC. Automated transfer from OCLC to SERHOLD has been tested and is scheduled for implementation in FY2004.

Effective with its April–June bills for interlibrary loan service, NLM will now bill via the Electronic Funds Transfer Service (EFTS) operated for the NN/LM by the University of Connecticut for libraries that request this billing method. NLM and the Regional Medical Libraries are encouraging NN/LM members to use EFTS as a mechanism for reducing the administrative costs associated with ILL billing. During FY2003, EFTS participation increased 23% to 830 libraries. Participants receive either a net consolidated bill or a net consolidated payment each month.

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 Services Section handles all initial inquiries with contract assistance and many of those requiring second-level attention. Staff from throughout LO and NLM assist with second-level service when their specialized expertise is required.

In FY2003, the Reference and Customer Services Section handled a total of 105,784 inquiries (excluding spam) from onsite and remote patrons, an 8% increase from the previous year. Onsite requests declined 14%, but remote requests increased 30%. The introduction of MedlinePlus en español has greatly increased the number of customer service inquiries in Spanish. In FY2003, Reference staff handled 2,225 Spanish language inquiries. NLM uses the Seibel customer service software, integrated with a telephone call system. In FY2003, the Reference and Customer Service Section replaced its existing call system with Aspect and began using data mining tools developed by NCBI to analyze and characterize customer service inquiries stored in the Seibel database. The NCBI tools were selected after consultation with several program areas about methods for mining the inquiries for complaints and suggestions that might identify or corroborate the need for enhancements to NLM products and services.

In February 2003, "Cosmo," a virtual customer service representative built with the NativeMinds software, was launched to answer frequently asked questions about NLM programs, products, and services. Reference staff review questions that Cosmo is unable to answer and add to his knowledge base routinely, thus gradually improving his performance for questions within his job description. By the end of FY2003, Cosmo was answering 72% of such questions appropriately, thus fielding a number of questions that would otherwise have been answered by people.

Outreach

LO manages or contributes to many programs designed to increase awareness and use of NLM's collections, programs, and services by librarians and other information professionals, historians of medicine and science, researchers, educators, health professionals, and the general public. LO coordinates the National Network of Libraries of Medicine (NN/LM) which works to equalize access to health information services and information technology for health professionals and the public throughout the United States; serves as the secretariat for the Partners in Information Access for the Public Health Workforce; participates in NLM-wide efforts to develop and evaluate outreach programs designed to improve information access for under-served minorities and the general public; produces major exhibitions and other special programs in the history of medicine; and conducts a range of training programs for health sciences librarians and other information professionals. 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 NN/LM works to provide timely, convenient access to biomedical and health information resources for U.S. health professionals, researchers, and the general public, irrespective of their geographical location. It is the core component of NLM's outreach program and its efforts to reduce health disparities. The network includes 5,298 full and affiliate members. The regular members are libraries with health sciences collections, primarily in hospitals and academic medical centers. The affiliate members include some small hospitals, public libraries, and community organizations that provide health information service, but have little or no collection of health sciences literature. LO's NN/LM Office oversees the network programs that are administered by eight Regional Medical Libraries (RMLs), under contract to NLM. (See Appendix 1 for a list of the RMLs.)

In addition to the basic NN/LM contracts, NLM funds subcontracts for five national 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 outlined 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 FY2003, the Center assisted NLM and the RMLs in defining national, measurable outcomes for outreach to public libraries and public health departments, using a "logic model" structure often used in designing and measuring the effectiveness of public health interventions. The Center also provided evaluation training in many regions, expanded the evaluation resources accessible from the NN/LM Web site, and consulted on the design of individual outreach projects.

The National Outreach Mapping Center at Indiana University in Indianapolis assists NLM and the RMLs to display the geographic distribution and impact of NN/LM programs and services and therefore to identify gaps in services that should be addressed. During FY2003, staff at NLM and in the RMLs were trained in the use of mapping software, and base maps were designed for monitoring distribution of outreach activities. The main emphasis, however, was on defining and collecting uniform data from the RMLs for the NLM-wide outreach database, which also includes data for NLM grants and NLM outreach activities not supported via the NN/LM contracts, and assisting in defining and testing the features of the database. In FY2003, a new pilot Web-Services Technology Operations Center (Web-STOC) was funded at the University of Washington, which has led NN/LM Web site management for a number of years. In addition to ongoing technical management of the NN/LM Web sites, Web-STOC will investigate, recommend, and direct the implementation of additional Web technology for teleconferencing, distance education, Web broadcasting, online surveys, etc. A committee of representatives of each type of RML employee, i.e., a Director, an Associate Director, an Outreach Coordinator, etc., and the NLM Network Office oversees the activities of this pilot Center.

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 libraries, professional associations, public health departments, schools, churches, and other community groups. Some projects are identified by individual RMLs through regional solicitations or ongoing interactions with regional institutions; others are identified by periodic national solicitations for outreach proposals issued simultaneously in all NN/LM regions. In FY2003, the RMLs issued 68 subcontracts for outreach projects, 37 as a result of a national solicitation. The projects target a range of populations in rural and inner city areas in 28 states and the District of Columbia and involve partnerships between libraries and many different community groups.

With the assistance of other NN/LM members, the RMLs 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 of medicine and science. In FY2003, NLM and NN/LM services were displayed at 327 exhibits at national, regional, and state association meetings across the country. These exhibits highlight not just the databases and services to which LO contributes, but also other NLM products relevant to attendees at each meeting. In FY2003, National Network Office (NNO) and BSD staff worked with the Office of the Director, NLM and OCCS to develop a new exhibit database to track this activity.

The current NN/LM contracts call for site visits to all the RMLs near the mid-point in the 5-year contract cycle to assess progress toward Regional goals, discover best practices, discuss challenges and problems, and obtain input from the RMLs and NN/LM members about what NLM can do to assist them reaching Network objectives. Each site visit team includes at least one health professional, one academic health sciences librarian, one hospital librarian, and an RML Associate Director—all from other Regions than the one being visited. NLM attendees include the Head, NNO, the Associate Director for LO, and sometimes the NLM Deputy Director and/or an NLM Contracts Officer. Four sites visits were conducted in FY2003. It was a pleasure to learn more about the excellent work in the Regions and to hear from so many NN/LM members. Based on suggestions received during these visits, NLM established an NN/LM Hospital Internet Access Task Force to identify: (1) barriers to access to the Internet in hospitals, (2) best practices for achieving the twin goals of easy access to the Internet and appropriate security for hospital patient data systems, and (3) recommended actions that the NN/LM and NLM might take to assist hospital libraries in overcoming identified barriers. Comments heard during the site visits confirmed that electronic journal licensing and its impact on interlibrary loan and service to unaffiliated health professionals were significant concerns across the country. As previously mentioned, NLM and the RMLs are currently possible actions to address this issue.

Partners in Information Access for the Public Health Workforce

The NN/LM is a key member of the Partners in Information Access for the Public Health Workforce, a collaboration that also includes NLM, the Centers for Disease Control and Prevention (CDC), several other federal agencies, and a number of public health associations and organizations. The Partnership was initiated by NLM, the CDC, and the NN/LM in 1996 to

help the public health workforce find and use information effectively to improve and protect the public's health. The National Information Center on Health Services Research and Health Care Technology (NICHSR) coordinates the Partners for NLM; staff members from NNO, SIS, and the Office of the Associate Director for LO also serve on the Steering Committee. The Partners develop new information and training resources for the public health workforce; sponsor meetings, workshops, and satellite broadcasts geared to promote access to information services and technology as a means to improving public health practice; foster outreach and distance learning partnerships; and disseminate information on relevant funding and training opportunities for the public health workforce. Through participation in the Partnership, LO and the NN/LM have identified specific outreach opportunities and increased the number of outreach project and grant proposals involving the public health community.

As part of NLM's participation in the Partners, NICHSR funded and organized a Public Health Grand Rounds satellite broadcast, "Wired Communities: Putting the *e* in Public Health" on January 31, 2003. The broadcast demonstrated the utility of the Internet to local public health departments and emphasized services available from NLM and NN/LM. In FY2003, staff from the NNO, NICHSR, and PSD worked with other Partners to completely redesign the Partners Web site. The redesign, which shifted the focus from the Partnership itself to information sources useful to the public health workforce, was guided by usability testing and greeted by universal praise and substantially increased use. One of the key information sources featured on the Web site is the Healthy People 2010 Information Access Site, developed by NLM and the Public Health Foundation to provide access to information that can assist in developing strategies to meet public health goals. The site continues to expand to include evidence-based PubMed search strategies and links to relevant MedlinePlus topics for additional Healthy People 2010 objectives. Librarians from throughout the NN/LM assist in the developing strategies, which are then reviewed for relevance and utility by public health experts.

With assistance from an NLM Associate Fellow, members of the Partners used the Partnership Self-Assessment Tool developed by the Center for the Advancement of Collaborative Strategies in Health at the New York Academy of Medicine to examine the current strengths and weaknesses of the Partnership. The Self-Assessment identified a number of issues to be addressed by the Partners in FY2004.

Special NLM Outreach Initiatives

LO participates in 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 FY2003, the Office of the Associate Director, LO worked with other NLM components, the American College of Physicians, and some NN/LM members to launch and monitor the project to test the use of "information prescriptions" for MedlinePlus in physician's offices in Georgia and Iowa. BSD assisted in developing the prescription pads and other materials that play a key role in this project. The project will be expanded to other states in FY2004, with modifications based on the experience to date and information obtained from focus groups with participants in Georgia and Iowa. In the second phase, there will be more explicit efforts to link the physicians' offices to local libraries willing to serve their patients and to enlist the assistance of physicians in obtaining feedback from patients about their experiences in filling the information prescription.

The Office of the Associate Director, LO, the NNO, and BSD continued to work with the Public Library Association, a Division of the American Library Association, to improve public library awareness of MedlinePlus and MedlinePlus en español. A repeat direct mail campaign was conducted in February 2003 to 20,000 public and health sciences libraries to introduce them to MedlinePlus en español. The libraries were offered free bookmarks and posters. Over the two and a half years of this project, 6,120 libraries responded to NLM, and 2 million bookmarks have been distributed. Work also continued on increasing the number of links to MedlinePlus from professional association Web sites. To date, 75% of the 350 associations that were invited to link to MedlinePlus have done so. In FY2003, an additional 350 organizations to which MedlinePlus links were contacted about linking back to MedlinePlus. The response rate from this effort has been low.

LO staff members are actively involved with NLM's partnership with Wilson High School in the District of Columbia. In FY2003, LO provided summer employment and training opportunities for several students and teachers.

Historical Exhibitions and Programs

HMD directs the development and installation of major historical exhibitions in the NLM rotunda, with assistance from LHC and the Office of the Director. Designed to appeal to the interested public as well as the specialist, these exhibitions highlight the Library's historical resources and are an important part of NLM's outreach program. *Dream Anatomy*, an exhibition focused on anatomy, medicine, and the artistic imagination, opened October 9, 2002 and closed in July 2003. Widely reviewed and praised, this exhibition featured rare anatomical books and illustrations from the NLM collection, as well as 20th and 21st century art and interactive displays that drew upon the Visible Human datasets. Despite the sniper attacks, Code orange alerts, and snowstorms, more than 10,000 visitors came to see the Dream Anatomy exhibition. Many others got glimpses of the exhibition from channel 4 local news and

a 10-minute segment on Voice of America television, which was dubbed in French, Spanish, Farsi, Urdu, Arabic, Chinese, and Hindi.

The *Dream Anatomy* Web site received more than 1.5 million its in FY2003. The Web site includes "A Learning Station" designed for teachers and students in grades 6-12. The Library hosted a number of ancillary programs related to the exhibition: an opening program and reception; a film series, *Cinematic Dream Anatomy;* and a June 12, 2003 symposium, "Visionary Anatomies," on the history and meaning of anatomic imaging. In conjunction with the June symposium, NLM installed a smaller lobby exhibit on *Anatomical Revisioning: Art as Applied to Medicine*, which was designed and fabricated by LHC and HMD staff in collaboration with the Association of Medical Illustrators. HMD also used the occasion of the symposium to launch *Historical Anatomies on the Web,* with an initial group of 120 digitized images from important early anatomical atlases. LHC is developing a DVD version of *Dream Anatomy.*

While *Dream Anatomy* was on display, work continued on the next major exhibition, *Changing the Face of Medicine: Celebrating America's Women Physicians,* which will open in October 2003. The new exhibition will feature more than 300 women physicians, living and dead, selected with advice from an advisory committee of eminent physicians (both women and men), chaired by Tenley Albright, M.D., former chair of the NLM Board of Regents. One of the principal intended audiences for the exhibition is girls who might be encouraged to pursue a medical degree. The exhibition has been designed to illustrate the range of possible careers open to women physicians and to show that women from all segments of U.S. society have excelled in the field. The NIH Office of Research on Women's Health has provided partial funding for a future traveling version of the *Changing the Face of Medicine.*

In October 2002, the traveling version of a previous exhibition, *Frankenstein: Penetrating the Secrets of Nature,* began a two-year tour to 80 public, academic, and medical libraries across the country. Developed by the American Library Association, in conjunction with HMD, and funded by the National Endowment for the Humanities and NLM, the traveling exhibit has received substantial publicity and hundreds of thousands of visitors across the country. Each library that hosts the exhibit also presents related public programs related to science, medicine, and the humanities. A printed catalogue of the original *Frankenstein* exhibition, as it appeared at NLM, was published by Rutgers University press in conjunction with the opening of the traveling exhibit. The American College of Allergy, Asthma, and Immunology provided funding to the Friends of the National Library of Medicine to print the catalog of a previous exhibition, *Breath of Life.* Copies were distributed to College members.

In addition to the major exhibitions in the rotunda, HMD installs "mini-exhibits" generally in cases near the entrance to the HMD Reading Room, but sometimes also in other locations. A small exhibit on *Donald S. Fredrickson* was installed in the lobby outside the LHC auditorium in conjunction with an NIH program held in his memory in October 2003. LHC and HMD also produced a film on his life and scientific accomplishments and expanded and upgraded the Fredrickson Profiles in Science Web site. *Here Today—Here Tomorrow*, an exhibit of AIDS ephemera, curated by staff from the HMD and the NLM Office of Communications and Public Liaison, was installed outside the HMD Reading Room from November 2002 to June 2003. Originally planned for World AIDS Day, the exhibit of posters, buttons, and brochures was selected from donations to the Library by William Helfand. *Hortus Sanitatis: The Universe of Medicinal Plans in the Late Middle Ages* by Visiting Scholar Alain Touwaide opened in June 2003 and will be on display until December 2003. This exhibition of medicinal plants from medieval manuscripts and early Renaissance printed books shows the legacy of Greek sources in medieval and early modern medical botany. Some previous mini-exhibits also live on as touring poster exhibits: in FY2003 the *Elizabeth Blackwell* and *Oriental Medicine* posters were installed at several academic institutions across the country.

"Turning the Pages," a remarkable program originally developed by the British Library that uses computer-animation, high-quality digitized images, and touch-screen technology to simulate the action of turning the pages of rare books, is on permanent display in the NLM Visitor Center and the HMD Reading Room. In FY2003, HMD staff worked with the LHC to add two more books, Conrad Gesner's *Historia Animalia* and Ambrose Pare's *Oeuvres Complets*, bringing the total to four.

In addition to the programs associated with *Dream Anatomy* already mentioned, HMD organized a regular series of seminars by historical scholars and several special historical lectures in conjunction with the NLM Diversity Council and the EEO Office. These included the first public lecture in celebration of Hispanic American History Month, "Rights, Recognition, and Revolution: The USPHS and the Mexican Border" by Professor John McKiernan-Gonzalez. In FY2003, a number of visiting scholars came to work at NLM for brief periods. In addition to those mentioned previously as contributing to the use of the historical image collections, HMD also hosted the first group of scholars in a new program designed to encourage the incorporation of HMD research materials in university classrooms. Under this program, HMD is also creating a collection of Web syllabi and course materials.

HMD staff members continued to present historical papers at professional meetings and to publish the results of their scholarship in books, chapters, articles, and reviews. HMD continued to play a lead role in preparing the recurring features "Voices from the

Past" and "Images of Health," which usually feature materials from the NLM collections, for the *American Journal of Public Health*.

Training and Recruitment of Health Sciences Librarians

LO develops online training programs in the use of MEDLINE/PubMed and other databases for health sciences librarians and other search intermediaries; oversees the activities of the National Online Training Center and Clearinghouse (NTCC) 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 and others 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 FY2003, the MEDLARS Management Section (MMS) and the NTCC taught PubMed, the NLM Gateway, and/or TOXNET searching to 848 students in 69 face-to-face classes. In response to feedback received during site visits to the RMLs, LO examined alternatives for providing more distance learning opportunities for health sciences librarians in FY2004 and beyond. In FY2003, MMS reorganized the online training class structure and student workbooks to cover the varied features and nuances of each database and retrieval system more effectively. MMS worked with LHC to create a videotape, "Branching Out —The MeSH Vocabulary," to be used in conjunction with the MeSH module of the online training course. The heavily used Web-based PubMed tutorial was updated several times to reflect new PubMed features. The training workbooks were updated on the Web three times in FY2003.

In FY2003, LO and the NTCC assisted NCBI in setting up, publicizing, and providing administrative support for four NN/LM offerings of the Introduction to Molecular Biology Information Resources class, the first time that this class was offered away from NLM and not in conjunction with a professional meeting. The NTCC also handles registration for the new basic UMLS course for medical librarians, which includes hands-on exercises with the UMLS Knowledge Source Server and MetamorphoSys. During FY2003, the NTCC conducted an online evaluation of its Educational Clearinghouse database, which is intended to reduce duplicative effort by providing access to existing syllabi, handouts, search examples, etc. developed by NN/LM members for different audiences.

The size of the Associate Fellowship increased in FY2003. There were six first-year and three second-year fellows. All six who finished the first year at NLM in August elected to continue on the optional second year at another health sciences library. The second year sites are: University of California, Los Angeles, University of Arizona, Vanderbilt University, Virginia Commonwealth University, Georgetown University, and George Washington University. Eight new fellows entered the first year program in September 2003.

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. Following up on a 2002 LO-funded MLA conference which explored the concept of the "informationist," LO assisted the NIH Library in planning and funding an evaluation of the impact of its expanding informationist program and also supported efforts by NLM's Extramural Programs Division to establish a new informationist fellowship. In FY2003, NLM expanded its offering of courses that may assist librarians in preparing for specialist roles. In addition to the molecular biology and UMLS courses previously described, NICHSR added *Health Economics Information Resources: A Self-Study Course* to its suite of courses on health services research, health policy, and public health.

NLM collaborates with several 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 under-served groups is a high priority. A high percentage of current health sciences library directors will retire over the next 5–10 years. LO has provided support to establish or increase the stipend for scholarships for minority students available through the Medical Library Association and the Association of Research Libraries. LO also supports the pilot NLM/AAHSL Leadership Development program. In FY2003, the first cohort of five fellows completed the program, which involves leadership training, mentorship, and site visits to the mentor's institution. AAHSL contracts with ARL for the formal leadership training courses. The program was extremely popular with both mentors and fellows, three of whom have since been selected for higher level positions, and there were many well-qualified applicants for the second year of the three-year pilot.

Table 1

Growth of Collections

Collection	*Previous Total (9/30/02)*	*Added FY 2003*	*New Total (9/30/03)*
Book Materials			
Monographs:			
Before 1500	584	4	588
1501-1600	5,922	16	5,938
1601-1700	10,211	10	10,221
1701-1800	24,613	24	24,637
1801-1870	41,379	45	41,424
Americana	2,341	0	2,341
1871-Present	712,170*	15,292	727,462
Theses (historical)	281,794	0	281,794
Pamphlets	172,021	0	172,021
Bound serial volumes	1,252,480	17,061	1,269,541
Volumes withdrawn	(78,127)	(2,356)	(80,483)
Total volumes	2,425,388*	30,096	2,455,484
Nonbook Materials			
Microforms:			
Reels of microfilm	132,544	4,898	137,442
Number of microfiche	445,794	1,580	447,374
Total microforms	578,338	6,478	584,816
Audiovisuals	70,739	2,226	72,965
Computer software	2,138	105	2,243
Pictures	56,962	1,048	58,010
Manuscripts	4,120,882	202,825	4,323,707**
Total nonbook	4,829,059	212,682	5,041,741
Total book & nonbook	7,254,447*	242,778	7,497,225

* Corrected figure
**Equivalent to 2,471 linear feet.

Table 2

Acquisition Statistics

Acquisitions	*FY 2001*	*FY 2002*	*FY 2003*
Serial titles received	20,314	20,350	20,476
Publications processed:			
Serial pieces	142,642	133,908	134,579
Other	21,338	22,274	24,523
Total	163,980	156,182	159,102
Obligations for:			
Publications	$5,155,054	$5,802,023	$6,217,417
(For rare books)	($279,710)	($446,039)	($297,894)

Table 3

Cataloging Statistics

	FY 2001	*FY 2002*	*FY 2003*
Completed Cataloging	19,024	21,419	19,927

Table 4

Bibliographic Services

Services	*FY 2001*	*FY 2002*	*FY 2003*
Citations published in MEDLINE	463,014	502,056	526,338
For *Index Medicus*	445,041	459,558	492,911
Journals indexed for Medline/PubMed	4,538	4,538	4,697
Journals indexed for *Index Medicus*	3,707	3,834	3,994
Abstracts entered	345,624	398,885	465,975

Table 5

Web Services

Services	*FY 2001*	*FY 2002*	*FY 2003*
NLM Web Home Page			
Page Views	36,248,000	40,607,752	37,166,023
Unique Visitors	4,490,000	5,300,363	4,792,482
MedlinePlus			
Page Views	62,069,000	116,335,454	214,127,932
Unique Visitors	4,409,000	9,594,429	16,356,444

Table 6

Circulation Statistics

Activity	*FY 2001*	*FY 2002*	*FY 2003*
Requests Received	682,777	705,069	653,916
Interlibrary Loan	338,627	373,292	363,352
Onsite	344,150	331,777	290,564
Requests Filled:	535,594	539,274	511,032
Interlibrary Loan*	251,525	268,816	268,714
Onsite	284,069	270,458	242,318

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

Table 7

Online Searches—All Databases

	FY 2001	*FY 2002*	*FY 2003*
Total online searches	313,000,000	382,000,000	452,000,000

Table 8

Reference and Customer Services

Activity	*FY 2001*	*FY 2002*	*FY 2003*
Offsite requests	59,634	49,153	64,010
Onsite requests	51,287	48,395	41,774
Total	110,921	97,548	105,784

Table 9

Preservation Activities

Activity	*FY 2001*	*FY 2002*	*FY 2003*
Volumes bound	31,625	25,609	15,646
Volumes microfilmed	5,131	5,255	2,795
Volumes repaired onsite	1,403	1,542	1,285
Audiovisuals preserved	225	283	500
Historical volumes conserved	128	66	111

Table 10

History of Medicine Activities

Activity	*FY 2001*	*FY 2002*	*FY 2003*
Acquisitions:			
Books	314	424	314
Modern manuscripts	1,340,150	840,000	498,750*
Prints and photographs	3,324	3,176	1,000
Historical audiovisuals	1593	1,361	97
Processing:			
Books cataloged	510	368	215
Modern manuscripts cataloged	190,750	984,025	203,000**
Pictures cataloged	20	0	1,048
Citations indexed	285	846	856
Public Services:			
Reference questions answered	15,718	14,898	14,693
Onsite requests filled	4,844	6,870	16,163

*Equivalent to 285 linear feet
**Equivalent to 116 linear feet

SPECIALIZED INFORMATION SERVICES

Jack Snyder, M.D., J.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 NLM 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 sources. 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 FY2003 SIS continued to balance efforts to enhance and re-engineer existing information resources with efforts to provide new services in emerging areas. We further developed various prototypes that rely on geographical information systems, innovative access and interfaces for consumers, and graphical display of data from information sources. Highlights for 2003 include

- SSEUS (SIS SQL Entry Update System), a new data creation and maintenance system for TOXNET, our premier collection of databases on toxicology, hazardous chemicals, and related areas;
- Listserv, Automated Indexing, and other refinements of the interface and multi-database search capability for TOXNET;
- Household Products Database, which links the ingredients in over 4,000 consumer brands to health effects described in Material Safety Data Sheets provided by manufacturers;
- Several new Toxicology & Environmental Health Special Topic Web resource pages, including Environmental Justice and Asian-American Health;
- TOXMAP, a prototype system that uses maps of the United States 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 encountered in everyday life, in everyday places;
- Continued support of PAHO/NLM Disaster Preparedness Information Centers in Honduras, Nicaragua, and El Salvador;
- Expanded Native American outreach initiatives; and
- 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 SIS resources related to toxicology and environmental health information, HIV/AIDS information, and special populations information includes many databases that are created or acquired as well as other services and projects.

Haz-Map database, released in 2002 at http://hazmap.nlm.nih.gov, is an occupational toxicology database designed to link jobs and hazardous job tasks to occupational diseases and their symptoms. It is a relational database of chemicals, jobs, and diseases that averaged nearly 20,000 queries per month in 2003. The Haz-Map *jobs table* is based on the 1997 Standard Occupational Classification (SOC) system. The *industries table* is based on the Standard Industrial Classification (SIC) system. The *diseases table* is based on the International Classification of Diseases (ICD-9). Information from textbooks, journal articles, and electronic databases was classified and summarized to create the database. A user may search this occupational database by chemical agent, occupational disease and by job type.

ChemID*plus* (Chemical Identification File) is an NLM online chemical dictionary that contains nearly 370,000 records, primarily describing chemicals of biomedical and regulatory importance, and available to users on the Internet at http://chem.sis.nlm.nih.gov/ chemidplus. ChemID*plus* features include chemical structure search and display for over 177,000 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 ChemID*plus*. During FY2003 over 75,000 queries per month were made of this database. To assist with spelling errors, a chemical spell checker was released to help users retrieve substances more efficiently by chemical name. The checker, which can be instantly revised using the SIS DBMaint2 online update system, contains new spelling indices for more than 1.3 million chemical names and synonyms. The database was enhanced by the addition of various new locators pointing to international resources, including coverage of agents found in NIAID ChemDB, ATSDR Medical Management Guidelines, and ClinicalTrials.gov. In FY2003, new test versions for the new ChemIDplus "Light" and "Heavy" systems were ported to a UNIX-based chemserver for testing and further development. The new system capabilities include a simpler Web front end that does not require plug-ins for structure display, and an advanced version that allows numeric searching by acute toxicity data and effect, and chemical/physical properties.

The **Hazardous Substances Data Bank (HSDB)** continues to be a highly used resource, averaging 50,000–60,000 searches each month (a 5% increase over FY2002). 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 2003, there has also been a continued emphasis on adding to HSDB new chemicals with the potential for high toxicity and high human exposure. Over 135 new chemicals were added in 2003, including new pesticides, drugs, and environmental pollutants. The emphasis on the addition of new chemicals will continue in the coming year. Newer sources of relevant data are being examined for incorporation into new and existing data fields within the current 4,757 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. As part of this effort, SSEUS (a relational HSDB database using the MySQL database application) was created, and a new client-server interface was programmed to allow easier updates. The new maintenance system is now poised for integration with other new features, including numeric searching and automatic indexing.

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. SIS continues to integrate this configuration with other database creation systems and Web access to them. Further refinements of the SIS search interface enhance the ability of 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. Feedback from TOXNET user online surveys has provided a basis for current and future planning, and as result, SIS will implement a chemical spellchecker, automated indexing, and a virtual meta-search tool during the coming years.

Alternatives to Animal Testing (ALTBIB)—SIS continues 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, through a new online resource named ALTBIB, which allows search access to all of the 7,595 citations organized from previous bibliographies. ALTBIB 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 FY2003, the database included over 3 million citations to toxicology literature dating back to 1965. In 2003, users accessed standard journal literature in toxicology and environmental health as part of the enlarging MEDLINE database, while NLM continued to add 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, SIS further enhanced 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 improved TOXNET user interface.

DIRLINE (Directory of Information Resources Online) is NLM 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 (nearly 7000 searches per month). The 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 continues to improve as duplicates have been eliminated through changes in policy and streamlining of maintenance. *Health Hotlines*, the always popular publication of health-related toll-free telephone numbers, has a recently updated Web version which also indicates the availability of Spanish speaking customer service representatives and Spanish language publications from the resources listed.

The **Toxics Release Inventory (TRI)** series of files now includes online files TRI86 through TRI2001. These files remain an important resource for environmental release data and are a useful complement to other SIS 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 (NCI) 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 540 chemicals.

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

The **Registry of Toxic Effects of Chemical Substances** (RTECS) is a data bank based upon a National Institute for Occupational Safety and Health (NIOSH) file by the same name which NLM restructured and made available for online searching. With our move to free Internet access to all databases, NIOSH requested that we no longer include RTECS on our system. SIS continues 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 FY2003, next generation DART consisted of 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) in this subject area. In FY2003, a new contract was awarded, thousands of new records were added, and easy access to DART Special and to DART Core was maintained at the new TOXNET interface. DART is funded by NLM, the Environmental Protection Agency, the National Institute of Environmental Health Sciences, and the FDA's National Center for Toxicological Research, and is managed by NLM.

The **Environmental Mutagen Information Center** (EMIC) database contains over 24,000 citations to literature on agents that have been tested for genotoxic activity. A backfile for EMIC (EMICBACK) contains over 75,000 citations to the literature published from 1950-1991. The Environmental Protection Agency, the National Institute of Environmental Health Sciences and NLM, collaborating partners in this effort, stopped compiling this special collection as of December 1999, but 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. A Webcast is available at on the CDC Web site under public health training network. In addition, a children's environmental health resource sampler was developed and put on the National Network of Libraries of Medicine Web site.

AIDS Information Services

NLM has continued its successful AIDS Community Information Outreach Program with 15 new awards in FY2003, bringing the total number of awards made to 172. In addition to these awards, NLM continues to work with other organizations to raise awareness of HIV/AIDS information resources among small community organizations at a grassroots level. A new "HIV/AIDS and Older Adults" webpage is at http://sis.nlm.nih.gov/HIV/HIVOlderAdults.html.

NLM remains as the project manager for the multi-agency AIDS Clinical Trials Information Service (ACTIS) and the HIV/AIDS Treatment Information Service (ATIS), which now have been merged into a new service entitled "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, is now updated by the University of Alaska at Anchorage; the newly released Asian-American Health web site was announced by HHS Secretary Thompson during the celebration of Asian and Pacific Islanders Health Month, and the forthcoming American Indian Health Web site is now undergoing review. These Web sites include relevant policy, legislative, and organizational information as well as organized links to health and environmental issues of concern to the designated population.

NLM-Tox-Enviro-Health-L listserv was created in June 2003 to send announcements about SIS toxicology and environmental health resources. Messages sent to the nearly 500 subscribers include lists of new chemicals added to Hazardous Substances Databank, announcements about the new Household Products Database, and new environmental health topics for consumers added to Tox Town or MedlinePlus.

In FY2003, the Toxicology Information Outreach Panel (TIOP) evolved a new strategic plan and was renamed the Environmental Health Information Outreach Panel (EnHIOP). Dr. Henry Lewis, Dean of the School of Pharmacy at Florida A&M University, became Chair of the new group. The new EnHIOP includes representation from additional Historically Black Colleges and Universities (HBCU's) as well as from Tribal Colleges and Hispanic Serving Educational Institutions. The panel will address a broader spectrum of environmental health issues in the coming years.

SIS continued its health information training programs at national and regional meetings of the National Medical Association. These programs cover all NLM online resources, including TOXNET, PubMed, ClinicalTrials.gov, and Medlineplus.

A more recent addition to NLM's outreach programs is one to improve access to health-related disaster information in three disaster-prone Central American countries: Nicaragua, Honduras, and El Salvador. In FY2003, SIS continued its support of the Regional Disaster Information Center for Latin America and the Caribbean to strengthen the capacity of these countries to collect, index, manage, store, and disseminate public health and medical information related to disasters. The 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 40 conferences in this fiscal year. Several of these provided opportunities for presentations or workshops about NLM information resources. In addition, NLM-SIS hosted the National Congress of American Indians President's Task Force on Health Information, and a national conference on Refugee Health Information. NLM also sponsored the e-health track at the Technology Partnerships Conference (formerly the HBCU/MI Technology Expo) held at the Georgia Centers for Advanced Telecommunications and Technology, Atlanta Georgia.

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. For example:

The **Household Products Ingredients Database** (http://householdproducts.nlm.nih.gov) provides a Web resource for consumers that links brand name household products (more than 4,000) with their ingredient chemicals (more than 2,000) and potential adverse health effects. Information derived from manufacturer's Material Safety Data Sheets and from SIS databases can provide answers to various questions, including: a) what

chemicals are contained in specific brands and in what percentage; b) which products contain specified chemicals; c) who manufactures a specific brand and how can that manufacturer be contacted; d) what are the potential acute and chronic health effects of the chemical ingredients found in a specific brand; and e) what other information is available about such chemicals in the toxicology-related databases of the NLM.

The **ToxTown** (http://toxtown.nlm.nih.gov) project explores how best to provide environmental health information to a general audience**. ToxTown** is an interactive guide to commonly encountered toxic substances, your health, and the environment. It uses color, graphics, sounds and animation to convey connections between chemicals, the environment, and the public's health. Tox Town is designed to provide:

- Facts on everyday locations where toxic chemicals might be found;
- Information about how the environment can affect human health;
- Non-technical descriptions of chemicals;
- Links to authoritative chemical information on the Internet; and
- Internet resources on environmental health topics.

Tox Town helps users explore an ordinary town or city or farm to identify its common environmental hazards. The city, town, or farm can be toured by selecting Location or Chemical links. Locations, like the school, home or office building, can be opened for cutaway views and for detailed information about potentially hazardous chemicals that might be found there, as well as for links to environmental health resources. Tox Town also offers some resources in Spanish.

In FY2003, SIS began beta-testing of **TOXMAP,** a prototype system that uses maps of the US to help users visualize data about chemicals released into the environment. TOXMAP integrates data from the EPA's Toxic Release Inventory 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.

WISER (Wireless Information System for Emergency Responders) is designed to provide critical chemical information quickly and conveniently on a 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, using ATSDR Medical Management Guidelines for Acute Chemical Exposures, which were developed to aid emergency department physicians and other emergency health care professionals who manage acute exposure following chemical incidents. The WISER prototype has focused on approximately 400 agents found in the Hazardous Substances Data Bank, and current deployment plans include a user's guide, a tutorial, evaluation methodology, and "in-field" testing.

ToxSeek provides a virtual meta-search tool for simultaneous searching of target information systems, displaying search results from targeted systems, and harvesting related concepts. This tool can be configured to define a set of target information/search tools, which for SIS are T&EH 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 FY2004.

Other Interagency Initiatives

SIS personnel continued their leadership of the Interagency Tox-to-Consumer Workshop, which convened in June 2003 to finalize plans for an Inventory of Federal Government Consumer Environmental Health Resources.

Evaluation Activities

In FY2003, SIS completed ten Reviews of PDA (Personal Digital Assistant) Applications in Toxicology and Environmental Health. Each review typically covers the following topics: General Information, Intended Users, Data Source/Authorship, Contents, Navigation, System Requirements, License Type/Price, Availability, and Useful Web Links.

With support from an NIH Evaluation Express Award, SIS established online professional and consumer focus groups to evaluate various information products, including HAZMAP, Household Products, and ToxTown.

The NIH Office of Evaluation also provided set-aside funds to NLM for the evaluation of TOXNET, AIDSinfo, MedlinePlus, and the NLM Web site using the American Customer Satisfaction Index.

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, established by a joint resolution of the United States Congress in 1968, is a research and development division of the National Library of Medicine. Seeking to improve access to high quality biomedical information for individuals around the world, the Center continues its active research and development in support of the NLM's mission. The Center conducts and supports research and development in the dissemination of high quality imagery, medical language processing, high-speed access to biomedical information, intelligent database systems development, multimedia visualization, knowledge management, data mining and machine-assisted indexing. An external Board of Scientific Counselors meets biannually to review the Center's research priorities. The most current information about Lister Hill Center research activities can be found at http://lhncbc.nlm.nih.gov/.

Lister Hill Center 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 the NIH, and academic and industry partners. Staff regularly publish their research results in the medical informatics, computer and information science, and engineering communities. The Center is often visited by researchers from around the world.

Lister Hill Center research activities fall into several broad categories. Our training program brings 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 develop and continue to support a number of information systems, all of which are informed by our basic research activities.

The Lister Hill Center is organized into five major components. The work of each is described below, and an organization chart, with the names of Branch and Office Chiefs, is on the inside back cover of this report.

Organization

Computer Science Branch

The Computer Science Branch (CSB) applies techniques of computer science and information science to problems in the representation, retrieval and manipulation of biomedical knowledge. CSB 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. CSB research staff include the team that has developed NLM's Gateway, the team that annually produces the Unified Medical Language System Metathesaurus, and the staff who coordinate the Center's training programs. Staff members participate in the meetings of the Internet Engineering Task Force. CSB staff also coordinate the many training activities of the Center. The most current information about the Computer Science Branch can be found at http://lhncbc.nlm.nih.gov/csb/.

Cognitive Science Branch

The Cognitive Science Branch (CgSB) 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, as is the newly released Genetics Home Reference, which provides 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/.

Communications Engineering Branch

The Communications Engineering Branch (CEB) 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 developed 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 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. CEB 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/.

Audiovisual Program Development Branch

The Audiovisual Program Development Branch (APDB) conducts media development activities with several specific objectives. As its most significant effort, the branch participates in the 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 image creation, editing, enhancement, transfer and display. Consultation and materials development are also provided by the branch for the NLM's other information programs. From applications of optical media technologies and teleconferencing to support for world wide Web distribution, the requirement for graphics, video, and audio materials has increased in quantity and diversified in format. In addition to the development by the staff of new techniques and processes, the facilities and hardware infrastructure must reflect state-of-the-art standards in a rapidly changing field. Included within the Branch is the Office of the Public Health Service Historian. The office preserves and disseminates information about the history of Federal efforts devoted to public health. 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 (OHPCC) serves as the focal point for 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 coordinates the interagency HPCC R&D program. Office staff serve as NLM's liaison to scientific organizations at all levels of 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 OHPCC can be found at http://lhncbc.nlm.nih.gov/ohpcc/.

Training Opportunities at the Lister Hill Center

Working towards the future of biomedical informatics research and development, the Lister Hill Center provides training and mentorship for individuals at various stages in their careers. The Medical Informatics Training Program, ranging from a few months to more than a year, is available for visiting scientists and students. Fellowship programs may be as short as eight weeks or as long as one year, with the possibility of being renewed for a second year. Each fellow is matched with a mentor from the research staff. At the end of the fellowship period, fellows prepare a final paper and make a formal presentation open to all interested members of the NLM and NIH community.

In FY2003 we provided training to 48 participants from 18 states and 6 countries. 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 information systems, knowledge based research, natural language processing, ontology research, palm technology, semantic Web research, text mining, distance education, and visualization. We continue to offer a successful NIH Clinical Elective in Medical Informatics for third and fourth year medical students. The elective provides an overview of the state-of-the-art of medical informatics in a lecture series by nationally and internationally known speakers, and offers an opportunity for independent research under the preceptorship of expert NIH research staff. We maintain our focus on diversity through our participation 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.

Established in 2001, the NLM Rotation Program continues to grow. The eight week rotation program for existing NLM Medical Informatics Trainees provides fellows an opportunity to learn about National Library of Medicine programs and current Lister Hill Center research. The rotation includes a series of lectures and the opportunity for students to work closely with established scientists and meet fellows from other NLM funded programs.

Additional information about Lister Hill Center training opportunities is available at the Center's Web site under "Training Opportunities." Interested individuals will find descriptions of each of the training programs including specific application procedures.

Language and Knowledge Processing

The Lister Hill Center conducts and supports research in language and knowledge processing to extract usable and meaningful information from biomedical text. Natural Language Processing Research investigates how to build and improve systems that understand the meaning of human language in order to mediate between the questions of users and the information systems they seek to use. The successful integration of Lister Hill Center developed research 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.

Developing SPECIALIST, an experimental natural language processing system for the biomedical domain, is the focus of the Center's natural language processing work. 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 addresses 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

The Lexical Systems group builds and maintains the SPECIALIST lexicon, a large syntactic lexicon of medical and general English terminology released annually with the UMLS Knowledge Sources. New lexical items are continually added to the lexicon using a lexicon building tool developed and maintained by the lexical systems research team. LexBuild allows researchers to enter items directly into a central database via a Web browser. As items are created, the Java-based system eliminates time consuming uploading and unnecessary errors. New items are flagged for review when they are ready for release. The SPECIALIST Lexicon increased by over 12% to 183,000 lexical items in the FY2003 release.

The SPECIALIST lexicon records the spelling variation inherent in English orthography, and this, together with a set of spelling suggestion rules and techniques, has been incorporated into the lexical access tools, which are freely available to the research community. Several additional modules developed by the Lexical Systems group have been completed recently and are now available independently as tools for a variety of natural language processing projects. These modules include a tokenizer, a lexical look-up utility, and a noun phrase extractor. Lexical access tools, including LVG, warding, and norm, are also distributed with the UMLS.

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 used to address a variety of problems in biomedical informatics. MetaMap maps noun phrases in free text to concepts in the UMLS Metathesaurus. The MetaMap Technology Transfer program (MMTx) is an exportable, Java-based version of MetaMap that runs under Windows, Mac OS X or Unix/Linux and is provided as a resource to the bioinformatics community. MMTx allows users to exploit the UMLS MetamorphoSys program to exclude or reorder specific Metathesaurus vocabularies. Users also are able to create MMTx data files independently of the UMLS. MetaMap Technology Transfer source code is included in the MMTx release, and an error reporting and tracking system ensures that problems reported by users are effectively addressed.

The development of SemRep, a tool that uses the Semantic Network to determine the relationship asserted between concepts developed in MetaMap, underlies the increased understanding of viable strategies for effective natural language processing. SemRep serves as the basis for ongoing research initiatives in biomedical information management such as projects for extracting medical and molecular biology information from text, processing clinical data in patient records, and research in knowledge summarization and visualization. Recent enhancements to SemRep's linguistic coverage include the addition of a mechanism for interpreting hypernymic propositions. A modification of SemRep, called SemGen, is being developed for identifying and extracting semantic propositions on the causal interaction of genes and diseases from MEDLINE citations. Project staff are also

developing methods for automatically suggesting appropriate images as illustrations for anatomically oriented text.

Word-sense ambiguity in language constitutes a major impediment to accurate management of biomedical text through automatic strategies. 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 to 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 project seeks to identify genes, gene products, and gene functions in abstracts and compares this information to that found in the Gene Ontology. Another project supports comparison of protein function by identifying protein-protein interactions in text. A third project uses semantic information to support text-based knowledge discovery systems in molecular biology.

Indexing Initiative

The Indexing Initiative investigates concept-based indexing methods for the automatic selection of subject headings in both semi-automated and fully automated indexing environments at the NLM. The goal of the Indexing Initiative is to obtain retrieval performance equal to or better than performance of systems using manually assigned index terms. A prototype indexing system for testing indexing methods, the Medical Text Indexer (MTI), is being tested by NLM indexers. MTI is based on three core indexing methodologies. The first methodology calls on the MetaMap program to map citation text to concepts in the UMLS Metathesaurus. The second methodology, the trigram phrase algorithm, uses character trigrams to match text to the Metathesaurus. The third methododology uses a variant of the PubMed related articles algorithm to find MeSH headings by using existing indexing terms on articles similar to the input text. Results from the three methods are restricted to MeSH and combined into a ranked list of recommended indexing terms. Substantial progress has been made in applying the MTI system to both semi-automated and fully automated indexing environments at the NLM. MTI recommendations are now available to all indexers. In addition, results of the MTI system have recently been assigned as keywords for AIDS/HIV, health sciences research, and space life sciences collections of meetings abstracts that will not be manually indexed. These collections are accessible via the NLM Gateway. As part of the research underlying the Indexing Initiative, the Journal Descriptor (JD) project investigates fully automated indexing based on the 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 journal descriptors with words in journal titles and abstracts in a two-year training set of approximately 910,000 MEDLINE records. Each record "inherits" the JDs from the journal title in the record, and then each word in the training set can be described by a list of JDs ranked according to the number of co-occurrences between the word and the JDs.

Unified Medical Language System

Unified Medical Language System research regularly develops and distributes multi-purpose, electronic knowledge sources and associated lexical programs. Products such as the Metathesaurus, Semantic Network and SPECIALIST Lexicon are used by system developers to enhance patient data, create digital libraries, retrieve Web and bibliographic data, apply natural language processing, and improve decision support. The Metathesaurus represents multiple biomedical vocabularies organized as concepts in a common format providing a rich terminology resource in which terms and vocabularies are linked by meaning. The Semantic Network allows users to investigate relationships among semantic types and relations and retrieve a list of Metathesaurus concepts assigned to a particular semantic type. Finally, the data in the SPECIALIST Lexicon provides users with the syntactic and morphologic information about each of its lexical items.

The Metathesaurus continues to grow in size, scope, and mission. As of FY2003, there are more than 900,000 concepts with 2.5 million names from 102 source vocabularies in 15 languages. The scope of the Metathesaurus is also growing to include the current and candidate HHS standard vocabularies under the Health Insurance Portability and Accountability Act (HIPAA) in a common format and with increasing interconnections. The Metathesaurus' mission has grown to include the distribution of HIPAA vocabularies for clinical use in the United States.

In July 2003, DHHS Secretary Tommy Thompson announced the government-wide license for the SNOMED Clinical Terms (SNOMED CT), a key clinical vocabulary. Under the government-wide license, SNOMED CT will be distributed in the Metathesaurus and will be freely available for all U.S. health care systems. SNOMED CT contains 344,000 concepts with 913,000 names and 1.3 million relationships. The resulting increased visibility and national expectations for the Metathesaurus have added new demands for quality, currency, and the incorporation of additional vocabularies and mappings. In order to uphold NLM quality standards, all new and updated vocabularies are required to be demonstrably complete, contain full attribution, and be

correct in the Metathesaurus. The new "Source Transparency" requirement has led to a major redesign of the Metathesaurus editing, production, and release management systems. A new Rich Data Format (MR+) for transparent releases has been defined and is targeted for a first release with SNOMED CT early in 2004.

Work on the NLM's new clinical drug vocabulary RxNorm and on the Gene Ontology will continue until the 2003AC release of the UMLS Metathesaurus. Significant efforts in FY2003 have made possible the creation and editing of RxNorm within the Metathesaurus editing system.

Collaborating with researchers from the University of Amsterdam, staff have developed an interactive editing interface for the International Classification of Primary Care (ICPC) medical vocabulary. The ICPC contains concepts in 20 languages including Hebrew, Japanese, Russian, and Greek with their character sets represented in Unicode. A platform-independent Web-based system using the open source tools Apache/PHP/MySQL has been developed. The Web-based ICPC system allows multilingual display and editing for clients that have no Unicode capability by means of a Java applet and server-side Unicode manipulation.

The UMLS Knowledge Sources are made available over the Internet through the Knowledge Source Server (KSS). The KSS incorporates several features that allow fast and direct access to UMLS data. For example, users can request information about a particular concept in the Metathesaurus, including definitions, semantic types, and synonyms as well as other concepts that are related to the input term. The KSS 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, data in the SPECIALIST lexicon provides users with syntactic and morphologic information about each lexical item. The most recent release of the KSS 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 architecture includes a Web server implemented as a collection of Java servlets that provide quick and easy access to UMLS data.

The KSS server software connects through the Internet to a backend Remote Method Invocation 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 to develop the system.

In addition to enhancements to the user interface, XML has been incorporated into the design of the KSS 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 an object-oriented fashion within their programs. The object model provides a mechanism for representing concepts and related data consistently among developers.

The Terminology Server provides tools to manage diverse medical vocabularies for various purposes. Throughout FY2003, the project achieved a significant milestone in providing customized vocabulary data sets to ClinicalTrials.gov, Profiles in Science, and Genetics Home Reference. An important function of the Terminology Server is to allow individual users to customize terminologies from the UMLS and other sources. Filters are being developed that will help users select subsets of medical terms. The first filter will identify UMLS term variants suitable for natural language processing. Another task of the Terminology Server is to develop models for handling UMLS data retrieval, data maintenance and periodic data updates. In addition, research is focused on developing tools to create and edit "local," non-UMLS terminologies. The project will continue to integrate tools with existing applications and provide updates to the application data sets corresponding to the latest releases of the UMLS.

Medical Ontology Research

Medical Ontology research focuses on the development of a medical ontology that will enable various knowledge processing applications to communicate with one another. Creating a usable ontology requires the definition, organization, visualization, and utilization of semantic spaces created from biomedical knowledge processing applications. Although the UMLS is used as the primary source of medical knowledge, OpenGALEN, Gene Ontology, and WordNet are being explored as well. During the past year, research focused on two subdomains of biomedicine: ontology and terminology. In one project, the representation of anatomical concepts in the Foundational Model of Anatomy and GALEN were compared. In other projects, the research team studied terminological differences across clinical records and biomedical literature. Redundancy in hierarchical relations in the UMLS was also investigated. Finally, the team developed methods for navigating between phenotype and genotype information as well as visual methods for exploring the UMLS semantic groups.

Image Processing

The Lister Hill Center performs extensive research and development in the capture, storage, processing, retrieval, transmission, and display of biomedical documents and medical imagery. Areas of active investigation include image compression, image enhancement, image recognition and understanding, image transmission, and user interface design.

Visible Human Project

The Visible Human Project (VHP) 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 VHP data sets are available through a free license agreement with the NLM. Data sets are distributed to licensees over the Internet at no cost and on DAT tape for a duplication fee. Worldwide use of the data sets continues to grow as they are applied to a wide range of educational, diagnostic, treatment planning, virtual reality, virtual surgeries, artistic, mathematical, and industrial uses by over 1800 licensees in 47 countries. FY2003 saw the continued maintenance of two databases to record information about Visible Human Project use. The first database logs information about VHP license holders and records their plans for using the images. The second database records information about the products that licensees are developing.

FY2003 has seen the development of certain anatomical applications that are focused on allied health, undergraduate medical and dental education, and continuing medical education. The concept has been to develop applications that bring visible human data into the classroom and laboratory teaching environments. A second focus has been the final period of performance of the Anatomical Methods contracts designed to address mitigation of method-based artifact generation, and to increase the resolution for detection, discrimination and identification of minute neurovascular structures. These studies have been conducted at University of Colorado and Brigham and Women's Hospital.

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, extra-ocular 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.

The Visible Human Dissector was added to the atlas Web site in FY2003. The Dissector provides access to 3D renderings of Visible Human anatomy as a virtual cadaver. The virtual cadaver includes identified anatomical structures and the cross-sections from which they were derived. The Dissector can be used as a free-form reference or navigated with user-created lesson plans. This unique application provides the ability to approach the human body from any combination of traditional views, including cross-sectional, regional, systemic, clinical, surface and surgical anatomy perspectives. Additional content in the basic anatomical structures involving small dimension neuromuscular connections, clinical cases, and surgical approach discussions to certain relevant procedures was also added to the atlas Web site. Certain muscles that were segmented in the Mastication module were deformed using spline technology to mimic normal musculoskeletal system function.

Two groups are investigating advanced anatomical methods using the Visible Human data with research support from the NLM. 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 de-mineralized 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 appear 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.

The Insight Toolkit (ITK), a research and development initiative under the Visible Human Project, began official releases of the software during FY2003. ITK makes available a variety of open source image processing algorithms for computing segmentation and registration of high dimensional medical data on a variety of hardware platforms. Platforms currently supported are PCs running Visual C++, Sun Workstations running the GNU C++ compiler, SGI workstations, Linux based systems and Mac OS-X. A consortium of university and commercial groups is executing this work. The consortium includes General Electric Global Research, Kitware, Insightful, the University of North Carolina Chapel Hill, the University of Pennsylvania, the University of Utah, Harvard University, the University of Pittsburgh, and Columbia University.

FY2003 saw explosive growth of the ITK research community. Accompanying the official ITK 1.0 software release, NLM made additional awards to exercise and integrate the software infrastructure into clinical and research applications. Research institutions including the Mayo Clinic, the Carnegie Mellon University Robotics Institute, Georgetown University Medical School, Imperial College London and Guys Hospital, have joined the development team. Non-funded researchers from across the world are now testing, developing and contributing to ITK in over 30 countries. At the end of FY2003, ITK v1.4 was released, including the components developed as part of the funded research supported by NLM. By the end of this fiscal year, we will have attained our primary goal of creating a strong, usable, public, open-source software infrastructure to support medical imaging research.

3D Informatics

During FY2003 the 3D Informatics Program has continued to mature and develop its in-house research efforts around problems encountered in the world of 3-dimensional (x,y,z and x,y,t) imaging. Research is continuing in the areas of image-based implicit rendering, research and systems trials for ITK, and haptic latency analysis for surgical simulation. We have extended and enhanced our pilot project for creating the framework for an archive of volume image data, the National Online Volumetric Archive. This project includes the physical implementation of the pilot archive for volume image data, as well as a tutorial for data submission, metadata structure management tools using XML, and Web page structure.

Research is continuing on template guided interventions, including joint work with the USUHS Orthopedics Department and the Bethesda National Naval Medical Center Radiology Department on total hip resurfacing arthroplasty and the planning and fabrication of surgical templates with NNMC. A surgical planning workstation and custom-built drill template was devised from CT scans to accurately place guide pins in the femur head during total hip resurfaces. This work is related to our previous work in Patient Specific Surgical Instrumentation for spine surgery. Efforts are also continuing in the area of data-driven modeling with implicit surfaces with colleagues at the University of Maryland, Baltimore County (smooth surface generation from binary volumes) and the University of North Carolina, Charlotte (reconstructing implicit surfaces from contours from arbitrary ultrasound slices). Finally, we have begun explorations in artistic and non-photorealistic rendering of digital models. A project in laser scanning of physical artifacts was undertaken in FY2003 as well as the software design of a layered architecture for implementing medical illustration techniques using computer graphics technologies.

AnatQuest

AnatQuest provides widespread access to the Visible Human images for a broad range of users, including the lay public, who are frequently limited to low speed Internet connections. It builds on earlier projects which focused on developing an object-oriented database for the images, establishing an FTP server for access to the high-resolution version of the images, and developing tools for processing the images. The AnatLine system developed earlier allows access through anatomic terms to high resolution cross-sectional images and segment masks (useful for rendering anatomic objects). The tools developed to use AnatLine are VHParser and VHDisplay. The first is for unpacking the data files into their individual components (cross-section images, byte masks, coordinate and label tables, etc.), and VHDisplay is for displaying both cross-sectional and rendered images.

The new system, AnatQuest, is a Web system based on a 3-tier architecture in which the first tier consists of Java applets for displaying thumbnails of the cross-section, sagittal and coronal images of the Visible Human Male, from which detailed (full-resolution) views are accessed. The second tier is a set of servlets that process user requests and compress the requested images prior to shipment back to the user. The third tier is the object-oriented database. 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. In addition to its main purpose, AnatQuest serves as an access point for AnatLine as well as for about 300 surface-rendered objects, the majority of which were created at the Lister Hill Center and the rest acquired from outside sources (e.g., VoxelMan). Also through AnatQuest, the public can access the FTP server for bulk transfer of high resolution image files.

An initial prototype of a system linking MedlinePlus to the anatomic image database has recently

been developed. This system relies on a proxy server developed to intercept user requests to MedlinePlus. The proxy server first retrieves the MedlinePlus page that satisfies the user query. In parallel, the proxy server sends the user query to the AnatQuest image server which uses the UMLS Knowledge Source Server and a term-mapper module to map the query terms (mostly disease terms) to the corresponding anatomical structures. The term-mapper module addresses the problem of identifying appropriate anatomical terms corresponding to the biomedical terms in the document. These biomedical terms are likely to be disease terms rather than explicitly anatomical ones. The module uses the location of concept relationship in the UMLS Metathesaurus to map a biomedical term to a related anatomical term. For example, the term "pneumonia" (a disease) could be mapped to "lung," the underlying organ for the disease. The links to these images are then inserted by the proxy server as hotlinks in the image section of the MedlinePlus page, which is then returned to the user.

In addition to the Web-mediated version of AnatQuest, a kiosk version was developed for the Dream Anatomy exhibit at NLM as a Java application suitable for onsite patrons using a touchscreen monitor. To eliminate dependency on the network for the retrieval of VH images, and to speed up the kiosk operation, the images were also stored in the local machine. The effort to redesign the GUI in the AnatQuest Web-mediated system for the kiosk application has been to tailor the displayed icons, buttons and other graphical elements to allow convenient human interaction via a touchscreen.

WebMIRS

The Web-based Medical Information Retrieval System (WebMIRS) allows 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 during the years 1976–1980 and 1988–1994, respectively. The NHANES II database, accessible through WebMIRS, contains records for about 20,000 individuals, with about 2,000 fields per record; the NHANES III database contains records for about 30,000 individuals, with more than 3,000 fields per record. In addition, 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 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 allows the user to save the returned data to the local disk drive, where it may be analyzed with appropriate statistical tools such as the commercially available SAS and SUDAAN software. The WebMIRS NHANES II database also contains vertebral boundary data that was collected by a board-certified radiologist for 550 of the 17,000 x-ray images in WebMIRS. These data consist of *x,y* coordinates for approximately 20,000 points on the vertebral boundaries in the cervical and lumbar spine images. Users may do queries for both radiological and/or health survey data. An example of this type of query is: "Find records for all persons having low back pain (health survey data) AND fused lumbar vertebrae (radiological data)". The boundary data points are displayable on the WebMIRS image results screen and may be saved to the user's local disk.

WebMIRS enhancements done this year include collaborative work with Texas Tech University to develop an advanced compression capability custom tailored to the image characteristics of the x-ray images, to allow delivery of the WebMIRS images in compressed form rather than in the low-resolution form as at present. Software written in Java has been developed for decompression at four different levels. Project staff have begun the implementation of new design architecture to provide a software framework for the incorporation of new text/image databases in a much more general way than the current WebMIRS, and to provide new features for the database end user that extend current WebMIRS capabilities.

Digital Atlas of the Cervical and Lumbar Spine

This 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 17,000 images collected in the NHANES II survey. For the cervical spine images, the Atlas contains numerical interpretations or "grades" for anterior osteophytes and disc space narrowing, on a scale from 0-3, with 0 being "normal" and 3 being "most abnormal"; and also interpretations for subluxation, on a 0-1 scale, with 0 being "normal" and 1 being "abnormal." Similarly, for the lumbar spine images, the Atlas contains interpretations for anterior osteophytes and disc space narrowing, on a scale from 0-3. The Atlas user may display single or multiple images in order to view, for example, all grades from normal to most abnormal of anterior osteophytes in the cervical spine. Image processing capability is provided to assist in contrast enhancement for viewing of detail. The Atlas may be accessed either as a Java applet, or downloaded as a Java application, from the project Web site. In addition, we provide a version of the Java application on CD that allows the user to add his/her own images (either grayscale or color) in a special "My Images" section, and to annotate and title those images for later use. This year the Atlas was enhanced by the addition of capabilities to display color images, add extensive text annotations, and import/export sets of images and annotations as a package.

Online X-ray Archive

The complete set of 17,000 NHANES II x-ray images in the full-resolution form in which they were digitized are available by FTP and have been accessed by researchers from around the world. For viewing the x-rays, we have created the ImViewJ software, a Java application that may be downloaded from our Web site and which allows the viewing of the images at their full spatial resolutions (1463x1755 for the cervical spine images, 2048x2487 for the lumbar spine images). For 550 images we also have coordinate data collected under the supervision of a radiologist at Georgetown University. This coordinate data defines landmark points for each vertebra in a manner commonly used in the field of vertebral morphometry, and serves as reference data to aid in creating and evaluating the performance of image processing algorithms for segmentation of the vertebrae. This coordinate data is publicly available on the FTP site along with TIFF 8-bit versions of the corresponding x-ray images. Users may access this coordinate data either through the FTP archive or through the WebMIRS system.

Content-Based Image Retrieval

The Content-Based Image Retrieval (CBIR) project develops methods for effective extraction of biomedical information from digital images of the spine. CBIR focuses on the computer-assisted indexing of image data, as well as the ability to search image data. Computer-assisted image searching is a potential enabler of enhanced information extraction from a database that has already been indexed. The most popular 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 from a set of choices provided by the system or by providing a new image, and queries the database with respect to one or more characteristics of the example image (e.g., shape, histogram, or texture). In query by sketch, the input image is replaced by a sketch of the image made using drawing tools provided by the system. In either case, the system analyzes the input into component features and searches the database for images with similar features. Results are usually returned as a similarity ranking.

Developing a computer-assisted indexing system poses many challenges. For example, the only indexing data available for the NHANES II images is the collateral (alphanumeric) data collected in questionnaires and examinations. There is no indexing information available that has been derived directly from the images. The prohibitive cost of employing radiological experts to compile and interpret indexing data means that it is unlikely that NHANES II indexing information will ever be acquired manually. However, indexing data might be acquired if reliable, biomedically validated software could automatically produce image interpretations. Even the development of semi-automated methods could sufficiently reduce labor costs to allow the creation of databases of significant biomedical information. CBIR research seeks to develop computer-assisted image indexing to acquire data, and at the same time reduce the costs of indexing.

An initial prototype Content-Based Image Retrieval system (CBIR1) was implemented in FY2001 for the retrieval of images based on simple vertebral shape models. The program allows users to specify a search for up to nine control points and the geometric configuration of these points to define an approximate vertebral shape. The prototype database contains 100 cervical and lumbar images with the ability to rotate and scale every 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. In FY2003, a second CBIR prototype (CBIR2) was implemented. CBIR2 is significantly enhanced, including an *indexing function* with the capability to perform active contour segmentation, create detailed representations of vertebrae boundaries, and to convert boundaries into multiple shape representations (e.g., global shape descriptors, invariant moments, polygon turn functions, and Fourier descriptors). In addition, a retrieval function supporting the retrieval of shapes by any of the shape representations was implemented. CBIR2 also includes NHANES text data and supports query by sketch, image example, or text, in addition to hybrid text and image-based queries. The MySQL database system was incorporated into the retrieval function for the storage and retrieval of text data. Current CBIR work is directed towards the completion of the 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 the creation of a database of segmented vertebrae of significant size and accuracy to serve as test-bed data for ongoing CBIR 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 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 test-beds 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 archives a variety of image content. Most of the machines housed in the laboratory are equipped with multiple networking ports (e.g., FDDI, ATM, Ethernet, fast Ethernet) which allow, in addition to standard networking capabilities on the local Ethernet, the capability of alternate physical communications channels. ATM switches connect the Ethernet and FDDI networks to other local area networks throughout the Lister Hill Center, the Internet, experimental ATM, Abilene, and the infrastructure for the Next Generation Internet and Internet2 initiatives.

The Document Image Analysis Test Facility is an off-campus facility containing high-end Pentium workstations and servers for the MARS production system. While routinely used to produce bibliographic citations for MEDLINE, this facility also serves as a laboratory for research into techniques that are fundamental to the automated extraction of descriptive metadata for the long term preservation of document images. Techniques include automatic zoning, labeling, and reformatting of bibliographic fields from document images, as well as intelligent spell-check by pattern recognition and other key elements of MARS. Besides real time performance data, the Document Image Analysis Test Facility also collects and archives 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

Multimedia research and development efforts concentrate on the engineering of technical improvements applied to issues such as image quality and resolution, color fidelity, transportability, storage, and visual communication. In addition to developing new methods and processes, Lister Hill Center facilities and hardware infrastructure reflect state-of-the-art standards in the rapidly changing field of multimedia research and development. High definition video, for example, is being used as the future for improved electronic image quality. Multimedia systems, scientific visualization and networked media are being pursued for their performance, educational, and economic advantages. Project staff conduct research in three dimensional computer graphics, innovative animation techniques, and photorealistic rendering. Research into digital video and image compression techniques is being used in projects requiring the storage of large images and rapid data transmission. Project staff completed an evaluation of one of the leading digital media asset management systems designed for the video environment in FY2003. The system includes a video database fully integrated with a logging and digitizing system. Based on the evaluation results, the system vendor's engineers have been working with Center staff to redesign the software.

CD-ROM, DVD and DVD-ROM technology for capturing media assets including video, audio, Web information, and computer text slides continue to be explored. Web links within these assets are used for updating program content and providing links to additional information tools (e.g., PubMed). A template allowing the simultaneous viewing of multiple interactive windows, including speaker video, slides, and an interactive index was developed to improve access to program content on CD-ROM, DVD and DVD-ROM technology. By selecting any one slide from the index, two other windows immediately synchronize to that point in the presentation. Using the new template, project staff developed two prototypes: a Board of Scientific Counselors meeting as a CD-ROM and the 2002 Leiter Lecture ("Genomics, Medicine and Society" with Dr. Francis Collins) as a CD-ROM and as a DVD.

Three prototype DVDs representing the Once and Future Web Exhibition were developed in FY2003. The initial prototype DVD demonstrated overall design concepts, the implementation of Web-enhanced DVD technology, a navigable text viewer and virtual object manipulation within the program. Source media for the DVD included still imagery, three-dimensional video graphics and high definition video. Feedback on the content, organization and design was incorporated into the second prototype. A new segment of the program focusing on the NLM Telemedicine Program, including video content, has been added. MPEG encoding, DVD authoring and Web DVD programming were ongoing throughout the development process. The third DVD, Prototype version 1.1, was demonstrated in FY2003. Additional feedback on interface design, navigation and content will be integrated into the final design and production of the DVD. Additional enhanced video graphic animations have been completed and included in the "Show Me How it Works" section of the program. The identification and integration of Web sites to enhance fixed DVD content was finalized and incorporated into the program's companion Web portal. Investigation of Web-enhanced DVD programming for delivery on multiple platforms has been a key element in the overall development of all prototypes.

In consultation with the Office of Communications and Public Liaison and the HMD Exhibition Program, project staff have been working with MacNeil/Lehrer Productions on planning the developmental phases of a Web-enhanced DVD for the NLM exhibition, Changing the Face of Medicine: Celebrating Americas Women Physicians. The DVD is to serve as the prototype for the subsequent Local Legends DVD, a collaborative project between the NLM and the American Medical Women's Association. Video interviews of 12 physicians identified for inclusion in the first DVD have been conducted in 12 cities, and they include Dr. Tenley Albright, Dr. Julie Louise Gerberding, Dr. Donna Christian-Christensen, Dr. Nancy Snyderman, and Col. Rhonda Cornum. These interviews are part of an

up-close and personal interactive video profile of each doctor. A youth mentoring program in California was video recorded and will be included in the prototype. All was recorded on high-definition video to assure high quality for production and archive purposes. Overall interface and navigational designs were developed, and include transitional segments featuring young adults.

Information Systems

The Lister Hill Center performs extensive research in developing advanced computer technologies to facilitate the access, storage, and retrieval of biomedical information.

Digital Library Research

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

The Profiles in Science Web site uses innovative digital technology to make available the manuscript collections of prominent biomedical researchers, medical practitioners, and those fostering science and health. Database content is created in collaboration with the History of Medicine Division, which processes and stores the physical collections. Most collections have been donated to the NLM and contain published and unpublished materials, including books, journal volumes, pamphlets, diaries, letters, manuscripts, photographs, audio tapes and other audiovisual resources.

The collections of Donald S. Fredrickson, Fred L. Soper and Florence R. Sabin were added in FY2003, bringing the total number of archives for prominent biomedical researchers, medical practioners, and those fostering science and health to eleven: Christian B. Anfinsen, Oswald T. Avery, Julius Axelrod, Donald S. Fredrickson, Joshua Lederberg, Barbara McClintock, Marshall W. Nirenberg, Linus Pauling, Martin Rodbell, Florence R. Sabin, and Fred L. Soper. The Reports of the Surgeon General (1964–2000) and the history of the Regional Medical Programs (1964–1976) are also available on Profiles in Science.

In FY2003, project staff continued to enhance the effectiveness of Profiles in Science. A new user interface was developed to standardize consumer navigation, a link to the Reports of the Surgeon General was added, and three new categories were established, Biomedical Research, Health & Medicine, and Fostering Science & Health. Enhancements to the underlying Profiles in Science digital library infrastructure include improvements to the back-end database and the development of new methods for viewing database data, detecting and correcting errors, and automatically updating data. Finally, the development of an XML-based front end and transition to a new XML-based search engine continue to be pursued.

The Lister Hill Center collaborates with the History Office of the Food and Drug Administration and the NIH Historical Office on preservation efforts for the Public Health Service. An exhibit on the history of smallpox, on display at the NLM in 2002, was developed into an online exhibit for the NLM Web site in FY2003. Staff worked to develop a database of resources on the history of African-Americans in medicine and also conducted research on the history of the PHS's involvement in National Negro Health Week. Staff continue to answer historical queries about the history of the PHS and actively work to preserve documents and artifacts related to PHS history.

MARS

Document image analysis and understanding research combined with database design, GUI design for workstations, image processing, string pattern matching, lexical analysis, speech recognition and related areas underlie our 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, institutional affiliations and the abstract. Research investigations center on the identification of rules for algorithms for page segmentation, zone labeling, OCR error correction, affiliation ranking and other essential functions. Manual input is limited to entering fields other than the ones automatically extracted, as well as verifying the text before the records are made available to indexers. In FY2003, research focused on improving the operation of MARS, developing a system to extract bibliographic data from online journals, and enhancing the processing of publisher-supplied citations in XML format. The work in this project also contributes to the automatic extraction of metadata from electronic resources that need to be archived for preservation purposes.

Anticipating an increasing availability of online (Web-based) journals in the future, we conducted research toward the design of a system that automatically extracts MEDLINE citation data from such journals. Areas of investigation included such techniques as breadth first search algorithm and constraint satisfaction methodology, fuzzy rule-based methods, format conversion methods (to convert PDF to HTML) and Web-based GUI design tradeoffs. Based on this research,

modules were created to download issues and articles from the Web, zone and label the relevant text, implement MEDLINE conventions in reformatting the title and author fields, and selecting the correct affiliation from the many that usually appear. Specific developments included a journal-specific learning algorithm to automatically search for and download journal issues and articles, and to classify them as HTML or PDF documents; fuzzy rule-based algorithms to automatically zone and label the bibliographic data; automatically convert PDF documents to HTML for processing; and a stress-test study to investigate the adequacy of the existing cluster-based failover system in MARS to include future online journal processing.

The modules outlined above were integrated to create a prototype system called WebMARS, which was tested to automatically extract data from online journals, with and without publisher-supplied XML data. The latter function is of interest because the publishers who send NLM their XML-coded citations usually leave out Grant Numbers, Databank Accession Numbers and other fields requiring manual entry at NLM. This function requires matching the XML journal title with the title in the Web page, looping through each issue to match articles by author names and article titles, and computing a confidence value. Tests with 3,000 XML citations and 300 journal issues on the Web are planned toward identifying a threshold for this confidence value, beyond which the online article and the XML citation can be reliably linked for further processing. Performing this function in addition to creating citations automatically from Web journals, WebMARS augments the citations sent in by the publishers, and thereby reduces the labor required in production. The prototype system was tested under realistic conditions to estimate the labor required to create complete citations as a comparison to the labor required in the alternative approaches: keyboarding, MARS, and publisher-supplied XML data, and found to be significantly more efficient than these.

Ground truth data released to the public

In August 2003, a database named Medical Article Records Groundtruth (MARG) was released for research in document image analysis and understanding techniques by the computer science and informatics communities. The data consists of over 1,000 bitmapped images of the first pages of articles from biomedical journals indexed in MEDLINE falling into nine layout types encountered in MARS production. Included in addition to the page images are the corresponding segmented and labeled zones, OCR-converted and operator-verified data at the zone, line, word and character levels, all in XML format. Also available from this Web site is Rover, an analytic tool that may be used to compare the results of a researcher's program with the ground truth data. A paper describing MARG was presented and published at the Symposium on Document Image Understanding Technologies (SDIUT 2003) held in April. In the first two weeks after it was announced at an international conference in document analysis and recognition (ICDAR 2003, Edinburgh), MARG was accessed by more than 1,750 registered users.

DocView

The DocView project enables users to store, view, manipulate, copy, paste, email and print bitmapped images delivered through the internet. The program also serves as a TIFF viewer for compressed images. Users may receive document images via Ariel FTP or Multipurpose Internet Mail Extensions protocols. Library patrons, for example, often use DocView to receive scanned journal articles from libraries that use Ariel software for interlibrary loan services. The number of DocView registered users increased 21% in FY2003 (14,500 users in 181 countries). DocMorph provides additional functionality for DocView by providing the technology for users to convert files into various formats. The system enables users to convert more than 50 different file formats to PDF. Also, by combining OCR with speech synthesis, DocMorph assists the visually impaired in using library information. DocMorph continues to be used by librarians for the blind and physically handicapped to convert documents to synthetic speech that is recorded onto audio tapes. The number of DocMorph registered users increased 40% to 8,000 in FY2003. Research on DocMorph usage and the availability of new technology have pointed out new opportunities for improvements and innovations. The availability of Simple Object Access Protocol (SOAP) that combines XML with HTTP has allowed us to create a Web service that significantly improves the DocMorph function used 75 percent of the time, viz., the conversion of files to PDF. This Web service (MyMorph) consists of a Windows-based client software and modifications to DocMorph for accommodating SOAP. Inhouse testing has shown that MyMorph significantly improves user productivity compared to the (conventional) use of DocMorph through a Web browser, particularly for users who need to convert large numbers of files to PDF. This is accomplished by reducing the time required for users to interact with the software. Test results show that MyMorph reduces the user interaction time from hours to seconds for all users regardless of their Internet connection speed. This new capability is finding frequent use by document delivery librarians, and also by organizations that have used it for mass file migration.

Turning The Pages Information Systems

Turning the Pages Information Systems research seeks to design more efficient methods to translate paper volumes from the NLM historic collection to electronic form, extend the virtual books into information systems, and to increase the accessibility of historical documents for the public. In 2001–2002, the NLM and the British Library collaborated in the production of two virtual books, Blackwell's *Herbal* and Vesalius's *Anatomy in*

Photorealistic to create the "Turning the Pages (TTP)" format. The pages of the two books were scanned into the computer as high quality color images. The images were manually processed by Adobe Photoshop, animated by Macromedia Director, and displayed on a touch screen monitor. Consumers were able to "touch and flip through" each book on a touch screen monitor. After the initial development of the TTP format, research began to transform the initial TTP design into a usable information system (TTP+). Research focused on a "discovery" and a "storyline" model as directions for TTP+. The TTP+ version of Blackwell's *Herbal* uses the "discovery" model, retaining 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, users can go to various search engines (PubMed, ClinicalTrials.gov, USDA, etc.) and obtain citations or general information on St. John's Wort. The TTP+ version of Vesalius' *Anatomy in Photorealistic* uses the "storyline" model and contains images from other sources (e.g., rendered Visible Human images, pictures of Italian cities, etc.). Images are interlinked to present the consumer with several multimedia "stories," including *Man of Padua* and *Modes of portraying anatomy*.

Paré's *Surgical Treatise* and Gesner's *Animalium* were selected for TTP conversion in FY2003. Pages were digitized and enhanced to remove artifacts, edge effects and lighting non-uniformity. Compared to the initial processing of the first two books, project staff have improved the development procedures for Pare's and Gesner's books. A 3D wireframe model was developed in Maya, a modeling and animation system. The 3D wireframe model texture-maps each pair of page images to both sides of the wireframe of a turning page. A multisource lighting model provides diffuse lighting, specular highlights and shadows. For each flip of the page, 12 intermediate animation frames are generated, rendered and then imported into Director. Another improvement exploits the characteristics of the wireframe model. The node attributes within the model can be adjusted, allowing different rates and styles of curvature to be expressed during page flipping. For example, there is a choice of three flip behaviors depending on where the finger is placed on the page to start flipping (e.g., the upper right corner of the page will flip over if a finger is placed at the top right of a page). Both new books have been completed in TTP form, with Paré's book extended to the TTP+ format using the "storyline" model with explanatory visuals.

NLM Gateway

The NLM offers a number of Internet-based information resources, each with its own user interface. The NLM Gateway provides an easy to use, "one-stop" search method that allows users to simultaneously searching nine document collections using five retrieval methods from a single interface. Several enhancements of NLM Gateway occurred in FY2003. One of the most significant enhancements is the result of a collaborative effort with the Indexing Initiative. In order to improve retrieval results, all of the Gateway's meeting abstracts have been automatically indexed by the Indexing Initiative systems. Other enhancements 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. User selectable search subsets for AIDS, Bioethics, History of Medicine, and Space Life Sciences in PubMed have been added to the NLM Gateway. The subsets are available through a pull-down menu on the Limits page or by using a new field qualifier in advanced searches. Phrase detection has been added so that users no longer have to put search phrases in double quotes. Modifications to accommodate a new PubMed applications program interface, a new version of the Voyager integrated library system underlying Locatorplus, and new XML output from NLM's Document Creation and Maintenance System have been incorporated. An applications program interface for the NLM Gateway has also been completed and is being evaluated.

PubMed on Tap

PubMed on Tap is a research and development project to develop accessible biomedical information at the point of care through handheld devices used by clinicians and other mobile health care providers. User interface, content selection, content organization, and system performance are necessary for effective access to information. Initial research is focused on the design of a user interface for search and retrieval of MEDLINE bibliographic citations through PubMed. Initial content selection is involved with categorizing citations returned in response to a query, creating multi-document summaries for clusters of highly related documents, and single-document descriptions containing features specific only to a given document in the cluster. System performance research is focused on discovering design factors that ensure the speed and reliability of the hardware and software required for accurate and timely retrieval of data. Areas of investigation include choice of parsers, efficient use of a database to store recent queries and citations, and load testing. A prototype system, developed for Personal Digital Assistants (PDAs) running the Palm operating system, was built and tested in FY2003. The software uses the PDA's wireless communication interface and HTTP protocol to communicate with a servlet residing on a proxy server. The proxy server communicates with PubMed through the Entrez programming utilities (e.g., Esearch, Efetch and Elink). The proxy server stores queries, results, and citations to provide a quick response to recurring queries and fast delivery of frequently requested citations. The proxy server also monitors performance measures and accumulates aggregate statistics to help in developing

clustering and ranking tools. The client program is responsible for the user interface and for storing user-specific information, such as preferred search strategies or recurring queries.

Consumer Health Informatics Research

Exploring consumer information needs, information-seeking behavior, and cognitive strategies, consumer health informatics uses medical informatics and information technologies to study methods to develop, organize, integrate, and deliver accessible health information to consumers with all levels of health literacy.

In the spring of 2003 we launched the Genetics Home Reference, an integrated Web-based information system designed for consumers and others to learn about specific genetic conditions and the genes that are associated with those conditions. The research results made possible by the Human Genome Project are increasingly being made available in scientific databases on the Internet, but, because of the often highly technical nature of these databases, they are not readily accessible to the lay public. Our goal is to provide a bridge between the clinical questions of the public and the richness of the data emanating from the Human Genome Project. The Genetics Home Reference provides basic information, in a question and answer format, on the nature of genes and how they give rise to various conditions and diseases. For each condition, the site provides information about the specific genes linked to it, how common the condition is, and what its symptoms and available treatments are. For each gene, the site provides information about the normal function of the gene, its chromosome location, the conditions linked to the gene, and whether any gene therapy is available. Each description includes a glossary as well as alternative names for the gene or condition being described. In addition, each condition or gene description links directly to pertinent information available on a variety of other resources, including MedlinePlus, ClinicalTrials.gov, PubMed, Gene Tests, Gene Reviews, LocusLink, and Online Mendelian Inheritance in Man.

As a further guide to users of the site, we have developed a resource called, "Help me understand genetics," which explains, together with diagrams and other visuals, some basic concepts in genetics. This resource offers, for example, easy to understand explanations of DNA, genes, proteins, chromosomes, and how genes control the growth and division of cells. In addition, the resource has sections on the nature of genetic disorders, genetic consultation and testing, gene therapy, and genomic research.

The system architecture includes three primary modules. The first module is for content collection and work flow management, the second is a "publisher" module that retrieves data from the content manager and relevant external sources, and the third is the public Web site that presents the gene and condition descriptions, interlinking these with related resources, such as MedlinePlus, LocusLink, and ClinicalTrials.gov. An important aspect of the content manager is that it tracks the status of each description, including whether it has yet undergone expert review. No description is released to the public until it has been reviewed by one or more external experts who are, in most cases, board-certified medical geneticists or molecular biologists. The system makes extensive use of the Unified Medical Language System and its constituent vocabularies. The UMLS, in most cases, provides definitions and synonyms for the glossary, and MeSH terms are used to facilitate linking to other NLM resources. The Gene Ontology is used for browsing genes by their function, by the biological processes in which they are involved, or by their cellular structure. The Genetics Home Reference currently focuses on single gene or polygenic conditions that are also topics on MedlinePlus. As knowledge of genetics expands, the interrelationships between genes and diseases will continue to unfold, and the site will continue to reflect these developments.

ClinicalTrials.gov provides comprehensive, up-to-date information about federally and privately supported clinical trials throughout the U.S. and many other parts of the world. The system grew out of 1997 legislation requiring the U.S. Department of Health and Human Services, through the NIH, to establish a registry for both federally and privately funded trials "of experimental interventions for serious or life-threatening diseases and conditions," thereby broadening the public's access to information on potential interventions for a wide range of diseases. ClinicalTrials.gov was launched in February 2000 and provides patients, families and members of the public easy access to information about the location of clinical trials, their design and purpose, criteria for participation and, in many cases, further information about the disease and intervention under study. There are also links to individuals responsible for recruiting participants to each study.

Because clinical trials bridge biomedical research conducted in laboratories and applied clinical research in humans, information in this area is often difficult for non-specialists to read. ClinicalTrials.gov is designed to help members of the public make sense of the information. The site includes general resources to help people understand what clinical trials are, including a glossary of common terms used to describe clinical trials, and a list of frequently asked questions about human research. In addition, each study is presented in a standard format that helps readers quickly identify important elements of a study, such as its purpose, criteria for participation, locations of the trial sites, and contact information. Furthermore, to provide additional context, study records also point users to relevant health topics at the NLM's consumer health Web site, MedlinePlus. Some study records also contain links to published literature, either for background information or study results.

The number of daily visitors to the site increased by over 50% from 8,000 daily visitors in 2002 to 12,000 daily visitors in 2003. The site increased the number of protocol records by over 25% from 6,600 protocol records in 2002 to 8,300 records in 2003.

Research Infrastructure and Support

The Lister Hill Center performs extensive research in developing and advancing infrastructure capabilities such as high-speed networks, nomadic computing, network management, and improving the quality of service, security, and data privacy.

Next Generation Networking

The NLM completed its program to define Next Generation Internet (NGI) capabilities that will allow the NGI to be used routinely in health care, public health and health education, as well as biomedical, clinical and health services research. Collaborative capabilities include quality of service, security and medical data privacy, nomadic computing, network management, and infrastructure technology. Principal investigators were invited to the NLM for a reverse site visit as a conclusion to the NGI projects. Each of the 15 sponsored projects was given 45 minutes to present an overview. A video of each talk and its associated PowerPoint presentation will be posted on the NGI Web site. The NGI networks are being used for multimedia applications involving voice and video. The Abilene network supports full Internet Protocol multicast. In this mode, NLM can receive and transmit multicast voice and video sessions.

NLM's Internet2 connection to the MAX GigaPOP (Mid Atlantic Exchange Gigabit Point of Presence) was increased from 155 megabits per second (OC-12) to gigabit speed (gig-E) in FY2003. A full, native multicast is broadcast via the Center's Internet2 connection allowing the Center to implement a multimedia node on the Internet2 Access Grid. Multimedia nodes are a collection of nodes that transmit and receive a variety of audio and video media that may be used for teleconferences and meetings. The high bandwidth and Quality of Service (QoS) characteristics of Internet2 permit the Access Grid to pass high quality audiovisual signals between nodes.

NLM continues to collaborate on the Multilateral Initiative on Malaria in Africa. Working with Infinite Global Infrastructures personnel, Lister Hill Center staff investigated approaches to overcoming the drawback of limited bandwidth in the satellite links to Africa. For example, a review was conducted of Redwing Satellite Solutions, the space segment and Internet access provider located near London. A presentation was given on the NLM's communications work and the performance measurement task at an NLM-sponsored event in Kenya.

The purpose of the Scalable Information Infrastructure (SII) initiative is to encourage the 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 will 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 use test-bed networks linking one or more of the following: hospitals, clinics, health practitioners' offices, patients' homes, health professional schools, medical libraries, universities, medical research centers, laboratories, or public health authorities. Eleven SII research contract awards were made at the close of FY2003.

Applications of smart card technology continue to be explored at the Lister Hill Center. Smart cards are credit-card-sized plastic cards with an embedded circuit chip. Cards may be used for security authentication and for data storage. Recent applications involve the use of biometrics, the storage of biomedical information (e.g., thumbprint, iris scan), in order to increase smart card security. The Lister Hill Center continues to co-sponsor the Western Governors' Association Health Passport Project, one of the largest health-oriented smart card pilot programs in the United States. The Health Passport Project stores data from multiple Federal, State and local agencies on cards used by clients receiving health benefits such as well-child care, checkups, immunizations and food benefits. Phase II of the Health Passport Project is under way in the San Diego area. Phase II will incorporate biometric authentication on the smart card, digital certificates, and trusted third party systems to facilitate the safe, encrypted transfer of private medical and demographic information over the Internet.

Utilizing its technical expertise, Lister Hill Center staff provide technical consultation and representation in a variety of environments. Project staff provided technical consultation and coordination for the LHC's participation at the Radiological Society of North America's conference in Chicago in FY2003. Primary responsibilities included the implementation of a Gigabit Ethernet connection from the Internet2 backbone to the McCormick Place Convention Center. Engineering staff provided technical advice and cost options for the telecommunications link that will be developed between the NLM and the site for the NLM's backup databases. Staff also represented the NLM and NIH at the Joint Engineering Team, the Internet2 Applications Strategy Council and Health Sciences Advisory Group.

The NLM continues to sponsor the Telemedicine Information Exchange (TIE), a Web-based resource of telemedicine and telemedicine-related activities maintained by the Telemedicine Research Center in Portland, OR. During FY2003, approximately 526 non-NLM bibliographic citations and other records were received by the TIE. Staff continue to participate in the monthly meetings of the multi-agency Joint

Telemedicine Working Group. Participating in this group, LHC staff made a formal presentation to Congress and the Administration on state-of-the-art telemedicine and e-health projects and solutions.

The Collaboratory for High Performance Computing and Communications

The Collaboratory for High Performance Computing and Communications investigates innovative means for assisting health science institutions in their use of online distance learning technologies. The Collaboratory also explores advanced computer and network technologies for distance interactivity, including wireless technology and virtual reality research.

Major upgrades to existing videoconferencing codecs were done in FY2003 and new codecs were added. Several significant demonstrations were performed using the technology, both at NLM and at national meetings. Demonstrations of streaming and wireless Webcasting were done and videoconferencing and Webcasting were employed routinely in program activities. One significant upgrade was the conversion of the MPEG2 high bandwidth, high quality videoconferencing codecs from Litton to those of StarValley. Several upgrades were made to the Wavelet videoconferencing codec as it continued to be refined. Finally, and perhaps most importantly, an Access Grid node was installed allowing NLM to experiment with multicast videoconferencing and participate in this form grid technology in collaboration with others in the Internet2 research community. The new MPEG2 codec and the Collaboratory's traditional h.323 codec were used in demonstrations of differences in Internet2 and commodity Internet capabilities in a week-long tutorial at the Radiological Society of North America's annual meeting in Chicago. The MPEG2 codec also was employed in a demonstration of collaboration and virtual reality distance learning technology between NLM and Stanford University for the Library's Board of Regents. NLM used its Access Grid technology in multiple demonstrations with Project TOUCH, a collaboration between the Universities of New Mexico and Hawaii. East Carolina University, the University of Arkansas, and the University of Utah also participated. Wavelet technology was demonstrated at the annual meeting of the American Society of Clinical Pathology/College of American Pathologist annual meeting in Washington. Videoconferencing technology used by NGI contractor George Mason University was demonstrated at the interagency CENDI meeting held at NLM and the MACAW workstation using the technology in the collaboratory was employed for the demonstration.

Experiments were started testing the use of conventional h.323 videoconferencing technology with NLM's Adopt-A-School Partner, Wilson High School, in Washington, DC. Some preliminary tests were also done with the Drew Medical Magnet School in Los Angeles. Additional tests are planned with the aim of doing a pilot distance learning program for minority students interested in health sciences. Work continues in experimenting with new codecs. Digital video compression technologies are being acquired for testing with members of the Internet2 community, since the DV format is being considered as a compression format for the Access Grid. Finally, an assessment was made of alternative display technology to accommodate both the Access Grid's multi-screen displays and stereo images.

All videoconferencing and collaboration efforts have been encumbered significantly this year due to firewall policies. The firewall continues to plague current efforts to use h.323 and other videoconferencing codecs. Technologies identified for penetrating firewalls require at least one end point to be outside a firewall. Operators of the Internet2 Commons videoconferencing service recommend placing collaboration tools outside of firewalls and every NGI project funded by NLM using collaboration tools identified contending with firewalls as a critical problem during the reverse site visit.

The EtherMed database of Web accessible health professions educational materials continued to be expanded through collaborations with colleagues at the University of Utah, UCLA, and the University of Oklahoma. Another major review of the database was conducted. Upgrades were made to the hardware, Web server, SQL server and ColdFusion server software needed to run EtherMed. Collaboration with the Heal Education Assets Library (HEAL) program funded by NLM and NSF continued, focusing on ways the HEAL program could regularly harvest EtherMed records for inclusion in their database. Several improvements to EtherMed's search methods were identified. Search terms are now highlighted in retrieval and a contract has been made to prioritize search results based on number of search term hits. A decision was made to delay the research study with the University of Alabama at Birmingham and to implement HEAL harvesting technology until these improvements were made.

System Security and Advanced Network Planning

System Security and Advanced Network Planning research focuses on computer security, the NLM network, the Next Generation Internet (Internet2 and NGI), and the upgrading of Lister Hill Center systems. A gigabit/second capacity firewall system, installed in FY2003, has helped reduce security problems. Work on the LHC Network has continued to improve its performance and reliability. Two core routers for the LHC network are redundantly attached both to the OCCS network and to the edge routers throughout the Lister Hill Center. In addition, gigabit/second links have been provided to some desktop workstations and some servers. FY2003 security improvements have made Lister Hill Center systems less vulnerable to external security attacks. However, the increasing prevalence of worms and viruses necessitates constant vigilance in order to keep systems up to date.

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 15 years, the ability to integrate vast amounts of complex and diverse biological information created the 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 flood of genomic data, most notably gene sequence and mapping information. NCBI will meet the challenge of collection, organization, storage, analysis, and dissemination of scientific data 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, bibliographic citations, and other related resources. Such features allow GenBank to serve as a critical research tool in the analysis and discovery of gene function. In FY2003, approximately 9 million sequences were added to GenBank, and the base count rose from 22.6 billion in August 2002 to 33 billion in August 2003. The 27 million sequences in GenBank represent data from over 130,000 organisms. During the first quarter of FY2003, GenBank release 132.0 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 those 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 ten years of processing direct submissions, NCBI has issued over 1.4 million accession numbers, with approximately 23% of these assigned in FY2003. There are now over 936,000 direct submission accession numbers that are publicly available and approximately 70,000 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/PubMed bibliographic information, including abstracts and publishers' full-text documents. GenBank provides links to outside sources when direct links to publishers are not available. This latter service, called LinkOut, points to 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.

The Third Party Annotation (TPA) 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. TPA also accepts submissions from Whole Genome Shotgun (WGS) sequencing projects. Annotations are allowed in these assemblies and are 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 FY2003. In this new version, improvements were made to facilitate ease of TPA sequences—an Assembly Tracking pop-up allows users to input the primary accession number(s) from which the TPA sequence is derived; and the alignment view in the Record Viewer displays the nucleotide sequences contained within the alignment rather than a graphical view of the alignment. The GenBank submission tool, Sequin MacroSend, was designed to upload large Sequin files to avoid problems that occur when large submission files are sent via e-mail, 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. Guides for WGS sequence and bacterial genome submission were added to the GenBank site to help submitters with these types of specialized submissions.

BankIt, another sequence submission software tool, is now in its ninth 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.

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 2003 there were 18,050,373 public EST entries stored in dbEST.

Another 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 2003 there were 6,412,085 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 2003 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. 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 2,000 species including bacteria, archaea, and eukaryotes, complete microbial genomes, a number of viroids, mitochondria, a broad host range of plasmids, and over 1,000 viruses. The genomes represent both completely sequenced organisms and those for which sequencing is in progress. Some of the 38 new complete genomes added to the database in FY2003 include: Escherichia coli CFT073, Clostridium tetani E88, Mycoplasma penetrans, Ciona intestinalis, Streptococcus pygenes SSI-1, Salmonella enterica subspecies enterica serovar Typhi Ty2, Enterococcus faecalis V583, Bacillus anthracis str. Ames, Chlamydophila pneumoniae TW-183, Bacteroides thetaiotaomicron, Vibrio parahaemolyticus, Aspergillus terreus, and SARS

Coronavirus.

The Human Genome

NCBI is responsible for collecting, managing, and analyzing 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. In FY2003, the DNA reference sequence of *Homo sapiens* was completed and released. The assembly is considered a finished sequence in that it is highly accurate and highly contiguous. The only remaining gaps correspond to regions whose sequence cannot be reliably resolved with current technology. The annotation of genes and other genomic features is still under development. This resource is truly an international public sequencing effort resulting from the cooperation of scientists and sequencing centers from around the world.

Assembling and Annotating the Human Genome

A team of NCBI scientists is 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 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, "Human Genome 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 linkage, 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 "Component" map), the location of genes, exons, STSs, FISH mapped clones, ESTs, GenomeScan models, SAGE tags, 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. The results table includes links to a chromosome graphical view where the gene or marker can be seen in the context of additional data. The Evidence Viewer is a feature that provides graphical biological evidence supporting a particular gene model and the Model Maker allows users to build a gene model using selected exons.

In FY2003, NCBI continued to improve its Map Viewer. A new Map Viewer home page was released showing all organisms for which map information is available. The ability to query across multiple species was added, as well as standardization of display and function across species, tremendously enhancing the power of this tool. In response to public request more maps were added for all organisms, particularly human. In addition, connections from LocusLink to the Map Viewer were added for the organisms Drosophila melanogaster and Caenorhabditis elegans.

NCBI's Human-Mouse Homology Map is designed to allow navigation between the human and mouse genomes using NCBI's FLASH homology browser. Links to numerous mapping resources as well as a view of various sequence alignments is also provided. Various maps can be compared to produce slightly different overviews of conserved synteny between humans and mice.

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 Web page, now a part of the NCBI Books site, contains descriptions for over 150 genetic diseases and provides links to databases and organizations that can supply additional information. For each disease-causing gene there is a link to the PubMed literature, the Online Mendelian Inheritance in Man database (OMIM), and LocusLink. In FY2003 links were added for a genome view that graphically displays gene locations, BLink links to related sequences in organisms other than human, and related external resources.

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 14,500 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. OMIM also contains two maps showing the cytogenetic location of disease genes. The "OMIM Morbid Map" is organized by disease, and the "OMIM Gene Map" is organized by chromosome. Both maps are searchable by gene symbol, chromosomal location, and disorder keyword.

LocusLink is a single-query interface to curated sequence and descriptive information about genetic loci. LocusLink presents information on official nomenclature, aliases, sequence accession numbers, phenotypes, EC numbers, OMIM records, UniGene clusters, map information, and related Web resources. LocusLink contains 216,321 records for the seven organisms represented in the database. In FY2003, LocusLink's organism list was expanded to include C. elegans (nematode) and Bos taurus (cow), bringing the total number of organisms to eight. LocusLink 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. Other LocusLink features include gene annotation; 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, Human-Mouse Homology Maps and BLink. This year, gene ontology (GO) annotation from GOA (SwissProt) was added to the database as well as connections from LocusLink records to SwissProt accessions.

The Reference Sequence (RefSeq) database provides a comprehensive, integrated, non-redundant set of sequences, including genomic DNA, transcript (RNA), and protein products for major research organisms. These standards serve as a basis for medical, functional, and diversity studies by providing a stable reference for gene identification and characterization, mutation analysis, expression studies, polymorphism discovery, and comparative analysis. RefSeqs are used as a reagent for the functional annotation of some genome sequencing projects, including those of human and mouse. Curated genomic annotations can be retrieved via LocusLink and the Map Viewer. In addition to the six organisms previously represented in RefSeq, Bos taurus (cow) was added in FY2003. Also, major updates occurred for the Drosophila melanogaster (fruit fly), and C. elegans collections. The first full release of all NCBI RefSeq records was released in FY2003 which includes over 785,000 proteins and 2,005 organisms. The full RefSeq release includes RefSeq sequences, a catalog of the release contents, statistics, and extensive documentation and is available on the NCBI FTP site.

The dbSNP database serves as a comprehensive catalog of common human polymorphisms for the international research community. SNP detection and discovery is expected to facilitate large-scale association genetic studies. dbSNP is available for search and retrieval in the Entrez retrieval system and each individual SNP record also contains links to other Entrez databases, LocusLink, genomic sequence data, and external SNP resources such as the SNP Consortium.

In FY2003, several enhancements were made to the dbSNP Web site to integrate new information on assay validation and marker variability in reference populations. New report formats were introduced to support detailed genotype data and to export the human subset of dbSNP to the International Hap Map project for use in haplotype reconstructions. The complete set of 11.4 million submissions were processed and reduced to a non-redundant set of 7.4 million refSNP clusters, representing variation in 14 different organisms including model organisms such as mouse, rat, zebrafish, nematode, Plasmodium falciparum (malaria parasite), and Arabidopsis. The largest subset of data is human sequence polymorphism, with 5.8 million records successfully mapped and annotated on human genome sequence. The genome annotation pipeline has been refined to support parallel processing of multiple organisms and repetitive or low complexity DNA sequence.

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 ultimate goals of the public mouse genome sequencing 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.

The mouse genome resources guide provides information on diverse mouse-related resources from multiple centers including sequence, mapping, and clone information. It allows easy navigation to mouse genome BLAST pages, mouse Map Viewer, trace repository, and the Human-Mouse 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. In FY2003 the mouse genome resources were improved and refined. NCBI Build 30 was released which is the first composite mouse assembly combining MCGS version 3 Whole Genome Shotgun assembly and HTGS sequence. A detailed explanation of the build process was also created to facilitate user understanding.

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 been enhanced over the last year. Full-text journals that link to PubMed have increased from 3,056 in September 2002 to 4,054 in September 2003. Approximately 60% of all PubMed citations from 1990 to 2003 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 1.5 million searches and approximately 230,000 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 August 2003, over 1,000 organizations have supplied links to their Web sites including over 700 libraries, 130 full-text providers, and 160 providers of non-bibliographic resources including biological databases. Together they provide links to 27 million Entrez records. The LinkOut resources page provides 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. Enhancements to the LinkOut program include an icon hosting service and statistics for Submission Utilities for Libraries. Free full-text links are marked with a special icon to facilitate access. The LinkOut SERHOLD interface facilitates SERHOLD libraries to display print holdings in LinkOut. As of August 2003, over 700 libraries were participating in this program.

System enhancements made to PubMed throughout the year include the display of an icon link indicating the availability of an abstract for a citation, as well as the availability of the full-text article in PubMed Central, on the PubMed Summary page. Additional system enhancements made to PubMed include a feature allowing users to receive search results via e-mail. A new cancer subset was created in cooperation with NCI along with several new indexes including: Corporate Author, Comments/Corrections, Place of Publication, Grant Number, and Ahead of Print.

In April the MeSH database was incorporated into the Entrez retrieval system. The MeSH database is NLM's controlled vocabulary for indexing articles in PubMed. Its addition to Entrez provides additional search, display, and linking features for PubMed.

The NCBI Bookshelf is a dataset of biomedical books adapted for the Web, available via the Entrez Retrieval system. Books may be searched directly or found through links in PubMed abstracts. The majority of links are between the books and PubMed, with more links in the future between books and other types of information such as gene and protein sequences. As of August 2003, there were 26 reference books available in addition to the recent inclusion of original monographs. The NCBI Handbook was added to the Books site in 2003, providing an in-depth description of NCBI's major information resources. Other titles added to the Books database this year include, The Human ATP-Binding Cassette (ABC) Transporter Superfamily, The KIR Gene Cluster, The NCBI C++ Toolkit, and chapters taken from the Eurekah Bioscience Collection.

PubMed Central (PMC) is a Web-based repository of life sciences journal literature providing barrier-free access of full-text articles to the public. This repository is based on a natural integration with the existing PubMed biomedical literature database of abstracts. As of August 2003, PMC included over 130 life science journals, and use of the system has grown to more than 300,000 unique users each month.

In 2003 PMC was added to the Entrez retrieval system allowing closer integration of full- text literature with genetic data and other factual databases. Back issues of journals are being scanned to create digital copies for online access. Also this year, an XML DTD (eXtensible Markup Language Document Type Definition) suite for journal archiving and publishing was developed and will be adopted by a number of archiving and publishing

organizations sending information to PMC.

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. This year, a new queuing system was put into production for better service. This system splits a BLAST request across multiple machines and reassembles the results to present them to the user. The results are then stored in an MSSQL database.

The BLAST suite of programs was refined in FY2003. At this time, version 4.0 of the BLAST database is fully supported. Structure linkouts and the ability to retrieve parts of large sequences found in the search were added in FY2003. 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. The BLAST homepage was redesigned and made public at the end of FY2003.

A new type of nucleotide search, Discontiguous MegaBLAST, is now available. This service is designed especially for comparison of diverged sequences, especially sequences from organisms, which have alignments with low degree of identity, where the original MegaBLAST is not very effective. It uses a greedy algorithm and concatenates queries to save time scanning the databases. Instead of using exact matches to databases sequences for initial hits, it finds matches to a discontiguous template made from the query which results in fewer hits, but more statistically significant matches.

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. Efforts of reorganization took place in order to allow easier access to the tools and database by users. There was a reorganization of the BLAST preformatted database files on the FTP site in order to make transfers of the large files more convenient. There was also a complete reorganization of the entire BLAST database FTP site to allow easer access. There was also a reorganization of the BLAST executable FTP site including Standalone BLAST Binaries and WWW BLAST server programs.

Other Specialized Databases and Tools

A new NCBI resource called dbMHC was introduced in FY2003. dbMHC serves as an open, publicly accessible reference database for human Major Histocompatibility Complex (MHC) related DNA typing reagents and for individual data of MHC related research projects. dbMHC is currently receiving data from the projects of the International HLA Working Group (IHWG) as well as the scientific community at large. dbMHC will also provide tools for further submission and analysis of research data linked to MHC.

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 August 2003, there were over 30 organisms included the Plant Genomes database.

The Viral Genomes Website provides a convenient way to retrieve, view, and analyze complete genomes of viruses and phages. This site now contains 1,176 viral genomes and over 1,500 viral genomic reference sequences. This year Genomes group renovated the existing NCBI retrovirus subtyping tool and new sequence reference sets for Hepatitis B virus, Hepatitis C virus and poliovirus were added.

A new Web resource devoted to Rattus norvegicus (rat) was made available in FY2003. This resource provides access to both the original sequence data deposited in GenBank as well as the assembled genome data. From this page, users can access information through the rat Map Viewer, rat-specific BLAST pages, the taxonomy database, and other NCBI resources. Additional links point to the Rat Genome Sequencing Consortium and other external resources. Version 3.1 of the rat genome build was released at the end of FY2003.

A SARS Coronavirus Resource was made available in FY2003. This resource provides data and information relevant to the newly discovered virus. It includes links to the most recent sequence data and publications, results of pre-computed sequence analysis, and other SARS related resources.

The SKY/M-FISH and CGH database provides a repository of publicly submitted data which are complementary fluorescent molecular cytogenetic techniques. Spectral Karyotyping (SKY), Multiplex Flourescence In Situ Hybridiation (M-FISH) and

Comparative Genomic Hybridization (CGH) are complementary fluorescent molecular cytogenetic techniques. SKY/M-FISH permits the simultaneous visualization of each human or mouse chromosome in a different color, facilitating the identification of chromosomal aberrations. CGH utilizes the hybridization of differentially labeled tumor and reference DNA to generate a map of DNA copy number changes in tumor genomes. A new tool, the CGH Case Comparison Tool, was released in order to compare profiles from multiple cases.

Microarray technology, a method for generating gene expression data, is another important experimental breakthrough in the field of molecular genetics. 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 FY2003 a new database was created to store curated, coherent dataset information for the GEO repository as well as a new interface for retrieving data. GEO data can be retrieved by GEO accession number, through the GEO current holdings page, GEO DataSets (GDS), or through the Entrez ProbeSet search interface. ProbeSet is deeply indexed and is reciprocally linked to Entrez Nucleotide, PubMed, and Taxonomy. The GEO database has been growing rapidly, and currently contains over 8,000 accessioned objects and approximately 4,000,000 gene expression profiles. GEO data is available via FTP as well as the Web.

Entrez GEO and Entrez GEO DataSet (GDS) were incorporated into the Entrez system in FY2003. Entrez GDS contains curated GEO DataSet definitions to facilitate identification of experiments of interest. A GDS record represents a collection of biologically and statistically comparable GEO samples and forms the basis of GEO's suite of data display and analysis tools. Entrez GDS can search all GEO submissions with any text found in the GEO or GDS databases.

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 and links to genomic sequence via the Map Viewer are also available.

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 August 2003, there were approximately 15,200 distinct human clones and 11,200 distinct human genes. There were also about 10,800 distinct mouse clones and 8,800 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, designed 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).

The MMDB sequences, retrieved via Entrez, provide links to related information such as: MEDLINE/PubMed citations, 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. In FY2003 taxonomy links were added to the structure summary page using PDBeast software. As of August 2003, MMDB served about 50,000 queries a day and contained over 23,500 structures, up from approximately 20,000 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. The latest version of Cn3D, 4.1, includes the addition of alignment algorithms for molecular modeling as well as structure-structure alignment options.

The Conserved Domain Database (CDD) 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. CDD is part of the Entrez retrieval system, labeled as the "Domains" database. Conserved Domains are indexed for retrieval by keywords, and links between Conserved Domains and Proteins, PubMed, and Taxonomy are

available. 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.

Over 11,900 domain models have been added to CDD between. Updates to CDD include alignments to COGs, KOGs, PFam updates, and CDD-curated alignments. Improvements to CDD also support structure-based alignment and evolutionary classification, and check-in of curated alignments to the CDTrack database. Version 1 of "CDTree," the new curator workbench software, was deployed in FY2003 to assist the Structure team with curation efforts.

VAST, or the Vector Alignment Search Tool, is a service that identifies similar protein three-dimensional structures of newly determined proteins. VAST compares new proteins to those in the MMDB/PDB database. VAST computes a list of structure neighbors, or related structures, which allows a user to browse interactively, viewing superpositions and alignments in Cn3D. The 'VAST summary' provides a series of controls for selecting and sorting the structure neighbors and alignments. In FY2003, the VAST search service was re-engineered for improved performance and the addition of a new Web server allows for graphical alignment-footprint displays. Over 25,000,000 new alignments/superpositions were added to the database this year, bringing the total of structure-structure alignments recorded in VAST to about 75 million.

Protein Reviews on the Web (PROW) an online resource that features PROW Guides, was included in MEDLINE indexing this year. PROW contains authoritative short, structured reviews of 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 more 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 for information on an organism or taxon's lineage. 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. Over the past year, the browser has been upgraded and supports a much richer set of links to the Entrez databases. Links are available to all related records within the Entrez system as well as genome information, when available. The full-text PubMed Central archive is now being scanned and indexed with links to organisms in the taxonomy database.

Taxonomy LinkOut has been expanded to include more linkout providers. The Taxonomy group has improved this service by providing a more prominent display for links to external resources in the search results page. A Taxonomy Name/Id Status Report page is available 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 also assists in keeping track of a set of taxnames or taxids in order 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 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 FY2003, UniGene was added to the Entrez retrieval system. Also this year, all organisms with over 70,000 ESTs were added to the database, bringing the total number of organisms represented in UniGene to 30. New organisms added this year include: Lycopersicon esculentum (tomato), Glycine max (soya), Sus scorfa (pig), and Medicago truncatula (barrel medic), C. elegans (nematode), Chlamydomonas reinhardtii (algae), Ciona intestinalis (sea squirt), Oryzias latipes (Japanese medaka), Sorghum bicolor (sorghum), Dictyostelium discoideum (slime mold), Solanum tuberosum (potato), Silurana tropicalis (western clawed frog), Oncorhynchus mykiss (trout), Pinus taeda (loblolly pine), Toxoplasma gondii, and Vitis vinifera (wine grape). As of August 2003, approximately 576,356 clusters (sets) were included in UniGene.

HomoloGene is a database of curated and calculated homologs for genes represented by UniGene, or by annotation of genomic sequences. HomoloGene

allows users to explore possible homology relationships among the less well studied genes. Computed 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 FY2003, a new build of HomoloGene was released with many improvements. For example, an internal HomoloGene ID is now used as the defining unit, rather than UniGene cluster IDs. In cases for which an organism does not yet have genome annotation, the HomoloGene ID is equivalent to a UniGene Cluster ID. Organism identifiers have changed from a two letter code, to a taxonomic ID (tax_id). New organisms added this year include rainbow trout, cow, tomato, and pig, with a total of 21 organisms represented in the database. As of August 2003, there were 143,625 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 nucleotide and protein sequences (GenBank), organisms (Taxonomy), and literature citation databases (PubMed). At this time, Entrez consists of 20 integrated databases and those recently added include: PubMed Central, GEO, GEO DataSets, MeSH, and NCBI Web Site. The NCBI Web Site option provides a search of the entire NCBI Web site and FTP sites.

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 choosing a link for citations or sequences that resemble a given citation 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 180,000 Entrez nucleotide and protein queries are handled per weekday and the number continues to rise.

Various updates were implemented in the Entrez system in FY2003. Presentation changes include a more efficient way to navigate result pages, a streamlined way to add records to the Clipboard, Order page, File, and Text display, and a feature to activate different Links menu displays. The Entrez Links menu was modified to allow the menu to appear on all search results instead of only the first 50 results.

Also new in FY2003 is an important feature in which a set of Entrez databases can be searched simultaneously. A new Web page was created for the "global query" interface allowing users to run a search against all databases or go directly to a single database search page. Short explanations of the databases were also added for quick information.

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 FY2003, NCBI's site was averaging over 30 million 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.

In FY2003 computer room space was added in the B2 level in Building 38A to accommodate the growing need for disk storage and compute servers. The Compute Farm's batch processing system was expanded to 168 Linux-based CPUs. NCBI's flagship information retrieval service, Entrez/PubMed, was converted to run on Linux computers using Network-Attached Storage (NAS).

Two new NAS clusters were acquired and storage was added to existing clusters, for a total of approximately 16 TB of new network storage. NCBI's entire BLAST server infrastructure was converted from Solaris x86 to Linux and substantially augmented with the addition of more than sixty 2-CPU Intel-based servers.

Research

Research is at the core of NCBI's mission. The Computational Biology and 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 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 strengthened applications and database work and has led to the development of many new

theoretical and practical models. The application of these methods to the life sciences has opened doors to new areas of research.

NCBI's basic research group is within the Computational Biology Branch and consists of 66 senior scientists, staff scientists, research fellows, and postdoctoral fellows. Research projects focus on new computer methods to accommodate the analysis of genome sequences and molecular sequence databases resulting from the rapid growth in large-scale sequencing efforts. Other projects focus on such techniques as the analysis of particular human disease genes and the analysis of the genomes of several pathogenic bacteria, viruses and other parasitic organisms. Topics of current research include: low-complexity amino acid sequences, sequence signals, mathematical models of evolution, virological modeling and statistical methods in virology, dynamical behavior of chemical reaction systems, comparative genome analysis, protein structure and function prediction, taxonomic trees, and population genetics.

The intramural group is engaged in many projects, some 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. A list of members is in Appendix 4.The high caliber of the work of this group is evidenced by the number of peer-reviewed publications, approximately 100 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 Health and Human Services. 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 Postdoctoral Fellows program is designed to provide training for doctoral graduates in a variety of fields including molecular, computational, and structural biology as well as graduates in other fields who elect to obtain additional training in computational biology. 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 FY2003, NCBI expanded its outreach and education programs to increase awareness of its myriad of 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 Website: About NCBI

The "About NCBI" Website is 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. This year, a "Getting Started" section was added to increase user knowledge of specific resources. The Getting Started sections focus on one specific resource and provide basic information on the data included and how to best utilize the resource.

Education: Mini-Courses and Lecture Presentations

Three new mini-courses, "Structural Analysis Quick Start," "Map Viewer Quick Start," and "LocusLink Quick Start" were developed to provide a practical introduction to the various NCBI programs. These and the other mini-courses, "BLAST Quick Start," "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 to over 2,000 participants to date.

Education: Bioinformatics Training

The intramural Core Bioinformatics Facility (CoreBio) was designed to help NIH researchers make optimal use of computer science and technology to address problems in biology and medicine and construct a network of bioinformatics specialists serving individual institutes within the NIH. The Institutes and Centers select participants who are trained on the use of 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.

The CoreBio has trained representatives from 14 research institutes at NIH and conducted six courses since 2001, including two in the past year. Thirteen update sessions and two special topic sessions for the institute representatives have also been held. One-on-one consultations are available on an ongoing basis for NIH scientists with NCBI faculty as well.

Education: Extramural Educational Collaborations

The educational collaboration program was established to train a network of bioinformatics support specialists who provide local educational and user support services for a wide range of users and needs. The university medical library is becoming a centralized point for providing these services at the local level, and members of the collaboration are based in institutions that are leading this trend.

The foundation of the program is a five-day course, "NCBI Advanced Workshop for Bioinformatics Information Specialists," first offered in 2002. The course continues to be developed and taught as a collaborative project among NCBI staff and university-based biologists and librarians who currently provide support at their institutions. It is designed for library based, full-time bioinformatics support specialists, who provide user training and support for NCBI information resources. This course is offered yearly in August. Course developers and participants are added to the bioinformatics support network (BSN), which facilitates continued learning and communication among the group. The BSN currently consists of 30 members and participants are added to the BSN upon completion of the course.

A three-day introductory course, "Introduction to Molecular Biology Information Resources," is also being offered at NLM each fall and spring. It is designed to give medical librarians the background needed to provide introductory support to novice users across campus. A total of 34 people have completed the course on-site at NLM in FY2003. The three-day course is a revised and expanded version of the original one-day introductory course that was taken by 500-600 medical librarians from 1997 through 2001.

Education: NCBI Courses

The course, "A Field Guide to GenBank and NCBI Resources," was designed to help researchers and other users keep abreast of enhancements and the increasing diversity of NCBI molecular biology resources. This free course is offered to all those who work with biological sequence and structure data and provides a useful introduction and survey of the available NCBI tools and databases. The course consists of three components: a three-hour lecture, a two-hour hands-on practicum, and a one-on-one session if requested. The six-member teaching staff has presented 45 courses to over 4,000 people throughout the country in FY2003.

Outreach: User Guides for NCBI Resources

NCBI has continued to develop a comprehensive list of fact sheets that outline the services and databases offered by NCBI. These fact sheets and guides are available for printing via the "About NCBI" site. 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 be found for a number of databases and search and retrieval tools such as Entrez, PubMed, Structure, and BLAST. The 'About NCBI' site has an education page that serves as a comprehensive resource for all tutorials and courses.

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 FY2003, 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.

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 a 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 FY2003.

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

- Resource Grants for information management, often involving medical libraries
- Training and fellowship grants in support training of informaticians and information specialists
- Research Grants in informatics, information science, and biomedical computing
- Research Resource grants to support unique tools for informatics and bioinformatics
- Publication and Conference grants to enhance scientific and scholarly communication
- SBIR/STTR grants to support informatics innovations in small businesses
- Special Projects and collaborations with other agencies

Resource Grants (MLAA)

Resource Grants, authorized by the Medical Library Assistance Act, support access to information, connecting computer and communications systems and promote collaboration in networking, integrating, and managing health-related information. There are several 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 Resource Grants are open to public and private, nonprofit health institutions engaged in health education, research, patient care, and administration. All strongly encourage health science library involvement in the project.

Internet Connection Grants. This program was terminated with the October 1, 2002 application deadline. The program provided $30,000 to a single organization or $50,000 to a consortium, to support technology infrastructure and communication costs of internet connectivity. Five applications for Internet Connection grants were reviewed in FY 2003, two were funded. The average priority score was 169 for new grants funded.

Sources of Internet Connection Grants FY 2003

Type of Organization	# of Applications	# of Awards
Institution of Higher Education	0	0
Research Org/Institute/Foundation/Lab	0	0
Independent Hospital	2	0
Educ Org other than Higher Education	0	0
Other Health/HR/Env/Cmnty Srv	3	2
Other	0	0
Business	0	0

Information Access Grants. This program was terminated with the October 1, 2002 application deadline. The program provided $12,000 for a single organization and an additional $12,000 for each partner, to support network development, and document delivery and information services across a network of sites. Eleven information access grants were reviewed in FY 2003, two were funded. The average priority score was 168 for new grants funded.

Sources of Information Access Grants FY 2003

Type of Organization	# of Applications	# of Awards
Institution of Higher Education	5	1
Research Org/Institute/Foundation/Lab	0	0
Independent Hospital	2	0
Educ. Org other than Higher Education	1	0
Other Health/HR/Env/Cmnty Srv	3	1
Other	0	0
Business	0	0
TOTAL	**11**	**2**

Internet Access to Digital Libraries (IADL) Grants. IADL grants enable organizations to offer access to health-related information, 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. On February 1, 2003, IADL became a regular resource grant program, accepting new applications three times per year. Sixty-three applications for IADL grants were reviewed in FY 2003, 20 new grants were funded. Eight continuations were funded. The average priority score of new IADL grants funded was 179.

Sources of IADL Grants FY 2003

Type of Organization	# of Applications	# of Awards
Institution of Higher Education	7	5
Research Org/Institute/Foundation/Lab	0	0
Independent Hospital	24	6
Educ Org other than Higher Education	1	1
Other Health/HR/Env/Cmnty Srv	26	8
Other	0	0
Business	0	0
TOTAL	**58**	**20**

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. Fifty-three information system grant applications were reviewed in FY 2003. Ten new grants and 10 continuations were funded. The average priority score for new information system grants funded was 166.

Sources of Information System Grants FY 2003

Type of Organization	# of Applications	# of Awards
Institution of Higher Education	20	5
Research Org/Institute/Foundation/Lab	2	1
Independent Hospital	7	2
Educ Org other than Higher Education	2	0
Other Health/HR/Env/Cmnty Srv	20	2
Other	2	0
Business	0	0
TOTAL	**53**	**10**

Integrated Advanced Information Management Systems (IAIMS) Planning Grants and Operations Grants The NLM provides IAIMS grants to health-related organizations that seek to plan, design, test and deploy systems and techniques for integrating data, information and knowledge resources into a comprehensive networked information management system that crosses organizational and disciplinary boundaries.

In March 2002, NLM introduced an updated IAIMS grant program, shifting emphasis from infrastructure to content, with emphasis on projects that bridge disparate information systems. The IAIMS program contains five different grants. Two IAIMS grants are resource grants. 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. There are two IAIMS research grants. The IAIMS Pilot Study Grant provides up to $50,000 per year for one or two years and the IAIMS Testing and Evaluation Grant provides up to $100,000 for one or two years. In addition, there is an IAIMS fellowship that gives $50,000 for one or two years. 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).

Eighteen IAIMS grant applications were reviewed in FY 2003 and two review site visits were performed for IAIMS Operations grant applications. All of the Planning, Pilot and Testing and Evaluation applications came from institutions of higher education. Fifty percent of the Operations grant applications came from non-academic organizations (1 hospital, 2 community-based organizations, and 1 research organization). One planning grant and two operations grants were funded as continuations.

Type of IAIMS Grant	# of Applications	# of Awards
Planning Grants	6	0
Operations Grants	7	1
Pilot Study	2	0
Testing & Evaluation	2	0
Fellowship	1	0
TOTAL	**18**	**1**

Publication Grant Program The Publication Grant Program provides short-term financial support for scholarly research that will lead to a publication. Studies prepared under this program include critical reviews or research monographs in the history of medicine and life sciences; special areas of biomedical research and practice; and medical informatics, health information science and biotechnology information. In certain instances, secondary literature tools and scientifically significant symposia are supported. In the past, a printed monograph was the most common outcome of a publication grant but, increasingly, projects in electronic publishing, video, and. and other media are being supported. Unique among NLM's grant programs, the publication grant program accepts applications from individuals without an organizational affiliation.

Thirty-two publication grant applications were reviewed in FY 2003. Nine new grants and 14 continuations were funded. Eight publication grants were made to Institutions of Higher Education and one to an individual. The average priority score of new publication grants funded was 152.

Training and Fellowships (MLAA)

Overview

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 200 trainees at predoctoral and postdoctoral levels. Eighteen training programs were funded for a new five-year period beginning July 1, 2002. Eleven of the previous twelve 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 receives some co-funding from NIBIB, the new NIH Institute for bioengineering and imaging. The National Institute of Dental and Craniofacial Research continues 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 in FY2002 that combined opportunities for both basic and applied training, and offered a new stipend scheduled created to facilitate recruitment of computer scientists, engineers, librarians, and nurses into informatics. In FY2003 seven applications were received of which five were awarded, one withdrawn, and one received an unfundable priority score.

Midway through FY2003 EP issued another, related informatics training opportunity, the Senior Individual Fellowship, intended for those who have had ten or more years of professional experience in some appropriate field and were interested in a career change into informatics. In keeping with the senior status of such fellows, a more generous stipend scale was offered. Three applications were received; none was funded.

A new program, Early Career Development Grants, was established to provide transition assistance for biomedical informaticians who are establishing their initial research programs. No applications were reviewed in FY2003 for this program.

Minority Support from EP Authorized Grant Funds

The new IADL 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. Three of the resource grants were awarded to HBCU organizations. Ten grants were awarded to organizations considered to be Hispanic-serving organizations.

Research Support (PHS301)

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

Medical Informatics

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.

Bio-informatics 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 the 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.

Small Grant Program

To complement its traditional R01 grants, in 2003 NLM issued a program announcement for small project research grants, a mechanism used by most of the NIH Institutes. These grants provide $50,000 per year for one or two years, and are designed to help researchers who are just starting out in an area of inquiry. Feasibility and proof of concept studies, and the gathering of preliminary data that might support a subsequent R01 study are typical uses of the R03 grant. Originally published as offering $75,000 per year, the grant program was brought into conformance with suggested NIH funding guidelines in June 2003. Twenty R03 grants were reviewed in FY2003. Three new grants were funded. Seventeen applications came from institutions of higher education, two from research organizations, and one from a community-based organization. The average priority score of funded R03 grants was 155.

Informatics for Disaster Management

In 2002, NLM issued a program announcement for research grants exploring the application of informatics approaches in natural and man-made disasters. Initially an R01 mechanism was proposed for this program, but in 2003 the mechanism was changed to R21 to better accommodate projects that are more akin to engineering research and development than to hypothesis-testing experimental research. During the formal change process for this program, two other institutes (National Institutes of Mental Health and National Institute of Biomedical Imaging and Bioengineering) signed onto NLM's program announcement. Beginning with the Feb 1, 2004 deadline, disaster management applications will be triaged among the three institutes. Seven new applications in this program were reviewed in FY2003. One award was made. Five applications came from institutions of higher education, one from a community-based organization, and one from a business.

NLM Exploratory/Developmental Grants.

NLM's new Exploratory/Developmental grant fills a niche between Resource and Research grants. Announced in April 2003, the R21 grant supports high risk/high yield projects, proof of concept, and work in new interdisciplinary areas. Preliminary data are not required for these grants, and emphasis on hypothesis testing methods is relaxed. No FY 2003 awards were made in the R21 program.

Pan-NIH Projects

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 is funded. NLM participates in a number of other multi-institute projects including bioengineering, pharmacogenetics, imaging, and nanotechnology.

NLM and Roadmap Activities

A major pan-NIH enterprise initiated by the NIH Director is resulting in a battery of programs related to three themes: New Pathways to Discovery, Research Teams of the Future, and Reengineering Clinical Research. NLM is a participant in all of the Roadmap initiatives but by the end of FY2003 had made no specific awards.

Overlap Areas

In FY2001 Congress created a new Institute, the National Institute of Biomedical Imaging and Bioengineering (NIBIB). Several overlap areas between the interests of NLM and those of NIBIB have been identified, and are under discussion. The welcome upsurge of interest in biomedical computing across NIH has also resulted in some uncertainties about which Institutes cover areas of informatics. Experience and ongoing discussions are expected to clarify the situation in time.

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. Of four applications received in FY2003, two were funded.

Biomedical Ethics

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

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 FY 2003 follow.

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

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

The project is interested in electronic information in a broad spectrum of fields in arts and science. Improving network-based information access for health care consumers is an important goal of the project for NLM, although all aspects of digital libraries as applied to health domains may compete for funding. NLM, as have the other sponsors, contributed funds to NSF, which will manage the project. NLM's commitment for FY2001 was $1,000,000 as it had been in the previous year. The DL-2 project is an arm of the HPCC initiative. Target for total budget from all sources is $50 million over 5 years. The last installment of NLM commitment to this program was in FY2002, but several of the funded projects continued through FY2003.

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

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

In response to a Phase 2 RFP, five Phase 2 contracts were awarded during late FY1999 and FY2000. A Phase 2 RFP for the high-tech projects was issued in late FY2000. Awards were made in FY2001 to two of the Phase 1 high-tech applicants. Although the original RFP contemplated the possibility of a Phase 3 for this program, neither NHLBI nor NLM is planning to proceed with another Phase. Although some small supplements were added to several of these projects in FY2003, funding for the NHAAP informatics program is essentially complete. Work is still on going. A contractors' conference is planned for spring of FY2004.

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. Organizations that received NLM funds in FY2003 include the National Human Genome Research Institute, National Center for Research Resources and the National Science Foundation (NSF).

NLM received co-funding for NLM grants from other organizations, including the National Center on Minority Health and Minority Health Disparities, National Cancer Institute, National Institute of Dental and Craniofacial Research, National Human Genome Research Institute, National Institute of Mental Health, National Institute on Aging, National Institute of Biomedical Imaging and Bioengineering, and the Department of the Army.

SBIR/STTR (PHS 301)

All NIH research grant programs, including NLM's, by Congressional mandate allocate a fixed percentage of available funds every year to Small Business Innovation Research (SBIR) grants. These projects may involve a Phase I grant for product design, and a Phase II grant for testing and prototyping. Two such awards were made in FY2003 and another highly rated application was transferred to another Institute for funding when NLM funds were exhausted. 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.

Extramural Programs Operating Units—Highlights

Grants Management Office

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 five employees, issued a total of 220 awards in FY2003, including grants, administrative supplemental awards, and fellowships. Details of the grants are provided in Table 11.

Committee Management Activities

The Board of Regents (BOR) met three times in FY2003. The Extramural Programs Subcommittee was held prior to each of these meetings. Two new BOR members joined the EP Subcommittee in 2003: Ernest Carter, M.D., Ph.D., Howard University, and Thomas Detre, M.D., University of Pittsburgh.

The Board of Regents approved 387 grant applications, including special reviews made by the EP Subcommittee. These special reviews are conducted when 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; or when an application is from a foreign institution. The Board Operating Procedures were reviewed and approved without change at the February 11-12, 2003 meeting.

Scientific Review Office

NLM's initial review group, the Biomedical Library Review Committee (BLRC), evaluates grant applications for scientific merit. BLRC met three times in FY2003 and reviewed 202 applications. The Committee (see Appendix 5 for roster of members) operates as a "flexible" review group. It is composed of three standing subcommittees: eight members on the Medical Library Resource Subcommittee, nine members on the Medical Informatics Subcommittee; and four 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: 16 Special Emphasis Panels were held during FY2003. 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 283 applications during FY2003. Two site visits to evaluate IAIMS applications were also carried out by ad hoc panels.

A new mechanism, the Loan Repayment Program, was initiated and the applications were discussed using the Internet Assisted Review process, as part of an NIH-wide deployment of e-government business practices.

Three contract Special Emphasis Panels were convened during FY2003. One panel reviewed a contract from the University of Pennsylvania School of Dental Medicine. One panel provided first-level review of 69 proposals for Network Infrastructure Health and Disaster Research, and the final panel provided second-level review of 31 proposals for that same activity.

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.

The BOR, serving as second level of the appeals process, heard one appeal of review during FY2003. The appeal was denied.

Administration and Operations

EP has had several personnel changes over the past year. The recruitment of a Committee Management Assistant was completed. The Program Office experienced two losses, one through retirement (Dr. Susan Sparks) and one through transfer to another Institute (Dr. Carol Bean). In July, Dr. Charles Friedman came to EP as Senior Scholar. Dr. Friedman is on a one-year sabbatical from the University of Pittsburgh.

The NIH A-76 competition for grants management review and program support staff resulted in four EP staff members being identified as people whose functions fall 50% or more into the A-76 work statement. Although NIH won the competition, the "Most Efficient Organization" will remove these people/functions from direct supervision by EP staff. The functions of these staff will be performed through a new organization that reports to the NIH Office of Extramural Programs. Transition planning will begin in early FY2004.

Information Technology Support

IT support for EP continues to be provided by a contractor on site. Staff received several in-house training sessions in preparation for moving to a new operating system and the planned replacement of GroupWise by Outlook, the NIH-designated e-mail program. At the end of the year, new staff workstations were ordered.

EXTRAMURAL PROGRAMS FY 2003

($ in 000s)

MLAA	NON COMPETING NO	NON COMPETING AMT ($)	COMPETING NO	COMPETING AMT ($)	TOTAL NO	TOTAL AMT ($)
IAIMS						
Planning & Operations (G08)	3	$1,190	1	$400	4	$1,590
TOTAL IAIMS	**3**	**$1,190**	**1**	**$400**	**4**	**$1,590**
TRAINING						
Training Programs (T-15)	18	$14,554	0	$0	18	$14,554
Fellowships (F37/F38)	4	$188	6	$332	10	$520
TOTAL TRAINING	**22**	**$14,742**	**6**	**$332**	**28**	**$15,074**
Publication Grants (G13)	**21**	**$1,426**	**11**	**$834**	**32**	**$2,260**
RESOURCE						
IADL(G07)	13	$421	19	$1,074	32	$1,495
Internet Connection (G08)	1	$50	2	$25	3	$75
Information System (G08)	9	$1,269	15	$2,153	24	$3,422
TOTAL RESOURCE	**23**	**$1,740**	**36**	**$3,252**	**59**	**$4,992**
BIOETHICS (N01)*	**1**	**$982**	**0**	**$0**	**1**	**$982**
Loan Repayment Program (L30	**0**	**$0**	**4**	**$291**	**4**	**$291**
NN/LM Contracts (N01)	**8**	**$12,673**	**0**	**$0**	**8**	**$12,673**
TOTAL MLAA:	**78**	**$32,753**	**58**	**$5,109**	**136**	**$37,862**
PHS 301						
BIOMED-INFORM. RESEARCH						
R01/R03/R13/R21/R24/P41	25	$6,810	24	$6,956	49	$13,766
Protein Sequence Databank (IA	1	$188	0	$0	1	$188
Chairman's Grant (U09)	1	$150	0	$0	1	$150
BIOMED-INFORM. RESEARCH	**27**	**$7,148**	**24**	**$6,956**	**51**	**$14,104**
BIOINFORM. RESEARCH						
(R01/R03/R21)	13	$3,503	5	$1,373	18	$4,876
Bioinformatics Resource (P41)	5	$2,255	0	$0	5	$2,255
BISTI (R21/R33/P20/P41)**	6	$2,769	2	$1,160	8	$3,929
BIOINFORM. RESEARCH TOT	**24**	**$8,527**	**7**	**$2,533**	**31**	**$11,060**
SBIR/STTR(R43/R44/R41/R42)	**2**	**$1,047**	**0**	**$0**	**2**	**$1,047**
TOTAL PHS 301:	**53**	**$16,722**	**31**	**$9,489**	**84**	**$26,211**
TOTAL EP:	**131**	**$49,475**	**89**	**$14,598**	**220**	**$64,073**
*Bioethics Contract is awarded as an extension with						
**Program funded in part with PHS 301 funds						
Budget excludes NIH Taps						

TABLE 11

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 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 that make up 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: OCCS implemented 'Go Local' into MedlinePlus this year. This initiative is for the citizens of North Carolina who can 'Go Local' from any MedlinePlus health topic and browse a list of web links on local services in their county or city. This is the first step toward bridging the gap between health information and local health services needed by patients and their families. Five major releases were deployed which included:

- Redesign of the Home Page for both English and Spanish MedlinePlus.
- Incorporation of the Merriam Webster Medical Dictionary.
- Implementation of an e-mail feature for English and Spanish health topics pages, encyclopedia pages, news pages, and drug pages.
- Implementation of Section 508 compliance on all MedlinePlus public pages.
- Addition of 25 new English/Spanish Patient Education Institute (PEI) modules.

NIH Consolidated Collocation Site (NCCS): OCCS led the effort this year on the NIH Consolidated Collocation Site Project. NIH components were pursuing individual efforts to obtain collocation facilities and services for purposes of disaster recovery and Continuity of Operations Program until the NIH Chief Information Officer created an NIH Consolidated Collocation Site Action Team to plan and implement the NCCS. This project is a joint effort by NLM and the NIH Center for Information Technology (CIT) and consists of various phases including planning, procurement, deployment, and maintenance.

High-Speed Communication Network: OCCS implemented a fiber optic communications Gigabit Ethernet network for Specialized Information Services (SIS) and Extramural Programs (EP), two NLM Divisions located offsite. This effort resulted in a reliable, high performance, secure network. Also:

- OCCS improved and expanded broadband access (DSL and cable modem) to NLM contractors and employees, including indexing contractors of the BSD Indexing Section. The Citrix remote access system was enhanced to be a fully redundant system, and continued as the primary method of access to NLM services for NLM contractors and staff.
- Connectivity to the Internet II is provided via an OC12 fiber circuit to the MAX located at the University of Maryland. Both Internet and Internet II connections converge at the NLM network perimeter firewalls, which provide security controls to protect NLM IT resources.
- OCCS enhanced the storage area network central network storage system and implemented a switch in order to allow more servers to be connected to the unit. The system currently provides 800 Gigabytes of storage. The network central network is the primary data storage system for many NLM staff, providing a storage area that is automatically backed up each night. It also provides the basic storage for clustering of servers to provide high availability of services.

Multi-faceted IT Security Program: OCCS continued its multi-faceted and multi-layered IT security program that successfully prevented over 1.5 million virus attacks this year and detected more than 240,000 probes, scans, denial of service attacks and other security events on a monthly basis. Also:

- OCCS implemented an intrusion detection program to examine malicious network traffic and installed two Intruvert 400 gigabit sensors that resulted in not only detecting malicious traffic but also discovering any systems that have been infected by worms or viruses.
- OCCS implemented a vulnerability assessment program in an effort to mitigate risks associated with network systems. Vulnerability scanning is performed monthly and has resulted in a reduced number of vulnerabilities and improved security infrastructure that can withstand cyber

attacks of unknown scope and complexity.

Enhanced Senior Health/Accent Project ("Talking Web"): OCCS successfully applied innovative technology to the NIHSeniorHealth.gov Web site added narration, text magnification and color contrast features in support of blind and low vision users. These features are produced at the server so users need no additional desktop software. OCCS also:

- Compatibility tested for over 40 combinations of browsers, operating systems and platforms.
- Tuned pronunciation for medical terms and drug names.
- Redesigned the Home Page and added seven additional health topics with corresponding videos.

OCCS Help Desk Consolidation: Many of the core functions of the OCCS Help Desk were consolidated into an expanded NIH IT Help Desk. The Help Desk provides first level (Tier-1) IT problem tracking and support for NLM staff. As one of several functions mandated for consolidation across NIH, Desktop and PC networking support requests are now channeled to the NIH IT Help Desk for initial ticket entry into the call tracking system and attempted resolution by the NIH Tier-1 support staff.

Process Improvement Initiative: To establish a baseline of defined and repeatable standards for all stages of the software life cycle, OCCS implemented a full range of template documents this year. For system designers, developers, and QA specialists, the template suite was accompanied by sample documents and training sessions, plus commented PowerPoint presentations for later reference and future use in training. For the Systems staff, standard operating procedures for the daily practices were created. At year end, the Process Improvement (PI) team produced over 70 policies and documents on topics such as building standard desktop PCs, monitoring the NLM network, creating user accounts, and many other activities. The PI group has also taken over formal management of the OCCS Change Control Board.

Automated System Patch Deployment: OCCS leveraged NLM's new Active Directory's group policy functionality to trigger desktop systems to automatically apply patches and updates from an OCCS server as each PC is started up but before the user logs in. The specialized patch server automatically downloads the latest operating system updates and security patches from Microsoft. Over a five month period, OCCS scheduled the automatic application of over 35 service packs and security hotfixes and has applied these to approximately 1,200 PCs with low manpower costs.

OCCS UNIX Architecture Upgrade: The UNIX architecture upgrade continued this year to expand storage and provide added security. More systems and services were configured with highly available network attached storage and converted to gigabit subnet connectivity. Private and isolated subnets were deployed to provide robustness and additional security. Additional enhancements will be made in FY2004.

Enhanced DOCLINE: OCCS 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 1.5 was released in March and Version 1.6 in July 2003 and included a total of 30 enhancements. A new region code for Mexico, Region 21, was added to DOCUSER search and the ability to request a document from OLDMEDLINE from 1953 to 1965 using the NLM Gateway was added.

Enhanced OLDMEDLINE: OCCS transitioned the OLDMEDLINE database from the mainframe to an Oracle environment. As part of the overall directive to consolidate NLM derivative databases and make them available via PubMed, the OLDMEDLINE data was evaluated and transformed. Extensive coding was developed to accept the data in its original format and process it through the various stages to produce XML for export.

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, provides 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 Activities: The computer facility experienced numerous power surges during Hurricane Isabel yet maintained all systems operational during the unstable weather. OCCS expanded monitoring coverage to include the UPS/Environmental alarm box for the NCBI B2 level computer facility; and an outside company, CCG, was contracted to audit the B1W17 computer facility leading to various environmental improvements within the facility.

The following describes in more detail OCCS accomplishments in FY2003.

Customer Services

The Help Desk entered and tracked over 5,100 requests for IT support from its NLM customers this year. Many of the core functions of the OCCS Help Desk were consolidated into an expanded NIH IT Help Desk that is

taking over first level (Tier-1) IT problem tracking and support for NLM staff. As one of several functions mandated for consolidation across NIH, Desktop and PC networking support requests are now channeled to the NIH IT Help Desk for initial ticket entry into the Remedy call tracking system, with initial resolution attempted by the NIH Tier-1 support staff. What cannot be resolved by phone, including all PC/networking fieldwork, is then routed back to NLM for resolution by the current Tier-2 IT support engineers. OCCS Tier-1 staff will be realigned to serve a liaison role within NLM for the NIH Help Desk, and will continue to perform other NLM internal customer service roles such as reviewing and releasing broadcast messages. Library Operations' public-facing NLM Customer Service function will continue to use the Siebel problem reporting system, and is not included in this consolidation of internal IT support Tier-1 help desks.

Desktop Support

OCCS facilitated the NLM migration of several Microsoft NT network domains into a Microsoft Active Directory (Windows 2000) networking model in the first quarter, FY03. While this was a complex technical transition from several autonomous NT domains to a single Active Directory (AD), this project was also the culmination of two years of planning among all NIH Institutes and Centers to develop an NIH-wide federated networking environment based on Active Directory. Subsequently, the NIH Active Directory was adopted for practical use as the "authentication domain" for NIH enterprise applications, such as the New Business System (NBS), ITAS and others.

Virtually all NLM desktop systems are now members of the NIH Active Directory, and can be more carefully managed and updated using Active Directory management tools. The NLM AD team also worked to prepare for the introduction of NIH enterprise New Business Systems modules brought online this year, such as Budget and Travel. With AD resources, a single logon can be applied across many participating applications, reducing the burden for users of multiple systems. The NIH Active Directory also offers a more secure and accurate way of passing username/password information to participating applications than earlier approaches.

Identifying and applying Microsoft Windows operating system patches and security hotfixes is a complex and labor-intensive activity that has grown dramatically over the last couple of years. This year OCCS leveraged NLM's new Active Directory's group policy functionality to trigger desktop systems to automatically apply patches and updates from an OCCS server as the PC is starting up, but before the user logs in. The specialized patch server automatically downloads operating system updates and security patches from Microsoft. Then, as the Windows desktop systems join or reconnect to the Active Directory domain, they receive instructions to install these hotfixes and updates. These AD group policies enforce the installation of all of the tested and approved hotfixes, and produces logs of successes and failures that can be reviewed for completeness.

Over the past five months, OCCS scheduled the automatic application of over 35 service packs and security hotfixes, and has applied these to approximately 1,200 PCs with low manpower costs. Because these patches can be quickly and efficiently deployed, we have substantially improved the security posture of NLM desktops, while maintaining a trim support workforce. On Command and ZenWorks continue to provide active tools for the deployment of new PC builds and applications, but Active Directory group policies now have a firm role in support of the Windows operating systems on NLM networked PCs.

Network Support

OCCS continued to fulfill its mission to provide reliable LAN and Internet communications services, meet the communication needs for new IT systems, provide security services, provide 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. OCCS/NES took steps to increase the capabilities and reliability of network services and storage, by providing for:

- Enhanced monitoring and management
- Increased security
- New services
- Increased performance and throughput for networks,
- Additional redundancy
- Enhanced backup
- Expanded, centralized and efficient storage

Public Internet connectivity services continued to be provided through a contract with Genuity. Internet connectivity was provided via 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.

OCCS implemented redundancy of equipment and network paths this year so that single points of failure in the network could be eliminated. The NLM network perimeter connections to external networks provide an aggregate of 2 gigabits per second (Gbps) while the interconnection between NLM and the NIH/CIT campus backbone operates at 1 Gbps.

OCCS implemented a new software system, iTRACS, for documenting the LAN cabling and infrastructure. The process of documentation and complete labeling of the infrastructure will be a

continuing process.

Most 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. These consoles are centrally located in the NOSC and are shared by multiple systems administrators.

OCCS continued to support Novell and Microsoft network operating systems for NLM staff and contractor use of services such as file and print sharing, directory services, software operating system and application updates, remote access, etc. Many of these systems are now implemented in a clustered configuration in order to achieve a high level of availability.

Network support continues to provide 56K dial-in access, cable modem, DSL, and ISDN access for a wide range of NLM users.

In addition to supporting the indexing system, the Citrix terminal server solution has been implemented as a good solution for flexiplace 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 continued to improve and enhance two e-mail security mechanisms at the NLM network perimeter that connects NLM to external networks such as the Internet and the NIH Campus backbone: Trend Micro VirusWall anti-virus software and SpamAssassin. The Trend Micro anti-virus software scans all e-mail to and from NLM and the Internet for viruses and attachments with disallowed file extensions, deletes offending content, and notifies the sender and recipients accordingly. SpamAssassin has been configured to tag e-mail messages containing spam content so as to make it easy for e-mail users to take appropriate action against these spam messages by using automated rules within their e-mail viewer.

NLM cooperated with NIH efforts to consolidate wireless LAN networks. NLM representatives participated in the committee that studied the requirements and issues, and participated in the design of the NIH-wide architecture for the wireless network. The initial wireless capabilities were implemented, and further expansion of the wireless systems will continue in selected areas of NLM.

Systems Support

OCCS led the effort this year on the NIH Consolidated Collocation Site Project. While the NIH Community dependence on NIH Enterprise computer and communications capability increases daily, the growing complexities of providing that service has escalated the need for a backup capability that addresses the needs of all NIH Institutes and Centers (ICs). The widespread demand for disaster recovery capabilities by government and industry alike has resulted in commercially available "collocation sites." A collocation site is a computer center where floor space, power, air-conditioning, physical security, Internet communications, and optionally, various equipment and services, are obtained via contract with a commercial provider.

In late 2002, the NIH Chief Information Officer (CIO) created an NIH Consolidated Collocation Site (NCCS) Action Team to plan and implement the NCCS in a cost-effective manner for all ICs. This effort is jointly led by NLM and the NIH Center for Information Technology (CIT), and consists of various phases including planning, procurement, deployment and maintenance. The planning and procurement phases were completed in FY2003. A highly qualified facility was selected in the Washington metropolitan area at a distance of approximately 25 miles from NIH. This facility, not publicly advertised, has redundant power and cooling systems, excellent security, extensive Internet access, and abundant floor space capacity. The deployment phase will commence in November 2003. NIH services and applications will be present at the NCCS in both "active" and "near-active" conditions to ensure continuous ability to operate.

The NCCS must satisfy or exceed the present and predictable future IT requirements of the NIH Enterprise ICs and accommodate the wide variation of business needs of the individual ICs, including differences in missions, degrees and types of internet presence, types of services and applications, quantities and qualities of technical staff, and requirements and desires in terms of IT capacity, IT COOP, etc. Both NLM and CIT are highly confident that the NCCS will meet these NIH requirements, and look forward to commencing the deployment phase.

The UNIX architecture upgrade continued this year to expand storage and provide added security. More systems and services were configured with highly available network attached storage as well as converted to gigabit subnet connectivity. Private and isolated subnets were deployed to provide robustness and additional security. Additional enhancements will be made in FY2004.

OCCS built and deployed a secure server for the Malaria Research Project (ADRN). This work was coordinated with Library Operations and included creating restricted user accounts, customizing Web pages, and isolating the system itself. The researchers may use either FTP or the Web to access their research data and may use FTP to deploy their research data. Both LO and the researchers at WHO and in Africa are very enthusiastic.

OCCS achieved a milestone in NLM history this fiscal year: MEDLINE XML baseline data was delivered to licensees via FTP, a quicker alternative delivery by DLT tapes.

IT Security

Throughout the year, NLM continued to assess and strengthen its security posture based on current business requirements and risk assessment. A number of security improvements were implemented in FY2003.

OCCS established perimeter e-mail security to scan more than 10,000 monthly e-mails. The scanning mechanisms include anti-virus, spam tagging, attachment blocking for dangerous file extensions, message size limiting and blocking widely-spread viruses such as SoBig and MS Blaster. Evaluation of various mechanisms to fight virus-infected and spam e-mail continues.

OCCS implemented Web URL filtering this year in accordance with NIH Policy. All NIH Institutes have been mandated to filter out usage of inappropriate Web sites while simultaneously not affecting NIH business activity.

OCCS continued strengthening the NLM network and Internet security position by contracting with Symantec to perform an independent vulnerability assessment of the NLM network infrastructure. The resulting report provided extremely valuable focus for our security efforts. Regularly scheduled monthly vulnerability assessments were subsequently run for critical NLM assets throughout 2003. A vulnerability trend analysis showed a marked decrease in red (high priority) alerts that now allows NLM to focus attention on yellow (lower priority) alerts.

OCCS worked closely with the NIH Center for Information Technology (CIT) to facilitate the consolidation of all information systems NIH-wide by the end of FY03. All related NLM security activities were accomplished within the required time frame. In addition, security policies for remote access and wireless communications were developed with NIH in support of the consolidation efforts. These policies were reviewed and approved by the Department.

This year, OCCS underwent a formal security certification and accreditation (C&A) analysis of MEDLARS and TOXNET, conducted by an independent contractor for Department. Under the Federal Information Security Management Act, Federal Agencies are required to conduct annual IT security reviews to assess the risks to information and systems. The C&A analysis found both MEDLARS and TOXNET to be securely designed and operated, resulting in a recommendation for "Authority To Operate."

OCCS organized computer security awareness training for all NLM employees. Since a security program is only as strong as its weakest link, NLM staff are required to take mandatory annual online Computer Security Awareness training. This course, for both employees and contractors, provides an overview of basic computer and information security practices.

Computer Facilities

NLM systems continue to be supported in a safe environment in NLM's computer facility, which is available 24x7x365. The Network Operations and Security Center (NOSC), which was established in 2002, continues to serve as a central point in 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 from the corridor outside the computer room. The intended audience is the general public and NLM staff. The displays contain statistical charts and near-real time activity monitors with explanatory text. Each panel focuses on a particular NLM service or IT infrastructure component. Near-real-time counters show utilization levels for MedlinePlus and PubMed/MEDLINE and near-real-time utilization of NLM's Internet-1 and Internet2 communications links. NLM services as seen by remote users around the world are also shown.

Major computer facility accomplishments this year include the following:

- An outside company, CCG, was contracted to audit the B1W17 computer facility resulting in various environmental improvements within the facility.
- The computer facility experienced numerous power surges during Hurricane Isabel, yet the Uninterruptible Power Supply (UPS) maintained all systems operational during the unstable weather. A project is under way to procure a third UPS module to provide increased protection and redundancy for future equipment growth.
- Several staff of the Facilities Management Section (FMS) are working with the OCCS Process Improvement Team to create new Standard Operating Procedures and convert existing ones to the OCCS format. This documentation will assist the staff in the performance of their duties as well as provide procedures for new staff.
- OCCS expanded monitoring coverage to include the UPS/Environmental alarm box for the NCBI B2 level computer facility. It was installed in the B1W17 computer room in order for OCCS FMS staff to monitor the UPS on a 7X24 basis.

Consumer Health Information

MedlinePlus: In 2003 OCCS supported an aggressive schedule of major MedlinePlus releases which included the Go Local initiative and a Spanish e-mail announcement list. The Go Local initiative debuted in December 2002 allowing users in North Carolina to search for local medical service providers while viewing descriptive material in MedlinePlus. OCCS developed production methods to insert and maintain the Go Local links and also created a management infrastructure to

support other Go Local projects in the future. To strengthen outreach to Spanish speakers, a Spanish e-mail announcement list now disseminates biweekly listserv messages about MedlinePlus en español.

Service improvements include a new link to the Merriam Webster Medical Dictionary to help users correctly spell terms when performing searches. A new e-mail feature enables users to send articles of interest to friends and family members from the English and Spanish health topic pages, encyclopedia pages, news pages, and drug reference pages. A new printing capability for those same pages generates printer-friendly hardcopy. For advanced investigators, search results can now be exported in spreadsheet format. The page templates for MedlinePlus were analyzed and modified for Section 508 compliance, text layout was changed from two-column to three-column to enhance legibility, and A-Z organization pages were added to streamline navigation.

Senior Health Project: Throughout 2003, OCCS completed numerous enhancements to NIHSeniorHealth.gov, including text magnification and audio performance. This is a joint NLM and National Institute on Aging project that provides health information on the Web using modes of delivery—video and narration—appropriate for older Americans with access limitations (low vision and low hearing, etc.). The interface was redesigned to fully comply with Section 508 requirements, and page templates were extensively modified to integrate the Accent ("Talking Web") features. Specific challenges included improving legibility with text magnification and achieving the necessary audio performance.

Virtual Customer Service: Virtual customer service for the NLM Web site was released to production on January 31, 2003. The system's artificial intelligence (AI) software responds to customer questions in a conversational mode. For example, if a user types "when is the library open?" the system returns the hyperlink to a Web page with library hours. Content development required importing hundreds of topics from MedlinePlus. OCCS wrote programs to automate the task and also created a batch job to generate the pattern lists for Health Topics and Drugs needed to respond to user queries. OCCS enhanced the NativeMinds implementation with a user feedback mechanism. If a NativeMinds session fails to yield the desired results, the user is prompted to fill out a Siebel help desk ticket. The ticket thus created is e-mailed to the Siebel system with details on the user's virtual customer service session. The session information is used by OCCS and Library Operations to improve the customer service system for similar queries in the future.

Professional Health Information

NLM Classification System: This year OCCS implemented substantial improvements to the NLM Classification System, which now includes consistency checking between index entries in the classification system and terms in the MeSH database. Discrepancy resolution is automated to a great extent. A batch mode is provided for use when the new MeSH is released at year-end processing. Since medical library catalogers refer to both the classification system and MeSH, the timely reconciliation of these two databases will substantially reduce customer support time. The batch mode eliminates a major manual effort at the end of the year. Downloadable zip and PDF versions of the catalog are now available on the system site for catalogers with slow Internet connections who prefer to work from hard copy.

Mesh Browser: Since OCCS upgraded the MeSH Browser build process in 2002, 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: This year, Mexico was added as a participating country to DOCLINE. DOCLINE, the NLM interlibrary loan system, supports 3,200 domestic and international libraries in processing over 3 million interlibrary loan transactions a year. General DOCLINE service improvements include the capability to e-mail service requests to NLM about Interlibrary Loan issues. A new online form further simplifies problem reporting. DOCLINE users can now recognize and order documents from OLDMEDLINE citations. Beginning this year, DOCLINE displays easy-to-see links to PubMed Central when a user requests an article that can be downloaded free. OCCS is also working with the NLM Free Electronic Access Team to identify sources in PubMed and Locatorplus that are free and accessible so that appropriate messages can be displayed to users requesting DOCLINE copies. This will save the public money and help manage DOCLINE request increases. DOCLINE invoicing will soon be compatible with the Electronic Funds Transfer System (EFTS), a widely used electronic billing system, as well as with the National Technical Information Service. Testing of EFTS is in progress.

Relais: Early in this fiscal year, OCCS added delivery by e-mail and e-mail post-to-web to Relais. In addition, Relais servers were ported to Windows 2000 (from Windows NT) to reduce maintenance costs, and a more cost-effective printing solution, using three HP 9000 series laser printers, was installed for hard copy requests. New black and white scanners were installed in September 2003, and a color scanner was added to fill requests for color copies.

Voyager: A two-year effort to merge OCLC data into the Voyager database was completed. The project involved updating close to 100,000 records to ensure

synchronization of the Voyager, OCLC, and SERHOLD systems. OCCS also developed a new XML distribution of Voyager bibliographic data. The XML allows data sharing between Library Operations and NCBI's PubMed Entrez search and retrieval system.

Literature Selection Technical Review Committee (LSTRC): OCCS defined a number of Impromptu reports for the LSTRC production system to increase the Committee's ability to manage its increasingly complex functions and transactions. In addition, the OCCS support team tuned and modified the database configuration to improve performance.

OLDMEDLINE: As of Version 1.6 of DOCLINE and Loansome Doc, released in June 2003, users can order OLDMEDLINE documents (1953 to 1965) using the Gateway's search capabilities. OLDMEDLINE consists of journal citations from 1966 and earlier that were not included in MEDLINE. Numerous rounds of cleanup have been required to meet current NLM citation data standards. While the cleanup is on-going OLDMEDLINE has been stored in the NLM Gateway environment. To better integrate OLDMEDLINE citations with MEDLINE and improve public access, the OLDMEDLINE citations are currently being further cleaned up and loaded into the PubMed Entrez system. So far over 1.5 million citations have been sent to NCBI for loading. Availability to the public via PubMed is targeted for the end of the fiscal year. OLDMEDLINE will remain accessible from the NLM Gateway but via link to the MEDLINE/PubMed collection.

Medical Subject Headings (MeSH): This year the OCCS MeSH team supported development of two major upgrades: the MeSH Translation Maintenance System (MTMS) and the Global Change Maintenance System (GCMS). MTMS is an interlingual database of translations that permits automatic updating of the MeSH terminology tree in all languages. The German database and front-end application will be tested later this year. In the past, changes to the MeSH database have waited until year-end processing to be propagated to existing MEDLINE citations. GCMS allows propagation on demand. The MeSH component of GCMS (called MHGCMS) has entered production and is in use for propagating MeSH term changes. The Keyword maintenance system (called KWGCMS) provides a similar capability for citations in SPACELINE and other specialty areas managed by NLM's collaborating partners. Development of the KWGCMS is in progress with production anticipated for late 2003.

Data Creation and Maintenance System (DCMS): This year OCCS imported POPLINE citations into DCMS and produced an XML POPLINE data set for export to the public. The team also added a new "PubMed—not-Medline" review stream to allow processing of citations that go to PubMed without indexing. An important new function expedites the release of Publisher data while in process so that citations needing minor corrections can get released quickly. In the past, dissemination often waited for weeks until after indexing. A Daily Update batch job now updates the bibliographic data in existing citations when modifications are made in the Voyager ILS, thus keeping MEDLINE citations consistent with Voyager on a daily basis. An Indexing Consistency Study function was added for analysis of the indexing process. The study function forces an issue marked by the administrator to be indexed four times so the various rounds can be analyzed.

Serials Extract File (SEF): Enhancements this year increased the effectiveness and manageability of this important system component, which is essential to the performance of many NLM custom applications. OCCS added a serial viewer and a MEDLINE update approval function to the Serials Extract File. The Serials Viewer allows users to select a serials title and display the data for that title in the LSTRC and SEF databases, as an aid to debugging inconsistencies. A MEDLINE update approval function allows system administrator to authorize automatic MEDLINE updates triggered by changes in SEF data. The program displays the number of records affected and queries the user for approval. Work is under way to extend the SEF design to contain bibliographic data for Meeting Abstracts. These documents, originally cataloged as monographs, will be imported for indexing into DCMS in the near future.

Research and Development

Advanced Search Engine: The new search engine significantly increases the ability of public users seeking information with inexact knowledge of medical terminology and concepts to usefully interact with MedlinePlus and other NLM consumer databases. Early this year, the RecomMind concept-based search engine was added to MedlinePlus English and later added to the NLM Intranet and the NLM Gateway. The new technology has required configuration changes, tuning, and a host of improvements in collaboration with the vendor. The RecomMind patented Probabilistic Latent Semantic Analysis engine analyzes search terms entered by the user to infer meaning from the context. For example, a user researching work-related lung diseases might type in the terms "occupational lung diseases." The string contains no exact matches in the database, but the search engine recognizes an associated concept and locates articles with information on occupational asthma, occupational cancer, etc. The Library's implementation includes another important benefit, the "See References" for MedlinePlus health topics. For example, if a search contains "abnormality" the response will suggest searching on "birth defects" as well. In all cases, matches are sorted into category folders for easier assessment by

the user. Searches are diacritic-neutral producing relevant matches even if users enter international terms without the diacritics.

Accent (Accessibility Enhancement) Project ("Talking Web"): OCCS developed and integrated a technology package called Accent into the Senior Health Project. Accent enables a Web server to provide content to vision-impaired users through text magnification, color contrast, and content narration. The visual and auditory enhancements are produced at the server so the user requires no additional hardware/software. Team members researched and modified the audio player to improve performance and audio quality. Audio file compression ensures suitable voice quality for modem connection users and maintains general quality levels. In the fourth quarter senior citizens throughout the country used a demonstrations version. The positive results confirmed the appropriateness of the design and feedback provided guidance for further development. Audio pronunciation tuning will be an on-going effort with Library Operations content experts through the remainder of the year.

Digital Archive: NIH has chartered a Working Group to explore cost-effective measures to ensure that electronic information labeled as "permanent" remains accessible without adversely impacting searches for more current material. In response, the Library is currently seeking an archive solution for its main web site. OCCS has researched the feasibility of adapting NLM's TeamSite document management system for this task. The Digital Archive component of the OCLC Connexion product, as well as the DSpace software from Massachusetts Institute of Technology, were also evaluated. After extensive analysis and testing, TeamSite appears to offer the most promising basis for adaptation. Work in the second half of the year focused on defining the initial project phases and expanding TeamSite's capabilities. Phase 1 and 2 will be in production in late 2003; Phase 3 is now in specification and analysis.

NLM Web Support

Web Content Management: NLM's content management software, TeamSite, was deployed in April 2002 to strengthen quality assurance for Web projects. New systems placed under TeamSite control in this past year include Locatorplus, NativeMinds, and the NLM Technical Bulletin. The OCCS team configured templates appropriate for each project, providing tailored workflow and configuration management rules. The NLM Technical Bulletin is still in progress. The team integrated PDF and Microsoft Office files into the validation libraries and added PDF and Microsoft Office file validation to the workflow.

Outreach

Consumer Outreach and Health System: OCCS developed this system to support the Library's Consumer Outreach and Health System (COHS)**.** Work began in early 2002 and after extensive testing by Regional Medical Libraries, the system entered production in spring 2003. Important organization procedures, security levels, and automated functions were implemented, together with a Report and Analysis module to support inter-organizational operations. The OCCS team also enhanced the user interface and designed the XML and database schema for data sharing and distribution

Web Exhibits Project: OCCS created an exhibit activity tracking system this year to coordinate and manage NLM's presence at national meetings. The database tracks activities initiated by internal NLM staff and the staff of the Regional Medical Libraries. It provides a full set of user-friendly data entry forms and an evolving report generating capability for management oversight and coordination.

HSRProj: OCCS supported the periodic loading and indexing of updates from the content provider, the University of North Carolina, and substantially enhanced the interface to improve the productivity of site indexers. The indexers can now invoke the MeSH browser for a selected MeSH term or subheading and navigate more easily with an improved Web page design and additional user controls. HSRProj is an online database accessible through the NLM Gateway that provides information on current research funded by federal and private grants and contracts. Its users are policy makers, managers, clinicians and other decision makers.

Health Services Research Resources (HSRR): In 2003, OCCS upgraded HSRR back-office modules used by system administrators and content managers. New functionality includes online help, a link checker program, advanced search limited by fields and record type, operations management reports, and a "Suggest a Tool" page where users can recommend online research tools for inclusion on the HSRR site. HSRR is used by the National Information Center on Health Services Research & Health Care Technology to post information on datasets, instruments, and software frequently used in health services research and in the behavioral and social sciences.

Administrative Support Systems

Customer Service Support System: OCCS added an e-mail interface between the Siebel Service Request and Change Request System and the NativeMinds virtual customer service system. If a customer is unsuccessful finding an answer in NativeMinds, the system prompts the user to send e-mail to Siebel. In addition, the Siebel system has been extended to provide change control for OCCS development groups. Developers register change requests in the system, which tracks the status of the change implementation. The system is programmed to enforce workflow rules that ensure the prescribed authorizations and procedural steps are followed.

ADMINISTRATION

Jon G. Retzlaff
Executive Officer

Table 12

Financial Resources and Allocations, FY 2003
(Dollars in Thousands)

Budget Allocation:	
Extramural Programs	$66,670
Intramural Programs	228,341
Library Operations	(88,163)
Lister Hill National Center for Biomedical Communications	(59,452)
National Center for Biotechnology Information	(66,730)
Toxicology Information	(13,996)
Research Management and Support	11,034
Total Appropriation	306,045
Plus: Reimbursements	12,281
Total Resources	$318,326

Personnel

In October 2002, **Mr. Aaron Ucko**, joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. He received his B.S. degree in computer science and theoretical mathematics from the Massachusetts Institute of Technology in 2000. In 2001 Mr. Ucko came to NCBI as a contractor. Since then he has been responsible for the support and improvement of the NCBI C++ Toolkit development framework. Mr. Ucko has developed the IDXE2INDEX, a port of the current C-based sequence indexer to the C++ Toolkit and contributed to the development of the ID1_FETCH port of the Toolkit. His current position at NCBI will entail the support and improvement of the infrastructure of the Toolkit framework on multiple platforms.

In November 2002, **C. Carl Jaffe, M.D.**, joined the staff of LHNCBC as a Visiting Faculty member. Dr. Jaffe received his bachelor's degree from MIT and his M.D. degree from Columbia College of Physicians and Surgeons. Dr. Jaffe is Professor of Diagnostic Radiology and Medicine (Cardiology), at Yale University School of Medicine, New Haven, CT. Dr. Jaffe is exploring opportunities and future directions for the Visible Human Project and its related efforts. In addition Dr. Jaffe is collaborating with other LHNCBC staff to create visualizations from coronary artery intravascular ultrasound data using the Visible Human Project's Insight Toolkit (ITK).

In November 2002, **Mr. Viatcheslav Khotomlianski**, joined the Information Engineering Branch, NCBI as a Staff Scientist. Mr. Khotomlianski obtained his M.S. degree in computer science from the Moscow Institute of Economics and Statistics in 1973. Subsequently, he worked in various companies designing, implementing, testing and evaluating databases and assisted in developing and implementing the Soviet relational database management system (RDBMS) *UNISON*. In 1999, he joined NCBI under contract as Senior Database Administrator. Since then, Mr. Khotomlianski has developed the system to allow flexible access to the several ASE servers NCBI uses and designed the database and system procedure. In his current position, Mr. Khotomlianski's extensive experience will continue to assist him in developing NCBI's database as it rapidly expands in size and complexity.

In November 2002, **Leonardo Mariño-Ramírez, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as a Research Fellow. Dr. Mariño-Ramírez received his Ph.D. in biochemistry in 2002 from Texas A&M University where he was a Fulbright Fellow. After an introductory course in Computational Biology in Trieste, Italy, Dr. Mariño-Ramírez taught himself UNIX, PERL, and SQL. He quickly learned how to store his growing dataset on the computers in the laboratory and also how to analyze the predominantly sequence data. Dr. Mariño-Ramírez has become proficient as a PERL and SQL programmer and has a significant skill set to extend his training in computational biology. At the NCBI, Dr. Mariño-Ramírez will primarily initiate a new project on protein-protein interactions.

In December 2002, **Lori Black, Ph.D.**, joined the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Black received her Ph.D. in biology (emphasis genetics) from Johns Hopkins University in 1995. Subsequently, Dr. Black completed a brief postdoctoral fellowship at Johns Hopkins and later she joined the National Cancer Institute as an Intramural Research Training Award Fellow in the Laboratory of Genomic Diversity. In November 1998, Dr. Black began working at GenBank as a contractor for ComputerCraft. As a Scientific Data Analyst, Dr. Black has been involved in all phases of annotating and validating nucleotide sequences electronically deposited to GenBank. In addition, she is part of a rotation schedule responsible for acting upon and replying to the e-mail correspondence received by GenBank. At NCBI, she will be communicating with various facets of the scientific community to maintain the integrity of the GenBank database while striving to increase general indexing efficiency and quality.

In December 2002, **DeAnne Cravaritis, Ph.D.**, joined the Information Engineering Branch, NCBI as a Staff Scientist. She earned a Ph.D. in molecular biology from Vanderbilt University in 1998. In 2002, Dr. Cravaritis joined ComputerCraft Corp. to work in the GenBank indexing group at NCBI as a Scientific Data Analyst. She has been involved in processing incoming

GenBank direct submissions and has gained extensive experience in the use of BLAST to validate nucleotide and protein sequences as well as to determine if additional relevant biological information should be added to records. In her current position, Dr. Cravaritis will continue to apply her strong biological background and communications skills necessary for building and maintaining GenBank.

In December 2002, **Bisharah Libbus, Ph.D.**, joined the staff of LHNCBC as a Visiting Scholar. Dr. Libbus received his bachelors and masters degrees from the American University of Beirut, Lebanon. He received a Ph.D. degree in genetics from the University of Missouri, Columbia. At LHNCBC, Dr. Libbus is investigating the interaction of molecular biology and medical informatics, including the expansion of existing knowledge sources with genomic phenomena and the automatic identification of such concepts in text.

In January 2003, **Ms. Hsiu-Chuan Chen**, joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Ms. Chen received her M.S. degree in computer science from the University of Maryland in 1988. Following this, she worked for four years at the American Type Culture collection (ATCC) in Rockville where she was involved in the determination of systems requirements for relational databases. Later Ms. Chen joined Management Systems Designers as a contractor working on NCBI's projects. As a Staff Scientist, Ms. Chen will continue her work on genome projects and the presentation of genome data on the NCBI Website.

In January 2003, **Laura C. Dean, M.D.**, joined the staff of the Information Engineering Branch, NCBI as a Visiting Research Fellow. Dr. Dean received her M.D. from the University of Cambridge, England in 2000. Alongside her clinical work, Dr. Dean also played a significant role in research based on the Cambridge Biomedical Computing Unit. She developed the first clinical CoffeeBreak and assisted with new content on the Genes and Disease Website. The Bookshelf is in a phase of content growth and requires someone with a biomedical background to work with authors to explore effective ways to achieve the creation of online content. Dr. Dean has an ideal background for this role.

In January 2003, **Song Mao, Ph.D.**, joined the LHNCBC staff as a postdoctoral fellow in medical informatics. Dr. Mao, a native of the People's Republic of China, received his bachelors and masters of science degrees in engineering from Tianjin University. Dr. Mao received his Ph.D. degree in Electrical and Computer Engineering from the University of Maryland, College Park. At LHNCBC, Dr. Mao is doing research on page segmentation algorithms.

In January 2003, **Teresa Przytycka, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as an Investigator (Tenure Track). Dr. Przytycka received her Ph.D. in computer science from the University of British Columbia in Vancouver, Canada in 1990. Most recently, Dr. Przytycka worked as an Associate Research Scientist in the Department of Biophysics at Johns Hopkins University where she conducted research in computational biology with particular emphasis on protein structure and classification. Dr. Przytycka will continue to work on a variety of problems addressing the classification and identification of protein structures. Dr. Przytycka's research is at the forefront of protein structure prediction, comparison, and classification.

In January 2003, **Mohammad Al-Ubaydli, M.D.,** joined the staff of the Information Engineering Branch, NCBI as a Visiting Research Fellow. Dr. Al-Ubaydli received his M.D. from the University of Cambridge, England in 2000. During his clinical training, Dr. Al-Ubaydli developed Palm Pilot software for an electronic records system. Dr. Al-Ubaydli has the rare combination of a medical background and programming experience in areas highly related to NCBI's Bookshelf project and its future development. His experience in developing modeling tools and online content systems will assist related projects here at NLM.

In January 2003, **Ming Xu, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as a Post-Doctoral IRTA. Dr. Xu received his Ph.D. in chemistry from Florida State University in Tallahassee in 1997. Subsequently, he worked as a Research Associate with the Department of Medicine at Brigham and Women's Hospital/Harvard Medical School, Boston, MA. While at NCBI, Dr. Xu will develop algorithms for protein sequence database search based on peptide mass spectra. Dr. Xu will also explore the bioinformatics issues associated with distributing search databases to mass spec labs and developing a protein identification service at NCBI.

In February 2003, **Mehmet M. Kayaalp, M.D., Ph.D.**, joined the staff of the Cognitive Science Branch of the Lister Hill Center as a Staff Scientist. Dr. Kayaalp, received his M.D. degree from the Istanbul School of Medicine in Turkey in 1989 and his Ph.D. in intelligent systems (medical informatics track) from the University of Pittsburgh in 2002. Since 1998, he has served as a researcher in Identifying Patient Subsets (IPS) of Interest in Electronic Medical Record Repositories, Center for Biomedical Informatics, University of Pittsburgh. Dr. Kayaalp brings his skills of knowledge representation, modeling, inference and connecting specific queries of clinical questions to the team being assembled to focus on biomedical knowledge discovery. This will include working on knowledge-based representation and data mining.

In February 2003, **Robert Logan, Ph.D.**, joined the staff of LHNCBC as a Visiting Scholar. Dr. Logan received his doctorate degree in Mass Communication from University of Iowa. Prior to coming to NLM, Dr. Logan was Associate Dean of the Science Journalism Center at the University of Missouri School of Journalism. At LHNCBC, Dr. Logan is working on

research projects in the area of system usability and user satisfaction for NLM's public information systems, such as MedlinePlus and ClinicalTrials.gov.

In March 2003, **Tanya Barrett, Ph.D.**, joined the staff of the Information Engineering Branch, NCBI, as a Staff Scientist (Visiting Program). Dr. Barrett received her Ph.D. in molecular and cell biology from the Institute of Medical Science, University of Aberdeen, U.K. in 1998. At NCBI, Dr. Barrett has taken a lead role in curating data deposited into NCBI's Gene Expression Omnibus (GEO) project. This role required Dr. Barrett to analyze data, and design and implement custom data curation methodologies. Dr. Barrett will continue to lead the development and implementation of high-throughput data curation methodologies that will greatly increase the utility of this data resource.

In March 2003, **Dmitriy Beloslyudtsev**, joined the Information Research Branch, NCBI as a Staff Scientist. Mr. Beloslyudtsev received his B.S. degree in engineering physics with a major in electronics in 1989 from the Moscow Institute of Physics and Technology. In 1998, he joined Informax, Inc. and began work as an on-site contractor at NCBI. Prior to joining the NCBI, he held successively more responsible positions in software programming and systems administration resulting in an appointment as Manager of the Internet Service Providers Department of Euro Intersoft, where he worked in the development of Internet and intranet communications. Mr. Beloslyudtsev will continue to assume a critical role as a liaison between the system development and support activities of the NCBI.

In March 2003, **Denis Vakatov**, joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Mr. Vakatov received his M.S. degree in mathematics and physics in 1992 from the Moscow Institute of Physics and Technology. From 1992 to 1996 he worked as a researcher in the Joint Institute for Nuclear Research in Dubna, Russia providing software support for several research groups, both national and international. In 1996, Mr. Vakatov joined NCBI as a contractor with Informax, Inc. Mr. Vakatov is now participating in the development of the C++ toolkit and its application to more specifically biological and genomic problems.

In April 2003, **Janet Heekin, MLS**, joined the LHNCBC staff as a Senior Systems Librarian. Ms. Heekin received her bachelor's degree from the University of Cincinnati and her master of library science from Indiana University. She has worked as a health librarian since 1989, most recently as a biomedical librarian at the NIH Library. At LHNCBC, Ms. Heekin has joined the consumer health project team and is working on data quality assurance methods.

In May 2003, **Bhanu Rajput, Ph.D.**, joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Rajput received her Ph.D. in 1992 from the University of British Columbia where she studied tRNA genes in Drosophila melanogaster. In February 2000, Dr. Rajput joined NCBI as a contractor with Computer Craft Corporation to work as a RefSeq scientist curating RefSeq sequence records, contributing to the LocusLink database, and performing QA review of data. In addition to her curation work, Dr. Rajput participates in other tasks necessary for the building and maintaining of the RefSeq collection and the LocusLink database.

In June 2003, **Melissa J. Landrum, Ph.D.**, joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Landrum received her Ph.D. in human genetics in 2000 from the Johns Hopkins University, where she focused on recombinant adeno-associated vector integration sites in the human genome. In September 2000, Dr. Landrum joined the Computer Craft Corporation to work at the NCBI as a RefSeq Scientists. As a RefSeq Scientist, Dr. Landrum is involved in curating RefSeq sequence records, contributing to LocusLink database content, and QA review of the data.

In June 2003, **Wei Ma**, was appointed Chief, Applications Branch, Office of Computer and Communications Systems (OCCS). Ms. Ma received her Master's degree in Computer Science from Utah State University in 1988. She joined NLM in 1998 as Head, Software Support Section, Applications Branch, OCCS. During her tenure with OCCS, she has served as technical project manager of various important applications development projects such as: MedlinePlus, NIHSeniorHealth.gov, the NLM Classification system, and others.

In June 2003, **Michael R. Murphy, Ph.D.**, joined the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Murphy received his Ph.D. in molecular and cellular biology in 1991 from the University of Massachusetts, Amherst where he developed a novel approach to study centromere function in Saccharomyces cerevisiae. In January 2000, Dr. Murphy joined NCBI as a contractor with Computer Craft Corporation to work as GenBank indexer checking the accuracy of sequence submissions. In December 2001, Dr. Murphy began working as a RefSeq scientist involved in curating RefSeq sequence records, contributing to LocusLink database content and performing QA review of the data.

In June 2003, **Michael North**, was appointed Head, Rare Books and Early Manuscripts Section, History of Medicine Division, Library Operations). He first joined HMD in October 2000 as a Rare Book Cataloger. Mr. North received his M.S. in Theoretical Linguistics from Georgetown University in 1992 and his M.S.L.S. from The Catholic University of America in 1995. Before coming to NLM, he worked in the historical collections of Georgetown University, The New York Academy of Medicine, and The Grolier Club of New York, where he was Curator of the Library. Mr. North will work to expand access to the rare book and early manuscript collections and to raise the Division's profile,

especially through digitization projects and public programs.

In June 2003, **Mr. Yan Raytselis**, joined the Information Engineering Branch, NCBI as a Staff Scientist. Mr. Raytselis, a native of the Ukraine, received his M.S. degree in physics from Odessa Institute of Technology in 1986 and a M.A. in psychology from Newport University in California in 1994. Before coming to NCBI, he developed software systems for the Bank of America. Since March 2002, Mr. Raytselis has been working at NCBI as a contractor with Management Systems Designers on the BLAST project. Mr. Raytselis has extensive experience in software engineering of distributed systems such as the new BLAST queuing system. Mr. Raytselis will be involved in tuning and improving the reliability of the database, and analyzing usage patterns to improve the BLAST system.

In June 2003, **Barbara J. Ruef, Ph.D.**, joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Ruef received her Ph.D. in molecular biology in 1992 from the University of California, Irvine where she focused on molecular parasitology. She later joined the Seattle Biomedical Research Institute as a postdoctoral scientist where her research focused on immunoparasitology. In September 2000, Dr. Ruef came to NCBI to work as a RefSeq scientist under a contract with Computer Craft Corporation. At NCBI, she will continue her curation work and participating in other tasks that are necessary for building and maintaining the RefSeq collection and the LocusLink database.

In June 2003, **Ms. Lorraine K.** Tanabe, joined the Information Engineering Branch, NCBI as a Staff Scientist. Ms. Tanabe received her B.S. degree in molecular biology in 1997 from San Jose State University. In July of 2000, Ms. Tanabe began working for NCBI as a contractor for Management Systems Designers while pursuing a Ph.D. in bioinformatics at George Mason University. While at NCBI she reprogrammed the Brill Tagger in the C++ language and incorporated data from the Specialist Lexicon into it. She is actively pursuing refinements of ABGene. Her goal is to develop a database of gene/protein names which matches the breadth of the literature and provides links to GenBank sequence records.

In June 2003, **Janet A. Weber, Ph.D.**, joined the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Weber received her Ph.D. in molecular and cell biology in 1994 from the Pennsylvania State University. She developed a ligation-mediated PCR approach to genomic footprinting using Drosophila embryos as a model system. Her system was used as the foundation for a Cold Spring Harbor course on genomic footprinting. In January 2001, Dr. Weber joined the Computer Craft Corporation to work at the NCBI as a RefSeq Scientist. Dr. Weber will continue her work curating RefSeq sequence records, contributing to LocusLink database content, and QA review of the data.

In July 2003, **Brent Bolin**, joined the Division of Specialized Information Services (SIS). Mr. Bolin is a member of the Emerging Leaders Program (ELP) of the Department of Health and Human Services. The ELP is a two-year career development program. Mr. Bolin received his J.D. and Master of Public Affairs degrees from the joint program at Indiana University–Bloomington, where he concentrated in environmental law, management and policy. During his rotation with SIS, Mr. Bolin will be working on the Public Health Law Information Project to develop a comprehensive database of public health legal information in the public domain.

In July 2003, **Michael N. DiCuccio, M.D.**, joined the staff of the Information Engineering Branch of NCBI as a Staff Scientist. Dr. DiCuccio received his medical degree from Duke University in 1996. He completed his residency there in internal medicine and at Thomas Jefferson University in radiology. At Jefferson, Dr. DiCuccio developed image processing algorithms to measure bone structural parameters. As a contractor with Management Systems Designers assigned to NCBI, Dr. DiCuccio was given sole responsibility for designing the software framework for NCBI's next generation sequence analysis and display toll for genome scale data, the "Genome Workbench" (GW). Dr. DiCuccio is assuming a leadership role to manage additional personnel to be assigned to the project and to represent NCBI in public presentations of GW.

In July 2003, **Pierre Ledoux, Ph.D.**, joined the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Ledoux received his Ph.D. in biology in 1996 from McGill University, where he investigated mutations responsible for prolidase deficiency, a metabolic enzyme deficiency. In March 1998, Dr. Ledoux joined NCBI as a contractor as a Scientific Data Analyst in the GenBank indexing group processing annotated nucleotide sequence records submitted to GenBank for online publication and ensuring the biological accuracy of these annotations. In addition, Dr. Ledoux has helped set up the protocols for processing reference sequence entries as part of NCBI's comprehensive reference genome collection. Dr. Ledoux will continue to be involved in all phases of GenBank submissions.

In August 2003, **David Gillikin**, was appointed Head, MEDLARS Management Section (MMS), Bibliographic Services Division, LO. Mr. Gillikin received his master's degree in library science from the University of Tennessee, Knoxville. Previously, Mr. Gillikin was responsible for project development and management of Web publishing projects at HighWire Press, including the daily operation and production processing for the 335 journal sites hosted by HighWire. From 1996 to 2001, Mr. Gillikin worked for the American Association for the Advancement of Science where he managed the transition of the content of Science and other publications to the Web. Earlier in his career he worked for BRS Technologies where he served as project manager for MEDLINE, EMBASE, PDQ, and other

databases. At NLM, Mr. Gillikin will lead MEDLINE-related work in system testing, quality assurance, usability testing, documentation, and training.

In August 2003, **Mr. Shiuan-Haur** (Howard) Lu, joined the Lister Hill National Center for Biomedical Communications (LHNCBC) staff as a Computer Scientist. Mr. Lu received an M.S. degree in Computer Science from Indiana University and a B.S. degree in engineering from the National Taiwan University. Mr. Lu holds two patents, one for a "Communication Resource Fault Detection" system and another for a "Method and System for Managing Communication Resources." At LHNCBC, Mr. Lu will coordinate Lister Hill Center computing resources, including existing computer systems, network architectures, and emerging technologies.

In August 2003, **Aaron B. Navarro, Ph.D.**, joined the LHNCBC staff as a Computer Scientist. Dr. Navarro received a Ph.D. degree in Computer Based Instruction from the University of Maryland and a M.S. degree in Systems Information Science from Syracuse University, and a B.S. degree in Applied Physics, magna cum laude, from the New York Institute of Technology. He has taught graduate level computer science courses at Johns Hopkins University and George Washington University. Dr. Navarro was on the NLM staff in the early 1990s, serving first as deputy and then as acting director of the Office of Computer and Communications Systems. At LHNCBC, Dr. Navarro will contribute to information technology research and development programs.

In August 2003, **Terry Wittig**, was appointed Head, Acquisitions Unit, Serial Records Section, Technical Services Division, LO. Ms. Wittig received her M.L.S. from the University of Pittsburgh in 1985. Previously she was Head, Collection Development and Preservation Officer at George Mason University Libraries and also served as Interim Head of Serials. Previous positions include working in Collection Management and Preservation at North Carolina State University Libraries and serving as Collections Coordinator for the Sciences in the Engineering and Science Library at Carnegie Mellon University. She also worked as Assistant Librarian for Technical Services at the Massachusetts Institute of Technology. At NLM, Ms. Wittig's responsibilities include management of serials acquisitions, including licensing electronic resources.

NLM Associate Fellowship Program

The NLM Associate Fellowship Program is a one-year training fellowship for recent graduates of Master's degree programs in library and information science. Fellows receive a comprehensive orientation to NLM programs and services during a structured five-month curriculum phase, and conduct individual projects over the remaining seven months. Projects are typically of a research, development, or evaluation nature. Eight new fellows began the program in September 2003.

Theodora A. Bakker received her M.L.I.S. in August 2003 from the University of Illinois. She has experience as a Graduate Assistant in the Library of the Health Sciences at the University of Illinois as well as additional experience in the Geology Library. Her undergraduate degree is in Philosophy.

Lonelyss B. Charles received her M.L.I.S. in August 2003 from the University of Pittsburgh. As a Highmark Fellow in the library school, Ms. Charles worked on a project to facilitate access to information resources for minority health consumers. She also has several years experience in business and public relations. In addition to the M.L.I.S., she holds a Master's in Education and B.A. in French and Liberal Arts.

Erinn E. Faiks received her M.S.I. in April 2003 from the University of Michigan, specializing in the Library and Information Services program within the School of Information. She has four years of library experience in the Public Health Information Services and Access unit of the University of Michigan. She also has experience at the Taubman Medical Library and the English Language Institute Library at Michigan. Her undergraduate training was in Spanish and Linguistics.

Barbara J. Few received her M.S.I. in April 2003 from the University of Michigan, specializing in the Human-Computer Interaction program within the School of Information. She has experience as an intern at Michigan's Taubman Medical Library. She holds B.S. and M.S. degrees in nursing, and comes to the program following a solid career in nursing practice and developing nursing performance improvement programs.

Julie K. Gaines received her M.L.I.S. in May 2003 from the University of South Carolina. She has experience as a Graduate Assistant at the Lexington Medical Center Library as well as Graduate Intern experience developing Web pages at the South Carolina State Library. Her undergraduate training was in Exercise Studies and Mathematics.

Jeffery L. Loo received his M.L.I.S. degree in May 2003 from the University of British Columbia in Vancouver. He has experience as a Graduate Academic Assistant in the Woodward Biomedical Library at UBC as well as varied experience in public and academic libraries. He also has experience as a Research Assistant in a chemistry laboratory, with the Canadian Biotechnology Secretariat, and with the Centre for Health Evaluation and Outcome Sciences. His undergraduate training was in chemistry.

Nancy Pulsipher received her M.S.I. in April 2001 from the University of Michigan, specializing in the Library and Information Services program within the School of Information. As Graduate Research Assistant, she worked

on projects related to information services for Native Americans. She has three years of reference and library instruction experience in the Public Health Information Services and Access unit of the University of Michigan. She also has experience in a Patient Education Resource Center and the Social Work Library at Michigan. Her undergraduate training was in Humanities and Business Administration.

Andrea N. Ryce received her M.L.I.S. degree in May 2003 from the University of British Columbia in Vancouver. She has experience as a Graduate Academic Assistant in the Woodward Biomedical Library at UBC. She has additional varied experience in public and academic libraries, including a project on digitizing and cataloging an editorial cartoon collection. She has a B.A. in English Literature.

Retirements and Separations

In January 2003, **Mr. Charles C. Herbert (Chuck)**, retired from his position as Assistant Director for Program Planning and Coordination with the LHNCBC and the Federal government after 41 years of service. Mr. Herbert began his federal career in 1961 as a Revenue Officer Trainee with the IRS. He transferred to the Federal Drug Administration in 1963 as a Management Analyst and then began his tenure with the NLM in 1966 as a Management Analyst Officer. He later became Assistant Director for Administration for the National Medical Audiovisual Center then located in Atlanta, Georgia. He accepted his most recent position of Assistant Director for Program Planning and Coordination with the Lister Hill Center in 1978.

In January 2003, **Susan Sparks, Ph.D.**, retired from her position as Senior Education Research Specialist with the Division of Extramural Programs and the Federal government with over 28 years of service. She joined the Library in February 1974 and spent her entire Federal career at NLM. With her strong background in nursing education and development of educational materials in the health professions, Dr. Sparks was responsible for reviewing the state-of-the art in educational technology, recommending types of research and appropriate development or demonstration, and designing grant and resource programs. As Senior Education Research Specialist, she monitored research grants and analyzed complex research projects. Her extensive background and experience at NLM will be greatly missed.

In April 2003, **Elizabeth Van Lenten, Ph.D.**, retired from her position as assistant Head of the Index Section, Library Operations. Dr. Van Lenten joined the NLM staff in 1968 shortly after receiving her Ph.D. in Biochemistry from Yale University. She became a Unit Head in 1980 and was named Project Officer for the Library's indexing contracts in 1986. In 1987, Dr. Van Lenten assumed the position of Assistant Head while remaining Project Officer. Hundreds of NLM staff and contract indexers have benefited from Dr. Van Lenten's expertise as a trainer, reviser, and mentor.

In April 2003, **Alex Lash, M.D.**, resigned his Staff Scientist position with NCBI. Dr. Lash came to NIH in 1994 as a resident in Anatomic Pathology in the Laboratory of Pathology, NCI and as a Lieutenant in the U.S. Public Health Service. Dr. Lash joined the Information Engineering Branch, NCBI, as a Staff Scientist in June 1998. He was involved in projects that link visual and molecular information on the CGAP project as a liaison with the NCI. Dr. Lash has accepted a position with the Computational Biology Center, Memorial Sloan-Kettering Cancer Center.

In August 2003, **Gabor Marth, D.Sc**., resigned his Staff Scientist position with NCBI. Dr. Marth received his Ph.D. in System Science and Mathematics from Washington University, St. Louis, in 1994. As a post-doctoral research associate at the Genome Sequencing Center of Washington University, he worked on a broad range of projects related to large-scale human genome sequence production. He joined the Information Engineering Branch, NCBI as a Staff Scientist in March 2000. Dr. Marth was involved in projects that focused on the theoretical and algorithmic improvement of sequence-based SNP detection. Dr. Marth accepted a faculty position in the Department of Biology at Boston College.

In September 2003, **Carol Bean, Ph.D**. left her Health Scientist Administrator position with the Division of Extramural Programs to join the staff of the National Center for Research Resources, NIH. Dr. Bean received her doctoral degree in biopsychology from the University of Georgia in 1985. Dr. Bean joined the NLM in January 2001 to provide technical expertise and scientific leadership and direction for major grants programs in the fields of informatics and biomedical computing. She played a lead role in determining the direction of program operations in the area of informatics as applied to health care delivery and to medical research.

Awards

The NLM Board of Regents Award for Scholarship or Technical Achievement was awarded to Dr. Stephen T. Sherry for his advisory role in employing DNA forensic methods to help identify victims of the World Trade Center Tragedy.

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 2003 award was Ms. Christa F. B. Hoffmann in recognition of leadership and vision in the reinvention of the production, dissemination, and maintenance of the NLM Classification.

The NLM Director's Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to three employees: Karen D. Riggs (OAM) for significant contributions in leading the award process and achieving a 59% increase in small purchases, while continuing to provide outstanding

customer service; Melanie A. Modlin (OCPL) for advancing the National Library of Medicine's Outreach Program through superior writing and editing and for exemplary management of the Visitors Program; and William R. Leonard (LHC) for outstanding work in video production crucial to the continuing development of a quality visual presence for the National Library of Medicine here and throughout the world.

The NIH Merit Award was presented to seven employees: Ms. Patricia A. Bosma (LO) for her dedicated management of the selection and acquisition of the modern biomedical literature for the NLM collection; Mr. Joseph P. Fitzgerald (LHC) for his organization, coordination and friendly leadership in effectively orchestrating the Turning of the Pages historical books program; Ms. Julia C. Player (LO) for her successful management of NLM's Interlibrary Loan Unit which provides document delivery to thousands of libraries in the U.S. and most countries abroad; and, as a group, the NLM IT Security Leadership Team—Jules P. Aronson (LHC), Rand S. Huntzinger (NCBI), Dar-Ning-Kung, Ph.D. (OCCS), and Phillip L. Thomas, Ph.D. (SIS) for superior work securing the existing network infrastructure and establishing a technological and procedural framework for the long-term protection of NLM's IT assets. In addition, Ms. Jane Bortnick Griffith (OD) was presented the NIH Merit Award as part of a group recognized by the NIH Office of the Director for outstanding leadership in assisting the biomedical research community to implement the Health Information Portability and Accountability Act.

The Philip C. Coleman Award recognizes significant contributions to the NLM by individuals who demonstrate outstanding ability to motivate colleagues. The recipient of the 2003 award was Judy C. Jordan of the Public Services Division.

The NLM EEO Special Achievement Award was presented to the NLM Reads Steering Committee: Cassandra R. Allen, Karen B. Casey, Dr. Keith W. Cogdill, James T. Dean, Dr. Stephen J. Greenberg, Yuen Yin K. Kwan, and Dr. Angela B. Ruffin.

Table 13

FY 2003 Full-Time Equivalents (Actual)

Office of the Director	12
Office of Health Information Programs Development	7
Office of Communications and Public Liaison	8
Office of Administration	46
Office of Computer and Communications Systems	53
Extramural Programs	18
Lister Hill National Center for Biomedical Communications	81
National Center for Biotechnology Information	138
Specialized Information Services	34
Library Operations	297
Total FTEs	**694**

NLM Diversity Council

The NLM Diversity Council began 2003 by welcoming five new members: Jason Donaldson, Tameka Gore, Donald Jenkins, Renee McLean-Banks, and Linda Tang. Each will serve a two-year term from January 2003 through December 2004. Continuing on the Council are: Tamar Clarke, Kathleen Cravedi, Felicia Derricott, James Knoben, and Michael Simpson. The Council continues to receive support from its ex-officio members: Jon Retzlaff, Executive Officer, David Nash from the Equal Employment Opportunity Office, and Nadgy Roey from the Office of Human Resources, as well as its distinguished alumni. Michael Simpson accepted the responsibilities of Council Chair and Tamar Clarke and Kathleen Cravedi became council Vice-Chairs.

FY2003 Accomplishments:

- *NLM Director's Employee Education Fund*: Continued coordination of the NLM Director's Employee Education Fund. In FY2003, the Fund enabled 46 staff to take 65 classes from 13 area schools. This is down from 57 staff taking 88 classes in FY2002. Of the NLM staff who have taken advantage of the Fund, 34 LO, 6 from the Office of the Director, 2 from OCCS, 5 from the NCBI, and 4 from SIS. There were no participants from the Lister Hill Center in FY2003. Undergraduate classes made up the majority of classes supported. The school with the largest number of NLM enrollees is the University of Maryland (17 attendees) with Montgomery College coming in second (15 attendees). Other institutions attended include: the American University, University of the District of Columbia, George Mason University, Bowie State University, Marymount University, Frostburg State, Coppin State, and Morgan State University. Course disciplines enrolled in included computer graphics, communications, business, Web database, English, law, 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. In fact, the Director's Employee Education Fund is currently featured on the Diversity Council bulletin board.
- *Facility Accessibility and Reasonable Accommodation:* The Council continued efforts to upgrade access at NLM for people with disabilities. 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 the addition of accessibility features in Conference Room B and in restrooms throughout the NLM, including:

  - LED Caption Display for Conference Room B. This device provides scrolling LED display of CART and real-time captioning to be seen by everyone in a large meeting room.
  - Accessibility features in many of the remaining bathrooms in NLM that have not yet been renovated to accommodate the disabled community.

- *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. The Council voted to have OCPL staffer Fran Sandridge attend meetings on an ex-officio basis to assist in the design of needed bulletin displays.
- *English Language Courses:* The Council began a new program to enable NLM employees to improve their proficiency in 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. English language instructors and students were selected by fall 2003. Two instructors have received training and the course is expected to be in place by the end of the year.
- *Ice Cream Social and "Laborless" Moment for NLMers.* Together with the Friends of the National Library of Medicine, the Diversity Council sponsored a "Laborless" Moment to honor NLMers. Ben and Jerry's supplied the ice cream and Alec Stone, Director of the FNLM, served up shaved ice cones from 1:00 to 3:00 p.m. on September 15 on the patio adjacent to the Lister Hill Auditorium. About 400 NLMers attended. All agreed that it was a great success and a nice way to thank staff for the great work. Many suggested it should become an annual event.
- *Reading Club.* The Diversity Council sponsors a reading club for interested employees. The club meets regularly and is attended by several dozen NLMers on a regular basis.
- *Looking Under Your Hood?* The Diversity Council piloted what is the first in a series of monthly lectures by Dr. Donald Jenkins. The lectures, as the title suggests, provide an overview of regions of the body, as shown in the David Bassett archive of images of human cadaveric anatomy as examples of human form. This lecture series is based on the belief that personal knowledge about the intricate structure of the human body is beneficial to health and well-being

Board of Regents

The Board of Regents (BOR) met three times in FY2003 on February 11-12, May 13-14, and September 9-10. The Extramural Programs (EP) Subcommittee and the Subcommittee on Outreach and Public Information were held during each of these meetings. Two working groups were established and were held in conjunction with the May and September BOR meetings. The Working Group on Biomedical Imaging and Bioengineering, which met on May 12 and September 8, was established to ensure that the NLM is collecting, organizing, preserving, and providing access to the literature which supports the mission of a new Institute, the National Institute of Biomedical Imaging and Bioengineering. The Working Group on Bioethics, which met on May 14 and September 11, was established to provide a comprehensive review of the collections, databases, and services that NLM supports in bioethics. This group will assess the breadth and quality, and indexing and cataloging coverage of the bioethics literature that has been collected and supported by the NLM.

During the FY2003 meetings, the Board heard reports from several NLM Divisions on their current activities, including the Lister Hill Center Board of Scientific Counselors, and the NCBI Board of Scientific Counselors, both of which are required to report to the Board as stated in their Charter. Grants-related activities are listed under the Extramural Programs section of this Annual Report.

A new Chair was elected to the Board of Regents, Ms. Eugenie Prime, Manager of Corporate Libraries at the Hewlett-Packard Laboratories in Palo Alto, California. Two new members also joined the Board in September: Dr. Vasiliki Karlis, Associate Professor, Department of Oral and Maxillofacial Surgery, New York University, and Dr. Holly Buchanan, Director and Professor, Health Sciences Library and Informatics Center, University of New Mexico.

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:

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

BUCHANAN, Holly S., Ed. D.
Director and Professor
Health Sciences Library & Informatics Center
University of New Mexico
Albuquerque, NM

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

Karlis, Vasiliki, D.M.D., M.D.
Associate Professor
Department of Oral and Maxillofacial Surgery
New York University College of Dentistry
New York, NY

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

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:

FULLER, Sherrilynne S., Ph.D. (Chair)
Professor of Biomedical & Health Informatics
University of Washington School of Medicine
Seattle, WA

CARTER, Jerome H., M.D.
Director, Division of Infectious Diseases
University of Alabama
Birmingham, AL

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

FERRIN, Thomas E., Ph.D.
Professor of Pharmaceutical Chemistry
University of California
San Francisco, CA

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

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

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

KWITEK, Anne E., Ph.D.
Assistant Prof., Dept. of Physiology
Human & Molecular Genetic Center
Medical College of Wisconsin
Milwaukee, WI

MACKAY, Trudy F., Ph.D.
Professor, Dept. of Genetics
North Carolina State University
Raleigh, NC

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

SALZBERG, Steven L., Ph.D.
Senior Director of Bioinformatics
The Institute for Genomic Research
Rockville, MD

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:

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

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
St. 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

CLAYTON, Paul D., Ph.D.
Chief Medical Informatics Officer
Intermountain Health Care
University of Utah
Salt Lake City, UT

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

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

OGUNYEMI, Omolola I., Ph.D.
Research Associate
Department of Radiology
Brigham and Women's Hospital
Boston, MA

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

TAIRA, Ricky K., Ph.D.
Associate Professor, Dept. of Radiology
University of California
Los Angeles, CA

TANJI, Virginia M.
Library Resource Center
School of Medicine
University of Hawaii at Monoa
Honolulu, HI

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

YOKOTE, Fail A.
Associate University Librarian
Peter J. Shield Library
University of California
Davis, 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:

TOLEDO-PEREYRA, Luis H., M.D., Ph.D. (Chair)
Director, Surgery Research & Molecular Biology
Borgess Medical Center
Kalamazoo, MI

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

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

FREY, John J., M.D.
Professor and Chair
Department of Family Medicine
University of Wisconsin
Madison, WI

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

SHARPS, Phyllis W., Ph.D.
Associate Professor
School of Nursing
Johns Hopkins University
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

STERNBERG, Esther M., M.D.
Director, Integrative Neural Immune Program
National Institute of Mental Health
Bethesda, MD

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

DELAMOTHE, Anthony P., M.D.
Editor, British Medical Journal
London, England

EISEN, Michael B
Genome Sciences
Lawrence Berkeley National Laboratory
University of California
Berkeley, CA

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

JOHNSON, Richard K.
Enterprise Director
Scholarly Publishing & Academic Resources Coalition
Washington, D.C.

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

KIRSCHNER, Marc W., Ph.D.
Professor and Chair
Department of Cell Biology
Harvard Medical School
Boston, MA

RUBIN, Gerald M., Ph.D.
Investigator
Howard Hughes Medical Institute
Chevy Chase, MD

THOMAS, Sarah E., Ph.D.
Carl A. Kroch University Librarian
Cornell University
Ithaca, NY

VARKI, Ajit P., M.D.
Professor of Cellular Biology & Molecular Medicine
University of California
San Diego, CA

WATSON, Linda A.
Director, Claude Moore Health Science Library
University of Virginia
Charlottesville, VA

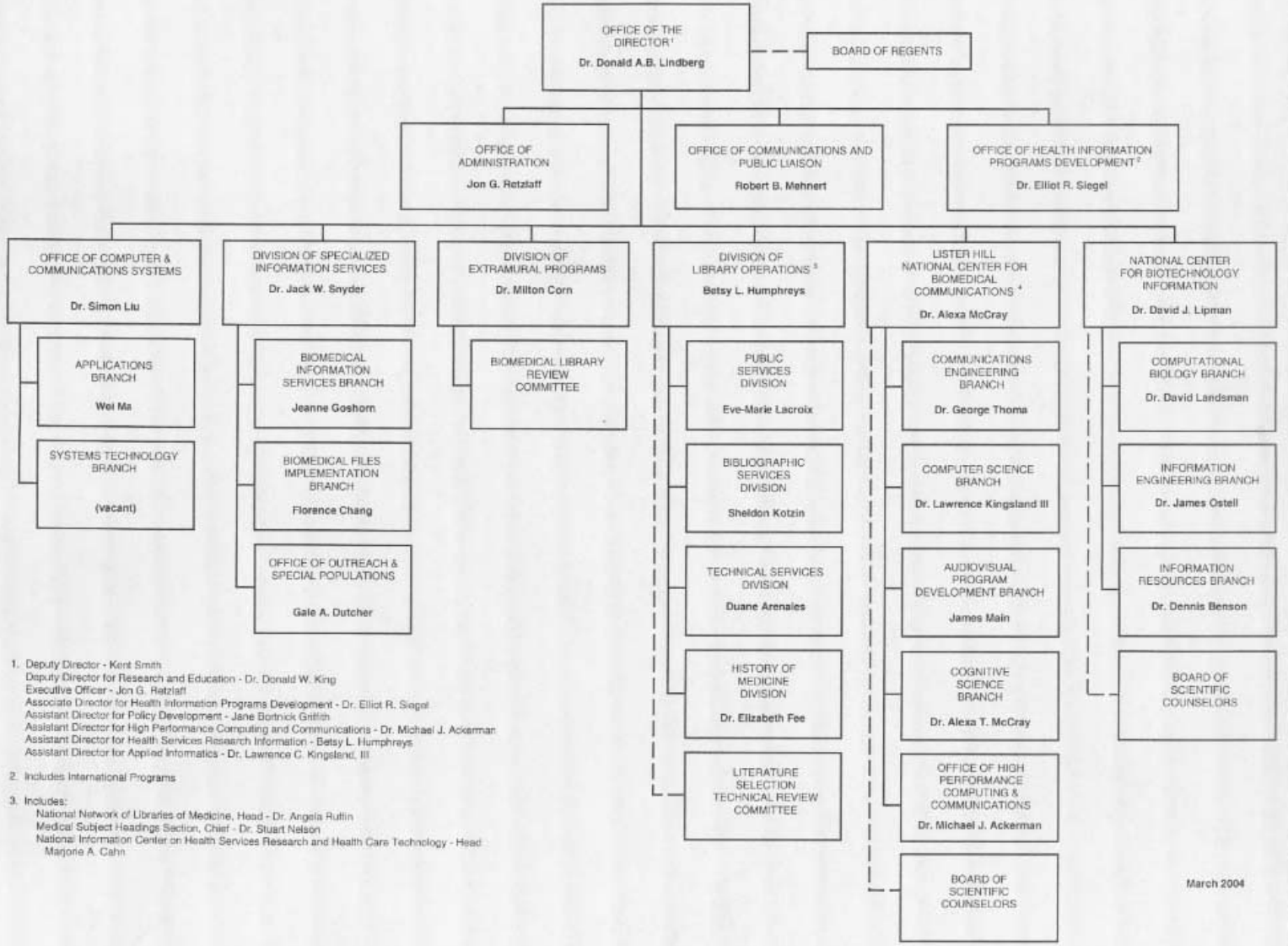
WILLIAMS, James F.
Dean of Libraries
University of Colorado
Boulder, CO

APPENDIX 8: ORGANIZATIONAL ACRONYMS USED IN THIS REPORT

AAHSL	Association of Academic Health Sciences Libraries
ACSI	American Consumer Satisfaction Index
ACTIS	AIDS Clinical Trials Information Service
ALTBIB	Alternatives to Animal Testing
AMPA	American Medical Publishers Association
APDB	Audiovisual Program Development Branch
API	Applications Programmer's Interface
ASQ	American Society for Quality
ATIS	HIV/AIDS Treatment Information Service
ATSDR	Agency for Toxic Substances and Disease Registry
BISTI	Biomedical Information Science and Technology Initiative
BLRC	Biomedical Library Review Committee
BOR	Board of Regents
BSD	Bibliographic Services Division
BSN	Bioinformatics Support Network
CBIR	Content-Based Image Retrieval
CCRIS	Chemical Carcinogenesis Research Information System
CDC	Centers for Disease Control and Prevention
CDD	Conserved Domain Database
CEB	Communications Engineering Branch
CgSB	Cognitive Science Branch
ChemIDplus	Chemical Identification File
CIT	Center for Information Technology
CSB	Computer Science Branch
DART	Developmental and Reproductive Toxicology
DDBJ	DNA Data Bank of Japan
DHHS	Department of Health and Human Services
DIRLINE	Directory of Information Resources Online
DTD	Document Type Definition
EBI	European Bioinformatics Institute
EEO	Equal Employment Opportunity
EFTS	Electronic Funds Transfer Service
EMBL	European Molecular Biology Laboratory
EMIC	Environmental Mutagen Information Center
EnHIOP	Environmental Health Information Outreach Panel
EP	Extramural Programs
EST	Expressed Sequence Tags
EPA	Environmental Protection Agency
ETICBACK	Environmental Teratology Information Center backfile
FDA	Food and Drug Administration
FNLM	Friends of the National Library of Medicine
GEO	Gene Expression Omnibus
GUI	Graphical User Interface
HBCU	Historically Black Colleges and Universities
HEAL	Heal Education Assets Library
HHS	Health and Human Services (Department)
HIPAA	Health Insurance Portability and Accounting Act
HMD	History of Medicine Division
HSDB	Hazardous Substances Data Bank
HSRProj	Health Services Research Projects
HSRR	Health Services and Sciences Research Resources
HSTAT	Health Services and Technology Assessment Text
IADL	Internet Access to Digital Libraries
IAIMS	Integrated Advanced Information Management Systems
ICPC	International Classification of Primary Care
ICs	Institutes and Centers (of NIH)
ILL	Interlibrary Loan
IRIS	Integrated Risk Information System
ITK	Insight Toolkit
JD	Journal Descriptor
KSS	Knowledge Source Server
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
MARG	Medical Article Records Groundtruth
MARS	Medical Article Records System
MEDLARS	Medical Literature Analysis and Retrieval System
MeSH	Medical Subject Headings
MHC	Major Histocompatibility Complex
MLA	Medical Library Association
MLAA	Medical Library Assistance Act
MMDB	Molecular Modeling Database
MMS	MEDLARS Management Section
MMTx	MetaMap Technology Transfer
MTI	Medical Text Indexer
NCBI	National Center for Biotechnology Information
NCI	National Cancer Institute
NCVHS	National Committee on Vital and Health Statistics

NGI	Next Generation Internet
NHANES	National Heath and Nutrition Examination Surveys
NHII	National Health Information Infrastructure
NHLBI	National Heart, Lung, and Blood Institute
NIA	National Institute on Aging
NIAID	National Institute of Allergy and Infectious Diseases
NIBIB	National Institute of Biomedical Imaging and Bioengineering
NICHSR	National Information Center on Health Services Research and Health Care Technology
NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
NIOSH	National Institute for Occupational Safety and Health
NLM	National Library of Medicine
NN/LM	National Network of Libraries of Medicine
NNMC	National Naval Medical Center
NNO	National Network Office
NOSC	Network Operations and Security Center
NSF	National Science Foundation
NTCC	National Online Training Center and Clearinghouse
OAM	Office of Administrative Management
OCCS	Office of Computer and Communications Systems
OCHD	Outreach, Consumer Health, & Health Disparities Coordinating Committee
OCPL	Office of Communications and Public Liaison
OCR	Optical Character Recognition
OD	Office of the Director
OHIPD	Office of Health Information Programs Development
OHPCC	Office of High Performance Computing and Communications
OMIM	Online Mendelian Inheritance in Man
PAHO	Pan American Health Organization
PDA	Personal Digital Assistant
PDB	Protein Data Bank
PHS	Public Health Service
PMC	PubMed Central
PROW	Protein Reviews on the Web
PSD	Public Services Division
QoS	Quality of Service
RefSeq	Reference Sequence Database
RML	Regional Medical Library
RTECS	Registry of Toxic Effects of Chemical Substances
SBIR	Small Business Innovation Research
SII	Scalable Information Infrastructure
SIS	Specialized Information Services
SNOMED CT	Systematized Nomenclature of Medicine Clinical Terms
SNP	Single Nucleotide Polymorphism
SOAP	Simple Object Access Protocol
SSEUS	SIS SQL Entry Update System
TEHIP	Toxicology and Environmental Health Information Program
TIE	Telemedicine Information Exchange
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
UPS	Uninterruptible Power Supply
USUHS	Uniformed Services University of the Health Sciences
VAST	Vector Alignment Search Tool
VHP	Visible Human Project
WebMIRS	Web-based Medical Information Retrieval System
Web-STOC	Web-Services Technology Operations Center
WGS	Whole Genome Shotgun
WISER	Wireless Information System for Emergency Responders

National Library of Medicine



1. Deputy Director - Kent Smith
   Deputy Director for Research and Education - Dr. Donald W. King
   Executive Officer - Jon G. Retzlaff
   Associate Director for Health Information Programs Development - Dr. Elliot R. Siegel
   Assistant Director for Policy Development - Jane Bortnick Griffith
   Assistant Director for High Performance Computing and Communications - Dr. Michael J. Ackerman
   Assistant Director for Health Services Research Information - Betsy L. Humphreys
   Assistant Director for Applied Informatics - Dr. Lawrence C. Kingsland, III

2. Includes International Programs

3. Includes:
   National Network of Libraries of Medicine, Head - Dr. Angela Ruffin
   Medical Subject Headings Section, Chief - Dr. Stuart Nelson
   National Information Center on Health Services Research and Health Care Technology - Head Marjorie A. Cahn

March 2004

NIH Publication No. 04-256