NATIONAL INSTITUTES OF HEALTH
NATIONAL LIBRARY OF MEDICINE
PROGRAMS & SERVICES FY 2006

CATCGAT
ATGTACCGTACC
CAATCCTAGAG
AGCATCCTCGAT
TCCTAGTCAGA
TGTACCATCGATGC
ACCATGTACCGTACC
TCGAATCCT

U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES

Cover: From "Visible Proofs: Forensic Views of the Body," a major exhibition opened at the Library in February 2006.

NATIONAL INSTITUTES OF HEALTH

National Library of Medicine

Programs and Services

Fiscal Year 2006

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

CONTENTS

Appendixes

Tables

Preface

A signal event in Fiscal Year 2006 was the completion of "Charting a Course for the 21st Century: NLM's Long Range Plan 2006–1016" and its approval by the Board of Regents at their September 2006 meeting. Over the years the Library has found that the signposts such roadmaps give us are of tremendous help in making program and budget decisions. The report is described in the chapter on the Office of Health Information Programs Development.

Another important event this year was the awarding of eight five-year contracts following a recompetition of the Regional Medical Libraries. The RMLs administer the many outreach and other programs carried out by the 5800 full and affiliate member institutions in the National Network of Libraries of Medicine. The Network is a vital component of NLM's efforts to ensure that medical information—scientific and consumer—is accessible by all. We at the NLM truly appreciate their work. A description of the work of the Network is in the Library Operations chapter and a list of the Regional Medical Libraries is in Appendix 1. Among other highlights in this report:

- The remarkable growth in genomic resources and tools created by NLM's National Center for Biotechnology Information;
- Opening of the exhibition "Visible Proofs: Forensic Views of the Body" in February 2006;
- The launch of a new quarterly *NIH MedlinePlus Magazine* in September 2006 and the addition of a weekly MedlinePlus PodCast;
- Introduction at the end of FY2006 of Tox Mystery, an interactive Web site for children; and
- The continuing evolution of NLM's Unified Medical Language System and its Metathesaurus, and our work in helping to coordinate clinical vocabularies.

This year I extend my thanks not only to the Library's outstanding staff and to the many advisors who give generously of their time to serve on Boards and Councils, but to the more than 100 experts whose views shaped the Long Range Plan that will guide us in the next decade.

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

OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT

Elliot R. Siegel, Ph.D.
Associate Director

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

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

Planning and Analysis

The NLM Long Range Plan 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.

At its September 2004 meeting, the NLM Board of Regents decided to develop a Long Range Plan for 2006–2016. A Subcommittee on Planning was appointed that was co-chaired by the Honorable Newt Gingrich and Dr. William Stead; members included Dr. Holly Buchanan, Dr. Wallace Conerly, Dr. Thomas Detre, and Dr. Kenneth Walker.

In April 2005, a Strategic Visions Working Group comprised of outstanding leaders from all sectors of NLM's diverse constituencies met to provide the broadest view of NLM's mission, current situation, and its potential future contributions to the health and well-being of America in the 21st century. A vision statement identified new scientific, medical, technical, social, and economic developments that might impact national and global needs for research, clinical and patient data and information. It formed the basis for the creation of four long range planning panels that met four times in 2005–2006:

- Resources and Infrastructure: Dr. Edward Shortliffe and Ms. Gail Yokote (Chairs)
- Health Information for Underserved and Diverse Populations: Dr. Louis Sullivan and Ms. Eugenie Prime (Chairs)
- Support for Clinical and Public Health Systems: Dr. Reed Gardner (Chair)
- Support for Genomic Science: Dr. Daphne Preuss (Chair)

Nearly 100 panelists worked to identify the forward-looking strategies and infrastructure that will enable NLM to maintain its role as a premier national library and positive force for change in the U.S. and abroad in the 21st century. The panelists considered, among many relevant issues and trends, exciting changes in genomic and computer science, scientific publication models, transformational changes in health care delivery (including electronic health records), and quality and safety made possible by new information technology. The promise of new research correlating genotype, phenotype and environmental data figured prominently in their deliberations, as did the challenges posed by a critical lack of space needed to house NLM's programs and collections. Other major factors considered were the existence of health disparities among the underserved, a lack of trust in societal institutions (including government), and the mitigation of threats to the public health from disasters and epidemics.

At its May 2006 meeting, the Board accepted with thanks the individual reports of the four planning panels, discussed their recommendations, and requested that NLM staff prepare a consolidated 10-year plan based on these reports along with appropriate staff input.

The Board approved *Charting a Course for the 21st Century: NLM's Long Range Plan 2006–2016* on September 19, 2006. The report includes the following chapters:

- Executive Summary
- Strategic Vision
- 1986-2006: Two Decades of Progress
- Plan for 2006–2016

Goal 1. Seamless, Uninterrupted Access to Expanding Collections of Biomedical Data, Medical Knowledge, and Health Information

Recommendation 1.1. Ensure adequate space and storage conditions for NLM's current and future collections to guarantee long term access to information and efficient service delivery.

Recommendation 1.2. Preserve NLM's collections in highly usable forms and contribute to comprehensive strategies for preservation of biomedical information in the U.S. and worldwide.

Recommendation 1.3. Structure NLM's electronic information services to promote scientific discovery and rapid retrieval of the "right" information by people and computer systems.

Recommendation 1.4. Evaluate interactive publications as a possible means to enhance learning, comprehension, and sharing of research results.

Recommendation 1.5. Ensure continuous access to health information and effective use of libraries and librarians when disasters occur.

Recommendation 1.6. Establish a Disaster Information Management Research Center at NLM to make a strong commitment to disaster remediation and to provide a platform for demonstrating how libraries and

librarians can be part of the solution to this national problem.

Goal 2. Trusted Information Services that Promote Health Literacy and the Reduction of Health Disparities Worldwide

Recommendation 2.1. Advance new outreach programs by NLM and NN/LM for underserved populations at home and abroad; work to reduce health disparities experienced by minority populations; share and actively promote lessons learned.

Recommendation 2.2. Work selectively in developing countries that represent special outreach opportunities, such as improving access to electronic information resources, enhancing local journal publications of high quality, and developing a trained librarian and IT workforce.

Recommendation 2.3. Promote knowledge of the Library's services through exhibits and other public programs.

Recommendation 2.4. Test and evaluate digital infrastructure improvements (e.g., PDAs, intelligent agents, network techniques) to enable ubiquitous health information access in homes, schools, public libraries, and work places.

Recommendation 2.5. Support research on the application of cognitive and cultural models to facilitate information transfer and trust building and develop new methodologies to evaluate impact on patient care and health outcomes.

Goal 3. Integrated Biomedical, Clinical, and Public Health Information Systems that Promote Scientific Discovery and Speed the Translation of Research into Practice

Recommendation 3.1. Develop linked databases for discovering relationships between clinical data, genetic information, and environmental factors.

Recommendation 3.2. Promote development of Next Generation electronic health records to facilitate patient-centric care, clinical research, and public health.

Recommendation 3.3. Promote development and use of advanced electronic representations of biomedical knowledge in conjunction with electronic health records.

Goal 4. A Strong and Diverse Workforce for Biomedical Informatics Research, Systems Development, and Innovative Service Delivery

Recommendation 4.1. Develop an expanded and diverse workforce through enhanced visibility of biomedical informatics and library science for K–12 and college students.

Recommendation 4.2. Support training programs that prepare librarians to meet emerging needs for specialized information services.

Recommendation 4.3. Continue support for formal, multidisciplinary education in biomedical informatics

to increase the supply of informatics researchers who can work at the intersections of molecular science, clinical research, health care, public health, and disaster management.

Charting the Course has been formally published in both print and Web (HTML) layouts. Print copies are available from the NLM Office of Communications and Public Liaison.

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

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. An NLM-wide Coordinating Committee on Outreach, Consumer Health and Health Disparities (OCHD) plans, develops, and coordinates NLM outreach and consumer health activities.

One activity, the Physician Information Prescription Project ("Information Rx"), reported in previous years, has been expanded to include the American Medical Association, American Osteopathic Association, hospital librarian members of the Medical Library Association, and disease-focused organizations such as the Fisher Center for Alzheimer's Research. An article reporting an evaluation of the Information Prescription project was published in Information Services & Use 26(2006) 1–10. Another outreach activity, reported in past years, is the Consumer Health Diabetes Project. Data from a controlled field experiment with patients enrolled in a diabetes program in Washington, D.C. are now being evaluated.

Web Evaluation

The Internet and World Wide Web 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 Web evaluation priorities of the OCHD include both quantitative and qualitative metrics of Web usage and measures of customer perception and use of NLM Web sites. During FY2006, 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: access to a syndicated telephone survey of the US public's online and offline health information seeking behavior; analysis of NLM Web site log data; and access to Internet audience measurement estimates based on Web usage by user panels organized by private sector companies.

Also during FY2006, OHIPD continued to work with other units of NIH to complete a trans-NIH online user survey project based on the American Customer Satisfaction Index (ACSI), with significant initial and supplemental funding support from the NIH/OD Office of Evaluation. The project extended the ACSI online user survey methodology to about 60 NIH Web sites at about 28 different ICs/OD units. The project contributed to: 1) strengthening participating IC/OD Web evaluation capability; 2) sharing of Web evaluation learning and experience on a trans-NIH basis; 3) aggregating ACSI results and learning on a trans-NIH basis; and 4) sponsoring several NIH-wide meetings and a final workshop that highlighted the contributions and challenges of the ACSI from the NIH perspective. The project was managed by a multi-institute ACSI Survey Leadership Team. The primary ACSI contractor was ForeseeResults Inc. via an arrangement brokered by NLM. The primary evaluation contractor was Westat Inc.

Tribal Connections

NLM has continued to focus 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, and a final project evaluation was published. Also, Phase 3, which involved more intensive community-based outreach and training at select Phase 1 and 2 sites, is complete. A Phase 3 evaluation report is available from the Pacific Northwest Regional Medical Library, University of Washington, Seattle.

NLM has funded 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). Phase 4 was completed in FY2005, and an evaluation report is available from the Midcontinental RML. NLM has also funded a Phase 5 that included outreach to Four Corners public libraries serving American Indians in that region, and the convening of an

NN/LM meeting to consolidate lessons learned from Native American outreach as of July 2006.

Other Native American Outreach

Also, in 2006 OHIPD again participated in the NIH American Indian Pow-Wow Initiative. This included exhibiting at seven pow-wows mostly in the Mid-Atlantic area. An estimated 4,500 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. In other parts of the country, in FY2006 OHIPD supported several projects in the Dakotas, Hawaii, and Alaska. These projects resulted largely from the Native American Listening Circles conducted in FY2003–2004:

North Dakota—Cankdeska Cikana Community College (via the Greater Midwest RML), Spirit Lake Nation, Ft. Totten, ND, continuing project to develop a health-related educational program at the Community College, and improvements at the tribal library;

North Dakota—MHA Systems Inc., a tribal enterprise of the MHA Nation, economic development outreach project to provide outreach assistance to a tribal information technology company that would ultimately result in jobs creation on the reservation (in this case, the Ft. Berthold Indian Reservation); the project is intended to improve the competitive capabilities of MHA Systems Inc., and also to refine, test, and strengthen the company's core scanning services.

Hawaii—Papa Ola Lokahi (via the Pacific Southwest RML), two Native Hawaiian Community Health Education Projects:

- Community of Miloli'i, Hawaii (The Big Island)—increased the knowledge of community members about health information and health resources by providing computer hardware and software to the community's library, training for the librarian and other community members, and by increasing multi-media resources at the Miloli'i Community Library; and supporting community-based initiatives founded on Hawaiian concepts of health (involving a balance between body, mind and spirit).
- Waimanalo Health Center, Oahu (Windward side)—increased the knowledge of community members about health information resources in order to better understand their own health conditions or health conditions of family members and to enable more effective self-management and more informed communication with health service providers; to be achieved by providing training and access to Web-based sources of health and medical information.

Outreach to Hispanics

The Lower Rio Grande Valley Hispanic Outreach Project was 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 was the beginning of an intensified NLM effort to meet the health information needs of the Hispanic population in Texas and elsewhere. In April 2006, NLM convened a workshop on Hispanic outreach strategies, with a group of academic, community, research, and private sector communications and outreach specialists, most of whom were Hispanic. The workshop developed a number of suggestions for NLM consideration in strengthening outreach to the Hispanic population.

International Programs

The focus of the office of International Programs is on outreach to researchers, physicians, and librarians in developing countries with an additional more recent emphasis on health workers and end users. This office continues to develop pilot programs, dissemination strategies, and training opportunities as well as evaluation, presentation, and publication of results. In addition, NLM has a Web site, Resources for International Librarians, Health Professionals and Researchers in Developing Countries at http://www.nlm.nih.gov/psd/ref/international.html .

MIMCom and Beyond

Since 1997, NLM/NIH participation in the Multilateral Initiative on Malaria has provided the core of NLM's work in the field of international outreach through the Africa-based Multilateral Initiative on Malaria Communication Network (MIMCom). MIMCom's goal was to identify the information needs of malaria researchers in Africa and enhance their access to the Internet and to medical information. NLM played a leadership role in getting these sites up and running and moving them to a place where IT has become a line item on the sites' grant proposals and annual budgets. In addition, NLM has trained African technical experts to become key players at their sites and, in two instances, on the continent. This activity supports the NLM objective of capacity strengthening, so that NLM is not always playing a central role.

In FY2006, there was much activity involving Uganda. The point of departure here began with the Ugandan medical and healthcare professionals themselves and what they are doing to further their medical and public health agendas. NLM has assisted researchers, physicians, and librarians in responding to needs such as the following:

- The Dean of the Faculty of Medicine, Makerere University in Kampala, through a new case-based curriculum, seeks to give medical students practical, positive experiences in and knowledge about the needs in rural areas.
- Third-year medical students at Makerere University are trying to effect behavior change in villages in the rural areas.
- The Serials Librarian at Albert Cook Library at the Faculty of Medicine, Makerere University, is trying to give access to her library's holdings catalog to libraries in Africa and around the world.
- The Editor of African Health Sciences, a new medical journal already indexed in MEDLINE, is upgrading its electronic management and publishing processes.
- The Vice-Chancellor at Kabale University in southwest Uganda is trying to establish practical and useful electronic links with the community where the university is based.
- A physician at Arua Hospital in northwestern Uganda is trying to get a reading of slides from a pathologist at Mulago Hospital in Kampala, 325 miles of rough road away.
- Researchers affiliated with the Ugandan Academy of Science would like to map the incidence of malaria from historical medical records at Mulago Hospital to climate change data from international databases.

These are widely varying needs. Will electronic communication tools prove useful? Will MedlinePlus for Africa interactive tutorials serve the Dean in his mission and the students in their quest? What form of telemedicine will be most convenient for the pathologist? Will the Serials Librarian be able to make use of the free NLM LinkOut feature? What sort of connectivity will best serve the goal of the Vice Chancellor? Will researchers be successful in harnessing the ability of information technology to carry out research that in the past would have been difficult if not impossible?

MedlinePlus Tutorials for Africa

This project is another effort by NLM to reach the consumer/end user, no matter where that user is located. MedlinePlus tutorials for Africa focus on tropical disease issues in developing country contexts. The first two tutorials are on malaria and diarrhea and were developed with the Faculty of Medicine at Makerere University in Uganda. In coordination with the Dean of the Faculty of Medicine, NLM worked with African doctors, artists, and medical students to create two original tutorials as well as guides for their use in the field. The tutorials were field tested as part of the medical students' curriculum and were translated into local languages of Luganda and Rukiga. The project leaders reported that the students enjoyed using the tools and were especially pleased on seeing the community's positive response.

4

MIMCom News and Web Sites

Every week, MIMCom Malaria News Update reaches more than 1600 malariologists around the world providing a broad spectrum of first class information. It offers an effective medium to the entire MIM community to communicate messages among a large group of professionals. The newsletter aims to be the most complete electronic malaria information resource, covering announcements, contributions from subscribers, scientific publications, reports, events, jobs, grants, training- and research opportunities, and news. From surveys, it is clear that this publication has become an invaluable resource to malaria researchers and managers in Africa and the rest of the world. In addition, NLM hosts a Web site for MIMCom at www.nlm.nih.gov/mimcom.

National Workshop in Medical Librarianship, Addis Ababa, Ethiopia

NLM organized this 3-day workshop, the first of its kind, on April 4–6, at the Ethiopian Civil Service College/Global Distance Learning Center in Addis Ababa. Librarians from five Ethiopian universities attended: Black Lion/Addis Ababa University, Debub, Mekelle, Jimma, and Gondor universities. The workshop was designed for librarians, working in health institutions or medical schools in Ethiopia, who are not yet trained in searching medical databases.

The workshop was co-led by an NLM reference librarian and the medical librarian at the University of Zambia (who is a former NLM Associate). The training was deemed a success by participating librarians. They learned how to search MEDLINE, MedlinePlus, and other NLM databases; how to find free full-text medical journal articles on the Web; how to use MeSH (Medical Subject Headings), NLM Gateway, and the NLM Catalog; and they participated in a two-way videoconference with NLM librarians. There was a follow-up workshop six months later via videoconference.

African Medical Journal Editors Partnership Program

This Partnership Program focuses on journals associated with MIM sites in Mali, Ghana, Uganda, and Malawi. The program comprises editors of these journals, editors of JAMA, BMJ, Lancet, EHS, and AJPH, and the Council of Scientific Editors. NLM contributed to technical capacity building by providing site visits by experienced IT experts from Africa and helping to purchase equipment, including computers, printers, scanners and software. Staff from each African journal visited the offices of its partner journal for one to two weeks. African editors reported these site visits to be extremely useful for observing the editorial and publishing practices of another journal.

With the support of the Partnership Project, African journal editors organized a series of training workshops for editors, authors, reviewers, researchers, and journalists. The workshops provided hands-on experience and lectures emphasizing international standards for writing and a systematic approach for reviewers. International trainers helped facilitate some of these workshops, and an element of training the trainers was incorporated into many of them. Workshops were well attended and feedback has been positive from both participants and facilitators. Some of the editors have already noticed improvements in the quality of their contributors' work.

Visitors

In FY 2006, the Office of Communications and Public Liaison and the History of Medicine Division's Exhibition Program arranged for 440 tours—108 regular daily (1:30 pm) tours and 332 specially arranged tours and programs. There were 8800 visitors in all. They came from the following 36 countries:

Argentina, Australia, the Bahamas, Brazil, Burma, Chile, China, Colombia, Croatia, Cuba, Denmark, England, Ethiopia, France, Germany, Guatemala, Guyana, Hungary, India, Japan, Mexico, Nigeria, Norway, Peru, the Philippines, Russia, Scotland, Singapore, South Africa, South Korea, Switzerland, Taiwan, Turkey, Ukraine, the United States and Vietnam.

International MEDLARS Centers

Continuing bilateral agreements between the Library and 18 public institutions in foreign countries allow them to serve as International MEDLARS Centers. As such, they assist health professionals in accessing MEDLINE and other NLM databases, offer search training, provide document delivery, and perform other functions as biomedical information resource centers. A list of the 18 centers is at http://www.nlm.nih.gov/pubs/factsheets/intlmedlars.html.

LIBRARY OPERATIONS

Becky J. Lyon
Deputy Associate Director

NLM's Library Operations (LO) Division is responsible for ensuring access to the published record of the biomedical sciences and the health professions. LO acquires, organizes, and preserves NLM's comprehensive archival 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 national backup document delivery, reference service, and research assistance; helps people to make effective use of NLM products and services; and coordinates the National Network of Libraries of Medicine to equalize access to health information across the United States. These basic services support NLM's outreach to health professionals, patients, families and the general public, as well as focused programs in AIDS, molecular biology, health services research, public health, toxicology, and environmental health.

Library Operations also develops and mounts historical exhibitions; carries out an active research program in the history of medicine and public health; collaborates with other NLM program areas to develop, enhance, and publicize NLM products and services; conducts research related to current operations; directs and supports training and recruitment programs for health sciences librarians; and manages the development and dissemination of national health data terminology standards. LO staff members participate actively in efforts to improve the quality of work life at NLM, including the work of the NLM Diversity Council.

The multidisciplinary LO staff includes librarians, technical information specialists, subject experts, health professionals, historians, museum professionals, and technical and administrative support personnel. 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 Information Center for Health Services Research and Health Care Technology (NICHSR); and a small administrative staff. The activities of all these components receive essential support from a wide range of contractors.

Most LO activities are critically dependent on automated systems developed and maintained by NLM's Office of Computer and Communications Systems (OCCS), National Center for Biotechnology Information (NCBI), or Lister Hill National Center for Biomedical Communications (LHC). LO staff work closely with these program areas on the design, development, and testing of new systems and system features.

Program Planning and Management

LO sets priorities based on the goals and objectives in the NLM Long Range Plan, and the closely related NLM Strategic Plan to Reduce Racial and Ethnic Disparities. In FY2006, LO contributed to the development of a new Long Range Plan for 2006–2016 under the auspices of the NLM Board of Regents. The plan was approved by the Board in September 2006 and will be implemented over the next 10 years. The plan includes a focus on developing a comprehensive program to ensure perpetual access to digital information beyond that which is being preserved in PubMed Central; development of effective systems and mechanisms for uninterrupted communications and access to vital information during and in the aftermath of disaster events; and continued development of information services that promote increased health literacy and decreased health disparities. LO will have a pivotal role in addressing all of these priorities.

In FY2006, LO continued to review and revise policies, procedures, services, and organizational lines to reflect shifting workloads; to use electronic information to enhance basic operations and services; and to work with other NLM program areas to ensure permanent access to electronic information. An LO-wide group which includes members from OCCS was created to develop functional requirements for an NLM Digital Repository as an initial step in moving forward with a plan for permanent access to digital information. Work also continued on the Indexing 2015 initiative, an NLM-wide research and development effort to improve indexing performance and productivity which is being led by LO. In the Technical Services Division (TSD) both the Cataloging and Serial Records Sections were reorganized to improve workflows and reflect changes brought about by ever increasing amounts of material published in electronic form. A Systems Organization and Activities Review Team (SOAR) in TSD recommended consolidating systems responsibilities and staff from the Sections to a Systems Office in the Office of the Chief.

Although many LO efforts are devoted to dealing with electronic information and supporting NLM's high priority outreach initiatives, LO must also devote substantial resources and attention to the care and handling of physical library materials and the space and environment for staff, patrons, and physical and electronic collections. As funds have not yet been appropriated for a new NLM building that will have increased space for its growing collections, some collections are already out of space and remaining space will be completely filled by 2010. In FY2006 LO with input from the Office of Administration developed plans for expanding existing space. The plan involves strengthening one of the floors in the collection area and installing compact shelving in phases.

Implementation of the plan will accommodate collection growth until approximately 2027.

New 5-year contracts for eight Regional Medical Libraries in the National Network of Libraries of Medicine for 2006–2011 were signed at the end of April 2006. The program continues to emphasize outreach to the public, network members, and health professionals with a focus on minority and underserved populations. The contracts also seek to build and improve collaborations with community-based organizations as an effective means of reaching these populations. Details about the new contracts are found elsewhere in this report.

In FY2006, LO's Administrative Office continued to assist managers, supervisors and staff with a wide range of administrative requirements, including transition to a new performance management system, which entailed creating new performance plans for all employees in the middle of the performance year. LO continued to encourage staff to take advantage of flexiplace work arrangements as appropriate. Nearly 100 LO employees work at home at least one day per week.

Collection Development and Management

NLM's comprehensive collection of biomedical literature is the foundation 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 relevant literature in all languages and formats; organizing and maintaining the collection to facilitate current use; and preserving it for subsequent generations. At the end of FY2006, the NLM collection contained 2,520,347 volumes and 6,665,946 other physical items, including manuscripts, microforms, pictures, audiovisuals, and electronic media.

Selection

Selectors worked on a variety of projects to enhance the breadth and depth of the NLM collection. The projects resulted in selection of African titles on HIV/AIDS, doctoral theses on Native healing and history of medicine, history of medicine titles listed on the Osler Library of the History of Medicine Web site, New Zealand Ministry of Health electronic resources, and new Slavic, Chinese and Korean works. TSD began a joint project with HMD to review a large gift of older Cyrillic medical publications donated to NLM from the Countway Library of Medicine at Harvard University. Many titles are in Old Russian and need special attention in transliteration.

Acquisitions

The Technical Services Division received and processed 156,388 contemporary physical items (books, serial issues, audiovisuals, and electronic media) which is slightly below last year's total. Electronic publishing has not yet had a significant impact on the number of physical items that NLM acquires. Net totals of 8,233 volumes and 612,727 other items (including nonprint media and manuscripts and pictures acquired by the History of Medicine Division (HMD) were added to the NLM collection. A major project to deaccession volumes from the Z collections resulted in 24,690 volumes being withdrawn which accounts for the significantly lower number of net volumes added.

LO uses subscription agents and book vendors to acquire current literature published around the world. In FY2006, TSD established an approval plan for Chinese monographs with the Beijing based firm China Publishing Industry Trading Corporation. MARC-formatted bibliographic records for titles supplied by this company will also be provided as part of the agreement.

HMD acquired a wide variety of important printed books, manuscripts and modern archives, images, and historical films during FY2006. Among the books were Francisco Martinez de Cartrillo, Coloquio Breve y Copedioso…(Valladolid, 1557), an early dental work about the anatomy of teeth, their development, extraction, diseases and the importance of dental hygiene; an unusual Cuban work on homeopathy: Honorato Bernard de Chateausalins, El Vademecum de los Hacendados Cubanos, o Guia Practica…(Havana, 1854); Johann Remmelin's Pinax Microcosmographicus (Amsterdam, 1667), the second edition of an anatomical atlas which used multi-layered flaps to show the organs of the human body; Aristotle's Master Piece Completed (New York, 1788), a pseudonymous work and one of the most popular manuals of pregnancy and childbirth ever published; William Harvey, De Motu Cordis (Padue, 1689), an Italian edition of Harvey's greatest work; and John C. Gunn, Gunn's Domestic Medicine (Pumpkintown, Tennessee, 1689), an early edition of one of this country's most literate and knowledgeable home-health guides.

Archives and modern manuscript collections acquired included the institutional records of the Hastings Center for Bioethics; the papers of Janette Sherman (an occupational health physician and environmental activist), Evarts Loomis (physician and proponent of holistic health), G. Octo Barnett (a physician and medical informatician at Harvard Medical School and Massachusetts General Hospital), Alan Peters (electron microscopist and neurologist), Theodore Puck (winner of the 1958 Lasker Award for his work on mammalian cell culture), and Carol Romano (nursing informatician).

New prints and photographs acquisitions included 4,300 pieces of medical ephemera donated by William Helfand and a 1980 watercolor of the NLM Building 38 donated by former NLM Director Martin Cummings. New videos and films included eight films on Telford Work's research on tropical arboviruses donated by Martine Work, 21 films from the Microcirculatory Society dating from the 1940s, and videotapes produced by the National Institute of Mental Health.

Preservation and Collection Management

LO carries out a wide range of activities to preserve NLM's archival collection and make it easily accessible for current use. These activities include: binding, copying deteriorating materials onto more permanent media, conservation of rare and unique items, book repair, maintenance of appropriate environmental and storage conditions, and disaster prevention and response.

In FY2006, the decision was made to end the microfilm preservation program to focus future efforts on digitization as the preservation format of choice. This 20-year program (1986–2006) resulted in the microfilming of 105,000 volumes (39 million pages) of brittle serials and monographs. Priority was given to serials indexed in *Index Medicus* and *Index Catalogue* and monographs at risk of text loss. After the development of extensive specifications, a purchase order to conduct a pilot digital preservation project was awarded to OCLC Preservation Services of Bethlehem, PA. A working group to develop priorities for digital formatting for preservation and access, led by PSD, submitted its report and recommendations to the Office of the Associate Director in September 2006.

A multi-year inventory of the NLM serials collection began in FY2006. Contractor CBase Solutions Inc. inventoried 8,142 *Index Medicus*/MEDLINE titles (351,747 volumes or unbound issues) and created or modified 7,901 missing item records. The Phase I inventory of all IM/MEDLINE titles should be completed by spring 2007.

A shift of serials published 1990–1994 from the B-1 to the B-3 level was completed. This involved moving 111,000 volumes and removing 47,307 duplicate unbound journal issues from the collection.

In FY2006, LO bound 19,317 volumes, microfilmed 1,509 volumes, repaired 1,814 items in the onsite repair and conservation laboratory, made 637 preservation copies of films and audiovisuals, conserved 77 prints and photographs and 170 other rare or unique items. A total of 756,295 items were shelved or re-shelved.

Permanent Access to Electronic Information

NLM's approach to addressing the unique challenges of preserving electronic information is to use its own electronic products and services as test-beds and to work with other national libraries, the Government Printing Office, 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 collaborates with other NLM program areas on activities related to the preservation of digital information.

PubMed Central (PMC), a digital archive of medical and life sciences journal literature developed by the National Center for Biotechnology Information, is NLM's vehicle for ensuring permanent access to electronic journals and digitized backfills. LO assists NCBI in soliciting participation of additional journals, particularly in the fields of clinical medicine, health policy, health services research, and public health.

LO's Public Services Division continued to work closely with NCBI to scan and add digitized backfiles of journals depositing newly published articles in the archive. PSD prepares back issues for scanning, ships them to the scanning contractor, and manages the human review portion of the quality control of scanned images, accompanying OCR data, and XML-tagged citations for articles that pre-date current MEDLINE/PubMed coverage. Since bindings are cut to make scanning more efficient, NLM does not use volumes from its archival collection in this effort, but solicits copies from publishers and other libraries. In FY2006 12 new titles were added including: *American Journal of Public Health*, *Annals of Surgery*, *Biochemical Journal*, *British Journal of General Practice*, *Environmental Health Perspectives*, *Journal of Neurology*, *Psychiatry*, *Skull Base*, *Skull Base Surgery*, and *Journal of Physiology*.

By the end of FY2006, more than 700,000 articles were in the PMC database. ScanTrac, a database to track journal titles as they are processed for PMC was launched in FY2006. The system, designed by the PMC teams in LO and NCBI, and programmed by OCCS staff, tracks the processing of newly accepted titles from signing of agreements to delivery of XML and donated source material. It is also used to track the quality assurance process and to pay invoices. Currently the system manages data on 331 titles.

A Digital Repository Working Group (DRG) was appointed comprised of members from PSD, HMD, TSD, and OCCS. It is charged with developing functional requirements for an NLM Digital Repository for NLM collection materials and to identify policy and management issues related to the creation, design and management of the repository.

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 biomedical literature and other types of information; develops, supports, or licenses for U.S. use vocabularies designed for use in electronic health records and clinical decision support systems; and works with the Office of Computer and Communications Systems and the Lister Hill Center to produce the Unified Medical Language System (UMLS) Metathesaurus, a large vocabulary database that includes many vocabularies, including MeSH and several others developed or supported by NLM. The Metathesaurus is a multi-purpose knowledge source licensed by NLM and many other organizations in production systems and informatics research. It 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 and promote use of standard clinical vocabularies in patient records and administrative transactions governed by the Health Insurance Portability and Accountability Act of 1996 (HIPAA). In this capacity, LO staff members serve on the Department of Health and Human Services Data Council, provide staff support to the National Committee on Vital and Health Statistics (NCVHS) Standards and Security Subcommittee, participate in the Consolidated Health Informatics e-government initiative, and participate in the Public Health Data Standards Consortium.

The 2006 edition of MeSH contains 24,357 main headings, 83 subheadings or qualifiers, and more than 163,000 supplementary records for chemicals and other substances. For the 2006 edition, the MeSH Section added 494 new descriptors, replaced 99 descriptors with more up-to-date terminology, deleted 22 descriptors, and added 1661 entry terms or "see" references. Areas receiving the most new vocabulary included stem cells, pituitary cells, urogenital diseases, hereditary diseases, viruses, viral vaccines, DNA damage, nutrition phenomena and processes, nanostructures, interleukin receptors, and keratins. Another project was undertaken to create a correspondence between MeSH terms and terms used in the latest edition of Goodman & Gilman's *Pharmacological Basis of Therapeutics*. The review and addition of Pharmacologic Actions (PAs) to commonly used ingredients in medications was begun. Production of 2007 MeSH descriptors ended on July 1, and work began on 2008 MeSH.

A document proposing revisions to the MeSH qualifiers was developed, posted to the Web with a request for comments from searchers and other users, and also sent to key NLM partners. A presentation was made about the proposed changes at the annual meeting of the Medical Library Association in May and a lively discussion followed. Many comments were received from librarians and from NLM staff. A final proposal will be developed following consideration of the comments. Any changes to qualifiers in MeSH resulting from this proposal will not occur until the 2008 MeSH.

Eight translations of MeSH were received and incorporated into the UMLS. Several of these translations are using the MeSH translation maintenance system, and some other translators have begun using it.

Clinical Vocabularies

The MeSH Section and its contractors also produce RxNorm, a clinical drug vocabulary that provides standardized names for use in prescribing. It is released through the UMLS. RxNorm was designated as a U.S. government-wide target clinical vocabulary standard by the HHS Secretary as one of a suite of standards for use in U.S. federal government systems for the electronic exchange of clinical health information. It represents the information that is typically known when a drug is prescribed, rather than the specific product and packaging details that are available at the time a medication is purchased or administered, and provides a mechanism for connecting information from different commercial drug information services. In FY2006, LO and OCCS prepared and released monthly updates to the RxNorm clinical drug vocabulary. The current coverage is over 17,000 clinical drugs currently active, most of them prescription drugs used in the United States.

Through LO's National Information Center on Health Services Research and Health Care Technology (NICHSR), NLM supports the continued development and free distribution of LOINC (Logical Observation Identifiers, Names, and Codes) by the Regenstrief Institute. LOINC is a clinical terminology important for laboratory test orders and results. In 2003, LOINC was designated by the Secretary of Health and Human Services as one of a suite of standards for use in U.S. federal government systems for the electronic exchange of clinical health information. In 2005 the Secretary also proposed adoption of LOINC as a HIPAA standard for some segments of the Claims Attachment transactions. In FY2006, the Regenstrief Institute began a collaboration with the HHS Office of the Secretary, Columbia University, and the Centers for Medicare and Medicaid for a special project to provide standard identifiers for elements of patient assessment instruments used in long-term care facilities.

NLM continues to pay the College of American Pathologists (CAP) annual license fees for the U.S.-wide distribution of SNOMED CT (Systematized Nomenclature of Medicine–Clinical Terms) through the UMLS Metathesaurus. SNOMED CT is a comprehensive set of clinical terminology. In 2004 SNOMED CT was designated by the HHS Secretary as one of a suite of standards for use in U.S. federal government systems for the electronic exchange of clinical health information. In 2005 the CAP and the UK National Health Service drafted a proposal to establish an international standards development organization to oversee the ongoing ownership, maintenance, and distribution of SNOMED CT. The proposal has been well received, and CAP intends to transfer the intellectual property rights for SNOMED CT to the new "International Health Terminology Standards Development Organisation" in 2007. Countries expressing their intent to participate in the new organization include Australia, Canada, Denmark, Lithuania, The Netherlands, New Zealand, Sweden, the United Kingdom, and the United States. NLM, in consultation with the HHS Office of the National Coordinator for Health Information Technology (ONCHIT), represents the U.S. federal government in these negotiations.

HHS has set a goal for the nationwide implementation of an interoperable health information technology infrastructure to improve the quality and efficiency of health care. Achieving this goal will require that key clinical data elements are captured or recorded in detailed, standardized form (using standard vocabularies, codes, and formats) as close to their original sources (patients, health care providers, laboratories, diagnostic

devices, etc.) as possible. If these standardized clinical data can also be used to generate HIPAA-compliant billing transactions automatically, this will provide another incentive for adoption of clinical data standards. For automated generation of bills from clinical data to become a reality, robust mappings from standard clinical terminologies to the HIPAA code sets must be created. HHS has given NLM the responsibility for funding, coordinating, and/or performing official mappings between standard clinical terminologies and HIPAA code sets. Several mappings are in various stages of development and technical validation. In September 2006 NLM released a draft set of "LOINC to CPT Mappings" for public review and comment, the first step toward a set of official mappings.

UMLS Metathesaurus

The MeSH Section and its contractors are responsible for content editing of the UMLS Metathesaurus, using systems developed by the Lister Hill Center. In FY2006, Library Operations assumed additional responsibilities for the production of the UMLS products. Transfer of the responsibility for production of the Metathesaurus from LHC to LO/OCCS was completed at the end of September. The Medlars Management Section (BSD) assumed a greater role in Quality Assurance and Documentation, and MeSH continued its supervision and training of Metathesaurus editing. An additional position was added to the MeSH staff, the incumbent assuming responsibility for monitoring vocabulary updates, the Metathesaurus production schedule, vocabulary licenses, and other agreements. Working with OCCS, a Metathesaurus production coordination group began meeting regularly to coordinate the production efforts between the two divisions. Regular review of inversions and insertions of updated and new vocabularies to the Metathesaurus was maintained and enhanced. Part of the inversion/insertion process has involved working with OCCS as they learn of the methodologies and issues arising as the vocabularies change over time.

Bibliographic Control

LO produces authoritative indexing and cataloging records for journal articles, books, serial titles, films, pictures, manuscripts, and electronic resources, 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 improve 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 by NLM to document what is available in the Library's collection or on the Web and to provide cataloging and name authority records that minimize the cataloging effort required in other health sciences libraries. Cataloging is performed by TSD's Cataloging Section, staff in HMD, and contractors. The Cataloging Section is responsible for the NLM Classification, coordinates the development and maintenance of the standard NLM Metadata schema for Web documents, and also performs name authority control for selected NLM Web services.

In FY2006, the Cataloging Section cataloged 21,662 contemporary books, serial titles, nonprint items, and cataloging-in-publication galleys, a slight increase from the previous year. The Section accomplished a major reorganization and adjusted workflows accordingly; completed a time and cost study for cataloging of monographic materials which successfully identified ways to decrease the cost without compromising the quality or usefulness of cataloging records; began processing and cataloging NIH Videocasts as part of the videocast archiving project; and continued to review and comment on *Resource Description and Access*, the proposed update to the current *Anglo-American Cataloging Rules*. NLM also co-chaired with the Library of Congress a working group of the Program for Cooperative Cataloging (PCC) charged with recommending appropriate encoding levels and authentication codes to be used in records for serials and for integrating resources, with the aim of providing clear and simple coding for PCC records. The report was submitted to the PCC Operating Committee in September. Implementation of the recommendations would result in a 20% decrease in the time required to create these records.

The Cataloging Section continued to work on producing a PDF version of the massive *NLM Classification*, which will be published in 3 parts in early FY2007. A PDF version would be of special interest to libraries in developing countries. The FY2006 online version, published in April, contained major updates to the bacteria and protein sections of the schedules, as well as the addition of many new 2006 MeSH terms to the index. Other improvements to cataloging included changes to the MeSH subject strings to streamline the subject analysis process and the addition of vernacular as well as Romanized bibliographic data for most in-house cataloging of Chinese, Japanese, Korean, Cyrillic, Arabic and Hebrew material.

Significant progress was made in providing cataloging records for NLM's historical and special collections. HMD cataloged 3,952 early monographs, an increase of 32% over the previous year. Also cataloged were 374 linear feet of manuscripts, 6,342 pictures, and 1,864 audiovisuals. HMD unveiled its *Rare Books and*

Early Printed Manuscripts Cataloging Manual for Early Printed Monographs which replaced a 20-year-old manual. It provides an outline of cataloging workflows, policies, and procedures, and will serve as a reference manual for catalogers.

Indexing

LO indexes 5,020 biomedical journals for the MEDLINE/PubMed database to assist users in identifying articles on specific biomedical topics. The indexing workload increases steadily due to the selection of additional journals to be indexed, increases in the number of articles published in biomedical journals, and the addition of elements to the MEDLINE record. In FY2006 Gene Expression Omnibus (GEO) databank accession numbers, International Standard Randomized Controlled Trial Numbers (ISRCTN), Wellcome Trust grant numbers, NIH grant numbers from the NIH Manuscript Submission System (NIHMSS), and RefSeq Databank accession numbers were added to the MEDLINE record for improved retrieval and links. A combination of Index Section staff, contractors, and cooperating U.S. and international institutions indexed 623,000 articles in FY2006, a 3% increase from the previous year. Previously indexed citations were updated to reflect 97 retractions, 7,175 corrections, and 31,685 comments found in subsequently published notices or articles.

In FY2006, indexers created 46,918 annotated links between newly indexed MEDLINE citations for articles describing gene function in selected organisms and corresponding gene records in the NCBI Entrez Gene database.

LO continues to work with other NLM program areas to identify, test, and implement ways to reduce or eliminate tasks now performed by human indexers. A successful new Web-based Indexer Training Tool has enabled training of new indexers as the need arises instead of at infrequently scheduled classroom training sessions. The time experienced indexers spend on training has also been reduced. One-on-one mentoring of trainees is done rather than classroom teaching. The Index Section has now installed dual monitor workstations for all in-house indexers and more than half of contract indexers. Dual monitors allow indexers to have simultaneous full-screen views of the online indexing system which already includes multiple windows for the MeSH vocabulary, PubMed, etc., and the text of the article being indexed, and it facilitates indexing from online journals. At the end of FY2006, 314 journals were indexed from an online version, including online-only journals and those with a print version. In addition to freeing the print version for immediate use onsite and for fulfilling interlibrary loan requests, indexing from an online version is part of LO's plan to maintain critical services to users in the event of a disaster.

Indexers perform their work after the initial data entry of citations and abstracts has been accomplished. Over the past ten years, great strides have been made in improving the efficiency of data entry. By the end of FY2006, more than 84% of all citation data entry consisted of XML-submitted data from publishers. The remaining citations were created by scanning and optical character recognition (OCR). A total of 60,000 more citations were received from publishers compared to the previous year for a grand total of 547,018 XML citations.

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 FY2006, LSTRC reviewed 434 journals and rated 78 of them highly enough for NLM to begin indexing them.

NLM continues to work with the Fogarty International Center and the editors of a number of prestigious Western medical and public health journals to assist editors in sub-Saharan Africa in improving the quality of their journals (See chapter on Health Information Programs Development.) NLM's role includes improving communications support so that the African editors can communicate with editors in other countries connected to the worldwide scientific journal community.

Information Products

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

Databases

LO managed the creation, quality assurance, and maintenance of the content of MEDLINE/PubMed, NLM's database of electronic citations; the NLM catalog, which is available to the public in two different databases; 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 several in the fields of health services research, public health, and history of medicine. These databases are richly interlinked with each other and with other important NLM resources, including PubMed Central, other Entrez databases, ClinicalTrials.gov, Genetics Home Reference, as well as Specialized Information Services' toxicological, environmental health, and AIDS information services. LO also participates in the testing and release of enhancements to the NLM Gateway.

Use of MEDLINE/PubMed, which now includes over 16 million citations, increased to 896 million searches in FY2006. LO extended the coverage of PubMed to include over 59,000 citations derived from the PubMed Central Back Issue Scanning Project, tested enhancements to the Gateway's searching of TOXLINE and Hazardous Substances Data Bank (HSDB) and the addition of six new

toxicological collections, and continued a project to map keywords in OLDMEDLINE records to MeSH for improved retrieval. BSD staff also assisted NCBI with the design, development, and testing of many enhancements to PubMed.

Use of MedlinePlus and MedlinePlus en español continued to increase dramatically. Nearly 100 million unique visitors viewed a total of 820 million pages. The number of page views grew by 28% and the number of unique visitors increased by almost 24%. More than 83,000 people subscribe to the weekly announcements of new additions to MedlinePlus content. MedlinePlus and MedlinePlus en español continue to receive high ratings from customers in the American Customer Satisfaction Index (ACSI), ranking in first and second place, respectively, among all government news/information sites. MedlinePlus was one of two U.S. winners of the 2005 World Summit Award. The award is part of the program of the World Summit on the Information Society, a United Nations effort organized by the International Telecommunication Union, the United Nations Industrial Development Organization, the United Nations Information and Communication Technologies Task Force, and UNESCO.

The Public Services Division (PSD) and OCCS continued to expand and improve the content and features of the English and Spanish sites. The site now features 741 health topics in English and 699 in Spanish. Among new features, a body map interface was added, allowing users to navigate to topics by clicking on one of 14 interactive body maps. Also a Director's PodCast, "What's New on MedlinePlus" was launched in May. Content expansion included the addition of Natural Standard, an evidence-based, peer-reviewed collection of information on alternative treatments, and 28 additional OR-Live surgical videos, including four in Spanish. Go Local was expanded with the release of 13 new sites, bringing the total to 19 sites in 17 states, with partial coverage in the states of Colorado, Ohio, and Texas. Go Local now covers 25% of the U.S. population.

Under the direction of NICHSR, NLM continues to expand and enhance its databases for health services researchers and public health professionals. In FY2006, NICHSR worked with NCBI to add the Centers for Disease Control and Prevention publication *Health US 2004* to the Entrez Bookshelf. Other additions to HSTAT (Health Services and Technology Assessment Text) on the Bookshelf included several evidence reports, produced by the Agency for Healthcare Research and Quality, as well as documents from the Substance Abuse and Mental Health Services Administration. 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 in Progress) to incorporate work funded by additional foundations and states. Organizations contributing data for the first time in FY2006 included the Illinois Department of Public Health, Michael Reese Health Trust, and Atlantic

Philanthropies. Two new HSRProj user guides were published: *Explore a Vital Link to Health Services Research* and *Cutting Edge Evidence for Policymakers*. The HSRR (Health Services Research Resources) database also continued to expand to cover additional datasets, survey, other research instruments, and software packages used with datasets. A new search interface and search engine were made available in February.

DailyMed, a Web site which presents the FDA-approved packaging information (labels) for drugs, was launched in FY2006. By the end of FY2006, there were approximately 1500 labels available through the DailyMed, either for viewing online or for download. The DailyMed site also provides links to several other sources of drug information, including information available through MedlinePlus, trial information from ClinicalTrials.gov, and literature searches through PubMed.

In FY2006, NLM made Endeavor's Voyager with Unicode release available on LocatorPlus. The Voyager with Unicode release allows users to view and search the non-Roman characters present in over 4,000 LocatorPlus records. Prior to this change, catalog records contained only transliterated title, author, publisher and other information. Now information is available for searching and display in the original language of the publication.

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 that vary depending on the database involved. The commercial companies, International MEDLARS Centers, universities and other organizations that obtain NLM data use them in many different database and software products for a very wide range of purposes.

Demand for MEDLINE/PubMed data in XML format continues to increase. At the end of FY2006, there were 406 licensees of MEDLINE data, a 19% increase from the previous year. The majority use the data for research and data-mining. To enable information sharing between licensees using the data for research purposes, the Bibliographic Services Division (BSD) developed a Web-based input screen to collect project information from those willing to provide it and posted the results on a new freely accessible Web site. Another Web site, the MEDLINE/PubMed Baseline Repository (MBR) was developed through BSD and LHC collaboration. The MBR contains resources derived from or pertaining to the MEDLINE/PubMed baseline files which are produced after the records have undergone annual maintenance.

At the end of FY2006 there were 3,532 UMLS licensees, an increase of 20% over the previous year. The transition from LHC to LO for quality assurance of the UMLS continued to make good progress; the OCCS/LO

UMLS Team assumed responsibility for Metathesaurus editing and UMLS release production on September 29.

Web and Print Publications

NLM's databases and Web sites are its primary publication media. In FY2006, NLM's main Web site use showed 9 million unique visitors using 61 million pages, an 11% increase in visitors and a 7% increase in pages viewed over FY2005. Publications available on the main Web site include recurring newsletters and bulletins, fact sheets, technical reports, and documentation for NLM databases. BSD's Medlars Management Section edits the NLM Technical Bulletin, which provides timely, detailed information about changes and additions to NLM's databases and related policies, primarily for librarians and other information professionals. Published since 1969, the *Technical Bulletin* also serves as the historical record of the evolution of NLM's online systems and databases. In FY2006, two new features were introduced—a Skill Kit and an RSS feed. Skill Kit articles provide search hints, review system features, and cover data and indexing issues for NLM databases with the goal of expanding a user's search skills and knowledge. The RSS feed allows users to receive updates about new articles via RSS readers.

In FY2006, LO staff was involved in the development of two new publications designed for patients, families and the public. On May 19, the Director's Comments PodCast was launched to bring users of MedlinePlus current health news. On September 20, the *NIH MedlinePlus Magazine* was rolled out at a press event on Capitol Hill attended by members of Congress and guest celebrity Mary Tyler Moore, who was featured on the cover. The magazine is initially being distributed free to 40,000 physician offices.

Direct User Services

In addition to producing heavily used electronic resources, LO is responsible for document delivery, reference, and customer service for both onsite users and remote users. LO provides document delivery to remote U.S. users via the National Network of Libraries of Medicine (NN/LM).

Document Delivery

LO retrieves documents requested by onsite patrons from NLM's closed stacks and also provides interlibrary loan as a backup to document delivery services available from other libraries and information suppliers. In FY2006, PSD's Collection Access Section processed 573,828 requests for contemporary documents. HMD handled 10,438 requests for rare books, manuscripts, pictures, and historical audiovisuals.

Although the number of onsite users registering to use the collection continued to decline by another 23% from last year, use of NLM's collection by users in the Main Reading Room was up 5% from the previous year's

total to 245,167. Users of the HMD Reading Room requested 9,273 items from the historical and special collections, a decrease of 22% from last year. Paid printing at Main Reading Room workstations, primarily from onsite use of electronic journals, remained virtually unchanged, totaling 496,337 prints and copies.

The Collection Access Section (CAS) received 328,661 interlibrary loan requests, a 3.7% decline from FY2005 with a fill rate of 81%, the highest every achieved. The number of requests processed in 12 hours rose to 96%, with 98% processed within one day of receipt. NLM now delivers 96% of interlibrary loan requests electronically. CAS and Lister Hill Center staff collaborated on a project to include LHC's DocMorph software in the ILL delivery workflow so that PDF documents can be converted to TIFFs, allowing articles to be sent from electronic journals by all delivery methods. Previously Ariel, fax, and mail were unavailable when delivering a PDF.

A total of 3,219 libraries use DOCLINE, NLM's interlibrary loan request and routing system. DOCLINE users entered 2,311,516 requests in FY2006, a 7% decline from last year; 92% of requests were filled. DOCLINE requests are routed to libraries automatically based on holdings data. At the end of FY2006, the holdings database contained 1,462,547 holdings statements for 54,909 serial titles held by 3,023 libraries. In FY2006, individuals submitted 544,273 document requests to DOCLINE libraries via the Loansome Doc feature in MEDLINE/PubMed and the NLM Gateway, a 33% decline from the previous year. Most of the decline is due to use of new document delivery software in lieu of Loansome Doc by the NIH Library. Document request traffic continues to decline in all Regions of the NN/LM due to expanded availability of electronic full text journals.

In an effort to assist libraries affected by Hurricanes Katrina and Rita, in October 2005, NLM began to provide free interlibrary loan services for as long as necessary to any library affected by these disasters. A total of 2,450 articles were provided to 65 eligible libraries before the program concluded at the end of September 2006.

NCBI and the staff at the Regional Medical Libraries continued to promote the use of PubMed's LinkOut for Libraries and Outside Tool as a means for libraries to customize PubMed to display their electronic and print holdings to their primary clientele. The number of libraries participating in LinkOut increased by 15% in FY2006 to 1,558; there are 296 libraries participating in the Outside Tool option.

NLM and the Regional Medical Libraries continued to encourage network libraries to use the Electronic Funds Transfer System (EFTS), operated for the NN/LM by the University of Connecticut, as a mechanism to reduce administrative costs associated with ILL billing. During FY2006, EFTS participation increased 5.5% to 1,128 participants. Participants receive either a single net consolidated bill or a net consolidated payment each month.

Reference and Customer Services

LO provides reference and research assistance to onsite and remote users as a backup to services available from other health sciences and public libraries. LO also has primary responsibility for responding to inquiries about NLM's products and services and how to use them effectively. LO's Reference and Web Services Section responds to initial inquiries and also handles the majority of questions requiring second-level attention. Staff from throughout LO and NLM assist with second-level service when their special expertise is required.

A total of 91,784 inquiries were received in FY2006, down 4% from FY2005. The number of onsite inquires declined 32% to 15,202, reflecting the decline in the number of onsite users. The number of remote inquiries increased 4% to 76,582, with the overwhelming majority arriving via e-mail.

PSD also continues to refine the knowledge base of "Cosmo," a virtual customer service representative built with software designed to answer frequently asked questions about NLM's programs, products, and services. In FY2006, Cosmo responded to 5,747 questions that were within his job description, up 34% from last year, and answered 88% of them correctly.

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 health information professionals, historians, researchers, educators, health professionals, and the general public. LO coordinates the National Network of Libraries of Medicine which attempts to equalize access to health information services and information technology throughout the United States; serves as secretariat for the Partners in Information Access for the Public Health Workforce; participates in NLM-wide efforts to develop and evaluate outreach programs for underserved minorities and the general public; produces major exhibitions and other special programs in the history of medicine; and conducts training programs for health sciences librarians and other information professionals. LO staff members give numerous presentations, demonstrations, and classes at professional meetings and publish articles that 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 for U.S. health professionals, researchers, and the general public irrespective of their geographic location. With more than 5800 full and affiliate members, the Network is the core component of NLM's outreach program and its efforts to reduce health disparities and to improve health information literacy. Full members are libraries with health sciences collections, primarily in hospitals and academic medical centers. Affiliate members include some smaller 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 (NNO) oversees network programs that are administered by eight Regional Medical Libraries (RMLs) under contract to NLM. In FY2006, the process of recompeting the five-year NN/LM contracts culminated in an award in each of the eight regions for the basic contract (see Appendix 1 for a list of RMLs). A new institution, New York University, was awarded the contract for service in the Middle Atlantic Region; all other contracts were awarded to incumbent organizations. By August, transition from the New York Academy of Medicine, which had provided distinguished service as the RML in that region since 1970, was completed. NLM also funded the Electronic Funds Transfer System, and awarded a contract for the National Training Center and Clearinghouse (NTCC) to the New York Academy of Medicine. The activities of the NTCC are described in the last section of this chapter. Subcontracts for the Outreach Evaluation Resource Center (OERC) and the Web-services Technology Operations Center (Web-STOC) were awarded to the University of Washington. The OERC provides training and consulting services throughout the NN/LM and assists in designing methods for measuring overall network programs and individual outreach projects.

RMLs and other network members conduct many special projects to reach under-served health professionals and to improve the public's access to high quality health information. Virtually all of these projects involve partnerships between health sciences libraries and other organizations, including public libraries, public health departments, professional associations, schools, churches, and other community-based groups. In FY2006, the NN/LM issued 125 subcontracts for outreach projects which target many rural and inner city communities and special populations in 42 states and the District of Columbia.

The Regional Medical Libraries assisted in identifying and initially funding several Go Local projects in 2006. Go Local launched 13 new sites during 2006— Utah, Wyoming, Maryland, Delaware, Texas-East/Northeast, New Mexico, Texas Gulf Coast, Ohio-Southeast, South Carolina, Nebraska, Vermont, Nevada, and Arizona. Each site received an award of $25,000 from its RML. The "MedlinePlus Go Local Guidelines" were updated and renamed, "Project Proposal Guidelines." An example of a Go Local proposal was also made available on the MedlinePlus Go Local Resources Web site.

The OERC produced step-by-step planning and evaluation methods within a series of three booklets. Each booklet contains a case study and worksheets to assist with outreach planning. The series supplements *Measuring the Difference: Guide to Planning and Evaluating Health Information Outreach* and supports NN/LM evaluation workshops.

- Booklet 1: *Getting Started With Community-Based Outreach*
- Booklet 2: *Including Evaluation in Outreach Project Planning*
- Booklet 3: *Collecting and Analyzing Evaluation Data*

The OERC facilitated the process of identifying Emergency Preparedness as the national collaboration for the 2006–2011 Contracts. The OERC will coordinate the development of the plan to guide the implementation of this initiative.

With the assistance of other NN/LM members, the RMLs do 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 in the Washington, DC area, and some distant meetings focused on health services research, public health, and history of medicine. In FY2006, NLM and NN/LM services were exhibited at 67 national and 192 regional, state, and local conferences across the U.S. These exhibits highlight all NLM services relevant to attendees.

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 12-member public–private agency collaboration initiated by NLM, the Centers for Disease Control and Prevention, and the NN/LM in 1997 to help the public health workforce make effective use of electronic information sources and to equip health sciences librarians to provide better service to the public health community. The NICHSR coordinates the Partners for NLM; staff members from the National Network Office, SIS, and the Office of the Associate Director for Library Operations serve on the Steering Committee, as do representatives from several RMLs.

The Partners Web site (PHPartners.org), managed by NLM with assistance from the New England RML at the University of Massachusetts, provides unified access to public health information resources produced by all members of the Partnership, as well as other reputable organizations. In FY2006, the Web site was expanded with more than 400 new links and two new categories: Fellowships and Upcoming Meetings. One of the most popular resources on the site is the Healthy People 2010 Information Access Project (HP2010 IAP). For every focus area of Healthy People 2010, the IAP resource includes four or more objective-specific evidence-based PubMed search strategies and links to MedlinePlus topics.

In addition, in FY2006 NICHSR awarded two new purchase orders under the Partnership to the: *Public Health Foundation for Public Health Performance Improvement Resource Access*; Association of State and Territorial Health Officials (ASTHO) for *Public Health Workforce Enumeration Strategy for Environmental Health and One Additional Public Health Field*; and an interagency agreement with the Health Resources and Services Administration for *Community Health Status Indicator (CHSI) Profiles*. The first purchase order will link public health workers with quality improvement information recast in a public health framework, the second will provide a specific enumeration strategy for two distinct fields within public health and make a more general enumeration possible. The interagency agreement will result in updated Web-based CHSI profiles for use by the public health workforce.

Special NLM Outreach Initiatives

LO participates in the Library's Committee on Outreach, Consumer Health, and Health Disparities and in many NLM-wide outreach efforts designed to expand outreach and services to the public as well as to address racial and ethnic disparities. BSD, NNO and the Office of the Associate Director, LO, participated in developing the agenda and made arrangements to host the National Commission on Libraries and Information Science (NCLIS) Libraries and Health Information Forum on May 3, 2006. The forum brought together librarians and health professionals interested in consumer health information as well as the ten finalists for the 2006 Health Information Awards for Libraries. The awards recognize library programs in each state that address one or more of the following: dietary choices; exercise; smoking cessation; alcohol and/or drug abuse prevention or cessation; immunizations and health screenings, and improved health literacy. Most of the awardees are members of the NN/LM; many have received NLM funding for their projects. Dr. J. Edward Hill, President of the American Medical Association was the Keynote Speaker; Ms. Eugenie Prime, former chair of NLM's Board of Regents was the forum moderator. Representatives of finalist libraries participated in panel discussions on "Effective Programs," "Health Literacy," and "Partnerships and Outreach." REACH 2010: Charleston and Georgetown Diabetes Coalition's Library Partnership (South Carolina) received the Grand Prize Award for their efforts to eliminate disparities for more than 12,000 African Americans diagnosed with diabetes by improving self-management and care. This partnership has received support from the NN/LM.

In FY2006, BSD continued to work with other NLM components and the NN/LM to provide ongoing support to the Information Rx Project by maintaining the InformationRx.org Web site for ordering materials and working on a redesign of the InformationRx Toolkit for librarians. Information Rx provides physicians with materials to write prescriptions for information from MedlinePlus for their patients. The Office of the Associate Director, LO, the NNO, and BSD continued to work with the American Library Association (ALA) and Public Library Association (PLA) to improve public library awareness of

MedlinePlus and MedlinePlus en español. The Office of the Associate Director also serves on an ALA/Walgreen's Advisory Committee for the "Be Well Informed @ Your Library" program which funded ten public library systems to conduct seminars on health education issues. BSD staff continued on several fronts to promote the onsite exhibition by planning programs for the Smithsonian Associates, professional groups and other special visitors. LO staff members continue to be involved in NLM's partnership with the SCIMATECH Academy at Wilson High School in the District of Columbia. In FY2006, LO provided summer employment and training opportunities for several students.

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. The current exhibition, *Visible Proofs: Forensic Views of the Body*, opened on February 16, 2006. Exploring developments in scientific methods that translate views of bodies and body parts into visible proofs, it tells stories of the people, sciences, and technologies that make visible the cause and manner of a death. Visitors to the exhibition observe, analyze, and decipher different forensic views of the body and examine important historical and contemporary cases and forensic techniques through the use of objects, graphics, and multimedia presentations. They also encounter experts whose contributions and discoveries have changed the field of forensic medicine. The exhibition opened with a ribbon cutting and special program featuring several of the persons portrayed in the exhibition.

Previous NLM exhibitions live on through heavily used Web sites, printed catalogs, DVDs, and touring traveling versions. The traveling version of *Frankenstein: Penetrating the Secrets of Nature*, for example, concluded its multi-year tour of public, academic, and health sciences libraries across the United States, under the auspices of the American Library Association. Traveling to 82 sites, it was seen by over one million visitors. Host libraries organized 1,309 public programs attended by more than 85,000 people.

The very successful exhibition, *Changing the Face of Medicine: Celebrating America's Women Physicians*, closed in November after a 25-month run. A traveling version, funded by the NIH Office of Research on Women's Health and NLM, began touring libraries in the U.S. through collaboration with the American Library Association.

Besides the major exhibitions mounted in the rotunda, HMD installed four mini-exhibits in cases near the entrance to the HMD Reading Room: *The Horse, a Mirror of Man: Parallels in Early Human and Horse Medicine; Animals as Cold Warriors; "I Swear by Apollo": Greek Medicine from the Gods to Galen;* and *From Monsters to*

Modern Medical Miracles: Selected Moments in the History of Conjoined Twins from Medieval to Modern Times.

NLM added four segments to the Profiles in Science Web site during FY 2006, bringing the total to 23. The new sites focus on Harold Varmus (a scientist who shared the 1989 Nobel Prize in Physiology or Medicine for discovery of "the cellular origin of retroviral oncogenes" and later became NIH Director); Michael Heidelberger (an immunologist who received two Lasker Awards); Virginia Apgar (best known for the Apgar Score, a simple, rapid method for assessing newborn viability); and Edward D. Freis (a cardiologist who demonstrated that medication could dramatically reduce disability and death from stroke, congestive heart failure, and other cardiovascular diseases).

Launched in September 1998, Profiles in Science promotes the use of the Internet for research and teaching in the history of biomedical science by making widely available archival collections of leaders in biomedical research and public health. Published and unpublished materials appear on the site, including books, journal volumes, pamphlets, diaries, letters, manuscripts, photographs, audiotapes, and video clips.

A symposium entitled *Global Health Histories* brought together scholars, scientists, administrators, and activists to examine global public health crises from an historical perspective. Convened in November 2005, the symposium was co-sponsored by the Institute of the History of Medicine at the Johns Hopkins University, the Fogarty International Center of NIH, and NLM's History of Medicine Division, in association with the Global Health Histories Initiative of the World Health Organization. As recent natural catastrophes and epidemics have shown, in a globalized world it is no longer possible to speak of public health crises as contained by local, regional, or even national boundaries. History provides a crucial tool to understand the response to disease on a global scale. The symposium was designed to initiate a series of conversations among historians, anthropologists, sociologists, policy makers, and practitioners in order to spark new understandings and collaborative relationships.

During May and June 2006, NLM presented a six-lecture series *Genomics in Perspective*, which explored this, complex and often confusing issue for an audience of scientists, physicians, policy makers, and the general public. Some welcome genomics as ushering in a golden age of new and more effective treatments, better diagnostic interventions, and more powerful means of biological investigation through bioinformatics, genetic analysis, measurement of gene expression, and determination of gene function. Others caution against over-optimism, and point to the importance of culture, society and history to an understanding of the complexity of interaction between biology, genes, and environment. Featuring a lecture by a historian or social scientist, a response by a physician or scientist, and an audience discussion period, the six sessions stimulated discussion of the social, historical, and cultural meanings and uses of genomics.

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. The Division also continued to prepare the recurring features, "Voices from the Past" and "Images of Health" for the *American Journal of Public Health*, which often features materials from the NLM collection.

Training and Recruitment of Health Sciences Librarians

LO develops online training programs to teach the use of MEDLINE/PubMed and other NLM databases to health sciences librarians and other information professionals; oversees the activities of the National Training Center and Clearinghouse (NTCC) at the New York Academy of Medicine; directs the NLM Associate Fellowship program for post-masters librarians; 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 American Library Association, the Association of Academic Health Sciences Libraries (AAHSL), 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 encourage specialist roles for health sciences librarians.

In FY2006, the Medlars Management Section (MMS) and the NTCC trained 829 students in 64 classes covering PubMed, the NLM Gateway/ClinicalTrials.gov, TOXNET, and the UMLS. Experimental use of the Breeze software for remote broadcasts of online training sessions was successful for delivering synchronous training in PubMed and Gateway/ClinicalTrials.gov classes. An average of about 22,000 unique users visited the Web-based PubMed Tutorial about 28,000 times each month. A new version of the PubMed Tutorial was implemented, along with ten new animated Viewlet QuickTour tutorials for targeted PubMed search features, bringing the total number of Web-based QuickTours to 23. MMS staff members were honored with an NIH Plain Language Award for the PubMed QuickTours. The average number of QuickTour page views was about 26,000 per month. A new Web instructional resource called The Basics of Medical Subject Headings (MeSH) was introduced to help searchers understand more about the use of MeSH for searching PubMed.

The UMLS for Librarians course and the UMLS Tutorial continue to represent some of the NLM training courses useful in preparing librarians for new and expanded roles. LO and the NTCC also assist NCBI in arranging network venues, scheduling, and publicizing the Introduction to Molecular Biology Information Resources class, which helps to prepare library-based bioinformatics specialists. NCBI also offers an advanced workshop for Bioinformatics Information Specialists at NLM. Both courses were developed and are taught by librarians who serve as bioinformatics specialists in universities and at NLM. NICHSR continues to add to its suite of courses on health services research, public health, and health policy.

The NLM Associate Fellowship program had 10 participants in FY2006: four 2nd year Associates at sites across the country and six 1st year Fellows, who completed their year at NLM in August 2006. Four of the latter also chose to participate in the optional 2nd year of the program at sites across the country: the University of North Carolina–Chapel Hill, George Washington University, Indiana University, and Johns Hopkins University. Seven new Fellows began a year at NLM in September, including one International Fellow from the medical school library at the University of Bamako in Mali. Efforts to recruit fellows from underrepresented groups have been successful in attracting diverse groups of fellows to the program, including African American, Asian American, and Native American representation in FY2006.

NLM works with several organizations on librarian recruitment and leadership development initiatives. Individuals from minority groups continue to be underrepresented in the library profession and a high percentage of current library leaders will retire within the next five to ten years. LO has provided support for scholarships for minority students available through the American Library Association, the Medical Library Association, and the Association for Research Libraries (ARL). LO also supports the NLM/AAHSL Leadership Development Program, which provides leadership training, mentorship, and site visits to the mentor's institution for an annual cohort of five mid-career health sciences librarians. AAHSL contracts with ARL for the leadership training portion of the program. Recruitment efforts have emphasized and been successful in attracting minority candidates. In addition to its continued support for this successful program for a second three-year period covering FY2006–FY2008, NLM also funded an evaluation of the program that will be completed in FY2007.

Table 1

Growth of Collections

Collection	Previous Total (9/30/05)	Added FY 2006	New Total (9/30/06)
Book Materials			
Monographs:			
Before 1500	591	0	591
1501-1600	5,976	6	5,982
1601-1700	10,245	22	10,267
1701-1800	24,684	52	24,736
1801-1870	41,582	103	41,685
Americana	2,341	0	2,341
1871-Present	758,150	15,436	773,586
Theses (historical)	288,091	0	288,091
Pamphlets	172,021	0	172,021
Bound serial volumes	1,307,834	12,233	1,320,067
Volumes withdrawn	(99,401))	(12,905)	(112,306)
Total volumes	2,512,114	14,947	2,527,061
Nonbook Materials			
Microforms:			
Reels of microfilm	145,794	2,400	148,194
Number of microfiche	452,709	3,053	455,762
Total microforms	598,503	5,453	603,956
Audiovisuals	77,344	2,685	80,029
Computer software	2,469	80	2,549
Pictures	63,255	5,739	68,994
Manuscripts	5,312,282	598,850	5,911,132*
Nonbook items added	6,053,853	612,807	6,666,660
Nonbook items withdrawn	0	(433)	(433)
Total nonbook items	6,053,853	612,374	6,665,227
Total book & nonbook	**8,565,967**	**627,321**	**9,193,288**

*Manuscripts equivalent to 3,378 linear feet

Table 2

Acquisition Statistics

Acquisitions	FY 2004	FY 2005	FY 2006
Serial titles received	20,769	20,989	20,815
Publications processed:			
Serial pieces	132,192	132,347	134,020
Other	24,323	24,659	22,368
Total	156,515	157,006	156,388
Obligations for:			
Publications	$6,942,747	$8,255,443	$8,683,250
(For rare books)	($300,831)	($324,398)	($337,386)

Table 3

Cataloging Statistics

	FY 2004	FY 2005	FY 2006
Completed Cataloging	21,238	21,238	21,662

Table 4

Bibliographic Services

Services	FY 2004	FY 2005	FY 2006
Citations published in MEDLINE	571,000	606,000	623,000
Journals indexed for MEDLINE	4,839	4,928	5,020
Total items archived in PubMed Central	343,998*	467,260*	823,148

*Amended figure

Table 5

Consumer Web Services

Services	FY 2004	FY 2005	FY 2006
NLM Web Home Page			
Page Views	48,335,875	56,332,002	61,008,251
Unique Visitors	7,934,966	7,996,621	8,129,906
MedlinePlus			
Page Views	498,702,940	662,421,882	820,041,002
Unique Visitors	51,724,958	74,290,591	95,539,948
ClinicalTrials.gov			
Page Views	33,651,851	61,303,796	105,131,193
Unique Visitors	3,190,813	3,499,091	5,242,218
Genetics Home Reference			
Page Views	6,298,501	13,633,547	20,181,044
Unique Visitors	365,231	841,999	1,434,722
Household Products Database			
Page Views	7,096,664	10,547,690	11,622,189
Unique Visitors	1,364,649	1,908,604	1,219,764
Tox Town			
Page Views	1,732,336	3,232,466	2,444,499
Unique Visitors	365,383	228,220	149,833
DailyMed			
Page Views	***	***	842,977
Unique Visitors	***	***	82,433

Table 6

Circulation Statistics

Activity	FY 2004	FY 2005	FY 2006
Requests Received	631,806	580,072	573,828
Interlibrary Loan	359,577	341,239	328,661
Onsite	272,229	238,833	245,167
Requests Filled:	510,751	475,623	467,143
Interlibrary Loan	281,543	273,870	265,562
Onsite	229,208	201,753	201,581

Table 7

Online Searches—PubMed and NLM Gateway

	FY 2004	FY 2005	FY 2006
Total online searches	678,000,000	754,000,000	896,000,000

Table 8

Reference and Customer Services

Activity	FY 2004	FY2005	FY 2006
Offsite requests	71,290	73,493	76,582
Onsite requests	36,649	22,298	15,202
Total	107,939	95,791	91,784

Table 9

Preservation Activities

Activity	FY 2004	FY 2005	FY 2006
Volumes bound	18,311	18,417	19,317
Volumes microfilmed	2,603	1,564	1,509
Volumes repaired onsite	1,652	2,095	1,814
Audiovisuals preserved	795	936	483
Historical volumes conserved	197	140	154

Table 10

History of Medicine Activities

Activity	FY 2004	FY 2005	FY 2006
Acquisitions:			
Books	498	1,070	1,401
Modern manuscripts	5,516,000*	1,107 (in ft)	783 (in ft)
Prints and photographs	1,591	11,252	9,945
Historical audiovisuals	757	176	1,864
Processing:			
Books cataloged	13,621	2,668	3,952
Modern manuscripts cataloged	740,250**	453 (in ft)	374 (in ft)
Pictures cataloged	2,758	2,974	6,342
Citations indexed	[now included in Table 4]		
Public Services:			
Reference questions answered	18,701	20,655	23,818
Onsite requests filled	8,618	12,738	14,140

*Equivalent to 3,152 linear feet
**Equivalent to 423 linear feet

SPECIALIZED INFORMATION SERVICES

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

The NLM Division of Specialized Information Services (SIS) creates information resources and services in toxicology, environmental health, chemistry, and HIV/AIDS. SIS also has an Office of Outreach and Special Populations that seeks to improve access to high quality, accurate health information by underserved and special populations.

The Toxicology and Environmental Health Information Program (TEHIP), known originally as the Toxicology Information Program, was established nearly 40 years ago within SIS. Over the years TEHIP has met the increasing need for toxicological and environmental health information by exploiting new computer and communication technologies to provide more rapid and effective access for a wider audience. SIS continues to move beyond the physical bounds of the Library, exploring ways to point and link users to credible sources of toxicological and environmental health information wherever these sources may reside. Resources include chemical and environmental health databases and Web-based collections. Development of HIV/AIDS information resources is also a focus of the Division and includes several collaborative efforts in information resource development and deployment, including a focus on the information needs of other special populations. The outreach program at SIS continuously evolves and reaches out to underserved communities through innovative information dissemination.

The SIS Web site provides a central point of access for the varied programs, activities, and services of the Division. Through this site (http://sis.nlm.nih.gov), users can freely access interactive retrieval services in toxicology and environmental health, HIV/AIDS information, and special population health information; find program descriptions and documentation; and be connected to outside related sources. Continuous refinements and additions to these Web-based systems are made to promote easier access to the wide range of information collected by SIS.

In FY2006 SIS continued to balance efforts to enhance and reengineer existing information resources with efforts to provide new services in emerging areas. SIS 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 2006 include the following:

Toxicology and Environmental Health Resources

The TOXNET (TOXicology Data NETwork) is a cluster of databases covering toxicology, hazardous chemicals, environmental health and related areas. These databases continue to be highly used resources, and in FY2006 customer surveys showed that 87% of responders would "return to this site" and "recommend it to others." In FY2006, enhancements to TOXNET were based on user feedback and routine upgrades of data and capabilities. Databases in TOXNET include:

- **LactMed** (Drugs and Lactation), a new database, provides information on drugs and other chemicals to which breastfeeding mothers may be exposed. LactMed includes information on reported levels of such substances in breast milk and infant blood, and on possible adverse effects in the nursing infant, with relevant links to other NLM databases.
- **HSDB** (Hazardous Substances Data Bank), a peer-reviewed database focusing on the toxicology of over 5,000 potentially hazardous chemicals.
- **IRIS** (Integrated Risk Information System), a database from the Environmental Protection Agency (EPA) containing carcinogenic and non-carcinogenic health risk information on over 500 chemicals.
- **ITER** (International Toxicity Estimates for Risk), a database containing data in support of human health risk assessments. ITER is compiled by Toxicology Excellence for Risk Assessment (TERA) and contains over 500 chemical records.
- **CCRIS** (Chemical Carcinogenesis Research Information System), a scientifically evaluated and fully referenced data bank, developed and maintained by the National Cancer Institute, with over 9,000 chemical records with carcinogenicity, mutagenicity, tumor promotion, and tumor inhibition test results.
- **GENE-TOX** (Genetic Toxicology), a toxicology database created by the EPA containing genetic toxicology test results on over 3,000 chemicals.
- **TOXLINE**, a bibliographic database providing comprehensive coverage of the biochemical, pharmacological, physiological, and toxicological effects of drugs and other chemicals from 1965 to the present. TOXLINE contains over three million citations, almost all with abstracts and/or index terms and CAS Registry Numbers.
- **DART/ETIC** (Development and Reproductive Toxicology/Environmental Teratology Information Center), a bibliographic database covering literature on reproductive and developmental toxicology.
- **Toxics Release Inventory** (TRI), a series of databases that describe the releases of toxic chemicals into the environment annually for the 1987–2004 reporting years.

- **ChemIDplus**, a database providing access to structure and nomenclature authority databases used for the identification of chemical substances cited in NLM databases. ChemIDplus contains over 380,000 chemical records, of which over 263,000 include chemical structures.
- **Household Products Database** provides information on the potential health effects of chemicals contained in more than 6,000 common household products.
- **Haz-Map**, an occupational toxicology database designed primarily for health and safety professionals, but also for consumers seeking information about the health effects of exposure to chemicals and biologicals at work. It links jobs and hazardous tasks with occupational diseases and their symptoms.
- **ALTBIB**, a bibliographic database on alternatives to the use of live vertebrates in biomedical research and testing.

WISER (**W**ireless **I**nformation **S**ystem for **E**mergency **R**esponders) is a tool developed for use by emergency responders during hazardous materials incidents, as well as during training sessions/exercises in preparation for such events. Based on user feedback, a Web-based version of WISER was developed this year, as an auxiliary resource to the downloadable stand-alone versions. Usage among first responders continued to grow with over 24,000 downloads of WISER onto PDAs (Palm and Pocket PC) and Windows-based desktop/laptops over the past 12 months. Total number of WISER downloads is over 44,000. Positive accounts from users about their applications of WISER continue to be received, including numerous accounts of its use during the Hurricane Katrina emergency response. New chemicals were added to WISER and design for inclusion of radiological substances began.

Tox Town was enhanced with new content (in English and Spanish) in the Tox Town "neighborhoods:" Tox Town, Tox City, Tox Farm, and a U.S.-Mexico Border scene. A search engine was also developed, which enables Tox Town to be searched from both within the application and by TOXNET and ToxSeek. To promote the use of Tox Town by educators, a teacher page was developed with sections on activities and discussion questions, interactive and illustrated resources, checklists and quizzes, career information and general resources for teachers.

TOXMAP, a Geographic Information System (GIS) system that uses maps of the U.S. to help users visually view data about chemicals released into the environment and easily connect to related environmental health information, underwent major enhancements in FY2006, including additional EPA Superfund sites and demographic layers (e.g., age, gender, income, cancer and disease mortality). Other enhancements included improved map projections, a

teacher resource page, and introduction of summary tables and graphs.

New **Enviro-Health Link** pages this year include *Dietary Supplements*, with links to many sources of relevant information, and *Pesticide Exposure*, with links to Web sites addressing acute and chronic exposure to pesticides.

ToxSeek is a meta-search tool that enables simultaneous searching of many authoritative information sources in environmental health and toxicology on the Web. ToxSeek provides integrated search results from the selected resources and displays related concepts for use in refining searches. ToxSeek was publicly released in FY2006 and has been well received by the user community.

ToxMystery, an interactive Web site for children between the ages of 7–10, was released at the end of FY2006. ToxMystery provides an animated, game-like interface that prompts children to find potential chemical hazards in a home, and then rewards them with "fun and interesting" sound effects when they successfully complete the task. Focus groups and feedback from the targeted user community have indicated that this innovative Web site provides kids with a fun and educational experience.

AIDS Information Services

NLM is the project manager for the multi-agency AIDSinfo service. This service provides access to federally approved HIV/AIDS treatment guidelines, AIDS-related clinical trials information (through Clinicaltrials.gov), and prevention and research information.

The American Customer Satisfaction Index (ACSI) continues to be used to evaluate AIDSinfo. The 2006 score for AIDSinfo was 84, which ranked it 2nd out of 30 governmental "Information/News Websites." The AIDSinfo score has improved by 4 points, which likely reflects improvements made in response to user feedback.

The NLM has continued its HIV/AIDS-related outreach efforts to community-based organizations, patient advocacy groups, faith-based organizations, departments of health, and libraries. This program supports the design of local programs intended to improve information access for AIDS patients, their caregivers, and the affected community. SIS outreach efforts emphasize provision of information or access in ways meaningful to the target community. Projects must involve one or more of the following information access categories: information retrieval, skills development, Internet access, resource development, and document access. In FY2006, in the 13th cycle of the program, NLM made 17 awards.

Evaluation Activities

In FY2006 several new and enhanced SIS Web products were professionally assessed via online surveys, focus groups, and online bulletin forums. Since 2003, the

following SIS Web products have been professionally evaluated: World Library, ToxSeek, LactMed, TOXMAP, TEHIP portal, WISER, Tox Town, Asian American Health, Arctic Health, American Indian Health, Household Products Database, TOXNET, and Haz-Map. Feedback has provided important input for enhancement, and for directions for new capabilities. The American Customer Satisfaction Index continues to be used to evaluate TOXNET and to guide future enhancements.

Outreach Initiatives

SIS outreach programs engage health professionals, public health workers, and the general public, with special focus on health issues that disproportionately impact minorities (e.g., environmental exposures and AIDS). Highlights from FY2006 include:

- United Negro College Fund Special Programs/NLM–HBCU Access Project awarded small grants to four HBCUs to develop projects to increase awareness and utilization of NLM resources both on campuses and in communities. This program is now in its fourth year, and evaluation reports from earlier grants provide evidence of successful implementations.
- Adopt-a-School program, in partnership with Woodrow Wilson Senior High School, Washington, D.C., encourages students to take an active interest in consumer health and promotes interest in science. Projects this year included online training about NLM databases, summer internships for students, donation of technological books and periodicals, tours of NLM, and guest lectures.
- Consumer Health Resource Information Service (CHRIS) Project is a faith-based pilot initiative designed by the Medical Education and Outreach group of the Oak Ridge Associated Universities, Oak Ridge, Tennessee. This project addresses minority health disparities through community level intervention and prevention measures in six predominantly African American churches in the inner city of Knoxville, Tennessee. NLM received a recognition citation on February 27, 2006, from Governor Phil Bredesen of Tennessee for the CHRIS Project. Success of the pilot program led to its adoption as a state-wide program in 2006. Surveys from the pilot indicate that 95% of the 170 church members who completed the survey felt the health information they received from the parish nurses resulted in positive changes in their health habits or lifestyle.
- The mission of the Environmental Health Information Outreach Program (EnHIOP) is to enhance capacities of minority-serving academic institutions to reduce health disparities through access, use, and delivery of environmental health information on campus and in the community. Two successful meetings were held in FY2006— "Rebuilding the HBCUs in the Gulf in the Aftermath of Hurricane Katrina" (Tuskegee University, Alabama); and "Forensic Science" (hosted at NLM). EnHIOP meetings included representation from 14 HBCUs, three tribal colleges, and three Hispanic-serving institutions.
- Improvements, updates, and expansions were made to SIS special population Web pages in FY2006. Average monthly hits included 104,000 (American Indian Health); 83,000 (Asian American Health); 40,000 (Arctic Health).
- The Sacred Root is an NLM-supported Native American Information Internship Program that provides opportunities for representatives from American Indian tribes, Native Alaskan villages, and the Native Hawaiian community to learn about NLM and the National Network of Libraries of Medicine, and to use that knowledge to improve access to health information and technology in their respective communities. Beginning in September 2006, NLM's Sacred Root Fellows come from the Navajo Tribe, where they work with the Health Promotion Program within the Tuba City Regional Health Care Corporation, Tuba City, Arizona.
- The Central American Network for Disaster and Health Information (CANDHI) is a group of health science libraries and information centers working together to enhance local health and disaster information management capacities with a goal of contributing to disaster preparedness in the region. CANDHI is a partnership between the NLM, the Pan American Health Organization, and the U.N. International Strategy for Disaster Reduction. CANDHI consists of centers in Honduras, Nicaragua, El Salvador, and Guatemala (with support from U.K. Department of International Development), and in Panama and Costa Rica (with financial support from the European Community Humanitarian Office or ECHO). The CANDHI libraries have acquired the knowledge, skills, and resources that promote delivery of reliable information, with more than 7,800 full-text documents now available online. During FY2006, a prototype Disaster Information Center toolkit was prepared that provides guidance for groups seeking to establish their own Disaster Information Centers. An independent assessment of CANDHI by a former head of PAHO's disaster program concluded that a majority of the participating centers not only achieved initial goals, but also improved the technology, networking, level of services, and visibility of their respective libraries.
- SIS is a partner in the Refugee Health Information Network (RHIN), a national collaborative partnership of several state refugee health offices,

NLM, and the Consumer Product Safety Commission. RHIN is committed to providing quality multilingual, multi-cultural health information resources for patients and those who provide care to resettled refugees. A Web site is being developed to improve access to medical information by state and local public health departments, refugee service providers, and refugees.

- National Medical Association (NMA) is a national professional and scientific organization representing the interests of more than 25,000 physicians of African descent and their patients. SIS continues to collaborate with the NMA to conduct online database training at the six NMA regional meetings held each year.

SIS exhibited at over 40 conferences in FY 2006. Several of these provided opportunities for presentations or workshops about NLM's information resources.

Research and Development Initiatives

To meet the mission of providing information on toxicology, environmental health, and targeted biomedical topics to the world, SIS continuously develops new ways to present the world of hazardous substances to a wider audience.

In an interagency collaboration, SIS and the DHHS office of Public Health Emergency Preparedness have developed a system for **Radiation Event Medical Management (REMM).** Intended for use by emergency physicians and related emergency health care providers, the system includes algorithm-based guidelines for evaluation and management of individuals exposed to radiation during accidental releases, use of radiological dispersion devices, and use of improvised nuclear devices. In FY2006, the REMM prototype was developed, reviewed by experts, improved based on user feedback, and prepared for

production. An official release of REMM is expected in the first quarter of FY2007.

The World Library of Toxicology, Chemical Safety, and Environmental Health is designed to provide a Web portal to global information resources in toxicology, chemical safety, environmental health, and allied disciplines. The World Library is being designed, developed, and maintained by SIS staff, and will provide a cyberhome for a project in which representatives from participating nations provide crucial input and feedback to assure credible and high-quality sources of information. The World Library has been populated with information resource sets from 32 countries and collaborations with many more in progress. With support from NIH's Fogarty International Center, this project is scheduled to release fully developed information resources in FY2007.

Another resource under development in FY2006 was the **Dietary Supplements Database**, a resource of comprehensive information on supplements used by U.S. consumers. Information on more than 1,000 dietary supplement brands will be available and searchable by brand name, active ingredient, target user, or manufacturer, with links to TOXNET and PubMed searches. SIS plans to release the database in second quarter FY2007.

The goal of the **Public Health Law Information Project (PHLIP)** is to create in the public domain a searchable database of public health law information that will be not only a guide for non-specialists (e.g., concerned citizens, attorneys, public health practitioners, academics, legislators), but also an excellent technical resource for those who are specialists in the field. In FY2006, a pilot project was established with the state of Delaware, the Widener University School of Law, the Delaware Academy of Medicine, and SIS to produce a searchable database containing statutes, regulations, and other information from Delaware that pertain to public health.

Finally, in FY2006, SIS led an NLM-wide collaborative initiative to produce a **Drug Information Portal** that will make it easier for consumers and health professionals to find drug information available from NLM and other federal governmental resources.

LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS

Donald W. King, M.D.
Acting Director

The Lister Hill National Center for Biomedical Communications (LHNCBC), established by a joint resolution of the United States Congress in 1968, is a research and development division of the NLM. 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 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 (Appendix 3) meets biannually to review the Center's research projects and 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. The Lister Hill Center is organized into five major components: Cognitive Science Branch, Communications Engineering Branch, Computer Science Branch, Audiovisual Program Development Branch (which currently includes the Office of the Public Health Historian), and the Office of High Performance Computing and Communications.

The Center's principal research activities and accomplishments are described in the remainder of the chapter.

Biomedical Imaging

The overall goal of this program area is to address fundamental questions that arise in the handling, organization, storage, access and transmission of very large electronic files in general and digitized biomedical images in particular. A special focus is research into these topics as applied to heterogeneous multimedia databases consisting of both images and text. Projects in this area have benefited from collaborators in several universities as well as at agencies such as the National Center for Health Statistics and the National Institute of Arthritis, Musculoskeletal and Skin Diseases. A great deal of effort in the past year has focused on a partnership with the National Cancer Institute in their research in cervical cancer caused by the Human Papillomavirus (HPV). Our biomedical imaging work may be broadly divided into Multimedia Database R&D and Content-Based Image Retrieval.

Multimedia Database R&D

Goals of this project are: (1) to research latest technological approaches for information retrieval and delivery for biomedical databases that include non-text data, with an emphasis on biomedical images; and (2) to develop prototype systems for the retrieval and delivery of such information for use by the research and, potentially, the clinical communities.

The Web-based Medical Information Retrieval System (WebMIRS) continues to provide access to images and text from nationwide surveys conducted by the National Center for Health Statistics. At the current time there are 444 users of WebMIRS in 54 countries. This Java application allows remote users to access data from 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.

The NHANES II database also contains vertebral boundary data collected by a board-certified radiologist for 550 of the 17,000 x-ray images. This data consists 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 such a 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.

The Digital Atlas of the Cervical and Lumbar Spine remains available for the public on a CD or on the CEB Web site either as a Java applet or a downloaded Java application. The Java application version allows the user to add his/her own grayscale and color images in a special

"My Images" section and to annotate and title those images for later use. The Atlas has capabilities to display color images, to add extensive text annotations, and to import/export sets of images and annotations as a package.

In addition, the FTP x-ray archive of 17,000 digitized spinal x-rays continues to be active, with 444 users worldwide. This archive allows access to the x-rays, available both in full 12-bit flat file format and also in TIFF 8-bit format which is easier for many researchers to use.

A suite of newer systems motivated by, but not restricted to, joint research with NCI, are at various stages of development:

- The Multimedia Database Tool, designed as the next generation WebMIRS system, provides a software framework for the incorporation of new text/image databases in a much more general way than the current WebMIRS, and new features for the database end user that extend current WebMIRS capabilities.
- The Boundary Marking Tool provides Web capability to manually mark boundaries on cervicography images, and to manage collected data with a MySQL database. It is in active use by NCI for multiple studies.
- The Virtual Microscope provides Web capability to view and collect information on histology images from expert observers.
- The Teaching Tool is a system for training medical personnel in cervix anatomy/pathology. It displays the uterine cervix images and quizzes an observer, and enables an NCI medical expert to tailor exams by specifying images and questions to use on an examination.

Content-Based Image Retrieval

The Content-Based Image Retrieval (CBIR) system provides capabilities for searching for spine vertebrae by shape and/or descriptive text, using a database of several thousand pre-segmented vertebral shapes and text data from the NHANES II database used by WebMIRS. The key characteristics of this system, developed in MATLAB and Java, are that it can operate in networked or standalone modes, uses XML for reporting, and allows the user to select either a more mature or an experimental version of the system.

A significant development is Relevance Feedback, a MATLAB experiment in utilizing feedback from an expert user for CBIR image retrieval. Work is under way to incorporate the capabilities of the CBIR3 and Feedback systems into SPIRS, our new, Web-based Spine Pathology and Image Retrieval System. In addition, development continues on the Pathology Validation and Collection (PathVa) tool, a Java-based system for segmentation review and editing. This tool will retrieve spine images that have been compressed with methods developed for NLM by Texas Tech, over the Internet, along with the segmented

boundary data. Three board-certified radiologists are collaborating with us to review several hundred images and record presence, type, and degree of severity of anterior osteophytes, disk space narrowing, subluxation, and spondylolisthesis. The recorded information is automatically transmitted to a database after validation of the segmented boundaries. Other work includes collaborations with several universities to develop or improve segmentation algorithms, segmentation systems or tools, and shape validation tools.

The Visible Human Project

The Visible Human Project image data sets are designed to serve as a common reference for the study of human anatomy, as a set of common public domain data for testing medical imaging algorithms, and as a test bed and model for the construction of image libraries that can be accessed through networks. The Visible Human data sets are available through a free license agreement with the NLM. They are distributed to licensees over the Internet at no cost; and on DAT tape for a duplication fee. The data sets are being applied to a wide range of educational, diagnostic, treatment planning, virtual reality, virtual surgeries, artistic, mathematical, legal and industrial uses by over 2300 licensees in 49 countries. The Visible Human Project has been featured in more than 850 newspaper articles, news and science magazines, and radio and television programs worldwide.

FY2006 saw the continued maintenance of two databases to record information about Visible Human Project use. The first, to log information about the 2300 license holders and record statements of their intended use of the images; and the second, to record information about the products the licensees are providing NLM in compliance with the Visible Human Dataset License Agreement.

During FY2006 an extensive search of the literature was completed, and a Visible Human Current Bibliographies in Medicine (CBM) was produced. This bibliography is an attempt to identify all publications in the scientific and technical literature that discuss the Visible Human Project and its derivative products. This includes citations to journal articles, books and book chapters, conference papers and meeting abstracts, audiovisuals, and literature discussing the use and applications of the VHP Insight Toolkit (ITK).

The Insight Toolkit, a research and development initiative under the Visible Human Project, is now in its sixth year with a recent official software release of ITK 2.9. 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. Support, development and maintenance of the software is managed by a community of

university and commercial groups, including OHPCC intramural research staff. The ITK continues to have an impact on the medical imaging research community. Researchers are testing, developing, and contributing to ITK in more than 38 countries, with 1200 active subscribers to the global mailing list for the project.

Across NIH, ITK is providing a foundation for new imaging investigations. The National Alliance of Medical Image Computing (NA-MIC), an NIH Roadmap National Center for Biomedical Computing (NCBC), has adopted ITK and its software engineering practices as part of its engineering infrastructure. NA-MIC is currently using medical imaging techniques to study the physiological sources of schizophrenia and other mental disorders. During FY2006, the NA-MIC organization held ten user-training workshops across the country and one in Lausanne, Switzerland. In addition, ITK software engineering practices and tools are having an impact outside the medical imaging community and are influencing some of the world's largest open-source software projects.

3D Informatics

OHPCC's 3D Informatics Program has expanded in-house research efforts around problems encountered in the world of three-dimensional and higher-dimensional, time-varying imaging. One of our most intense efforts is our project to create *PLAWARe* (Programmable Layered Architecture With Artistic Rendering), a software framework for artistic and non-photorealistic rendering of digital models. This entails the design of a layered, software architecture for implementing medical illustration techniques using computer graphics technologies. In FY2006, the project contributed to a technical exhibition, demonstrating some of its capabilities at the 2006 ACM SIGGRAPH Conference in Boston.

The 3D Informatics Group has continued work on image databases, including ongoing support for the National Online Volumetric Archive (NOVA), an archive of volume image data, as well as our continuing partnership with the NLM Specialized Information Systems Division and the U.S. Veterans Administration to study content-based retrieval methods for medical image databases. In the pharmaceutical identification project, we are assisting in the acquisition of imagery through digital macro-photography of the thousands of prescription pharmaceuticals dispensed routinely by the VA Centralized Mail-Order Pharmacies. Together we are creating a new, updated, visual database of all these products and developing techniques for automatically identifying any product in the inventory from a representative photograph. New OHPCC research has developed computer vision approaches for the automatic segmentation, measurement, and analysis of solid-dose medications.

In FY2006, the 3D Informatics group, along with representatives of the National Institute of Biomedical Imaging and Bioengineering and two directorates at the National Science Foundation, sponsored two workshops on Visualization Research Challenges. The group continues to publish and submit materials for scientific review at national and international venues and continues to organize and participate in regional, national, and international conferences as invited speakers, panelists, journal reviewers, and conference organizing committee members.

DocView: Document Imaging for the Biomedical End-user

This research area applies document image processing and digital imaging techniques to document delivery and management, thereby addressing NLM's mission of providing document delivery to end users and libraries. An additional focus is to contribute to the bulk migration of documents for purposes of digital preservation, also part of the NLM mission. The active projects in this area are DocView, DocMorph, MyMorph and MyDelivery.

DocView. This Windows-based client software was originally released in January 1998 and subsequently improved over several generations. It is widely used by libraries to deliver TIFF documents for interlibrary loan services. It currently has 17,997 users in 195 countries, an increase of 898 new users and two countries over last year. In September 2006 alone, there were 65 new users spread over 22 countries registering to use DocView. Although the use of DocView is expected to decrease, the changeover is likely to be gradual especially in foreign countries since their purchase of the new Ariel software may take longer.

MyDelivery. The MyDelivery project is seen as a successor to DocView. The goal of the project is to develop a new collaborative tool to improve the delivery and exchange of medical and health information, especially in very large files. MyDelivery provides users (researchers, administrators, librarians, physicians, patients, hospitals, and other health professionals) a fast, easy, and secure method to exchange medical information, regardless of the size of the electronic file in which it resides. The MyDelivery project seeks to overcome three significant obstacles: (1) transmission of large electronic files (e.g., document images, digitized photographs and x-rays, sonograms, CT and MRI scans, and digital video) over the Internet; (2) sending files reliably and securely; and (3) complying with requirements of the Health Insurance Portability and Accountability Act (HIPAA). To solve all three problems, the MyDelivery project focuses on the development of server-based software running on a cluster of Internet-based servers, and the development of client software for use by collaborators. MyDelivery allows two client computers to exchange large files through an intermediary server via a user interface similar to e-mail. In test conditions, the system permits the exchange of files ranging up to several gigabytes in size. Part of the development of MyDelivery has been to create a method of automatically recovering from communication failures due to reduced signal strength. This part of the project has been completed, and successfully tested over unreliable

networks. Additional work centers on using public domain software for security certificate generation and use.

<u>DocMorph and MyMorph</u>. The DocMorph system continued to serve both browser-based users (14,400 to date: 1900 more than last year) and MyMorph users (6500 users) this year. Most of the registered users are biomedical document delivery librarians. DocMorph allows the conversion of more than 50 different file formats to PDF, for instance, to enable multi-platform delivery of documents. Also, by combining OCR with speech synthesis, DocMorph enables the visually impaired to use library information. It has been used by librarians for the blind and physically handicapped to convert documents to synthetic speech recorded onto audiotapes for blind patrons. Most users continue to use it to convert files to PDF to enable multi-platform delivery of documents. DocMorph is available at http://docmorph.nlm.nih.gov/docmorph.

Document Image Analysis and Understanding

Research in Document Image Analysis and Understanding is directed toward developing production techniques in line with NLM's mission. The projects in this category are MARS and its various spin-offs.

Medical Article Records System

The Medical Article Records (MARS) production system has evolved through several generations of increasing capability. Its core engine consists of daemons (computer programs) based on heuristic rule-based algorithms that use geometric and contextual features derived from OCR output to automatically segment scanned pages of journal articles, assign logical labels to these zones, and to reformat zone contents to adhere to MEDLINE conventions. About a quarter of the total citations in MEDLINE now are created by MARS, the remaining coming in as XML-tagged data directly from publishers.

Changes continue to be made to the MARS production system to accommodate new requirements from indexers. Three MARS software modules (Edit, Reconcile, and Upload) and the validation library have been modified to automatically extract the clinical trial control and GEO (gene expression omnibus) databank numbers, and Wellcome Trust grant numbers. In addition, changes were made to the Reconcile and Upload modules to list author and corporate author names in the order they appear in the published article.

WebMARS

Efforts continue toward meeting goals of the Indexing 2015 Initiative through the continuing development of two systems relying on WebMARS to assist both operators and indexers. Initial versions of both systems, WebMARS Assisted Indexing (WAI) and Publisher Data Review (PDR) are currently under test.

PDR will provide operators data missing from the XML citations sent in directly by publishers (such as databank accession numbers, NIH grant numbers, funding sources, and PubMed IDs of commented articles), thereby reducing the burden on operators in creating citations for MEDLINE. In addition, incorrect data sent in by the publishers can be corrected by PDR. Correcting the publisher data is currently a labor-intensive process since the operators perform these functions manually by looking through an entire article to find these items, and then keying them in.

WAI will aid indexers in their search for terms in an article that correspond to biomedical terms in a predefined list. WAI will automatically search through the text and highlight these terms for the indexer to simply confirm and select, thereby reducing manual effort. An initial prototype was demonstrated to indexers who provided feedback for improvement. A pilot version of this system is currently being tested in the Indexing Section.

ACORN

This system is intended to extract bibliographic information from 60 volumes of the printed *Quarterly Cumulative Index Medicus* (QCIM) from 1927 to 1956 to populate the OLDMEDLINE database. The design of the system is rooted in research in document image analysis and pattern matching techniques. With the help of NLM's Preservation and Collection Management Section, the microfilm version of a particular volume (Vol. 59, Jan – June 1956) was scanned and the TIFF images subjected to OCR conversion. Currently, a module is being created to extract journal name abbreviations from the DCMS database to compare against the abbreviations in the microfilm images.

Text to Image Linking Engine

Text to Image Linking Engine (TILE) is designed to transparently link the print library of functional-physiological knowledge with the image library of structural-anatomic knowledge into a single, unified resource for health information, a long term NLM goal. An early prototype of the modular GUI interface to the system now called *Visual PubMed* (TILE-PubMed proxy server) was completed and demonstrated. This system allows a user to search PubMed and receive citations that are automatically augmented with anatomic images relevant to the article topic.

Research in TILE seeks the best alternatives for the functions needed to accomplish this linkage. These functions are: identifying biomedical terms in a document; identifying the relevant anatomical terms; identifying the images in the image database; and linking the identified terms to the images. Our main research focus is on the second function, the Term Mapper, which associates the biomedical terms in the document to appropriate anatomic concepts through the Metathesaurus concept relation table, and ultimately to images. Since this table typically yields

several relationships that can potentially map a biomedical term to multiple anatomical concepts, relevance ranking is then applied.

Medical Article Records Groundtruth

The Medical Article Records Groundtruth (MARG) database is available to the computer science and informatics communities for research in document image analysis and understanding techniques. 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 (http://marg.nlm.nih.gov/) is Rover, an analytic tool that may be used to compare the results of a researcher's program with the ground truth data. Rover has been enhanced to allow a visual comparison of researchers' algorithmic results with the ground truth data, as well as some statistical metrics. The MARG server has more than 9,600 unique IP visits from 96 countries.

Information Systems

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

Consumer Health Informatics Research

The Consumer Health Informatics research projects explore the needs, information seeking behavior, and cognitive strategies of health care consumers. The projects' principal goal is to apply medical informatics and information technologies to study ways to develop, organize, integrate, and deliver accessible health information to the members of the public at all levels of health literacy. These projects include the ClinicalTrials.gov and Genetics Home Reference Web sites and the Consumer Health Information Seeking research initiative.

ClinicalTrials.gov provides the public with comprehensive information about all types of clinical research studies, both interventional and observational. The site has over 34,000 protocol records sponsored by the U.S. Federal government, pharmaceutical industry, academic and international organizations in all 50 States and in over 130 countries. Some 47% of the trials listed are open to recruitment, and the remaining 53% are closed to recruitment or completed. ClinicalTrials.gov receives over 11 million page views per month and hosts approximately 29,000 visitors daily. Data are submitted by over 3,370 study sponsors through a Web-based Protocol Registration System, which allows providers to maintain and validate information about their trials.

ClinicalTrials.gov was actively involved in promoting the standards of transparency in clinical research through trial registration. These standards were communicated to a broad range of U.S. and international stakeholders via presentations and printed materials. As a result of increasing awareness of the importance of trial registration, over 12,000 new registrations were received over the last year. ClinicalTrials.gov also launched a comprehensive evaluation program, aimed at identifying and meeting user needs of various groups of ClinicalTrials.gov users. ClinicalTrials.gov continues to collaborate with other registries and professional organizations, working towards developing global standards of trial registration.

Genetics Home Reference (GHR) provides basic information about genetic conditions and the genes and chromosomes related to those conditions. Created for the general public, particularly individuals with genetic conditions and their families, the site currently includes summaries for more than 200 genetic conditions, more than 330 genes, and all the human chromosomes. On average, ten new summaries are added per month. In the past year, GHR's content was expanded to include information about disorders caused by mutations in mitochondrial DNA. This addition required development of new features such as a circular ideogram representing mitochondrial DNA. Companion tutorial materials were also developed for the *Help Me Understand Genetics* handbook. The new Handbook materials explain the role of mitochondrial DNA in cells and how changes in this DNA affect health. To support GHR's growing content, the site's search algorithm and display of search results were improved to help users find topics of interest. GHR's usage increased more than 60% in the past year, and the site is continually recognized as an important health resource.

New in FY2006 is a project that teaches first aid to Hurricane Katrina evacuees (conducted by Southern University and Louisiana State University), a weekly PodCast produced in conjunction with the NLM Office of the Director, Library Operations, and OCCS, and a follow-up study of the factors that make it difficult for consumers to understand a medical text.

The consumer health informatics team continues to publish at national and international venues. The team participates in national and international conferences and helped plan NIH's e-health national conference as well as the Surgeon General's National Meeting on Health Literacy in September. Additionally, members gave 12 invited lectures to universities and institutions around the nation and internationally.

The Consumer Health Information Seeking initiative focuses on understanding and improving access to online health information. One project explores the search and navigation behavior of consumers using health information systems. Another project investigates methods for developing readability assessment metrics to evaluate health-related text intended for consumers of varying health literacy. A third project examines different approaches for

using queries in one language (e.g., Spanish) to retrieve relevant documents in another language (e.g., English) to support access to health information for the Spanish-speaking community. A prototype system for providing basic information about clinical trials in Spanish is undergoing usability testing. Finally, the consumer health vocabularies project focuses on mapping words and phrases

Digital Library Research

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

The Profiles in Science digital library uses innovative digital technology to showcase digital reproductions of items selected from the personal manuscript collections of prominent biomedical researchers, medical practitioners, and those fostering science and health. The content of Profiles in Science is created in collaboration with NLM's History of Medicine Division, which processes and stores the physical collections. Most collections have been donated to NLM and contain published and unpublished materials, including manuscripts, diaries, laboratory notebooks, correspondence, photographs, journal volumes, poems, drawings, audiotapes and other audiovisual resources. The collections of Edward D. Freis, Virginia Apgar, and Michael Heidelberger were added this year. An additional 825 digital items composed of 6,500 image pages were also added to existing Profiles in Science collections. Presently the Web site features the archives of 19 prominent scientists. The 1964–2000 Reports of the Surgeon General, the history of the Regional Medical Programs, and Visual Culture and Health Posters are also available on Profiles in Science.

During this fiscal year, protocols for improving the quality of scanned images were developed and successfully used to disable automatic de-skewing of images and clean up speckled color/greyscale images. An early digital library, the Regional Medical Programs collection, was fully integrated into Profiles in Science. MeSH terms in use by the project were analyzed; obsolete terms were translated to MeSH 2006 terms, and "discontinued" MeSH terms were noted for future analysis. New queries and statistical reports were developed and implemented to detect unexpected patterns throughout the Profiles in Science data. Development of a new Profiles in Science XML-based Web front end and transition to a new XML-based search engine, particularly development of the

Annotations Server utilizing this new architecture, is ongoing.

MEDLINE Database on Tap

MEDLINE Database on Tap (MDoT) seeks to discover and implement systems and techniques to assist mobile clinicians in quickly finding relevant, high quality information addressing clinical questions that arise at the point of care. The primary goal is to present information to users so that they can quickly find the most pertinent parts, despite the limitations placed by the small screen and restricted bandwidth of handheld computers. MDoT explores display and navigation techniques, as well as information organization and content. MDoT also incorporates tools and systems from other LHNCBC projects, such as MetaMap and the Essie search engine. Essie and Google are offered as options, while the primary search engine is PubMed. We also seek to integrate semantic data from the search query and found citations to optimally rank results while maintaining real time response. A testbed system that supports MEDLINE search and retrieval from a wireless, Internet-connected PDA has been developed. Our client software for Palm OS and Pocket PC OS is freely available from the MDoT Web site, http://mdot.nlm.nih.gov/proj/mdot/mdot.php, which experiences between 5000 and 6000 hits every month. The Web site provides information about the project as well as the software, and allows us to solicit feedback from users and monitor aggregate user behavior. There are over 500 registered users of MDoT, and an unknown number of unregistered ones.

In the past year, MDoT has been evaluated in clinical settings at two institutions, the University of Hawaii and the VA Medical Center in Washington, D.C. In the first, medical residents enrolled in a Medical Informatics elective accompanied medical teams on morning rounds for four weeks, using MDoT to seek answers to clinical questions that arose at the point of care. They submitted daily summaries of each scenario and question in 44 rounds of about one hour each, with 187 clinical questions. Using a variety of MDoT options, they found relevant citations for 153 (82%) of the questions. This evaluation was observed by members of the MDoT team, who also conducted a number of workshops throughout the state, providing an overview of the system and hands-on training.

The second evaluation was conducted at the VA Medical Center in Washington, D.C. In this three-part collaboration among NIH/NLM/LHNCBC, the VA Center, and Universidade da Beira Interior, Portugal, a sixth year medical student rounded for 20 days with four different teams, using MDoT to search for answers to questions that arose on rounds. She recorded 144 clinical questions that were asked in context of 78 in-hospital clinical scenarios and 17 topic reviews. Because the VA Center is equipped with WiFi throughout, a significant difference between this study and the Hawaii study, in which residents used PDA/cell phones, is network data rate. One question to

address is whether this higher data rate has a notable effect on the ability to find useful information at the point of care. Results showed that answers were found in MEDLINE for 73% of the questions that arose on rounds, an unexpectedly high figure. On average, there were 5.2 queries per question, and less than four minutes per question was spent finding relevant citations. Results show that MDoT is fast and effective.

The MDoT evaluation plan calls for a "second opinion" of the selected citations by a senior investigator and expert MEDLINE indexer at LHNCBC. Based on a review of the scenario and clinical question, each selected citation is assigned a score of A (answers the question), B (contains a partial answer or is topically relevant and clearly indicates that the full text might answer the question), or C (does not answer the question).

As part of the MDoT project, outcomes research was conducted toward automatically finding patient outcomes (e.g., the population under study) from MEDLINE citations using knowledge extractors that rely upon NLM Unified Medical Language System and tools. The Extractor system identifies an outcome and determines whether a found outcome pertains to the topic of interest, the type of treatment studied, and the quality of the study.

Interactive Publications Research

The goal of this project is to create a comprehensive, self-contained and platform-independent multimedia document that is an interactive publication (IP), and to evaluate its value for better comprehension and learning. Following a study of existing open source formats and standards, a prototype interactive document was created containing many media objects: text, dynamic tables and graphs, a microscopy video of cell evolution, an animated spine in Flash, digital x-rays, and clinical DICOM images (CT, MRI, ultrasound). Both self-contained (embedded) and folder-type (linked) documents using all these media types were created in four formats: MS Word, Flash, HTML, and PDF. The IPs in these formats were compared in terms of ease of use and development effort.

While using such a document, the reader is able to: (a) view any of these objects on the screen; (b) hyperlink from one object to another; (c) interact with the objects in the sense of exercising control over them (e.g., start and stop video); (d) and importantly, reuse the media content for analysis and presentation. In light of the large sizes of such publications, possibly in the range of hundreds of megabytes, research is ongoing toward identifying techniques and protocols for rapid progressive download of the publications, and the development of a Download Manager based on this research.

To demonstrate the value of large tabular ("raw") datasets in an IP, some published articles were acquired from the American Psychiatric Institute for Research and Education, as well as the datasets underlying the tables appearing in the articles. The Institute also sent SAS scripts coding questions to the raw data. The datasets were loaded in SAS as well as CSV forms, and efforts are under way in linking the raw data to the published tables, and in creating hypothetical questions about specific age group and diseases that a reader might have (but which are not directly addressed in the paper). One of the articles is in the process of being converted to an interactive form.

NLM Gateway

The NLM Gateway provides an easy to use, one-stop search method that allows users to issue simultaneous searches in a number of NLM information resources from a single interface. The current version interacts with eight NLM search systems that provide results from 23 information resources. Changes to the underlying data structures or to the targeted search systems are carefully tracked and the Gateway modified accordingly. An example is the NLM Gateway release of October 2005 in which access to 100,000 meeting abstracts and health services research records was changed from the former Verity system to the new SE (Search Engine) system developed by LHNCBC staff.

While databases accessed by the NLM Gateway are regularly (sometimes even daily) updated, other resources incorporated into the Gateway itself are also regularly updated. New releases of the UMLS Metathesaurus, the UMLS mapping file, the 2006 MeSH update, and Year End Processing were incorporated during the year as they became available.

A new version of the Gateway accessing five additional toxicology-related resources was brought on line in November 2005. Later, changes in searching of the meeting abstracts retrieved search results from standardized XML data, allowing the display of diacritic marks and of additional fields from the records. In May 2006, the Bookshelf, a growing collection of full text biomedical books and other resources, was made accessible through the NLM Gateway. Access to the Household Products Database was also added. This database is a consumer's guide providing information on the potential health effects of chemicals contained in more than 5,000 common household products. In August 2006, access to the Profiles in Science collection was added. Profiles in Science contains archival collections of leaders in biomedical research and public health.

Feedback from users and statistical analysis of user actions helped to inform the planning of new functionality and new display options. Usability testing in the coming year will help in the testing of new ideas for facilitating user input and for creating better displays of system output. The intent is to create user-focused portals that help various categories of users quickly find what they need.

Digital Preservation Research

This project aims to investigate key issues related to the long term preservation of digital material, both documents and video. Our work in document preservation has matured and focuses on two processes: automated metadata extraction and file migration.

For document preservation, a prototype System for Preservation of Electronic Resources (SPER) was developed. SPER is a flexible, modular system that demonstrates key functions such as ingest, automated metadata extraction (AME) and bulk file migration. AME is implemented for the extraction of descriptive metadata from scanned and online journal articles as well as NLM's obsolete Web pages. Bulk file migration is implemented through an existing CEB system, DocMorph. While these functions are developed in-house, for the necessary infrastructure capabilities in SPER we have incorporated into the system, and customized, the latest version (1.4) of MIT's open source DSpace software. The Java client GUI for SPER was enhanced to incorporate batch metadata extraction and ingest for journal article TIFF pages, online journal articles and NLM Web pages (HTML). The GUI was also redesigned to display Web pages and online articles through Java Swing components.

SPER, in an abbreviated form, is being used in the preservation of a new collection at NLM consisting of over 65,000 historical FDA court records. Since the manual identification and entry of descriptive metadata from these records is labor-intensive, our focus is on automated extraction. In collaboration with the curator for this collection, we identified more than a dozen metadata items which could be extracted automatically. Our approach consists of: scanning the paper documents; auto-zoning the TIFF files using OCR output from the scanned documents; feature extraction; optimal feature selection; feature classification using a Support Vector Machine classifier; multi-class probability estimation; and statistical parsing using the Stolcke-Earley parsing algorithm.

Infrastructure Research

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

Advanced Biomedical Tele-Collaboration Testbed

The Advanced Biomedical Tele-Collaboration Testbed (ABC Testbed) project involves the use of open source, cross-platform technologies based primarily on grid technologies in general and the Access Grid (AG) in particular. The research is a collaborative effort with the University of Chicago, Argonne National Laboratory, the University of Illinois at Chicago, Northwestern University, the University of Rhode Island, and other institutions.

Among the scenarios that have been identified to test technologies: using the AG to link different patient safety and medical simulation; using AG with the daVinci surgical robot for distance education; using AG for wireless communication from mobile ambulances for patient treatment prior to arriving in the ER; the use of AG with handheld devices so residents can communicate more effectively; using the AG for 3D teleradiology; and using AG for volume rendering of patient image data in the operating room with wearable (e.g., eyeglass-like) environment. The latter allows surgeons to view the 3D data and to share it with colleagues and consultants while working on a patient.

In FY2006, the research team completed the substantial infrastructure required to test the scenarios. Several successful wide area wireless demonstrations of transmitting video and other patient data from ambulances using 3G and mesh cellular technology have been completed.

3D Telepresence for Medical Consultation

This project tests the efficacy of 2D versus 3D representations of video data transmitted in real time in remote clinical consultations. Although the research design could be undertaken without reference to the underlying computer algorithms for acquiring, transmitting, and displaying real time 3D video, the refinement and instantiation of these algorithms and related procedures for camera calibration, head tracking, and display in viable, transparent, user friendly 3D collaboration environments is the ultimate goal.

The research team made substantial progress in implementing the technology infrastructure. A prototype portable camera unit was added to the stationary one and calibrated. The PDA application was completed and all the basic components of the system proposed are in place. The current focus is on optimizing camera and sensor placement, refining calibration and rendering algorithms, and dealing with problems when perspective changes from different points of view, such as occlusion when an intervening object obstruct the view of interest. In addition, the team started experimenting with stereo displays of the 3D data rendered, and progress has been made in collecting data comparing the performance of paramedics in a simulation center working alone, working with a distant 2D video consultation, and with a 3D proxy consultation.

Scalable Information Infrastructure Initiative

NLM's Scalable Information Infrastructure (SII) Initiative is designed to establish testbed applications that demonstrate advanced network capabilities in health care, medical decision-making, public health, health education or biomedical, clinical or health research within the broad research agenda of the NLM. SII projects involve the use of testbed networks linking one or more of the following: hospitals, clinics, practitioners' offices, patients' homes,

health professional schools, medical libraries, universities, research centers and laboratories, and public health authorities. Among the applications:

- Wireless Internet Information System for Medical Response in Disasters (WIISARD) at the University of California, San Diego, the Advanced Network Infrastructure for Distributed Learning and Collaborative Research at the Stanford University School of Medicine, and the National Multi-Protocol Ensemble for Self-Scaling Systems for Health at Boston Children's Hospital.

- The Project Sentinel Collaboratory is a partnership involving Georgetown University and the Washington Hospital Center, among others. The project is tasked with building and deploying a data-centric Collaboratory to collect and analyze data from hospitals, clinics, weather services, satellite images of vegetation, mosquito collection, veterinary clinics and other sources in order to develop indicators and warnings of emerging threats to human health. During FY2006, all project infrastructure was completed and the hospital systems were linked. With the infrastructure completed and the advanced data visualization tools in place, the project team will continue to analyze Collaboratory data and explore open source strategies and methodologies in support of flexible access to biomedical data.

- SMART (Scalable Medical Alert and Response Technology) is a system for patient tracking and monitoring from the emergency site that continues through transport, triage, and transfer from external sites to the health care facility within a health care facility. The system is based on a scalable location-aware monitoring architecture, with remote transmission from medical sensors and display of information on personal digital assistants, detection logic for recognizing events requiring action, and logistic support for optimal response. Patients and providers, as well as critical medical equipment will be located by SMART on demand, and remote alerting from the medical sensors can trigger responses from the nearest available providers. The emergency department at the Brigham and Women's Hospital in Boston will serve as the testbed for initial deployment, refinement, and evaluation of SMART. This project will involve a collaboration of researchers at the Brigham and Women's Hospital, Harvard Medical School, and the Massachusetts Institute of Technology.

- The Tele-Immersive System for Surgical Consultation and Implant is aimed at developing a networked collaborative surgical system for tele-immersive consultation, surgical pre-planning, implant design, post operative evaluation and education. The Personal Augmented Reality Immersive System (PARIS) has been developed,

tested, and displayed publicly. The PHANTOM haptic device has been installed on the PARIS system. A Linux PC is used to drive the PARIS system. The PC controls two display devices at the same time. One is the projector on PARIS to display 3D stereo models. The other is an ordinary monitor to display the 2D user interface. The separation of the user interface and the sculpting working space allows much easier and smoother access to different functions of the application. The Physician's Personal VR Display was developed to facilitate consultation from the physician's desk without requiring that the physician go to a specialized facility. This system allows surgeons to do remote pre-operative consultation and post-operative evaluation, as the system enables all participants in a collaborative session to share their viewing angle, transformation matrix, and sculpting tools information over the network.

Wireless PDA PubMed Searching

Short Message Service (SMS) use in medicine is increasing with applications in monitoring of patients with chronic illnesses, appointment reminders and patient-doctor communications. Txt2MEDLINE is an application that provides access to MEDLINE/PubMed through SMS. A Web version of the application was presented at the 2006 American Telemedicine Association Annual Meeting in San Diego. Several clinical evaluations of the Txt2MEDLINE application are being done, including at the Prince Georges Health Center and at the University of the Philippines, to evaluate the application's effectiveness and usefulness in clinical practice.

Telemedicine Initiatives

OHPCC participated in talks and demonstrations of state-of-the-art telemedicine, e-health projects at the Dirksen and Hart Senate Office Buildings. The Congressional Steering Committee on Telehealth and Healthcare Informatics sponsored the talks, demonstration and roundtable discussion as part of its 2006 telehealth, e-health, and healthcare informatics projects and programs designed to address pressing healthcare issues. This program is intended to inform members of Congress and congressional staff, federal agency officials, healthcare and technology organization representatives, and the public. The NLM/OHPCC display featured wireless handheld access to PubMed.

A "Virtual Microscope" Website, http://erie.nlm.nih.gov/~slide2go/ was created to present the progress of the project. Teaching slides from the Department of Pathology medical student collection were digitized and archived work is available online at http://images.nlm.nih.gov/pathlab.

Videoconferencing and Collaboration

Major renovations were undertaken in FY2006, including rear screen and stereo projection and the installation of an Extron switch enabling multiple computing and video sources in the Collaboratory to be directed to various displays. The H.323 videoconferencing system and multipoint conferencing unit were upgraded to include a new Tandberg room system with multipoint conferencing, an improved H.264 video codec, and H.239 capabilities for application sharing. Major upgrades were made to the Access Grid (AG) and Conference XP computers and an Access Grid venue server was installed. A dual camera configuration of the AG was developed for 3D stereo video. Several software programs were installed and configured so the team could become familiar with newer collaboration tools. Web and streaming servers were upgraded.

A distance learning program in collaboration with SIS, coordinator of NLM's Adopt-A-School Program, continued to provide on-site and distance education about varied health science topics and information sources to students at the King Drew Medical Magnet High School affiliated with the Charles R. Drew University of Medicine and Science in Los Angeles. The link from an NLM-funded telemedicine study connecting the school to the university was re-activated to connect the school to Internet2. Programmatically, it eliminated the logistical problems of having to move students from the school to the university, enabled hands-on learning experiences in the school's computer lab, and allowed more classes to participate. The new Tandberg system at the Collaboratory and a new Polycom system at the school with identical capabilities improved the quality of the communication and the recordings made of the sessions considerably. The NIH Office of Science Education participated in the program and conducted several sessions on health science careers.

The Web casts of the bi-monthly Washington Area Computer Assisted Surgery Special Interest Group continued and videoconferencing was added so that there is now two-way interaction between those attending the meeting in the Lister Hill auditorium in Bethesda, where the presentations are made, and those in an auditorium at the Allegheny Hospital System in Pittsburgh. Attendees are now able to obtain continuing medical education credits because of this linkage. The team continues to do work with NCBI and the University of Puerto Rico to resolve technical problems in delivering distance education related to NCBI's information sources. Methods for providing application sharing and image manipulation with low latency have been identified and substantial progress has been made in enabling the instructor at NLM to view each remote student's desktop.

The Center for Public Service Communication (CPSC) was given a contract to pilot the use of video over IP to provide remote medical interpretation services. CPSC was open to collaborating with the team to do more formal assessment of the technology. As a result, team members have worked with CPSC staff to install and provide training on the technology in public health clinics in Florida and to develop the research methodology.

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. This research covers advanced library and terminology services, modeling and learning methods, medical ontologies, the indexing initiative, and semantic knowledge representation.

Advanced Library Services

Advanced biomedical information management applications exploit online information to support evidence-based medicine, enable scientific discovery, help translate discoveries into advances in patient care, and provide the basis for individual decision making. Some of the potentially exploitable information available online is in the form of text; examples include MEDLINE citations and associated full-text articles, ClinicalTrials.gov, and clinical narratives. Other online information is structured and includes biomedical vocabularies, clinical and molecular biology knowledge bases, and model organism annotation databases. The objective of the Advanced Library Services (ALS) project is to normalize and integrate biomedical information, both text-based and structured, into a repository of executable knowledge, a Biomedical Knowledge Repository (BKR), directly accessible by advanced applications including knowledge discovery, multi-document summarization, and question answering.

This project was launched during FY2006 and recently presented to the Board of Scientific Counselors. Two pilot projects were developed as a proof of concept. The gene information resource Entrez Gene was integrated into the Biomedical Knowledge Repository by converting it from XML representation to the Resource Description Framework (RDF). And Semantic Medline, an application that extracts knowledge from selected MEDLINE citations, creates a visual representation (graph) of salient assertions in those documents, and allows users to manipulate the graph interactively.

Work has begun to extract knowledge from the entire collection of documents in MEDLINE, as well as from structured databases, including the UMLS, and to include metainformation to the BKR. Future applications that exploit the repository will focus on a medical subdomain (e.g., cardiovascular diseases) and be based on user input.

Terminology Research and Services

LHNCBC research staff build and maintain the SPECIALIST Lexicon, a large syntactic lexicon of medical and general English that is released annually with the Unified Medical Language System (UMLS) Knowledge Sources. New lexical items are continually added using a

lexiconbuilding tool; the SPECIALIST lexicon contains over 330,000 records. The UMLS Lexical tools, including lexical variant generator (LVG), wordind, and norm are distributed with the UMLS as are text processing tools which analyze documents into sections, sentences, and phrases. The SPECIALIST lexicon, lexical tools, and text processing tools are released as open source resources and available under an unrestrictive set of terms and conditions for their use.

LexBuild is an evolving lexicon building tool designed to aid the lexicon building team by facilitating entry of lexical information and providing real time quality control. The SPECIALIST lexicon release tables are annually generated using the LexBuild tool. The SPECIALIST lexicon and tools are UTF-8 compliant and capable of dealing with non-ASCII characters. MMTx, the Java implementation of the MetaMap algorithm is a major application of the SPECIALIST lexical and text tools. A stochastic part-of-speech tagger is being developed for use in MMTx. The tagger will be specifically designed to exploit the SPECIALIST lexicon and will allow tagging of multi-word terms from the lexicon. It will be released as a freely available open source tool. LNHCBC researchers are engaged in porting the Journal Descriptor Indexing (JDI) tool to Java for future release as part of the UMLS lexical tools. The JDI should provide an element of context that can be useful for word sense disambiguation and other natural language processing tasks.

LHNCBC research staff also develop and maintain the UMLS Knowledge Source Server (UMLSKS) that provides Internet access to the UMLS knowledge sources through application programs and a user interface. UMLSKS is updated quarterly to accommodate quarterly UMLS releases. A grid/Web services implementation of the UMLSKS backend and an implementation of the user interface as a portal consisting of user-chosen "portlets" representing different parts and views of the UMLS data have been developed and will soon be released.

The goal of the Terminology Server (TS) project is to provide tools and data to manage diverse medical vocabularies for diverse purposes. Over the past year, the project continued to provide customized data sets using the released versions of the UMLS to several projects such as ClinicalTrials.gov and Genetics Home Reference for use in their operational environments. An important function of the TS is to support the customization of terminologies from the UMLS and other sources to satisfy individual project needs. A number of internal tools were developed to handle the data customization needs of the projects identified above, which resulted in periodic releases of data sets containing customized data. One new set of tools and processes added to the TS handles the generation of English-Spanish translation tables for the Spanish version of ClinicalTrials.gov. In addition, significant work was done on generating more efficient processing operations of existing vocabulary mapping algorithms, and producing the mapping tables for the current version of the UMLS Metathesaurus data.

A multilanguage search tool for non-English speakers for Medline/PubMed, (http://babelmesh.nlm.nih.gov), is continuing. It allows healthcare providers and researchers to search in their native language. Through international collaborations, including the WHO Eastern Mediterranean Regional Office in Cairo, more vocabularies were added to BabelMeSH. With the multilingual search portal, users can now search in Arabic, French, German, Italian, Japanese, Portuguese, Russian, Spanish, and English. Comments from international health liaison officers, including DHHS, were very encouraging after a presentation at a Fogarty International Center meeting.

PICO (Patient, Intervention, Comparison, and Outcome) Linguist is an application available through BabelMeSH that allows users to search Medline/PubMed in a more clinical and evidence-based manner. This work is significant because it is the only cross-language search portal on the Internet that allows the input in more than two languages. It is also unique because it allows the user to search in a character-based (non-Latin alphabet) language, transform it to an English language search and retrieve citations published in any language or language combination. Full-text articles may be linked to the result if published online and available without subscription requirements.

Modeling and Learning Methods

The Modeling and Learning Methods project is aimed at developing computational learning methods to enable scientists to utilize crossdisciplinary information effectively. Crossdisciplinary scientific information, associated always with uncertainty, comes with a multitude of overlapping but unidentical and sometimes conflicting perspectives. In order to cope with such information overload, scientists need assistance from computers to translate their mental models to computational models. Even with state-of-the-art computational tools, this translation is often very difficult due to the vagueness of the available information and its questionable reliability, forcing scientists to make artificially restrictive assumptions about the nature of their domain and information.

This project develops an information architecture called multifaceted ontological networks (muON), which is designed to cope with the aforementioned problems. In muON, every perspective of scientific information is captured in a different facet, which may overlap with other facets. Unlike the other ontological approaches, muON can cope with uncertainty via its underlying representation method called parameter interdependency networks (PIN).

PIN, a graphical modeling method being developed as part of this project, is based on probability theory and machine learning. The development and refinement of PIN are being driven by the needs of biomedical studies (e.g., Framingham Heart Study), of which parametric requirements directly determine the design, specifications and representational capabilities of the method.

Representing linguistic information on muON is also being studied in the context of information identification and labeling. The most fundamental information identification and labeling process in computational linguistics is tokenization. Each tokenizer makes a particular set of assumptions, which frequently fail, and the resulting errors are propagated to the subsequent steps of information processing. Experiments with different tokenization methods have strongly supported the necessity of the muON paradigm since it can preserve information in its entirety by representing both agreements and disagreements of different tokenizers concurrently.

Medical Ontology Research

While existing knowledge sources in the biomedical domain may be sufficient for information retrieval purposes, the organization of information in these resources is generally not suitable for reasoning. Automated inferencing requires the principled and consistent organization provided by ontologies. The objective of the Medical Ontology Research project is to develop methods whereby ontologies can be acquired from existing resources and validated against other knowledge sources. Although the UMLS is used as the primary source of medical knowledge, OpenGALEN, the Gene Ontology, and the Foundational Model of Anatomy are being explored as well.

During this fiscal year, the research team focused on relationships in biomedical ontologies. From a formal perspective, we studied dependence relations in the Medical Subject Headings showing how they correlate with statistical relations. While most efforts in biomedical ontology focus on organizing concepts, we analyzed how relationships from the UMLS Metathesaurus relate to relationships in the Semantic Network, paving the way for the development of an ontology of relationships.

Work was pursued on two particular domains: anatomy and molecular biology. New methods were developed for aligning anatomical ontologies, including complex rules to map groups of concepts. The Foundational Model of Anatomy was converted from its frame-based representation to the description logic language OWL, the Web Ontology Language used in Semantic Web applications. Similarly the gene information resource Entrez Gene was converted to the Resource Description Framework (RDF) in order to integrate it with other resources. Finally, semantic similarity in the Gene Ontology was used to compute similarity between genes and the results were used in several evolutionary biology studies.

The research team continues to work on the creation of an ontology of relationships as it is one critical element of a repository of biomedical knowledge supporting knowledge discovery and reasoning. Future work includes enhancing RxNav, the interface to the drug vocabulary RxNorm, integrating it with other drug information resources. We continue to participate in the progress of the Semantic Web for Health Care and Life Sciences and collaborate with leading ontology centers, including the National Center for Biomedical Ontology.

Indexing Initiative

The Indexing Initiative project investigates language-based and machine learning methods for the automatic selection of subject headings for use in both semi-automated and fully automated indexing environments at NLM. Its major goal is to facilitate the retrieval of biomedical information from textual databases such as MEDLINE. Team members have developed an indexing system, Medical Text Indexer (MTI), based on three fundamental indexing methodologies. The first of these calls on the MetaMap program to map citation text to concepts in the UMLS Metathesaurus. The second approach, the trigram phrase algorithm, uses character trigrams to also map text to Metathesaurus concepts. Finally, the third method uses a variant of the PubMed related articles algorithm to find previously indexed articles that are textually related to the input and then use some of the MeSH headings used to index them. Results from the three methods are restricted to MeSH, if necessary, and combined into a ranked list of recommended indexing terms.

The MTI system is in regular use by NLM indexers to create indexing terms for MEDLINE. MTI recommendations are available to them as an additional resource through the Data Creation and Maintenance System (DCMS). In addition, the indexing terms produced by MTI are being used as keywords to access collections of meeting abstracts via the NLM Gateway. These collections include abstracts in the areas of AIDS/HIV, health sciences research, and space life sciences. The Indexing Initiative staff continues with research efforts designed to improve MTI's accuracy by adding complementary methods of word sense disambiguation to the existing facility for reducing MetaMap ambiguity. They have also begun research to extend MTI's recommendations from unqualified MeSH headings to heading/subheading pairs.

Journal Descriptor Indexing

The Journal Descriptor Indexing (JDI) project investigates a novel approach to fully automated indexing based on NLM's practice of maintaining a subject index to journal titles using a set of 122 MeSH terms known as JDs (journal descriptors), that correspond to biomedical specialties. JDI was used as a broad filter to extract from a ten-year MEDLINE text collection of 4.59 million records, those likely to be of genomics interest (39% of the collection), as part of the NLM participation in TREC (Text Retrieval Conference) 2004.

Project staff also developed an algorithm used in a MeSH gene matcher program that contributed to the NLM TREC 2005 (Text Retrieval Conference) effort. This program takes as input names of genes in the topics for the TREC 2005 ad hoc retrieval task and returns MeSH

preferred terms and synonyms from 2004 MeSH, thereby functioning as a query expansion tool for query genes. The program was modified to return additional synonyms created in 2005 MeSH.

Current work involves efforts to use semantic type indexing based on Journal Descriptor Indexing for disambiguation in the MetaMap system, to produce a Java version of the JDI system including the semantic type indexing component to be distributed as an open source tool with the UMLS Natural Language Processing tools, participation in a study on just-in-time answers to clinical questions using MDoT on PDAs, and assistance in creating test data for the full-text collection against which retrieval tasks performed by participants in the Genomics Track of TREC 2006 are to be run.

Project staff continue to collaborate with researchers at Rouen Medical School and at the University and Hospitals of Geneva, who will perform the evaluation that compares automatic assignment of Metaterms by the CISMeF (Catalog and Index of French Language Health Resources on the Internet) system versus automatic assignment of Journal Descriptors by the JDI system, against the human gold standard finalized in May 2006. The group is also participating in a research project to enhance NLM's Medical Text Indexer to append MeSH qualifiers automatically to MeSH headings, using automatic indexing rules.

Unified Medical Language System (UMLS)

The mission, scope, and content of the Unified Medical Language System Metathesaurus continued to grow and evolve in FY2006. Most of the UMLS Metathesaurus group efforts have gone into continuing UMLS production operations while undergoing the transition of production operations from LHNCBC to the Office of Computer and Communications Systems and the Library Operations Division. A status report of the transition process for all phases following vocabulary inversion was presented to the Board of Regents at its September 2006 meeting. It is anticipated that transition of these phases will be completed in FY2007, while transition of vocabulary inversion and the LO transition continue.

The third quarterly release of the UMLS Metathesaurus in calendar 2006 contains more than 1.3 million concepts (an 18% increase over its predecessor) and 6.4 million concept names (a 20% increase). There are more than 100 contributing source vocabularies. The UMLS provides the only way for the U.S. health care community to obtain SNOMED CT, the largest HIPAA standard clinical vocabulary, under the U.S. government license.

The format and content of the underlying biomedical vocabulary files varies widely. Without unifying standards or common tools, it is difficult to understand and use any single vocabulary, and far more difficult to integrate multiple combined vocabularies. The UMLS customization and installation tool, MetamorphoSys, allows the selection of desired content from the Metathesaurus and generates the desired subset in Rich Release Format (RRF) or Original Release Format. MetamorphoSys includes an improved RRF Browser which allows users to view their own subsets in both Raw Data and Concept Report views. This means that any vocabulary in RRF may be reviewed, studied, or compared with views in other applications. This will make it much easier for users to make and then to see, understand, and verify their chosen Metathesaurus subsets in their own applications.

The Rich Release Format contains additional information allowing exact attribution of the sources for all its information. This allows specific mappings between vocabularies, correct inclusion and exclusion of specific sources, and simultaneous representation of a consistent UMLS view along with each source's own view, which may differ.

New development goals for MetamorphoSys include XML output, Section 508 compliance, a mapset browser, multiple instances to compare vocabularies, and a search by code function. In addition, the LHNCBC Metathesaurus group continues to work to refine and promote two standards for vocabulary exchange and UMLS submission: the Rich Release Format and the new single vocabulary Terminology Representation and Exchange Format (TREF).

Semantic Knowledge Representation

Innovative applications for providing more effective access to biomedical information depend on reliable representation of the knowledge contained in text. The Semantic Knowledge Representation project develops programs that extract usable semantic information from biomedical text by building on existing NLM resources, including the UMLS knowledge sources and the natural language processing tools provided by the SPECIALIST system. Two programs in particular, MetaMap and SemRep, are being evaluated, enhanced, and applied to a variety of problems in biomedical informatics. MetaMap maps noun phrases in free text to concepts in the UMLS Metathesaurus. SemRep uses the Semantic Network to determine relationships asserted between those concepts.

The MetaMap Technology Transfer program (MMTx) is an exportable, Java-based version of MetaMap that allows users to exploit the UMLS MetamorphoSys program to exclude or reorder the Metathesaurus vocabularies that MMTx uses. Users can also create MMTx data files independent of the UMLS, and the inclusion of source code with each release allows additional control of processing.

The development of SemRep is based on viable strategies for effective natural language processing and underpins foundational investigations in biomedical information management. At the core of this research is enhancement of linguistic coverage, and SemRep was recently expanded to address pharmacogenomics text. Syntactic and semantic mechanisms were added to

accommodate a range of semantic relations, including genetic (gene-disease), genomic (gene-gene), and pharmacogenomics (drug-gene, drug-genome); in addition, relations between genes and population groups and pharmacological relations (drug-disease, drug-pharmacological effect, drug-drug) are now identified.

Semantic predications produced by SemRep serve as the basis for continued work in biomedical information management. Application areas include automatic abstraction summarization and visualization of text from MEDLINE and ClinicalTrials.gov as well as cross-language summarization and question answering. A recent application, Semantic Medline, integrates PubMed searching, advanced natural language processing, automatic summarization, and visualization into a single Web portal. Semantic Medline is intended to help users manage the results of PubMed searches by normalizing core assertions in the citations retrieved. These normalized forms constitute computable knowledge accessible to further manipulation, including condensation by automatic summarization. The normalized and condensed output of Semantic Medline is visualized as an informative graph with links to the original MEDLINE citations. Convenient access is also provided to additional relevant knowledge resources, such as Entrez Gene, Genetics Home Reference, and the UMLS Metathesaurus.

Multimedia Visualization

The Lister Hill Center performs extensive research and development in the capture, storage, processing, retrieval, transmission, and display of multimedia biomedical data. Multimedia products include high quality video, audio, imaging, and graphics materials.

Turning The Pages Information Systems

The Turning The Pages Information Systems (TTPI) project brings rare books at the NLM to public view in a compelling way: as photorealistic volumes whose pages may be virtually "touched and turned." Visitors to the Library may experience this on kiosks, and those offsite may view the books online. The TTPI project investigates ways to efficiently produce and distribute the TTP books through the Web while maintaining high quality.

Originating as collaboration with the British Library in producing two virtual books, Blackwell's 18th century *A Curious Herbal* and Vesalius' 16th century anatomy book, we have made significant improvements on the original process. Our process consists of scanning the pages and book cover, enhancing these high quality color images by Adobe Photoshop, and creating animated 3D wireframe models of the pages using Alias Maya run on a computer by Macromedia Director software and displayed on a touchscreen monitor in kiosks. The library patron may "touch and flip through" each of these books in an intuitive manner that evokes the feel of a real paper volume.

In creating the 3D model using Maya, each pair of page images is texture-mapped to both sides of the wireframe model of a turning page, with a multisource lighting model that provides attractive diffuse lighting, specular highlights and shadows. For each flip, 12 intermediate animation frames are generated and rendered, and then imported into Director.

Three additional books from NLM's historic collection have been added to the first two: Paré's surgical treatise, Gesner's *Animalium*, possibly the earliest book in zoology, and Johannes de Ketham's *Fasiculo de Medicina* (1494). A sixth TTPI book is being prepared: Robert Hooke's *Micrographia*, the first book written about microscopes and in which reportedly the word "cell" was used for the first time. New technical challenges in converting this book include fold-out pages and the possible inclusion of images of historic and present day microscopes.

Visible Proofs Exhibition Preview DVD

In conjunction with the Office of Communications and Public Liaison and the History of Medicine Division's Exhibition Program, the Audiovisual Production Development Branch produced an event video featuring the new NLM exhibition, "Visible Proofs: Forensic Views of the Body." APDB provided pre-production planning, thematic treatment, and a production schedule to achieve a target delivery date for the overview video in time to be presented at the NLM's Board of Regents meeting. Interviews were conducted remotely and included Barry Scheck, JD, Innocence Project, New York City, NY; Marciello Fierro, MD, Virginia Medical Examiner, Richmond, VA; David Fowler, MD, Maryland Medical Examiner, Baltimore, MD; and Stephen Sherry, Ph.D., Staff Scientist, NCBI. The DVD has an original animated opening sequence as well as interstitial animations to support the DNA-based forensic science themes within the exhibition.

In addition, APDB produced a high definition (HD) video detailing the NLM major program accomplishments over the last year. Medical illustrators created 3D animations that brought clarity, understanding, and visual impact to these programs. An APDB producer compiled the research materials and coordinated with the spokespersons for each of the featured programs. On-camera interviews of the spokespersons were videotaped and, with the research information, incorporated into a production script that highlighted the programs and participants. Dr. Lindberg's comments regarding the programs and the accomplishments were also recorded and incorporated into the program. The video was shown at the May Board of Regents dinner. Also, APDB provided project management support and HD video recording of several events including the Information Rx press event held in Naples, Florida, and the NLM Diversity Council-sponsored Artificial Body Parts Symposium.

Training Opportunities

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 LHNCBC Informatics Training Program (ITP), ranging from a few months to more than a year, is available for visiting scientists and students. 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 which is open to all interested members of the NLM and NIH community.

In FY2006, the Center provided training to 46 participants from 13 states and nine countries. Participants worked on research projects including medical image processing, consumer health informatics, document analysis, grid computing, information retrieval, machine learning, medical illustration, micro-pathology, medical terminology research, natural language processing, medical ontology research, telemedicine, and ubiquitous computing.

The program maintains its focus on diversity through 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.

The Center continues to offer an NIH Clinical Elective in Medical Informatics for third and fourth year medical and dental students. The elective offers students the opportunity for independent research under the mentorship of expert NIH researchers. The Center also hosts the eight-week NLM Rotation Program which continues to provide trainees from NLM funded Medical Informatics programs with an opportunity to learn about NLM programs and current Lister Hill Center research. The rotation includes a series of lectures covering research being conducted at NLM and the opportunity for students to work closely with established scientists and meet fellows from other NLM-funded programs.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

David Lipman, M.D.,
Director

The National Center for Biotechnology Information (NCBI) was established in November 1988 by Public Law 100-607 as 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. Many experimental strategies in biomedicine now involve the integration and analysis of vast amounts of complex and diverse digital biological information. The flood of genomic data, most notably gene sequence and mapping information, has played a large role in the increased use of bioinformatics. NCBI meets 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-based discoveries of the 21st century.

The Center carries out its mission 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 has 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 and apply this research to the development of public information resources.

NCBI programs are divided into three areas: (1) creation and distribution of databases to support the field of molecular biology; (2) basic research in computational molecular biology; and (3) dissemination and support of molecular biology and bibliographic 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 allow seamless searching of the sequence data housed in GenBank and provide links to related sequences, bibliographic citations, and other related resources. Such features allow GenBank to serve as a critical research tool in the analysis of gene function and the identification of disease genes.

Important sources of data for GenBank are direct sequence submissions to NCBI from individual scientists and genome sequencing centers. Substantial staff and resources are devoted to the analysis and curation of these sequence data. 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. The database is comprised of two divisions, traditional nucleotide sequences and Whole Genome Shotgun (WGS) sequences, which are contigs (overlapping reads) from WGS projects. Annotations are allowed in these assemblies and records are updated as sequencing progresses and new assemblies are computed.

The Third Party Annotation (TPA) database, created in conjunction with international counterparts EMBL and DDBJ, supports third party annotation of sequence data already available in public databases. 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. This year, TPA records were divided into two sections—TPA:experimental and TPA:inferential. TPA:experimental includes annotation of sequence data supported by peer-reviewed, wet-lab experimental evidence, and TPA:inferential contains annotation of

sequence data by inference where the source molecule or its products have not been the direct result of experimentation.

The traditional and WGS divisions of GenBank combined reached a total of 100 billion bases this year, a milestone for the database and its collaborators. In FY2006, approximately 14 million sequences were added to the traditional GenBank division, and the base count rose from 50 billion in August 2005 to 65 billion in August 2006. The WGS division contained 80 billion bases and 17 million entries as of August 2006.

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.

When scientists submit their sequence data to GenBank, they receive an "accession number." The accession number serves as a tracking device and allows the scientist to reference the sequence in a subsequent journal article. Sequence data submitted in advance of publication are maintained as confidential, if requested.

NCBI is continuously developing new tools and enhancing existing tools to improve access to, and the utility of, the enormous amount of data stored in GenBank. Sequence data, both nucleotide and protein, are supplemented by pointers to corresponding PubMed bibliographic information, including abstracts and publishers' full-text documents as available. GenBank provides links to outside sources such as biological databases and sequencing centers. In addition to literature information, GenBank also 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 related biological data on the NCBI Web site.

Improvement of NCBI's sequence submission software continues to be a high priority. Sequin, NCBI's stand-alone submission tool, allows for updating as well as submission of a large number of GenBank sequences. The submission tool Sequin MacroSend 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 specialized submissions, such as genomes, batch sequences, and alignments, are also available. BankIt, another sequence submission software tool, is now in its twelfth year of use. BankIt is useful for small submissions that can be uploaded directly to NCBI.

GenBank contains 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. The Genome Survey Sequences (GSS) division of GenBank contains sequences that are genomic in origin, rather than cDNA. The Sequence Tagged Site (STS) division consists of short sequences that are operationally unique in the genome and used to generate mapping reagents. Expanded and comprehensive STS information can be found in the UniSTS database.

Genome Resources

Integrated Genome Data

NCBI has developed a suite of genomic resources to support comprehensive analysis of many genomes in addition to human. Specialized tools and databases have also been designed to facilitate researchers' use of data. NCBI maintains an expanding collection of specialized, yet integrated, database repositories that collectively capture and redistribute the biological relationships between genome sequences, expressed mRNAs and proteins, and individual sequence variations.

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, RefSeq, Map Viewer, Gene, Homology Maps, UniGene, HomoloGene, and GEO. Links to outside information are also available, such as Linkage and Physical Maps, TaxPlot, and chromosome-specific mapping information.

NCBI genome resource guides provide information on diverse organism-related resources from multiple centers including sequence, mapping, and clone information. Genome resource guides for 25 organisms provide easy navigation to organism-specific BLAST and MapViewer pages and other NCBI resources as well as to outside resources, such as documentation, maps and sequence information, annotation projects, assembly updates, and comparative genomic sites.

Assembling and Annotating Genomes

NCBI is responsible for collecting, managing, and analyzing genomic data generated from the sequencing and genome mapping initiatives of sequencing projects. NCBI also plays a key role in assembling and annotating genome sequences. These resources are truly an international public sequencing effort due to the cooperation of scientists and sequencing centers from around the world. The most recent updates to the human and mouse genomes (Build Version 36) have increased annotation, including access to alternate assemblies of the genomes previously unavailable and reference sequences for two alternate haplotypes. NCBI's past experience with curation of human and mouse

genomes has benefited the annotation pipelines for many other organisms.

A team of NCBI scientists is engaged in annotating, or characterizing, the biologically important areas of genomes. Genome builds are based on Gnomon, the gene prediction program developed by NCBI scientists. Gnomon uses RefSeq transcript sequences and EST alignments along with *ab initio* gene predictions, putting a greater emphasis on coding propensity and matches to existing proteins when predicting genes. NCBI's refined annotation pipeline allows annotation, in a single day, of microbial genomes submitted as whole genome shotgun sequences without annotation.

Gene Identification

The Reference Sequence (RefSeq) database is a comprehensive, integrated, non-redundant set of sequences, including genomic DNA, gene 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. In FY2006, the NCBI RefSeq database grew by over 900,000 proteins; the most recent full release of all NCBI RefSeq records, Version 19, includes over 2.8 million proteins from over 3,700 organisms.

NCBI is working with other groups to compare and evaluate genome annotation data and identify the set of proteins, as annotated on genomic sequence, that pass quality tests and are consistently identified by different groups. The Consensus CoDing Sequence (CCDS) database is a collaborative project between NCBI, European Bioinformatics Institute (EBI), University of California, Santa Cruz (UCSC), and Wellcome Trust Sanger Institute. The project identifies a core set of human protein coding regions that are consistently annotated and of high quality. Annotated genes included in CCDS are given a unique identifier similar to the GenBank system accession number. Over 400 CCDS gene records were updated in a preliminary review that primarily resulted in additional annotation for the coding sequences.

The Entrez Gene database offers a large scope of gene-specific data, integration with other NCBI databases, and enhanced options for query and retrieval from the Entrez system. Entrez Gene integrates information about genes and gene features annotated on RefSeq sequences and other model organism databases, making it easier for researchers to find and interpret gene-specific information. In FY2006, Entrez Gene continued to be enhanced in usability and content. The interactive display was modified to expose databases from which related information could be extracted. Documentation was also enhanced, indicating, for example, how connections are calculated between Gene and other key databases within NCBI. The number of species and genes represented in the database increased

significantly, to more than 3,500 taxa and almost 2 million genes.

NCBI's Map Viewer continues to be the primary resource for visualization of large genomes. Increased standardization of map features better supports cross-species comparison and multiple-species queries. Maps from other sequencing centers are also available. Genes or markers of interest can be found by submitting a query against a whole genome, or by querying one chromosome at a time. The results table includes links to a chromosome graphical view where a 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 FY2006, NCBI continued to enhance its Map Viewer with 38 organisms currently represented. New genomes added include *Macaca mulatta* (rhesus macaque), *Tribolium castaneum* (red flour beetle), and annotation updates were included for human, mouse, bee, and rat, among others. Support was added to provide information for current and previous builds for human and mouse genomes. In addition, links to related information were increased to improve marker navigation.

As the number of sequenced genomes continues to grow, there is increasing interest in comparative analysis of genes from represented species. The NCBI HomoloGene database of gene homologs performs such large-scale comparison, automatically presenting reports that include statistics on inter-species sequence and protein domain conservation with links to the genome-wide views available in MapViewer, Entrez Gene, and gene expression information in UniGene. A link was recently developed to enable downloading mRNA, protein, and genomic sequences of genes.

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. UniGene is continuously expanded to cover newly sequenced genomes.

The Probe Database, also part of the Entrez system, stores molecular probe data together with information on success or failure of the probes in different experimental contexts. Nucleic acid probes are molecules that complement a specific gene transcript or DNA sequence useful in gene silencing, genome mapping, and genome variation analysis. The database contained over 6 million probes as of August 2006.

The RNA interference (RNAi) resource stores the sequences of RNAi reagents and experimental results using those reagents, such as extent of gene silencing and a variety of phenotypic observations. It is fully integrated with the Probe Database, where the actual RNAi reagents are stored.

The dbSNP database of genetic variation is a comprehensive catalog of common human polymorphisms for the international research community. dbSNP continues to experience rapid growth, containing over 27 million submissions of human data, which have been processed and reduced to a non-redundant set of 12 million refSNP clusters. Thirty-four other organisms are represented in the SNP database. With the release of Build 125, dbSNP crossed the threshold of containing over 1 billion individual genotypes, representing the results of the International Hap Map Project, the Bovine HapMap Project (Phase 1), and numerous other medical genotyping studies. In FY2006, dbSNP released a genotype server product to provide this rapidly growing class of variation data to individual investigators and other bioinformatics groups that have been tasked with organizing the subset of public genome variation data relevant to their user constituencies.

dbSNP support for haplotype data was introduced in 2006 through a collaboration with UCSD to recompute haplotype structures of all individuals genotyped in each future build of dbSNP. This activity was also extended in 2006 through a collaboration with the HapMap analytical team in Oxford, UK to include updated chromosome phasing results for each build of dbSNP. dbSNP also completed several database redesign projects to respond to continued exponential growth in public genotype resources.

Comparative Genome Data

Entrez Genomes contains records representing over 3,000 species, including bacteria, archaea, and eukaryotes, complete microbial genomes, a number of viroids, mitochondria, a broad range of plasmids, and over 2,000 viruses. Links to related resources include Plant Genomes Central, SARS CoronaVirus Resource, and the Influenza Virus Resource.

The Entrez Genome Project database is based on cellular organism-specific genomic information, including but not limited to genome sequencing, such as whole genome shotgun or BAC ends sequencing projects, large scale EST and cDNA projects, and assembly and annotation projects. The database is organized into organism-specific overviews that function as portals from which all projects in the database pertaining to that organism can be browsed and retrieved. This design allows the collection of disparate data that all refer to a single organism, conveniently displayed for easy access with references to all subprojects.

Genome-specific resource pages have been added for *Macaca mulatta* (rhesus macaque) and *Tribolium castaneum* (flour beetle). These pages provide a comprehensive genome guide, including internal and external links to resources such as sequencing and annotation projects.

Fungal Genomes Central is a new portal to information and resources about fungi and fungal sequencing projects from NCBI and the fungi research community.

Plant Genomes Central is an integrated, Web-based portal to plant genomics data and tools. It provides access to large-scale genomic and EST sequencing projects and high resolution mapping projects. The Viral Genomes Website provides a convenient way to retrieve, view and analyze complete genomes of viruses and phages. This site now contains over 2,400 records for more than 1,600 viral genomes.

Model Organism Genomes

In FY2006, NCBI released Build 1.1 of the red flour beetle and rhesus macaque reference sequence genomes; annotation is also available in NCBI's Map Viewer. The red flour beetle (*Tribolium castaneum*) is a sophisticated model organism for higher eukaryotes and is a member of the largest and most diverse eukaryotic order, the Coleoptera. Studying the red flour beetle may yield insights into the genetic innovations that accompanied the evolution of higher forms with more complex development and facilitate the discovery of new pharmaceuticals and antibiotics. The rhesus macaque (*Macaca mulatta*) is an important model organism for biomedical research and behavioral studies. The genome sequence of the rhesus macaque will enrich our understanding of primate biology and evolution and will greatly facilitate the discovery of genetic traits that have uniquely arisen within the human lineage.

Also in FY2006, NCBI versions of the human and mouse genomes were released with significant improvements to the automated process of identifying genes within genomic DNA sequences. An update of the rat genome was released that includes re-annotation of the reference genome assembly and addition of an alternate genome assembly.

Genetic Disease Information

Genes and Disease is a collection of articles designed to educate the lay public and students on how genes are inherited and cause disease and how an understanding of the human genome will contribute to improving diagnosis and treatment of disease. For each gene description there is a link to PubMed, the Online Mendelian Inheritance in Man database (OMIM), the Map Viewer, Gene, and BLink for related sequences.

OMIM is an electronic version of Victor McKusick's "Online Mendelian Inheritance in Man," a catalog of human genes and genetic disorders. The database, produced at the Johns Hopkins School of Medicine, contains over 16,900 records and information from over 9,800 loci on the human gene map. 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. In FY2006, 748 new entries were added to the database.

Online Mendelian Inheritance in Animals (OMIA), authored by Dr. Frank Nicholas, is a database of genes,

inherited disorders and traits in animal species other than human and mouse. It contains textual information and references as well as links to other relevant records from OMIM, PubMed, and Gene. In FY2006, OMIA was added to the Entrez retrieval system.

The GeneTests database produced at the University of Washington is supported, as is OMIM, by contract from NCBI. GeneTests is used more than 25,000 times a day by genetics counselors and physicians for its comprehensive genetic testing information and genetic disease descriptions. GeneTests has been incorporated into the Books database under the title "Gene Reviews."

Genome-Wide Association Studies

The Genome-Wide Association Studies (GWAS) resource is a major NIH-wide initiative started in FY2006. It entails development of the systems for accepting, storing and providing access to the data from several whole genome association projects. GWAS involves linking up genotype data with phenotype information in order to identify the genetic factors that influence health, disease, and response to treatment. NCBI has developed the submissions, storage, and access mechanisms for GWAS data collected under several different programs.

Major near-term GWAS projects that are planned include the Genetic Association Information Network (GAIN), the Genes and the Environment Initiative (GEI), and the Framingham Heart Study. The planned GWAS Resource will organize the phenotype and genotype information in a publicly available database and provide unrestricted ability to browse and search projects and studies, protocols, questionnaires, and supporting documents, to view phenotype and genotype measures summary data, to identify studies of interest and to view precomputed associations. An authorization system has been developed for access to the de-identified individual phenotype and genotype data, which will require researchers to be approved by the Institute sponsoring the project.

Other Specialized Databases and Tools

The Red Blood Cell database (dbRBC) combines the well-established Blood Group Antigne Gene Mutation Database (BGMUT) with tools and interlinked NCBI resources. dbRBC provides publicly available genomic, protein, and structural information linked to red blood cell antigens and clinical data related to red blood cells. dbRBC provides such features as an alignment viewer, a sequence-based typing tool to facilitate analysis of sequences, a probe/primer resource that generates annealing predictions of probes based on sequence similarity, and a typing kit interface that provides a platform for submission and testing of genomic RBC typing kits.

The NCBI Trace Archive, which holds the raw single-pass reads of DNA sequence generated from large scale sequencing projects, surpassed 1 billion traces in FY2006. The trace data can be scanned using a rapid nucleotide-level, cross-species sequence similarity search tool called cross-species MegaBLAST. Using the visualization tools of the related Assembly Archive, researchers can examine an assembly of trace data from which a finished genomic nucleotide sequence has been derived and determine, for instance, whether a crucial nucleotide base change associated with a disease is well supported by the sequence evidence. In FY2006, the Archive also began taking data in a new format, short-read flow gram data (SSF).

The Influenza Virus Resource was created at NCBI with data obtained from the National Institute of Allergy and Infectious Diseases (NIAID) Influenza Genome Sequencing Project and NCBI's Influenza Virus Sequence Database, comprised of over 30,000 influenza sequences in GenBank and sequences from the RefSeq database. More than 11,000 new influenza virus sequences were entered into the database in FY2006. The Influenza Virus Resource was redesigned in FY2006 to improve the existing functionalities, such as the multiple sequence alignment tool. New features include an advanced database search tool and an influenza virus sequence annotation tool. A Flu Dataset Explorer provides an interactive tool for preliminary analysis of protein sequences from the NCBI Influenza Sequence Database or from a user's own file.

The Database for the Major Histocompatability Complex (dbMHC) contains variations found only in alleles of the Major Histocompatability Complex, a highly variable array of genes that plays a critical role in determining the success of organ transplants and is largely responsible for an individual's susceptibility to infectious diseases. The database now supports six major projects. One is a survey of Human Leukocyte Antigen (HLA) allele frequency distributions in various populations; this information is critical for establishing and searching bone marrow donor registries as well as for use in studies of HLA-associated disease susceptibility. Another project collects HLA genotype and clinical outcome information on hematopoietic cell transplants performed worldwide. Support for new projects related to Type 1 Diabetes, Rheumatoid Arthritis, and Natural Killer Cell Immunoglobulin-like Receptors, were added in FY2006.

The Gene Expression Omnibus database, or GEO, is a high-throughput gene expression/molecular abundance data repository, as well as a curated, online resource for storage and retrieval of gene expression data. GEO has grown to over 115,000 accessioned objects. GEO DataSets (GDS) was redesigned to allow full search capabilities to all GEO descriptive data, which facilitates search and retrieval of Platforms, Series and curated DataSets records. Full support was added for supplementary files, allowing researchers to reanalyze data recalculation of results based on newer algorithms as they become available. Data submission of microarray data has been enhanced to overcome problems with very large sets and to allow submission of supplementary files as an integral part of the main submission.

Taxonomy

The NCBI Taxonomy project provides a standard classification system used by the international nucleotide and protein sequence databases. The taxonomy group has continued to curate the rapidly growing Taxonomy database to include the names of species for which sequence has been submitted to the protein and nucleotide databases. This currently includes 150,000 formal binomial species names, which represent approximately 5% of the total number of described species of life on the planet. Tools have been developed for representing alternate, externally maintained taxonomies and cross-mapping them with Taxonomy database entries as well as a database of biological material collections used to enhance links between NCBI sequence entries and the corresponding specimen entries. The Taxonomy group is also working to expand the coverage of the systematic literature in PubMed.

The Taxonomy browser allows searches 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. The Taxonomy system also provides a "Common Tree" function that builds a tree for a selection of organisms or taxa.

PubChem and Protein Data

PubChem

The PubChem project is a key component of the NIH Roadmap project in Molecular Libraries and Imaging. The PubChem database is designed to be a repository for small molecule data and the foundation for the massive amounts of bioactivity data that will be produced by NIH-sponsored chemical genomics centers. Approximately 70 organizations are contributing substances to PubChem. PubChem's three databases—PubChem BioAssay, PubChem Compound, and PubChem Substance—contain information on millions of small molecules, including their structures, properties, and activities.

PubChem BioAssay allows users to examine descriptions of each assay's parameters and readouts and contains links to substances and compounds, enhanced through implementation of a queuing system and caching mechanism. The PubChem Bioassay database currently contains more than 220 bioactivity screens of chemical substances described in PubMed Substance and the number of new screens is rapidly growing. The database provides searchable descriptions of each bioassay, including descriptions of the conditions and readouts specific to a screen protocol. PubChem Compound searches unique chemical structures and validated chemical depiction information describing substances in the PubChem Substance division. PubChem Substance contains chemical substance records and associated information. Currently

PubChem contains over 10 million substance and 5 million compound records.

PubChem contains an extensive set of links within its own sets of data as well as to Entrez databases. Many compounds have literature citations to PubMed as well as links to the proteins and/or genes representing a protein to which they bind. Links between substances and compounds characterize chemical constituents. Links between substances and bioactivity indicate a substance was tested in a particular assay. Compound-compound links correspond to similarity relationships. Compounds are searchable by chemical structure, chemical properties, and bioactivity.

Improvements have been made to standardize submitted chemical information for inclusion in both PC Substance and PC Compound. Duplicate records in PC Compound are clustered and given a single compound ID so that the compound database is non-redundant. PubChem adds or modifies 120,000 records per day generating millions of links within the Entrez system.

A new Open Mass Spectrometry Search Algorithm (OMSSA) was released this year as well. OMSSA helps users to identify MS/MS peptide spectra by searching libraries of known protein sequences. OMSSA scores significant hits based on a probability score developed using classical hypothesis testing, the same statistical method used in BLAST. New versions were released in FY2006 with various improvements for better searching and results.

Protein Structure

NCBI's Molecular Modeling DataBase (MMDB) is the Entrez "Structure" database, a compilation of all the structures in the Protein Data Bank (PDB). PDB is a collection of publicly available, three-dimensional protein structures, nucleic acids, carbohydrates and a variety of other complexes experimentally determined by X-ray crystallography and NMR; it is maintained by the Research Collaboratory for Structural Bioinformatics (RCSB) and the European Bioinformatics Institute (EBI). MMDB continues to grow in size and currently contains over 37,000 unique, experimentally determined 3D structure records. MMDB is updated monthly, with the source PBD data checked for consistency in the chemistry, sequence, and 3D coordinates. The MMDB server program now displays small molecule structure together with a summary of the macromolecular chains and domains, helping users to better understand the nature of multi-molecule complexes.

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. Cn3D 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 Conserved Domain Database (CDD) is the Entrez database of sequence alignments and profiles defining protein domains as recurrent evolutionary modules. Identification of conserved domains within a protein sequence is also available via the CD-search service, which is run by default for each protein BLAST search. The CDD annotation staff produces curated hierarchies of models related by descent from a common ancestor, representing the ancient evolutionary history of protein and domain families. The staff uses 3D structure information, phylogenetic analysis, NCBI Entrez resources, and the published literature to enhance alignment quality, annotate functional sites, identify relevant links to PubMed and the NCBI Bookshelf, and to update domain family summary descriptions to reflect available knowledge of molecular function.

Imported models are being replaced with carefully curated representations of protein domain families, many organized in hierarchies of related domains. To date, more than 2,000 curated models are available on public services, with over 1,500 more in the curation pipeline. Models curated at NCBI make direct use of available 3D structure information and structure similarity. CDD records have been manually linked to relevant publications in Entrez and to chapters in the NCBI Books resource, and annotated functional sites on the majority of the models.

The CD-Search service, used to identify conserved domains present in a protein sequence, has been replaced with a new version that exclusively uses the NCBI C++ Toolkit. A new graphics library has been implemented, which is now employed in the visualization of MMDB data, VAST (Vector Alignment Search Tool) neighbors, related structures, and conserved domain annotation, enforcing a uniform display style. The Web service presenting conserved domain summaries has been enhanced, and it now presents "sequence trees," the results of simple phylogenetic analysis on curated alignment models. Those sequence trees are presented as evidence for sub-family hierarchies as built by the CDD curation staff.

CDTree, the main application used by CDD curators, was released in August 2006 together with a new version of Cn3D, NCBI's molecular structure viewer. CDTree functions as a helper application for Web-browsers, and enables users of the CDD resource to download and study curated hierarchies. It also allows users to "embed" arbitrary query sequences in curated CD hierarchies identified via CD-Search, and study evolutionary relationships, sub-family membership and more, facilitating accurate functional classification of proteins. An updated version of the Web-server displaying CD summary pages was also released together with CDTree.

VAST, or the Vector Alignment Search Tool, is a service that identifies similar three-dimensional structures of newly determined proteins. VAST compares new proteins to those in the MMDB/PDB database and computes a list of structure neighbors, or related structures, which allows a user to browse interactively, viewing superpositions and alignments in Cn3D.

Literature Databases

PubMed is a Web-based literature retrieval system developed by NCBI to provide access to citations and abstracts for biomedical science journal literature. PubMed is comprised of journals indexed in NLM's MEDLINE database as well as others beyond the scope of MEDLINE. It is the bibliographic component of 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.

During FY2006, PubMed added its 16 millionth citation to the database. Full-text journal PubMed links have increased from 4,774 in September 2005 to 5,750 in September 2006. A new AbstractPlus page is available which contains titles and links to related articles automatically, eliminating the need for the user to click on a "Related Articles" link. User help for PubMed was improved by the PubMed New and Noteworthy page as a Web/RSS feed. Also, the PubMed Help Document was added to the NCBI Bookshelf as a book.

The My NCBI tool allows customization of NCBI Web services featuring an option to automatically update and e-mail search results from user-saved searches. New features added this year include highlighting, saving search results, and institutional accounts for sharing filters, document delivery, and outside tool settings.

LinkOut is an Entrez feature 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. Close to 2,000 organizations have supplied links to their Web sites, doubling the number of participants from three years ago. Sources include over 1,500 libraries, 260 full-text providers, and 230 providers of non-bibliographic resources, including biological databases. Together they offer links to 43 million Entrez records. LinkOut usage jumped to more than 26 million hits per month, about 1 million hits per work day. Enhancements to the LinkOut program include a new homepage and help manual to provide easy navigation and quick access to all information for participation, and the ability to activate filter settings with a special URL and customize the filter name to help the use of LinkOut filters. A number of new training and promotional materials have been developed to support LinkOut providers and encourage more participation.

PubMed Central (PMC) serves to archive, index, and distribute peer-reviewed journal literature in the life sciences and provides free and unrestricted access to full-text journal articles. This repository is based on a natural integration with the existing PubMed biomedical literature database of indexed citations and abstracts. Currently, over 823,000 articles and related items are available free from the PMC journal archive, representing an increase of 65%

over the past 12 months. The additions have come from newly published material as well as from digitizing back issues that previously were only available in printed form. The PMC online archive includes material that was originally published as early as 1865.

Articles deposited by NIH-funded researchers under the NIH Public Access Policy are being processed efficiently and made available to the public via PMC. Articles from researchers funded by the Wellcome Trust in the UK are being handled similarly.

The portable PMC (pPMC) system, derived from the software that runs the U.S. PMC system, enables collaborating international archiving centers to replicate the PMC archive and service, thus increasing the viability of the archive. It has been tested successfully by five international archiving centers. These centers are expected to be the first members of a PMC International (PMCI) network. The long-term goal of PMCI is to create a network of digital journal archives that share content, leading to a more durable archive and better access to the biomedical literature.

The NCBI Bookshelf gives users access to the full text of over 60 textbooks in the clinical and research areas of biomedicine. Books may be searched directly or found through links in PubMed abstracts. In addition to textbooks from commercial publishers, the Bookshelf includes monographs authored by NCBI, NLM, and NIH staff. New books added in FY2006 include *Collective Expert Evaluation Reports, Ecology, Epidemiology, and Evolution of Parasitism in Daphnia, Approved Lists of Bacterial Names,* and various NCBI help documents. Other existing books and collections were updated and expanded, including the Eurekah Bioscience Collection, the HSTAT Collection, and Coffee Break. GeneReviews, derived from the GeneTests database, was added as well.

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 similar sequences and to statistically evaluate the resulting matches. In a matter of seconds, BLAST compares a user's sequence with up to a million known sequences and determines the closest matches.

The BLAST suite of programs is continuously enhanced for effectiveness and ease of use. BLAST genome pages allow for convenient searching of an organism of choice. The BLAST sequence comparison server is one of NCBI's most heavily used services and its usage has grown at a pace reflecting the growth of GenBank. Additional hardware and improvements in the BLAST code have enabled response times to decrease despite increases in the size of the database and number of users.

In FY2006, a new display option for BLAST results was designed and put into production. This "Distance Tree" displays the relative distance between sequences based on alignment to the query sequence. The distance tree is available for all nucleotide and protein BLAST results. The BLAST code and the Splitd architecture was modified and recoded to enable a second system to be implemented at the NIH Consolidated Co-location Site. Also, a system was designed and put in place to efficiently transfer and update BLAST databases between the two locations. Code was developed to provide users access through the MyNCBI authentication system to store and retrieve search strategies and results for BLAST. The option to sign in and save data should be made available later this year. Research and coding for an algorithm for multiple alignments of proteins was completed. This program is called COBALT and should be accessible as a Web application later this year.

Database Access

All of NCBI's Web servers and most of NCBI's public service, internal production, and non-Microsoft SQL Server database systems run on the Linux operating system. During the current reporting period a major project has been undertaken to migrate these systems from a public-domain version of Linux to a commercially supported version, Novell's SUSE Linux Enterprise Server. The planning and testing phase of this project is now largely complete and the initial phase of implementation is in progress. To date, the boot servers, a number of development servers, and a portion of the Platform LSF-based NCBI compute farm have been migrated.

Much of NCBI's computing infrastructure was moved from 32-bit hardware to 64-bit hardware in the past year. Specifically, an upgrade of the systems supporting NCBI's BLAST service that began in 2005 was completed in 2006, with a total of over 120 64-bit dual CPU servers dedicated for this purpose at the Bethesda site. In addition, approximately 60 of the older 32-bit systems were shipped to the Virginia co-location site to allow for development and testing of a version of the BLAST service that is distributed and load-balanced between the two sites. The development and testing phase of that project is complete, and 40 new 64-bit dual-core 2-CPU systems have been acquired to provide a substantial BLAST capability for that site. In addition, the storage infrastructure for BLAST has been upgraded. Eleven Network Appliance (NetApp) FAS270C NFS server clusters were replaced with six larger and more powerful FAS3020C clusters. The FAS270C clusters were repurposed as storage-area networks (SANs) to support other projects, including PubChem, PubMed and other database projects. NCBI's flagship service, PubMed, was also upgraded from 32-bit to 64-bit systems. An upgrade to storage for PubMed was largely completed, adding four new NetApp NFS servers.

Hardware to support PubChem, a component of the NIH Roadmap initiative, was significantly expanded. Three new SANs and one NFS server cluster were put in service, as were 10 SQL Servers, 12 back-end servers for PubChem public services, and 20 compute nodes dedicated to PubChem internal production processing. Twenty

additional compute nodes were added to the general compute farm complement. These 40 new systems are the first 64-bit servers in the compute farm and the first to run the SUSE version of Linux.

A new project this year is the Discovery Initiative. The object of the initiative is to provide users with richer and deeper information related to their queries, thereby facilitating the discovery process. To date, 50 new servers have been acquired to support this project. Another major infrastructure upgrade that began in the previous reporting period and greatly advanced in the current one is the replacement of directly attached storage for Microsoft SQL Servers with SAN storage, including NetApp clusters repurposed from the BLAST upgrade as well as new acquisitions, approximately 20 in all.

The proliferation of servers and storage in the past year has necessitated concomitant upgrades to our physical and network infrastructure. Approximately 50 new 20 Amp electrical circuits were provisioned for NCBI's computer room on the B2 level of Building 38A. In addition, NCBI participated with other NLM divisions in a plan to renovate the larger computer room on the B1 level. As of this writing, work to provide a 50% increase in the electrical power available to that room is about one month from completion.

On the network side, more than 20 FibreChannel switches were obtained and installed to support the new SANs described above and the number of GigabitEthernet ports was tripled to 288 ports. Six low-density (16-port) fiber GigabitEthernet cards in the core routers were replaced with four 48-port, fabric-enabled cards, and two 48-port, copper GigabitEthernet cards were added. The increase in disk storage has necessitated a corresponding increase in tape backup capability by upgrading and adding new tape drives.

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 application. The research program focuses on computational approaches to a broad range of fundamental problems in evolution, molecular biology, genomics, biomedical science and bioinformatics using theoretical, analytical, and applied mathematical methods.

NCBI's basic research group is within the Computational Biology Branch (CBB) and consists of 75 senior scientists, staff scientists, research fellows, postdoctoral fellows, and students. The basic research in CBB has strengthened NCBI applications and database work by providing innovative algorithms and approaches that form the foundation of numerous end-user applications. Moreover, researchers in the group continue to make fundamental biological and biomedical advances by applying NCBI databases and software to sequences, genomes, and other large-scale data, and by developing

concepts and proposing experimental strategies in collaborations with NIH and extramural laboratories.

Projects within CBB include new computer methods to accommodate the rapid growth and analytical requirements of genome sequence, molecular structure, chemical, phenotypic, and gene expression databases and associated high-throughput technologies. Computational analyses are also applied to human disease genes and to the genomes and functional biology of several pathogenic bacteria and viruses, often combining computational approaches with collaborations with experimental laboratories. Another research focus is the development of computer methods for analyzing and predicting macromolecular structure and function. Other recent advances include improvements to the sensitivity of alignment programs, analysis of mutational and compositional bias influencing evolutionary genetics and sequence algorithms, investigation of gene expression regulation and other networks of biological interactions, analyses of genome diversity in influenza virus and malaria parasites related to vaccine development and evolution of virulence, the evolutionary analysis of protein domains, comparative genomics and the development of theoretical models of genome evolution, genetic linkage methods, and new mathematical text analysis or retrieval methods applicable to full-text biomedical literature. Researchers also collaborate with other NIH institutes and several extramural and worldwide institutes on many of these research projects. New research projects were initiated in support of the PubChem molecular libraries project, a major component of the NIH Roadmap.

A Board of Scientific Counselors, comprised of extramural scientists, meets twice a year to review the research activities of NCBI. The high caliber of the work of the CBB group is evidenced by the number of peer-reviewed publications: over 80 this year with more in press. The staff also participate in numerous presentations at various scientific meetings, molecular biology research companies, and at universities worldwide as well as at other government institutes. They also give presentations at NCBI's weekly Computational Biology Branch Lecture Series, which often hosts outside speakers.

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.

Outreach and Education

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

Education: NCBI Courses

In response to an ever-increasing demand for education and training in the use of the growing diversity of NCBI's products and services, the course "A Field Guide to GenBank and NCBI Resources" was expanded; it is taught at NIH and throughout the U.S. upon request. The course consists of a three-hour lecture, a two-hour hands-on practicum, and optional one-on-one sessions. An expanded two-day course titled "Enhanced Field Guide" provides extended and in-depth coverage of BLAST, structure, and genomic resources with an advanced hands-on session included. The 11-member teaching staff presented 58 courses to over 5,000 people in the past year.

The course titled "Exploring 3D Molecular Structures Using NCBI Tools" premiered in FY2006. This course includes lectures and computer workshops on effectively using NCBI databases, search services, and analysis tools that focus on 3D macromolecular structure data. The course has been successful in its first year, being taught 14 times at various venues to over 350 participants.

Education: Mini-Courses and Lecture Presentations

NCBI offers 11 bioinformatics mini-courses at NIH and outside institutions to provide a practical introduction to various resources. The two-hour courses are both problem-based and resource-based and include a review and hands-on session. This year, over 100 mini-courses were offered to approximately 3,200 participants.

Education: Technical Workshop Series

The PowerTools NCBI Technical Workshop Series consists of three courses lasting three days each. "NCBI Power Scripting" includes lectures and workshops on effectively using the NCBI Entrez Programming Utilities (eUtils) with scripts to automate search and retrieval operations across the entire suite of Entrez databases. The "NCBI 4-Pack" course provides information on practical applications of bioinformatics resources. "Programming with NCBI BLAST" presents effective uses of BLAST within scripts, creating and maintaining local BLAST databases, and setting up a BLAST Web server. Each course is offered quarterly at the NCBI Training Center, with approximately 80 participants in FY2006.

Education: Bioinformatics Training

To help NIH researchers make optimal use of computer science and technology to address problems in biology and medicine, the NCBI has a network of bioinformatics specialists serving individual institutes within the NIH, called the Core Bioinformatics Facility (CoreBio). Individual CoreBio Members are trained over a nine-week period in the use of NCBI bioinformatics tools. The CoreBio Members, in turn, advise researchers within their respective institutes as to the best methods for conducting

their bioinformatics analyses. Information exchange among the CoreBio Members and the NCBI faculty is facilitated by regular meetings and e-mail forums.

CoreBio has trained representatives from 15 research institutes at NIH and has conducted eight nine-week training programs (two in the past year) since the program began in 2001. Thirty-three update sessions and two special topic sessions for the institute representatives have also been held. One-on-one consultations with NCBI faculty are available to NIH scientists on an ongoing basis in the NCBI Learning Center, the NIH Library, and the NCI-Frederick Cancer Research and Development Center.

Education: Extramural Educational Collaborations

The fifth "NCBI Advanced Workshop for Bioinformatics Information Specialists" was held during FY2006. The educational collaborators (Educollab) work with NCBI to develop and teach the course. The course alumni offer a variety of year-round services at their universities, including workshops on NCBI resources, individual research consults and support, and Web portals. Many of the workshops are based directly on materials presented in the Advanced Workshop, thereby extending the impact of those materials.

Together, the collaborators and course alumni form the growing Bioinformatics Support Network (BSN), a group supported by NCBI which has been established for the purpose of communication and continuing education among members. In FY2006, eight new members joined the BSN as a result of the workshop. BSN members, and in particular Educollab members, are leaders in their field and have significant influence on the growing movement of medical libraries to establish high quality bioinformatics support programs. The Educollab group began the process of converting instructional slides for the advanced workshop from HTML format to PowerPoint format, and the Educollab coordinator re-engineered parts of the Web site to adapt to the new format. This approach allows revisions of the course materials by a diverse group of contributors, yet preserves the essential components of the course Web site to facilitate access by library staff nationwide.

The regional training program for the three-day introductory course continued in FY2006. Eight of NCBI's educational collaborators served as regional instructors to present the course at five locations across the country. A total of 85 individuals attended these regional courses and the on-site course offered at NLM. The purpose of both the introductory and advanced workshop, as well as the Educollab program, is to train the trainers, who then provide assistance with NCBI resources to thousands of end-users across the country. The Web-based materials that exist for both courses also serve as a reference for those who have taken the courses.

Members of the Educollab group wrote a series of eight articles on building the role of medical libraries in bioinformatics, which were published in the July 2006

focus issue of the *Journal of the Medical Library Association*. The series concluded with invited papers by Roberta Pagon, M.D., and Joyce Mitchell, Ph.D., on genetics information resources for the clinical and consumer health communities, respectively. The series provides a synthesis of issues and a cross-section of models that can help libraries to establish bioinformatics end-user support programs, or to enhance existing ones.

Outreach: User Guides for NCBI Resources

NCBI continued to develop a comprehensive list of fact sheets that outline its services and databases. 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. Interactive tutorials may be found for a number of databases and search and retrieval tools, such as Entrez, PubMed, Structure, and BLAST. Many "Announce" e-mail lists give users the opportunity to receive information on new and updated services and resources from NCBI. For example, "NCBI Announce" provides updates on all NCBI services and education while "Books Announce" provides information regarding the Bookshelf.

     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 services and available genomes, NCBI investigator profiles, and a bibliography of recent staff publications. In FY2006, over 18,000 subscribers received printed copies of the *NCBI New*. Access to the newsletter and its archives via the NCBI Web site has increased dramatically as more people have become aware of its availability online.

Biotechnology Information in the Future

The Discovery Initiative is under way to enhance the integration of resources on NCBI's Web site, thus improving access and usability of its resources. Methods of linking are being redefined in order to provide more useful related information that is more easily visible.

     The commitment to providing the scientific community with both the resources and tools needed to fully explore 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 human and 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 that will enable discoveries that significantly impact health in the 21st century.

EXTRAMURAL PROGRAMS

Milton Corn, M.D.
Associate Director

The NLM Extramural Programs Division (EP) continues to receive its budget under two different authorizing acts: the Medical Library Assistance Act (MLAA, unique to NLM), and Public Health Law 301 (covers all of NIH). The funds are expended mainly as grants-in-aid, and in some instances as contracts, to the extramural community in support of the Library's goals. Review and award procedures conform to NIH policies. The EP Web site at http://www.nlm.nih.gov/ep/funded.html lists grants awarded since 1997, with links to abstracts provided in the NIH CRISP (Computer Retrieval of Information on Scientific Projects) database and, when available, links to project Websites.

Historically, EP has issued grants in a broad variety of programs, all of which pertain to biomedical computing, informatics, and information management with the exception of the Scholarly Works program. Some grant programs are issued by NLM, while others are multi-Institute initiatives or interagency partnerships. NLM makes awards in a number of different grant categories:

- Resource grants for information management, often involving medical libraries;
- Training, fellowship, and career development grants for informaticians and informationists;
- Research grants for fundamental research in biomedical informatics, information science, and bioinformatics;
- Research Resource grants to support unique research tools for informatics and bioinformatics;
- Scholarly Works and Conference grants to enhance scientific and scholarly communication;
- SBIR (Small Business Innovation Research) / STTR (Small Business Technology Transfer Research) grants to support informatics innovations in small businesses; and
- Multi-Institute grant announcements and interagency collaborations.

A significant feature of EP activities for FY2006 has been a retrenchment in the number and thrust of EP's offerings. In FY2006, EP focused on counteracting falling success rates, both by closing programs and by suspending participation in selected cooperative ventures. Falling success rates for research grant applications have become a serious problem for all NIH Institutes, despite the recent doubling of the NIH budget. Reasons for the decline are complex and vary among the Institutes, but for NLM, the principal cause is a significant increase in applications at a time of flat budgets.

Success Rates

In FY2006, success rates mirrored those in 2005, continuing to fall in NLM's core resource grant programs. Table 11 shows success rates between 2003 and 2006 for NLM's core grant programs. The success rate for research grants was stabilized. As they did in 2005, award decisions in 2006 continued to favor early career support. These applicants are most often trainees from NLM's informatics training programs who are now moving into independent research careers. Success rates are computed by dividing the number of awards by the number of applicants in a fiscal year.

Two major factors continue to shape success rates at NIH: the increased number of applications received in the face of essentially flat budgets. Table 12 shows the steep increase in the number of applications received for a five year period. (In 2003, the budget doubling period ended at NIH.) Success rates for some programs could improve in FY2007 if previously awarded grants end and are replaced by less costly projects. However, application trends toward
multiple principal investigator grants and larger consortium awards, as well as normal inflationary trends in research costs, make it difficult to lower average cost of awards. The combination of flat or even reduced budgets as applications increase inevitably reduces success rates to a level that could discourage the applicant community, and, over time, diminish academic support and health of the informatics professoriate.

Table 11. Success Rate, Core NLM Grant Programs

Grant Type	FY 2003	FY 2004	FY 2005	FY 2006
Research	22%	15%	9%	10%
Knowledge Management	32%	18%	12%	8%
Scholarly Works	32%	26%	19%	15%
Career Transition	N/A	32%	36%	37%

Table 12. Applications and Awards, FY 2002–2006
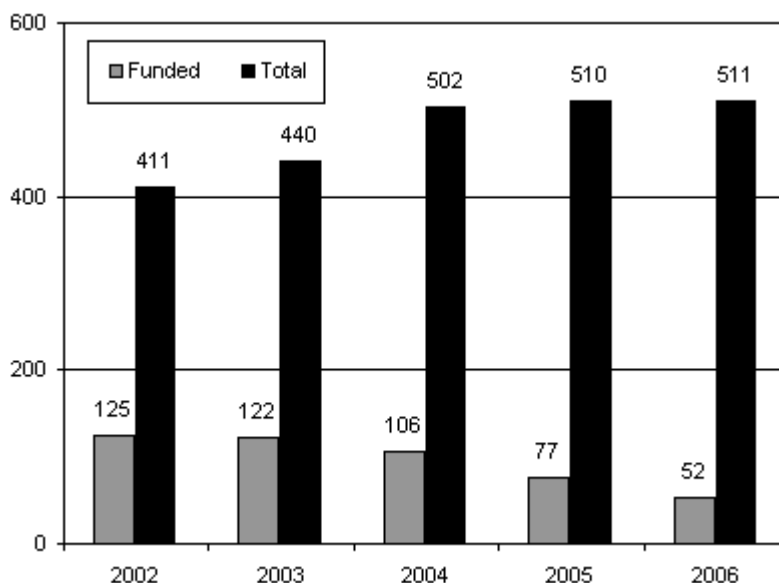
Research Support for Biomedical Informatics and Bioinformatics (PHS 301)

Extramural research support is provided through a variety of grant mechanisms that support investigator-initiated research. EP'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.

Research Grant Program

The R01 research grant program has two "branches": biomedical informatics (computing and knowledge management for clinical care, health services research, and public health), and bioinformatics (computing and knowledge management for basic biomedical research areas such as systems biology, genomics and proteomics). All research grants are funded from PHS 301 funds. EP reviewed 119 applications for this program in FY2006, compared to 107 applications in FY2005. Twelve awards were made. At the request of the applicant, one additional award was deferred for funding until FY2007. Example titles of new research grants awarded in FY2006: Vocabulary Support for Consumer Health Informatics; Beyond Abstracts— Issues in Mining Full Text; and Efficient Analysis of SNPs and Haplotypes with Applications in Gene Mapping.

The average score of awarded applications was 148, compared to 147 in FY2005. Achieving a stable award level is one of EP's strategies for stabilizing success rates in the research grant program. Another strategy is working more closely with applicants on new and amended applications. While many of EP's research grant applications come from the biomedical informatics research community, an increasing number come from computer science, engineering, and basic biomedical science fields.

In FY2007, EP will adopt a new strategy for seeking innovative, high-impact informatics research applications. First, as NIH moves to electronic receipt of applications for research grants, EP will abandon its NLM-only informatics research grant announcement and participate instead in the NIH "generic" research grant announcement. NLM's specific interests will be described on the EP Web site, allowing greater flexibility for changing priorities more quickly. Second, EP intends to issue the first of its NLM Challenge Grant announcements to support fundamental informatics research in focused high-priority areas.

Small Grant Program

In 2003, EP began offering the R03 small research grant, which provides modest support for start-up research projects and pilot studies. Twenty-seven R03 grants were reviewed in FY2006, compared to 36 in FY2005. Only three new grants were awarded. The average priority score of funded R03 grants was 156, compared to 163 in FY2005. The low success rate of applications to this program has been true for all the Institutes, and remains puzzling. Whether the applications are generally truly poor, or are, perhaps, being judged by overly strict criteria (that is, by criteria suited to the R01 full-blown research grant) is not known. Example titles of Small grants awarded in FY 2006: Identifying Genetic Factors for Predisposition in Polygenic Diseases; Optimizing Therapeutics for Hospitalized Patients with Impaired Renal Function.

Exploratory/Developmental Grants

EP's R21 exploratory/developmental grant supports high risk/high yield projects, proof of concept, and work in new interdisciplinary areas. This grant mechanism seemed better suited to informatics/engineering proposals than are the standard R01 research grants, which are judged in terms of hypothesis-based science. Success rates continue to be poor in this program. Many applicants treat it as a "small R01" grant, misunderstanding the purpose and, hence, suffering in review. In FY2006, 44 applications were reviewed, compared to 33 in FY2005. Success rate in FY2006 was 7%, compared to 10% in FY2005. The average score of awards was 156, compared to 168 in FY2005. Only electronic applications are now accepted in this program. For this relatively new grant mechanism, as for small grants, success rate has been low across NIH. Example titles of Exploratory/Developmental grants awarded in FY 2006: Extracting Reliable Information from Microarray Data; Managing ED Diversion with Information Technology.

Informatics for Disaster Management

Beginning in 2002, EP offered an R21 exploratory/developmental grant program exploring the application of informatics approaches in natural and man-made disasters. Budget constraints and poor success rates led to discontinuation of the program in FY2006. EP continues to receive grants relating to this topic through its other research and resource grant programs. In FY2007, EP will consider issuing a special Request for Applications on this topic, possibly in a narrow high-impact area. The final two awards in this program were made in FY 2006: Informatics for Empirically Supported Influenza Pandemic Control Strategies; Improving the Trauma System Response to Disaster.

Resource Grants for Biomedical Informatics/Bioinformatics

In August 2004, EP re-issued a P41 program announcement to support scientific research resources, a type of support grant offered by NLM for many years These resource applications are extremely expensive, often continue for decades as important community resources, and therefore tend to consume relentlessly the funds available for research grants.

EP believes the increasing requests for grants to support a myriad of essential research tools requires resources from multiple Institutes. Accordingly, the NLM program announcement was deactivated in FY2005. Only the six existing P41 awardees remain eligible to apply to the program for continuation funding. Two such continuation applications were reviewed in FY 2006, of which one was funded: REBASE: Restriction Enzyme Database.

Conference Grants

Support for conference and workshops is offered by almost all the Institutes, and for NLM is intended to provide relatively small amounts to scientific communities convening workshops and meetings in focused areas of biomedical informatics and bioinformatics. Applicants must obtain approval from EP program staff before they can apply. Of three applications reviewed in FY2006, one was funded: Annual UT-ORNL-KBRIN Bioinformatics Summit.

Integrated Advanced Information Management Systems Testing & Evaluation Grants

An updated program announcement was issued for this R24 grant program in March 2005. Although part of the IAIMS grant program, the Testing and Evaluation grant is considered a research grant and funded through PHS 301. Eight applications were reviewed in FY2006, compared to six applications in

FY 2005. One was funded: Evaluation of an Evidence-Based Program Planning System.

SBIR/STTR (PHS 301)

All NIH research grant programs allocate fixed percentages of available funds every year to Small Business Innovation Research (SBIR) and Small Business Technology Transfer Research (STTR) grants. These projects may involve a Phase I grant for product design, as well as a Phase II grant for testing and prototyping. SBIR and STTR applications are reviewed by the NIH Center for Scientific Review (CSR). In FY2006, there were 62 SBIR/STTR applications (58 of them Phase I applications) assigned to EP and reviewed by CSR. Of these applications, 46 were "unscored," indicating reviewer assessment that they were not competitive for funding. Two awards were made: Medical Emergency Disaster Response Network; Software Leveraging a Standards-Based Web Service Framework for Decision Support.

Resource Grants (MLAA)

Resource Grants, authorized by the Medical Library Assistance Act, support access to information, connect computer and communications systems, and promote collaboration in networking, integrating, and managing health-related information. Three of the four Resource Grant programs are centered on optimizing the management of health-related information; they are not research grants and are reviewed with relevant criteria. The fourth program, grants for Scholarly Works, supports the preparation of scholarly manuscripts in health sciences and health public policy areas.

Internet Access to Digital Libraries Grants

This grant program expired after the initial RFA in 2002 but was left open informally through FY2004. Because of overlap with other EP grant programs and with activities of the NN/LM, this program was officially deactivated in FY2005. However, several new applications and some amended applications were in the pipeline for review at the time of deactivation, and these were reviewed in FY2006. Of them, one was funded. No additional awards will be made in this program.

Knowledge Management and Applied Informatics Grants

This program is a refocused continuation of NLM's former Information Systems Grant program. The new program emphasizes knowledge management and application projects that translate informatics research into practice. EP reviewed 65 grants in FY2006, compared with 57 in FY2005. Five awards were made, and one was deferred for funding in FY2007. The average score of awards was 154.

Integrated Advanced Information Management Systems Planning Grant

Minor modifications, including an explicit outline of the expected deliverable, were made to the IAIMS planning grant announcement in 2005. Twenty-one planning grants were reviewed in FY2006 compared to 12 in FY2005. One new award was made, and two awards that were deferred from FY2005 were also made. Applications are expected to drop sharply in FY2007 due to the deactivation of the IAIMS operations grant program.

Integrated Advanced Information Management Systems Operations Grants

A new program announcement was issued for IAIMS operations grants in March 2005. Due to intense national interest in the installation of electronic health records, many calls were fielded, but no viable applications were received in FY2005. Due to the poor success rates, applicant confusion about the purpose of the program, and budget constraints, the IAIMS operations grant program was deactivated in FY2006. Eleven applications were in the pipeline and were reviewed, but none was funded. NLM's KM & AI grants can provide a small amount of funding to support IAIMS operations projects.

Grants for Scholarly Works

The Scholarly Works grant program continues to be very popular. EP reviewed 68 applications in FY2006, compared to 52 in FY2005. Ten new awards were made, and one was deferred for payment in FY2007. The average priority score of new scholarly works grants funded was 140, compared to 142 in FY2005. Because NLM, alone among the Institutes, is authorized to support book publications, this program

continues to play a key role in important areas of biomedical scholarship, particularly in the history of science and medicine. Example titles of Scholarly Works grants funded in FY2006: Handbook for Rural Health Care Ethics; Chinese Healing and the United States: 1849–2004; and Sending for Nurses: Importing Care to U.S Hospitals.

Training and Fellowships (MLAA)

Exploiting the potential of information technology to augment health care, biomedical research, and education requires investigators who understand biomedicine as well as fundamental problems of knowledge representation, decision support, and human-computer interface. NLM remains the principal source of support nationally for research training in the fields of biomedical informatics. EP provides both institutional and individual training support.

EP-Supported Training Programs

Five-year institutional training grants support pre-doctoral, post-doctoral, and short-term trainees in 18 programs across the country. Seven of these programs were funded for the first time in 2002, while 11 are continuations of previously funded programs. Three of the 18 are consortial programs that support training at multiple institutions. EP supported 274 informatics trainees at the 18 training programs with FY2006 funds: 161 pre-doctoral, 97 post-doctoral and 16 short term trainees. Collectively, the programs emphasize training in clinical informatics, bioinformatics and computational biology, and public health informatics.

EP receives some co-funding from the National Institute of Biomedical Imaging and Bioengineering to support training related to advanced imaging methods at UCLA, and the National Institute of Dental and Craniofacial Research supported training in dental informatics at Pittsburgh and Columbia.

After conducting site visits to all funded training sites to evaluate NLM-supported training across the nation, EP staff issued a Request for Applications (RFA) in January 2006. This program is re-competed every five years. A technical assistance teleconference was held on January 26, 2006. Following that meeting, an FAQ was posted on the EP Web site. Applications were due March 17, 2006, and were reviewed in May 2006. Thirty-six applications

were received. Of those, 17 received priority scores between 100 and 200. Although these grants will not be awarded until FY2007, award decisions were communicated to applicants at the end of FY2006, to inform recruitment plans. Eighteen awards will be made, two of them to new programs, at the University of Colorado and at the University of Virginia. Two existing programs at the University of Minnesota and the Medical University of South Carolina will close. By NLM custom, all trainees already matriculated into the programs facing sunset will be supported until completion of their training. Funded sites:

1. University of California, Irvine (Irvine, CA)
2. University of California, Los Angeles (Los Angeles, CA)
3. Stanford University (Stanford, CA)
4. University of Colorado Health Sciences Center (Aurora, CO) [NLM funding begins in 2007]
5. Yale University (New Haven, CT)
6. Indiana University - Purdue University at Indianapolis (Indianapolis, IN)
7. Harvard University (Medical School) (Boston, MA)
8. Johns Hopkins University (Baltimore, MD)
9. University of Minnesota Twin Cities (Minneapolis, MN) [NLM funding ends in 2007]
10. University of Missouri-Columbia (Columbia, MO)
11. Columbia University Health Sciences (New York, NY)
12. Oregon Health & Science University (Portland, OR)
13. University of Pittsburgh (Pittsburgh, PA)
14. Medical University of South Carolina (Charleston, SC) [NLM funding ends in 2007]
15. Vanderbilt University (Nashville, TN)
16. Rice University (Houston, TX)
17. University of Utah (Salt Lake City, UT)
18. University of Virginia (Charlottesville, VA) [NLM funding begins in 2007]
19. University of Washington (Seattle, WA)
20. University of Wisconsin–Madison (Madison, WI)

Every summer, all NLM-supported trainees attend a national informatics training meeting. On June 27–28, 2006, the two-day meeting took place at Vanderbilt University in Nashville, TN. Research projects were presented in plenary and semi-plenary sessions by 39 informatics trainees. An additional 44

trainees presented posters at the meeting. There were 350 attendees, including directors, faculty, and staff from all NLM-funded informatics training programs; holders of NLM individual fellowships; faculty and trainees from the Veteran's Administration informatics training sites; and NLM staff and guests. Aggressive actions by NLM during the 1990s to increase training for bioinformatics in these traditionally clinic-centric programs have been remarkably successful. Approximately half of the trainee papers presented at Nashville were on bioinformatics themes.

In 2005, NLM/EP and the Robert Wood Johnson Foundation (RWJF) formed a partnership to lend increased emphasis to training in public health informatics. Through a $3.6 million grant from the Foundation to EP (through the Foundation for NIH), four existing training sites received supplemental awards to develop formal training tracks in public health informatics and to support trainees in these tracks. The four selected sites were Columbia, Johns Hopkins, Utah, and Washington. Trainees in this initiative meet twice each year. The first meeting was held at the fall AMIA meeting in November 2005. In June 2006, the group met again in Nashville for a "Thinking Outside the Box" exercise. NLM and RWJF staff work collaboratively to plan these events.

Individual Fellowships

Informatics Research Training: For a number of years, EP offered two types of fellowships for informatics research training: an individual fellowship for basic or applied research (F37), which can be pre- or post-doctoral, and a senior fellowship intended for those with seven or more years of professional employment experience in an appropriate field (F38). As part of its strategy to stretch the flattened grant budget, and because its 18 university-based training programs perform the same function, EP closed both informatics fellowship programs in FY2006 Eight applications were received for the F37 informatics fellowship program, of which one was funded. Three applications were reviewed for the F38 program, and no awards were made.

Training for Informationists: In October 2003, EP issued program announcements for two new fellowships to support the training of in-context information specialists. These programs use the F37 and F38 mechanisms, but emphasize training for information specialist professional careers as opposed to research careers. In 2006, the F38 Informationist fellowship was closed due to budget constraints. One application was reviewed but not funded. The F37 Informationist fellowship program continues to receive and review applications, though the application rate is very low. Two applications were reviewed and one was funded. Since the informationist program's inception, EP has funded six informationist fellowships. Additionally, the Medical Library Association now has a program initiative in this area to provide continuing education and mentorship for librarians who wish to become in-context information specialists. If application numbers remain low, this program may be deactivated in FY2007.

Career Support (MLAA)

Early Career Development Awards: The K22 program was established to provide transition assistance with funds for salary and for research to biomedical informaticians who are establishing their initial independent research programs. EP reviewed 19 applications in FY2006, compared to 20 in FY2005. Seven awards were made. The average score of a successful application was 137, compared to 157 in FY2005, showing how keen the competition is for these awards. In January 2006, NIH announced a new career transition program, the NIH Pathways to Independence (PI) award (K99/R00), which combines a two-year mentored period with a three-year un-mentored research period (the latter being similar to NLM's existing K22 program). To make best use of funds and reduce confusion for interested applicants, NLM closed its K22 program and replaced it with the new PI award. Although applications to this new program are not restricted to NLM's informatics trainees, applicants from our own programs are particularly welcome. The first two PI awards were received and reviewed at the end of FY 2006; neither received a fundable score. Example titles of the K22 awards made in FY 2006: Protecting Genetic Privacy through Risk Assessment; Text Mining as a Translational Tool in Biomedicine; and Study & Development of a Comprehensive Decision Support System for Preventive Care.

Loan Repayment Program: EP participates in the NIH loan repayment program by identifying applications from informaticians involved in research related to

clinical medicine. These applications are reviewed for merit by a Special Emphasis Panel. For FY2006, EP funded six of 16 applications.

Pan-NIH Projects

EP and Roadmap Activities

A major pan-NIH enterprise initiated by the NIH Director is resulting in a battery of RFAs and RFPs related to three themes: New Pathways to Discovery, Research Teams of the Future, and Reengineering Clinical Research. EP is a participant in all of these Roadmap initiatives by fiat, but is functionally involved only with the subset that have biomedical computing components appropriate for the NLM mission. Generally speaking, these trans-NIH initiatives involve management teams of staff from all interested Institutes, but actual grants administration is done by only one or a few Institutes for each program. EP staff were particularly involved in NIH Roadmap teams for the National Centers for Biomedical Computing and the initiative related to interdisciplinary research centers. EP staff have served as preliminary reviewers for the NIH Director's Pioneer Awards, another NIH Roadmap initiative. In 2007, NIH Roadmap activities will be administered by the new OPASI office, which will administer the NIH Common Fund and decide on new initiatives that draw upon it.

NCBC and BISTI

Although conceptually related to the NIH Biomedical Information Science and Technology Initiative (BISTI) program, the National Centers for Biomedical Computing (NCBC) program is distinct and funded through NIH Roadmap grants under renewable cooperative agreements. The funds for the first five-year period of these grants came primarily from the NIH Roadmap initiative, but several ICs, including NLM, agreed to contribute an additional $800,000 per year for five years to the Roadmap pool so that more centers could be funded. Several other Institutes contributed additional funds as well. There are currently seven funded NCBC centers. Although NIH Roadmap grants are considered pan-NIH grants, each NCBC has a program officer and a lead science officer drawn from different Institutes, and these NIH staff serve on the NIH NCBC management team. EP administers one NCBC center, "Informatics

Integrating the Bench and Bedside," based at Harvard's Brigham and Women's Hospital. Roadmap funding for the first five years ends in FY2008; consequently, the NIH management team for the NCBC program has proposed a funding strategy for a second five-year period. It is not yet known whether Roadmap funds will be available and at what level of support. Planning is under way for a re-competition of the existing centers to provide funding for at least one additional five-year cycle.

Multi-Institute Grant Programs

In addition to its involvement in the NIH Roadmap, EP also participates with other NIH and federal organizations in a number of multi-agency grant announcements. Budget constraints and the importance of protecting NLM's own grant programs have increased EP's selectivity when considering the funding of applications to multi-institute initiatives; EP participation is confined to programs which do not duplicate its existing grant programs.

EP resigned from two multi-institute programs this year (NCBC Collaboration Grants and Continued Development and Maintenance of Software) to preserve funds for its core programs. As of 2006, the multi-institute programs NLM participates in are Understanding Health Literacy, several specialized SBIR/STTR programs, and both Research and Exploratory/developmental BISTI grants. The applications for these programs are reviewed by CSR, and then participating Institutes select grants for full or shared funding. These sources represent 5 to 10 percent of applications assigned to NLM. They are included in the listing for payline decisions. Links to the multi-institute initiatives in which EP participates are incorporated into the grant programs list on the EP Web site.

Shared Funding for Research & Training

EP contributed approximately $1 million under collaborative co-funding agreements with other NIH Institutes and Centers to 10 grants in FY2006. Three of them are new:
- Co-funding for the "Comparative Toxicogenomics Database (CTD)" ($200,000) which is being developed for public availability in order to promote understanding about the effects of environmental chemicals on human health.

- SBIR co-funding for the design, fabrication, and commercialization of a biological coating process, BioCoat^{TM} ($116,500). This system is important in the coating of biomedical devices such as implants, providing bioactive functions such as promoting bone growth or reducing bacteria.
- SBIR co-funding for the construction of a compact high-efficiency crossed-grid radiography machine that will control image scatter that provides higher detailed radio-images and reduce the number of x-ray images required of patients ($39,557).

NLM also receives co-funding from other Institutes for four of its active grants. In FY2006, the co-funded NLM grants included:
- Co-funding support received from the National Institute of Nursing Research (NINR) ($18,753) for R41 LM9051, an SBIR grant entitled "Software Leveraging, a Standards-based Web Service Framework for Decision Support."
- Co-funding support received from the National Institute of Allergy and Infectious Diseases (NIAID) ($83,945) for R01 LM9027, a research grant entitled "MSM: A Multi-Scale Approach for Understanding Antigen Presentation in Immunity." From the Multi-Scale Modeling Interagency initiative.
- Co-funding support received from the National Institute of Dental and Craniofacial Research (NIDCR) ($144,091) in support of dental informatics trainees for T15 LM7059, an institutional training grant to the University of Pittsburgh.
- Co-funding support received from the National Institute of Biomedical Imaging and Bioengineering (NIBIB) ($175,269) in support of imaging informatics for T15 LM7356, an institutional training grant to UCLA.

Interagency Agreements and Special Initiatives

EP continues to partner with the National Science Foundation (NSF) on a new grant program titled Dynamic Data-Driven Systems. EP provides funding for one grant in this program: Dynamic Data-Driven Brain Machine ($328,418 in FY2006). The project's goal is to develop computer architecture for computational modeling and training algorithms for robotic arm movement. A clinical application of this basic work relates to improving the brain-machine interface for paraplegics. For this partnership, the grant is administered by NSF, with EP providing adjunctive program officer oversight.

In the multi-agency Multi-Scale Modeling in Biomedical, Biological and Behavioral Systems, EP partnered with NSF, National Aeronautics and Space Administration, Department of Energy, and eight NIH Institutes. In addition to sharing costs for the grant review at NSF, EP funded one grant ($364,755) in this program, which was transferred from NSF to NIH. This grant, titled A Multi-Scale Approach for Understanding Antigen Presentation in Immunity, is administered by EP as an R01 grant. NIAID contributes 33% of the direct costs for this grant, which is now in its second of three years.

In FY2006, NLM finalized plans for a two-phase study titled Engaging the Computer Science Research Community in Health Care Informatics, to be undertaken by the National Academy of Sciences. NCRR and NIGMS contributed funds to the project in FY2006. The NIH portion of the project cost is estimated at $430,000 of which NLM contributed $320,000.

EP Operating Units

Grants Management Office

EP issued 191 grant awards in FY2006 for nearly $53 million. Staffing levels have been restored in FY2006. NLM provided key support for all grant co-funding agreements of the NLM, interagency agreements in support of grants, large scale training grants, and its general training and research grants split over two funding authorizations, the Medical Library Assistance Act and the PHS Act.

In addition, the NLM Grants Management Office provides full grants accounting support for the EP budget for all awarding mechanisms, updates the annual Catalog of Federal Domestic Assistance, and works closely with the NLM Freedom of Information Coordinator to provide timely response to FOI requests for grant information.

Program Office

Grant Program Development: Program activities in FY2006 were focused on (1) preparing and updating

EP's "proprietary" grant program announcements; (2) refining language for grant program announcements in which EP participates; and (3) initial planning for transitions to the new SF-424 electronic application form. By the end of 2007, all NIH grants will be submitted electronically, via Grants.gov. Each grant mechanism has a committee thinking through the features and restrictions for an online application. Some mechanisms (G08, G13, F37) are used solely by NLM and require extensive input to the committees.

Program Policy: EP program staff represent EP on various NIH standing program-related committees, including Extramural Programs Management Committee, Project Officer/Program Officer Forum, Training Advisory Committee, Human Subjects Protection Liaison Committee, Tracking & Inclusion Committee, Electronic Research Administration Program Officials Users Group, and Electronic Research Administration Population Tracking Users Group. In addition, EP program staff participate in NIH workgroups related to the trans-NIH Knowledge Management Disease Coding (KMDC) project. This initiative aims to create "fingerprints," collections of terms that can aid in the preparation of Congressional reports that show NIH's research investment in diseases, health disparities, and other high-priority themes. EP implemented processes for identifying which grants require human subject tracking. Program Class codes were updated twice.

Program Planning, Oversight and Evaluation: EP became the coordinating center for one of NIH's Government Performance and Results Act (GPRA) goals in FY2005. The three-year goal is titled CBBR-5: "By 2007, Expand by 5,000 the Pool of Researchers & Clinicians NIH has Trained in Biomedical Informatics, Bioinformatics, or Computational Biology." This activity requires three reports each year during the life of the goal, drawing together information from NLM extramural and intramural programs, NIBIB, NIGMS, and NHGRI. Having exceeded its targets in FY2005, EP and its partners set new goals for the FY2006 and 2007 reporting years.

EP program staff attended all working sessions of NLM Long-Range Planning committees and provided input on the final recommendations. On-site visits or reverse site visits were performed for the following funded activities: PDB (Protein Data Bank); BMRB (BiomagRes Databank); i2b2 National Center for Biomedical Computing; University of Cincinnati

IAIMS. EP initiated contracts with Humanitas, Inc. for two program assessment activities. One will compare and evaluate NLM's informatics training program graduates, and the other will evaluate achievements of NLM-funded fellows and R01-funded postdoctoral students. These contracts will begin in FY2007. In concert with the Training Activities Committee (TAC), EP prepared documentation on current evaluation strategies for its training programs.

Dissemination Activities: The following analyses were performed and presented to the NLM Board of Regents: Success rates of new vs. experienced investigators; five-year view of grant applications and awards; Summary of the K-22 program; Summary of SBIR and Conference grant programs. Formal presentations on NLM grant programs were made to the following groups: American Artificial Intelligence Association; Leadership program, AAHSL; Directors, NN/LM; SIS Native American Internship program; A panel presentation on the T-15 site visit evaluation was presented at fall AMIA in November, 2005. In January, EP again offered its three-day training curriculum to the NLM Associates. Program staff also participated in a funding opportunities roundtable and a panel presentation on national informatics research priorities at the Critical Issues in eHealth Research 2006 conference.

The EP Web site was updated with new grant awards for FY2006. All basic grants pages were restructured to provide a common look and feel across all programs, and to simplify maintenance of the site. Links to grantee project Web sites were added to many active grants.

The following EP grantees made presentations to the NLM Board of Regents and/or EP staff:

- Dr. Yves Lussier, University of Chicago: Collecting Phenotypic Information from Multiple Databases for Correlation with Genomics (February 2007)
- Dr. Patrick Jamieson, Logical Semantics, Inc: Medical Semantics Systems (March, 2006)
- Dr. Kathleen Walsh, University of Massachusetts: Pediatric Medication Errors and Computerized Orders (May, 2006)
- Dr. Benjamin Fregly, University of Florida: Computational Simulation of Joint Mechanics (September, 2006)

Scientific Review Office

BLIRC: EP's standing review group, the Biomedical Library and Informatics Review Committee (BLIRC), evaluates grant applications assigned to EP for possible funding for scientific merit. BLIRC met three times in FY2006 and reviewed 167 applications. The Committee (Appendix 5) reviews applications for most biomedical informatics and bioinformatics research applications, knowledge management/applied informatics, career support, and fellowships. BLIRC has two standing subcommittees: the Networked Information Access Subcommittee, and the Medical Informatics Subcommittee. The subcommittees review fellowship applications: informationist in the former committee, informatics in the latter. The charter of the BLIRC was amended to reflect 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.

Special Emphasis Panels (SEPs): Nineteen Special Emphasis Panels were held during FY2006. These panels are convened on a one-time basis to review applications for which the regularly constituted review group lacks appropriate expertise, when a conflict of interest exists between the applicant and a member of the BLIRC, or for certain special categories such as Conference grants. The Scholarly Works grants are routinely reviewed by a panel, due to the unique expertise needed. These panels reviewed a total of 285 applications during FY2006, including 36 T-15 informatics training center grant proposals.

The number of SEPs increased significantly in FY2006 in response to the increased number of applications assigned to NLM, and the limited ability of BLIRC to increase its already large load. The extra SEP burden markedly increased EP's review expenses partly because of the increased load, and partly because new business rules for reimbursing reviewers seem to be adding to the expenses of review.

Board of Regents/EP Subcommittee: A second-level peer review of applications is performed by the Board of Regents. 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 initial 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 Board also votes en bloc for all other applications that meet criteria for further consideration for funding.

Administration and Operations Office

In FY2006, EP installed its new electronic grants database, powered by software from ZyLAB. Throughout the fiscal year, EP successfully scanned its hard-copy grant files to electronic PDF files. The new electronic grants database will provide EP employees with quick access to past and future grant files from their desktops, as well as offset possible information loss due to natural or man-made disasters. It is hoped that the search capacities of the ZyLAB system will improve EP's ability to evaluate its applications and programs in detail.

Support for grants administration continues to be provided by four staff members from the NIH Division of Extramural Activities Support (DEAS) organization. In addition to staffing changes within DEAS, new practices were implemented for human subjects tracking, document scanning, adding program class code and program officer assignments. DEAS staff reorganized the grant files room and supply areas, and updated the administrative review listing to reflect current grant programs.

Table 13. Extramural Programs Budget FY2002–FY2006 (Dollars in thousands)
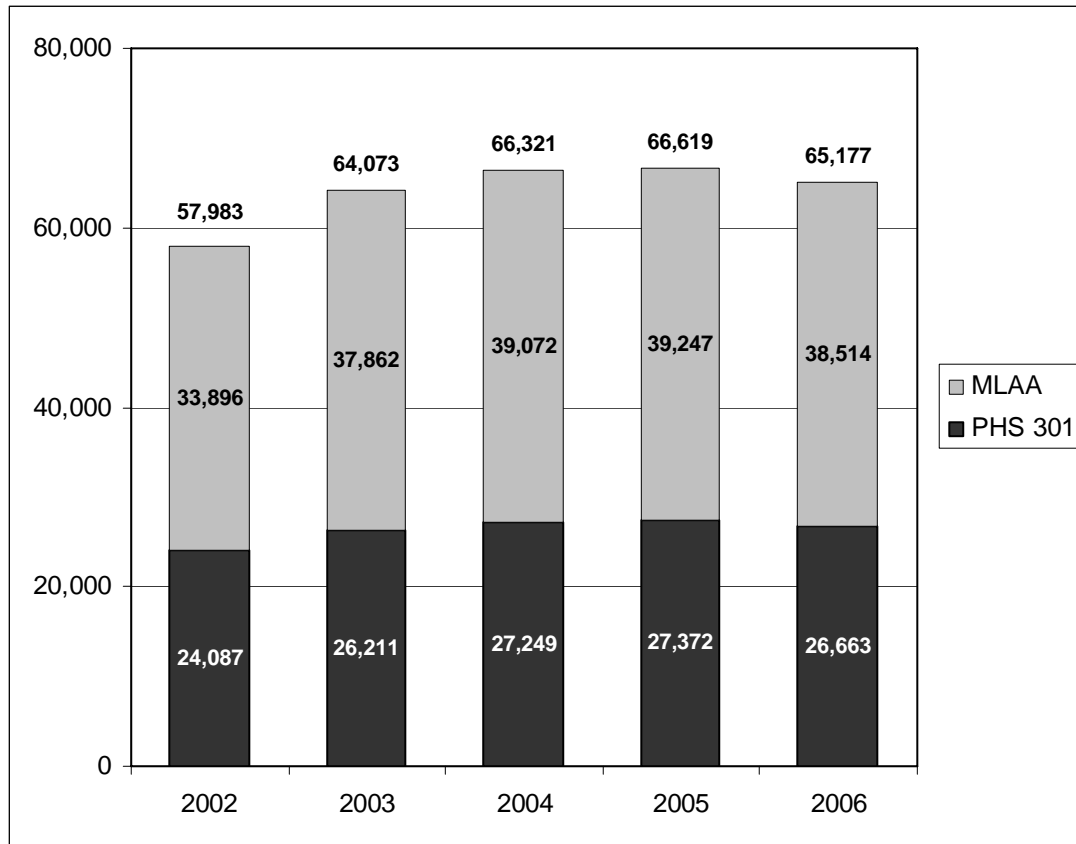
Table 14.   FY2006 Extramural Program Budget, by Function

Scholarly Works, $1,719,000 , 4%

Loan Repayment Program, $422,000 , 1%

Integrated Advanced Information Management Systems (IAIMS), $2,162,000 , 6%

Resource Grants, $3,324,000 , 9%

Training Programs, Fellowships and Career Awards, $18,186,000 , 47%

National Networks of Libraries of Medicine (NN/LM), $12,701,000 , 33%
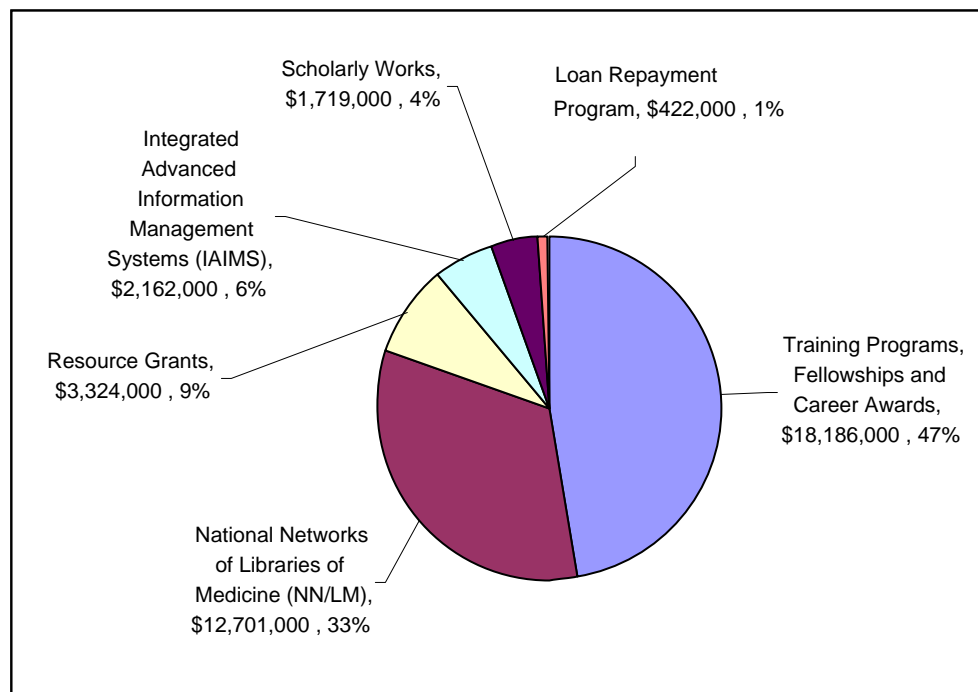
Table 15.   FY2006 Extramural Program Budget, Financial Resources and Allocations

SBIR/STTR - Small Business Grants, $784,000 , 3%

Bioinformatics Research Projects (incl. BISTI & Program Resource), $13,258,000 , 50%

Biomedical Informatics Research, $12,621,000 , 47%
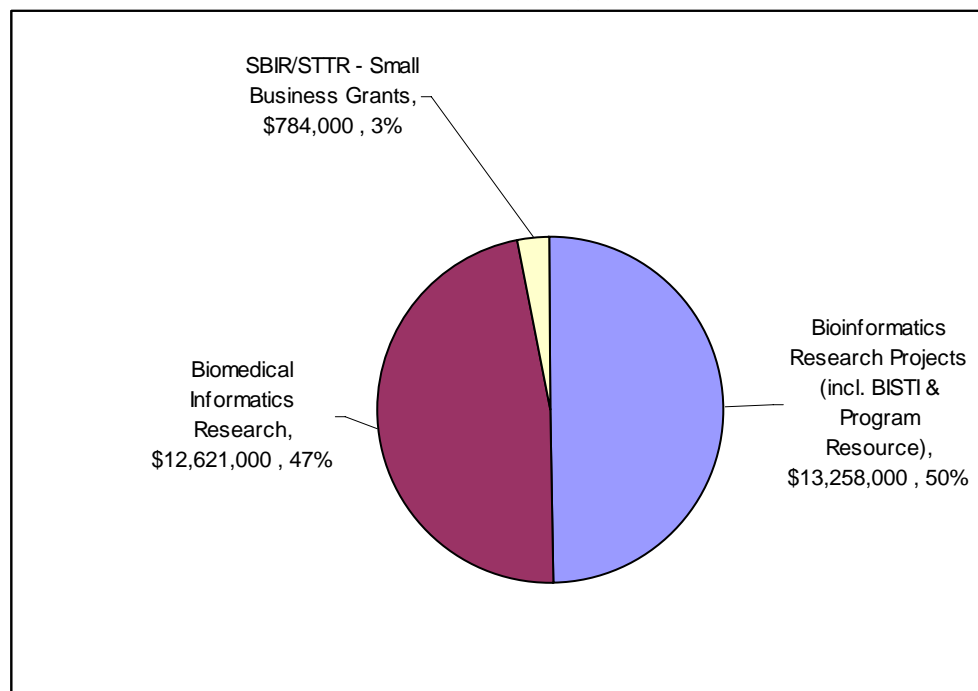
OFFICE OF COMPUTER AND COMMUNICATIONS SYSTEMS

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 Centers; 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 making up the 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.

The following describes OCCS accomplishments in FY2006:

Business Continuity and Disaster Recovery

In order to protect NLM's mission-critical systems, NIH's Center for Information Technology (CIT) and NLM have implemented an NIH Consolidated Colocation Site (NCCS) in Sterling, Virginia. The NCCS is operational with initial capabilities as a disaster recovery and load-balancing site. The NCCS serves as a disaster recovery/alternate computing site for NLM as well as CIT, NCI, NHLBI, NIDDK, NIAMS, OD/ORS and HHS/OS.

At present, all NLM mission-critical systems are either under active/active, active/passive or active/cold-backup mode depending on their business requirements The Business Continuity and Disaster Recovery Plan covers NCCS as the primary resource for system restoration and uninterrupted processing if the primary NLM computing facilities on the NIH campus are rendered unavailable by a disaster or other contingency. During this year, OCCS upgraded the load-balancers at the NCCS to provide increased network traffic capacity and more sophisticated traffic routing. OCCS also performed various other upgrades to the storage systems and servers located at this site.

The NLM computer facility has tripled its use of electrical power, cooling, and data transmission capacity over the last four years due to the dramatic growth in dependence on IT systems to deliver NLM's mission-critical applications. Recognizing that this rapid growth will continue in the years ahead, OCCS has begun a detailed reengineering process for evaluating the safety, reliability, and performance requirements of the computer facility. This year's reengineering efforts included:

- Designing and installing an overhead Ladder Rack in the computer facility, as a separate pathway for running data networking cables to improve the reliability, availability, and maintainability of data communication services. The Ladder Rack provides future cabling expansion while decreasing network cabling installation time.
- Securing an Onsite Alternate Computing Facility (OACF) to host redundant networking gear that will be used to provide an alternate path for connecting NLM networks to networks that are external to NLM (Internet, Internet 2, and NIHnet). The OACF will also be a redundant point for aggregating NLM's internal networks, and would come into play should a disaster strike the NLM computer facility.
- Developing plans for staging the installation of new cable trays for backbone fiber optic cabling throughout Buildings 38 and 38A. This is designed to provide alternate cable pathways in the event that a disaster destroys one path of the cabling.
- Developing plans for upgrading the High-Voltage Alternating Current system. This will increase power capacity by 50% from 480 kWh to 720 kWh.

Consumer Health

MedlinePlus: MedlinePlus's greatest achievement this year was the expansion of the MedlinePlus Go Local initiative, which brings local health services to the public by allowing users to search for healthcare providers in their localities while searching MedlinePlus for information on medical conditions or other health care issues. MedlinePlus Go Local implemented three releases this year, version 2.2 in January, version 2.3 in May and version 2.4 in July. These releases included the ability for Go Local users to have the capability of importing local data on their own as well as enhancements to the audit process. MedlinePlus added Go Local sites serving Utah, Wyoming, Maryland, East Texas, New Mexico, Texas Gulf Coast, Southern Ohio, Arizona, Nevada, Delaware, and Vermont.

There were also two versions of MedlinePlus released this year. Version 19 was released in January, and version 19.5 in May. These releases covered the public release of the new body map modules and the new LinkChecker module. A change in news providers and evidence-based herb and supplement information were also added in both English and Spanish.

This year NLM introduced MedlinePlus PodCasts, a weekly update by the NLM Director highlighting new materials on MedlinePlus. Users can listen to the audio or read the text version of the weekly transcripts. Since its implementation in June of 2006, there have been 23,864 unique visitors and 192,702 hits to the MedlinePlus PodCasts.

MedlinePlus realized a 27% increase in pages views with 820 million pages views and over 95 million unique visitors. Additionally, MedlinePlus and Spanish MedlinePlus received a score of 85 and an 83 respectively on the ACSI E-Government Satisfaction Index, a survey that tracks trends in customer satisfaction.

SeniorHealth Project: NIHSeniorHealth.gov is a joint project of NLM and the National Institute on Aging to provide health information on the Web using modes of delivery video and narration appropriate for older Americans. The system includes the Accent "Talking Web" module developed by OCCS to provide accessibility enhancements, including a selectable range of type sizes and spoken text.

With 30 topics now available in SeniorHealth, many new topics were added this year, including Chronic Obstructive Pulmonary Disease (COPD), Heart Attack, Heart Failure, High Blood Pressure, Osteoporosis and Paget's Disease of Bone. Videos on Cataract Surgery and Problems with Smell were added. SeniorHealth recognized over 1.1 million page hits and 850,700 unique visitors, with 22% being international visitors.

The Accent module received numerous enhancements, including better accuracy of unique phrases and words, such as medical terms or drug names, with their definitions stored in a lexicon dictionary file; the ability to concurrently have a page narrate while visual accessibility (contrast or font sizing) is changed by the user, and MPEG audio compression used to improve overall audio quality. Web site enhancements include new accessibility toolbar features which include asynchronous visual contrast and font size changes, Web page narration when speech is enabled, a new U.S. female voice with technical as well as voice quality improvements, and user preferences such as speech-enabling, contrast on/off, font sizing and their settings which are used when the browser is launched at the site for current and future sessions.

DailyMed Project: The DailyMed project is a partnership between the Food and Drug Administration (FDA), the Veterans Administration (VA), the NLM, medication manufacturers and distributors, and health care information suppliers. The project seeks to provide a standard, comprehensive, up-to-date, XML-based capability for labeling the contents of medications. This year, OCCS:

- Implemented a Really Simple Syndication (RSS) function which allows interested parties to subscribe to product label database additions and modification notifications.

- Integrated search results for PubMed, MedlinePlus, and ClinicalTrails.gov allowing users to see all published research materials for a given drug label.
- Integrated a Merriam Webster dictionary function providing user-friendly access.
- Added search results for the Lactation (LactMed) Database.
- Added more than 1,245 new drug labels.
- Recognized more than 96,500 unique visitors and over 210,000 page hits.

IT Security

NLM continued to assess and strengthen its security posture based on current business requirements and risk assessment. Security improvements continued throughout the year.

OCCS continues to perform a monthly cycle of vulnerability scanning, detection, and remediation thereby making concrete improvements in NLM's security posture.

Federal regulations mandate that systems be re-accredited every three years. This year, an independent contractor conducted the System Testing and Evaluation and developed Certification and Accreditation packages for MEDLARS and TOXNET. In August 2006, the project was completed and both systems were found to be securely designed and operated. The inspectors recommended "Authority To Operate" for both systems.

Due to the increase of new vulnerabilities and the rapid emergence of associated threats, OCCS must not only deploy more software patches than ever before, but must do so with a much greater degree of urgency. NLM's automated patch management program applied over 100,000 patches on commodity desktops this year fixing known vulnerabilities to software.

NLM implemented new HHS issued Minimum Security Configurations Standards for Departmental Operating Systems and Applications. The HHS Minimum Security Configuration Standards were created as part of the HHS Information Assurance and Privacy Program. Adhering to these standards will provide a baseline level of security, ensuring that minimum standards or greater are implemented to secure the confidentiality, integrity, and availability of NLM resources.

The combination of legislative requirements and the growing sophistication of attacks leaves OCCS with the unenviable task of protecting NLM Web applications and their associated data from compromise. Therefore, NLM acquired AppScan®, a Web application security testing suite, which provides solutions for all phases of security testing across the software development lifecycle. It will help ensure the security and compliance of NLM Web applications by discovering application vulnerabilities.

The Office of Management and Budget requires that HHS computer users complete annual IT security awareness training. NLM has completed 100% of the mandatory FY2006 Security Awareness Training for employees, contractors, and fellows.

Professional Health Information

Unified Medical Language System (UMLS) Project: The Unified Medical Language System provides a common, concept-oriented medical vocabulary and thesaurus based on more than 118 current medical source vocabularies in 17 languages. During FY2006, there were four versions of the UMLS Metathesaurus released. These releases included updates from 69 source vocabularies, six translations of significant vocabularies and a new UMLS licensing system. In Version 2006AA, NLM began changing how it represents the data it occasionally creates to provide consistency and usability of some sources. The UMLS Metathesaurus contains 1.35 million concepts, 5.3 million concept names and 21.9 million relationships. OCCS also:

- Transitioned ahead of schedule by four months;
- Reduced production time from six to four weeks;
- Streamlined QA process thereby reducing QA time by 25%;
- Established a mature software development environment; and
- Provided 24x7x365 information services.

RxNorm Project: Four versions of the RxNorm Editing System were released in FY2006: Version 3.5 in December, 4.0 in March, 4.1 in May and 4.3 in August. Enhancements in Version 3.5 included the ability to assign several terms to the same Semantic Branded Drug or Semantic Clinical Drug and system access control to prevent unauthorized changes to data. Enhancements to Versions 4.0, 4.1 and 4.3 consisted of preventing reformulation of drugs, providing type-ahead drop down for ingredient matches and the ability to mark NDC code attributes on source asserted atoms. There were 11 monthly releases to RxNorm this year and inversion and insertion activities consisted of including the Multum and Medi-Span data sources and the FDA electronic labels data. Four resynchronizations were completed which included the integration of several new scripts to accommodate the new functionalities created. RxNorm is resynchronized with UMLS data with each UMLS release. The RxNorm application currently contains 17,691 Branded Drug RxNorm Forms and 30,423 Generic Drug RxNorm Forms for a total of 48,114 RxNorm Form drugs.

Data Creation and Maintenance System: The major event this year for the Indexing Data Creation and Maintenance System (DCMS) was the baseline extraction. Due to an improved version of the Java-based DCMS extractor, the baseline extraction was completed in a record 17 hours, a 70% reduction from FY2005's 32 hours, even while realizing a 25% increase in citations. OCCS also:

- Upgraded the Citation Matcher to use the DCMS extractor;
- Added a new GEO (Gene Expression Omnibus) to the valid list of Databanks available in DCMS;

- Revised the Indexer Inventory and Reviser Inventory functions to provide users with more information;
- Expanded the Valid Publication Years range to include the eighteen hundreds; and
- Added the new ISRCTN databank to enhance productivity.

Medical Subject Headings (MeSH) and Related Systems: MeSH includes an interlingual database of translations and a system for extending and maintaining them, namely the MeSH Translation Maintenance System (MTMS). During FY2006, several new features were implemented for MTMS which included the generation and distribution of new XML files for the French, Italian, and German translations and reorganization of the MTMS Archive System which will preserve historical versions of various translations. Additionally, modifications implemented for the GCMS System included new search filters and changes in the subsystem that provides YEP reporting, a new Archive Report System for viewing previous years' data and the development of the Publication Type Maintenance System.

DOCLINE: DOCLINE, the NLM interlibrary loan (ILL) system, supports approximately 3,600 domestic and international libraries in processing more than 2.5 million interlibrary loan transactions a year. Four versions of DOCLINE were released this year, Version 2.6 in November, 2.7 in March, 2.7.1 (a maintenance release) in April and 2.8 in June. Version 2.6 gives libraries using ILL management software the option to use the ISO ILL Protocol to communicate between their ILL management software and DOCLINE. Version 2.7 introduced several major enhancements for the users which included a new field to indicate whether library supports urgent patient care requests and added a routing limit for EFTS Participants. Version 2.8 added validation on identification field to prevent libraries from prompting users for confidential information including VISA, Social Security number, credit card, etc. Over 45 DOCLINE enhancements were implemented during the fiscal year.

Voyager Integrated Library System (ILS): This year, baseline extracts of all data in Voyager were produced in XML format. Authority data, also produced in XML format, was implemented. This data is used by NCBI to load their Entrez Life Sciences Search system. Also implemented in January was the Unicode version of ILS Voyager 2003. This required extensive changes to the Impromptu catalog and reports. Base pulls containing all data in Voyager were produced in MARC21 format.

CIT Video-Cast Project: The NIH Video-Cast subsystem provides for bi-directional interchange of data between the NLM Voyager system and the CIT video streaming system. This year, an enhanced search capability was implemented through Voyager providing users with the mechanism for

accessing over 1,600 unrestricted past-events videos. Over 60 new videos are added each month.

Relais: NLM uses the commercial off-the-shelf Relais system for electronic document delivery and Interlibrary Loan management. Documents requested via DOCLINE are scanned and automatically delivered using the borrower's requested delivery method. The Internet address of the Relais Ariel standalone computer has been changed in order to improve security. Ariel is a system that delivers ILL requests as files over the Internet.

Scan Track (PubMed Central Inventory): PubMed Central (PMC) is NIH's free digital archive of biomedical and life sciences journal literature. Several enhancements were made this year to include, increased functionality for invoicing, new fields in "Create Lots," a way to add journals to ScanTrac before they appear in the Serials Extract File, and the ability to search and retrieve only certain years of titles selected in the Create Lots module. In addition, data was entered for 310 journals and issue tracking information for 51,000 issues.

Literature Selection Technical Review Committee (LSTRC): In FY2006, several modifications were made to the MEDLINE Review application, which is used to review journals for inclusion in MEDLINE. Program modifications were added to save more information from users' logins and submitted forms. New data fields and buttons were also added to aid in the retrieval of form data that has been removed.

Serials Extract File (SEF): Among numerous upgrades and fixes, the List of Serials Publication and the List of Journals Publication for the annual publication of SEF data was completed. Change requests for Citation Queue, Serial Name, PMC Holding, and Call Number were integrated; modifications were made in several SEF processes in order to compliment the new Voyager release; and a new translation process was implemented which converts MARC format to UTF-8 encoding in the SEF.

NLM Classification System: The NLM Classification System allows public and institutional access to the NLM Classification and related services and includes a Classification Editor. Publication of printed editions ceased with the 5th revised edition, in 1999. It contains approximately 2,400 pages. This year, a PDF version of this document was generated using Java/FOP programming.

Network and Systems Support

OCCS continued to provide reliable LAN and Internet communications services, meeting the data communications needs for new IT systems, providing security services as well as end user assistance and training, implementing new network-based applications and operating systems, and exploring new technologies and plans to meet NLM's continued growth in networking and communications. OCCS took steps to increase the capabilities and reliability of network services and storage by providing for the following:

- NCCS data communications services;
- Enhanced network monitoring and management;
- Increased IT and network security;
- New networked services to support the NLM user community;
- Additional redundancy to eliminate single points of failure;
- Enhanced backup for use in disaster recovery and daily recovery scenarios; and
- Expanded and efficiently centralized shared data storage.

Public Internet connectivity services to NLM are provided through a contract with Level3. Internet connectivity is provided via an OC3 (155 Mbps) circuit to the Level3 network node in McLean, VA. The contract also provides an OC3 link for CIT/NIH to the Level3 network. NLM and NIH collaborate in using these diverse connections to back up each other's Internet connectivity. The service features an automatic failover in the event of a scheduled or unscheduled outage of one Internet connection.

In response to the NIH Pandemic Flu Continuity of Operations Plan (COOP), NLM successfully participated in an NIH-wide test for staff working remotely. The remote access Citrix terminal server solution continues to be an effective solution for NLM flexi-place workers, as well as staff and contractors who need temporary or long term remote access to NLM IT systems and applications. It provides authentication into the NLM network, access to office and NLM business applications, network-based files, and the Internet. High-speed access is provided mainly through cable modems provided by COMCAST, but other high-speed access such as Verizon FIOS and DSL are also used.

Wireless services were implemented throughout most areas of Buildings 38 and 38A. Wireless access to the Internet and public services of NLM and NIH is provided for guests and typical users. Through a Virtual Private Network (VPN), authorized users can access internal applications in a secure manner.

Internet 2 has become an important resource for connection with NLM and the research community. Internet 2 connectivity is provided by a Gigabit Ethernet link to the Abilene high-speed backbone network via the Mid Atlantic Exchange (MAX) at the University of Maryland. LHC and OCCS work together to manage traffic to and from Internet 2. This year, a redundant, diverse fiber connection from NLM to the MAX was installed and provided by FiberGate. It provides for increased reliability for this critical network connection.

OCCS continued implementation of the High Availability Computing Solution to ensure that critical

applications and resources remain available to NLM users. OCCS deployed clustered Oracle server systems and clustered storage systems as NLM's high availability computing resources. A server or storage cluster is a group of independent computer systems working together as a single system thereby allowing multiple servers to deliver the same application services so that if one of the servers becomes unavailable as a result of failure or maintenance, another server immediately begins providing service. This initiative will increase the NLM mission critical database storage capacity eight times and will increase non-database storage capacity four times.

OCCS continued to make improvements to the UNIX and Wintel architectures. Various upgrades in additional servers, increased memory, and subnet reliability were performed.

Desktop Support

OCCS has shifted to perform most security, anti-virus, pest scanning, and software deployments after business hours. OCCS is prototyping the use of Wake-on LAN and scripted PC restarts to automate routine administrative and security tasks. This initiative has substantially reduced the interruption of a user's business time to apply these software changes and enforce the security of NLM systems.

OCCS assisted Library Operations with the prototyping and deployment of an upgrade to the Pharos Reading Room patron printing system. Along with this Windows 2003 Cluster Server migration, workstations and specialty stations were upgraded and reconfigured to offer updated printing services for NLM patrons, including adding touch-screen monitors for release stations, upgraded barcode scanners, and operating system upgrades.

PestPatrol, a centrally managed spyware and adware detector and remover, was deployed across OCCS-managed systems following market review of these tools. This product, on a weekly basis, automatically removes thousands of instances of nuisance or malicious desktop objects. The central reporting feature allows OCCS to monitor the state of spyware on user computers, and to track and remedy risks as they are detected.

Sixty-six new Microsoft operating system security patches that were released this year were applied to the roughly 800 OCCS-managed desktop computers on the NLM network. OCCS this year, migrated to the latest version of the Windows Software Update Server (WSUS) hotfix deployment solution. This new product also deploys patches for Microsoft applications such as MS-Office suite components, and deploys these more efficiently than with the previous generation of this product. These patches are deployed overnight to machines left in a "restarted" and not-logged-in state. Patches are then validated for effective application.

Outreach

Local Legends: Local Legends, the companion Web site to the *Changing the Face of Medicine* exhibition, features biographies and video clips of outstanding women physicians who have been nominated by members of Congress for their exceptional contributions in research and education, public health, military service, or patient care. Most notably, the NLM Web Team received the Director's Team Award for contributions to the Local Legends Web Site. In addition, six new video clips were added to complement the biographies already posted on the site and several biographies were added. Due to the extensive usability testing and research done on this site, no major enhancements have been necessary.

Health Services Research Projects in Progress (HSRProj): Version 2 of HSRProj Search was released in June. One of the major enhancements is the ability to search and retrieve archived projects. These are health services research projects that have been completed more than five years prior to the current year. With the additional search capability, all 14,803 projects are accessible. The HSRProj database was updated with over 1,100 new project records. Modifications were also made to the search engine to improve response time.

Health Services and Sciences Research Resources (HSRR) Database: Version 3 of the HSRR Search System was released in March and included a change to the display format of retrieved records for easier viewing. Version 2 of the HSRR Maintenance System was also released in March, which included the addition of a "Change Password" functionality. Changes were made to the function menu for easier use and viewing. In addition, changes were made to the view format for how records are displayed, which provides a more user-friendly approach.

Health Services Research Information (HSRInfo) Central: In FY2006, a "Suggest a Link" feature was added, allowing users to suggest links to government and nonprofit organizations that serve as gateways to information on topics related to health services research. An "About HSRInfo" feature was created which has a form to contact the program. Improvements were made to site navigation and user-friendliness in general.

Public Health Partners (PHPartners.org): Public Health Partners is a collaboration of U.S. government agencies, public health organizations, and health sciences libraries to present information for the public health workforce on a single Web site. New features added this year include Health Topic pages, improved navigation, and easier access to Healthy People 2010 information. Several enhancements were also made to the tutorial component of the Web site.

PHPartners and HSR Portal Input System: Version 3 of the system used to manage data for both PHPartners.org and HSR Info Central was released this year entitled "PHPartners and HSR Portal Input System." Enhancements included:

- Creation of News Archive for HSR Info Central;
- Automated display order of PHP conferences and meetings by date;
- Added country information for both PHP and HSR organizations; and
- Enhanced validation and protection from spam in the Suggested Links Page and the Comments Page for both PHP and HSRInfo applications.

Research and Development Initiatives

Voice Recognition: Two-way voice communication is an emerging technology that enables users to navigate Web sites by means of spoken user input and interact with Web applications that synthesize speech from the Web server. This year, OCCS implemented a voice recognition prototype that uses a WYSIWYS (What-You-See-Is-What-You-Say) approach to accurately recognize voice commands. The voice commands are compressed in real-time in the browser before being transported and decoded at a backend process. OCCS is researching Natural Language Processing technology in order to provide natural ways to voice a command.

Accessibility/Usability: The NIH Office of Evaluation Research provided funding to OCCS for a study titled "Matching Search Technology to User Expectations: Identification and Evaluation of End-User Search Goals and Behavioral Patterns When Accessing and Retrieving Health Information via the Web." The NLM Web Team sponsored interviews and a focus group to collect information from health care professionals regarding medical resource search scenarios and related issues. Analysis of data was conducted and a final report and briefing were presented detailing findings from various research efforts of the Search Feasibility Study.

Database: OCCS created a version independent, superset, database network of files on a shared UNIX Volume as part of the effort to manage database connections for all development, test, and production environments. This will provide seamless user connections to the databases in 9i Standalone environments as well as in the new 10g RAC environments. Also, a prototype 10g RAC database was created.

ReportNET Migration: OCCS will be migrating from Impromptu and Impromptu Web Reports to the COGNOS product ReportNET. ReportNET will provide decision-support reporting capability across the spectrum of NLM activities.

NLM Web Support

Web Content Management: NLM uses TeamSite to provide content and application management for Web sites. TeamSite was upgraded to Version 6.5 with major enhancements including the creation of the RSS template and modifications to several TeamSite templates, particularly the NLM Main Web templates.

Web Analytics: NLM uses the WebTrends software package to track the number of pages served over time by the sites being managed and to provide detailed analysis of trends in site usage, audience composition, and other matters. Enhancements included creating a test bed for the analysis of PubMed Web statistics and a profile for licensees of NLM products.

LinkChecker: Maintaining the validity of Web links across numerous hosted Web sites is an important challenge for Web administrators. This year, changes were made to the LinkChecker instance that is used for both HSR and PHP to extract extra database fields used for generating detailed reports. The LinkChecker instance for Consumer Health Topics was enhanced and LinkChecker was also added for HSR Project Supporting Agency and Performing Organization.

Computer Facilities Operations

NLM systems continue to be supported in a safe environment in NLM's computer facility, 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 plasma displays that are visible outside the computer room. The intended audience of this display system is the general public and NLM staff. The system consists of information "panels" with descriptive text, statistical charts, and near real-time activity monitors. Each panel focuses on a particular NLM service or IT infrastructure component. The panels include near-real-time utilization counters for MedlinePlus and for PubMed/MEDLINE, NLM services as seen by remote users around the world, and near real-time utilization data for NLM's Internet-1 and Internet-2 data communications links.

Customer and Administrative Support Systems

Since the 2003 Help Desk consolidation with NIH's IT Help Desk, NLM desktop and PC networking support requests are now channeled to the NIH IT Help Desk for initial ticket entry into the call tracking system. This year over 8,800 NLM ticket requests for IT support were entered and tracked. NLM IT staff resolved 60% of the calls; 40% were completed by NIH staff.

The Customer Service Support System (Siebel) was upgraded from 7.5.3 to 7.7.2.5 to take advantage of new functionalities and better performance including XML integration for inbound Web site form-based requests for the public and dynamic assignment of issues to Library Operations customer service personnel.

OCCS conducted 27 training courses this year, in topics such as "Outlook Calendar Tips," "Outlook FUNdamentals," "Breeze Demo," "Managing your Mailbox,." and "Citrix Remote Access." Focused training was provided in support of the Stay-In-School and NLM Associates programs.

ADMINISTRATION

Todd D. Danielson
Executive Officer

Table 16. Financial Resources and Allocations, FY 2006
 (Dollars in Thousands)

Budget Allocation:
Extramural Programs $69,252
Intramural Programs..................................230,826
     Library Operations............................ (85,594)
     Lister Hill National Center for
     Biomedical Communications............ (57,509)
     National Center for Biotechnology Information
     (73,537)
     Toxicology Information..................... (14,186)
Research Management and Support............11,643
Total Appropriation 311,721
Plus: Reimbursements.................................17,812

Total Resources $329,533

Personnel

In August 2005, **Ken Addess, Ph.D.,** rejoined NCBI as a Staff Scientist and will be working again with Dr. Stephen Bryant in the Structure group. Dr. Addess first worked for NCBI between 1997 and 2000. Before rejoining, Dr. Addess worked at a biotech company in California and then the Protein Data Bank at the San Diego Supercomputer Center. He earned a Ph.D. from UCLA, working on solving the structures of nucleic acids by NMR spectroscopy. Dr. Addess will be working on different aspects of MMDB.

In August 2005, **Joe Bischoff, Ph.D.,** converted to Staff Scientist. As part of the taxonomy group, Dr. Bischoff specializes in fungi. Dr. Bischoff received his Ph.D. in fungal systematics from Rutgers University in 2004. His dissertation work was focused on the phylogenetics and biodiversity of arthropod and plant pathogenic fungi in the family Clavicipitaceae (Hypocreales).

In August 2005, **Jerome Eastham, Ph.D.,** converted to Staff Scientist after having been a TAJ Tech contractor at NCBI since 2003. Dr. Eastman earned a Ph.D. in mathematics from the University of Tennessee for work on numerical partial differential equations. He has since taught at the University of Delaware and at Virginia Tech, and he has worked in industry and government positions as a systems consultant. He will continue doing primarily database programming for batch and interactive gene annotation of model organisms.

In August 2005, **E. Michael Gertz, Ph.D.,** joined NCBI as a Staff Scientist. He will be working on algorithms related to BLAST. Prior to this appointment Dr. Gertz completed his Ph.D. in Mathematics at the University of California, San Diego and postdoctoral appointments at Northwestern University and the University of Wisconsin, Madison. In his thesis and postdoctoral research he developed theory and software for nonlinear programming.

In August 2005, **Lakshminarayan Iyer, Ph.D.,** converted to a Staff Scientist after having been a Postdoctoral Fellow with Dr. Eugene Koonin since 2000. He will be working with Dr. Aravind Iyer on the evolution of proteins and genomes. Dr. Iyer completed his Ph.D. in biology from Texas A&M University working on plant virus resistance. His postdoctoral research involved the study of different problems in protein and genome evolution.

In September 2005, **Song Mao, Ph.D.,** joined the Lister Hill Center as a Staff Scientist. He holds Ph.D. and M.S. degrees in Electrical and Computer Engineering from the University of Maryland. Prior to his appointment, Dr. Mao conducted research at the IBM Almaden Research Center. At LHNCBC, Dr. Mao applies digital image analysis and understanding, pattern recognition, and computer vision to the automated extraction of descriptive and technical metadata from digital objects for long term preservation.

In September 2005, **Lon Phan, Ph.D.,** converted to Staff Scientist after having been a ComputerCraft contractor since 2000. Dr. Phan earned a Ph.D. in biological chemistry from the University of Minnesota for work on post-translational regulation of yeast Elongation Factor 2. He then did postdoctoral research at the National Institute of Child Health and Human Development at NIH. He will continue to process direct submissions to dbSNP, compute and annotate SNP data, and develop Web services for searching and retrieving SNP records.

In September 2005, **Eric Sayers, Ph.D.,** converted to Staff Scientist after having been a Kevric contractor since 2002. Dr. Sayers earned a Ph.D. in pharmacology from Yale University for work on the structures of protein-carbohydrate complexes using NMR and molecular dynamics simulations, and then spent three years at NIDCR working on the solution structures of ribosomal proteins using NMR. He will work at the NCBI service desk where he develops and teaches courses on NCBI resources, designs educational Web sites, and provides user support.

In September 2005, **Lowell Vizenor, Ph.D.,** joined the Lister Hill Center as a Postdoctoral Fellow. Dr. Vizenor received his doctorate degree in philosophy from University of Buffalo. At NLM, Dr. Vizenor's research will focus on creating a mid-level ontology of biomedicine and supporting the application of philosophically motivated ontology to biomedical terminologies.

In October 2005, **Patti Sherman, Ph.D.,** converted to Staff Scientist after having been a ComputerCraft contractor since 2000. Dr. Sherman earned a Ph.D. in human genetics from the University of Michigan. Her postdoctoral research at the Johns Hopkins University School of Medicine involved the cloning and characterization of a protein phosphatase. Prior to arriving at NCBI, Dr. Sherman was a science writer and editor for the Online Mendelian Inheritance in Man (OMIM) database. She will continue working on the RefSeq project.

In November 2005, **Kin Wah Fung, M.D.,** joined the Lister Hill Center as a Staff Scientist. He received his M.D. degree from the University of Hong Kong in 1984. He is a Fellow of the Royal College of Surgeons of Edinburgh and of the Hong Kong Academy of Medicine. He has a M.S. degree in Computer-based Information Systems from the University of Sunderland, UK and a Master of Arts degree in Medical Informatics from the highly regarded program at Columbia University. Dr. Fung came to NLM as a Postdoctoral Fellow in 2003. He is currently focusing on difficult knowledge representation questions important to the Unified Medical Language System (UMLS) project.

In November 2005, **Martha Gaie, Ph.D.,** joined the Lister Hill Center as a Postdoctoral Fellow. She has her doctorate degree from University of Wisconsin–Madison in Mass Communications. Dr. Gaie was a postdoctoral fellow at the Center of Excellence in Cancer Communications Research at the University where she conducted research in information-seeking, message evaluation processes and Internet uses and effects. At NLM, Dr. Gaie is developing a series of studies that detail how consumers seek health information on the Internet.

In November 2005, **Alan Graeff** converted to Staff Scientist after serving as the Chief Information Officer at NIH since 1998. He the NCBI Deputy Director, working with Dr. David Lipman, Director, on general program management issues and a broad range of NCBI branch components. Mr. Graeff earned a B.S. in distributed sciences from American University, and then spent several years working in a laboratory setting before joining the Information Technology Branch at the National Institute of Allergy and Infectious Disease. He later joined the Clinical Center at NIH, where he served as Chief Information Officer. As Deputy Director of NCBI, Mr. Graeff's responsibilities will include program development and evaluation, information policy formulation and direction, and coordination of all NCBI activities.

In November 2005, **Diane Howden** joined the SIS. She received her B.S. degree in chemistry from Old Dominion University in Norfolk, Virginia. Previously she was with the National Institute of Neurological Diseases and Stroke, where she managed a program that systematically tested a series of compounds for their anti-convulsive effects. Her experience there in using chemical nomenclature, chemical structures, and database creation and management make her a perfect fit for our chemical information group, where she is working on the creation and maintenance of ChemIDplus as well as other projects like the new drug information portal.

In November 2005, **Dr. Zoe Huang** joined NLM as a Health Scientist Administrator. Dr. Huang was trained at Shougang Medical School/Capital University of Medical Sciences, Beijing, China, specializing in pediatric endocrinology and diabetes. She received several NIH Fellowships before being employed by the National Heart, Lung, and Blood Institute where she contributed to over 40 review meetings. At NLM, Dr. Huang will be responsible for the administration of NLM's Special Emphasis Panels (SEPS) and will support the Scientific Review Administrator in organizing the activities of the Biomedical Library and Informatics Review Committee.

In November 2005, **Suresh Srinivasan** joined the Lister Hill Center as a Computer Scientist. He holds M.S. degrees in Transportation Engineering and in Operations Research and Statistics from Rensselaer Polytechnic Institute. He made significant contributions to the UMLS project and to the Center's Natural Language Systems group as a contractor. Most recently, he has been a project leader responsible for the design and development of the editing and workflow management systems for the UMLS. He is now responsible for the design, evaluation, and testing of the multi-platform MetamorphoSys software.

In December 2005, **Andrew Diggs** joined NLM as a Grants Management Specialist. Mr. Diggs previously served as a Grants Management Specialist for the National Center for Research Resources. Mr. Diggs will be responsible for the management of a complex grant portfolio that includes numerous Program Projects (P01s) and Cooperative Agreements (U01s). Mr. Diggs will also be responsible for compiling reports for the NLM Office of the Director, including House and Senate Subcommittee Reports that reflect grant expenditures for prior and future fiscal years.

In December 2005, **Mariana Dimitrov, Ph.D.,** joined the Lister Hill Center as a Visiting Scholar. She has her doctorate degree in Cognitive Science from the University of Sofia in Bulgaria. Prior to coming to NLM, Dr. Dimitrov was a guest researcher at the Bulgarian Academy of Sciences and before that was a Postdoctoral Fellow in the Cognitive Neuroscience Section of National Institute of Neurological Disorders and Stroke where she studied neuro-psychology and psychiatry. At NLM, Dr. Dimitrov is pursuing research in knowledge discovery using a semantic processing approach for hypothesis generation.

In January 2006, **Mary Hollerich** joined NLM as Head, Collection Access Section, PSD/LO. Ms. Hollerich's most recent position was Associate Director for Access Services at the Northwestern University, Pritzker Legal Research

Center. Her career of 17 years includes previous positions at Northwestern University and the University of Southern California. Ms. Hollerich is active in the American Library Association, where she is currently Chair of the Reference and User Services Association Committee. Ms. Hollerich holds a B.A. in Germanic Languages and Literature and an M.A. in Library and Information Science, both from the University of Illinois at Urbana-Champagne. Ms. Hollerich will manage NLM's interlibrary loan program, DOCLINE, document delivery related customer service questions, and onsite delivery of materials to patrons in the main reading room.

In February 2006, **Victor Cid** joined the Specialized Information Services as a Senior Computer Scientist. Mr. Cid has a M.S. degree in Information Technology Engineering from the University of Chile, and a M.S. degree in Telecommunications and Management from the University of Maryland University College. He was previously with NLM's OCCS, where he made significant contributions in the areas of network and systems performance and information accessibility. Since his arrival at NLM in 1991, Mr. Cid has been also supporting a number of NLM outreach activities involving national and international communities and partners. His new role with the Office of Outreach and Special Populations of SIS will allow him to continue his work in this area.

In February 2006, **Lisa Forman, Ph.D.,** joined NCBI as a Staff Scientist in the Information and Engineering Branch. Dr Forman has a M.A. and Ph.D. in Physical Anthropology from New York University. Dr. Forman comes to NCBI with extensive experience in forensic DNA testing. She started her career performing criminal and paternity casework and her role evolved into an involvement in forensic DNA national policy issues at the U.S. Department of Justice. She also has experience in the rare disease patient advocacy community by serving as a scientific director for a group participating in the NIH's "Rare Disease Clinical Network." Dr. Forman will be working on the development of an Open Source Independent Review and Interpretation System (OSIRIS) to assist in the evaluation of quality for high throughput STR DNA profiles used in the forensic and biomedical communities as well as in outreach activities enhancing the utility of NCBI tools for stakeholders in rare diseases research.

In February 2006, **Aurelie Neveol, Ph.D.,** joined the Lister Hill Center. Dr. Neveol received her Ph.D. in Computer Science from the Rouen Medical School in France in 2005. She is working on automatic MeSH indexing, particularly the evaluation of existing indexing software and the investigation of methods for automatic subheading attachment. Her research will aim at highlighting the strengths and weaknesses of the indexing system, and proposing solutions to enhance it. This work will be carried out in close collaboration with the indexing team.

In March 2006, **Carlos Evangelista, Ph.D.,** converted to NCBI Staff Scientist after having been a ComputerCraft contractor since 2004. Prior to this appointment, Dr. Evangelista was a Postdoctoral Fellow at Johns Hopkins University, where he did experimental and computational work on calcium signaling in yeast. He was also involved in several genome assembly projects, worked on the development and genetics of the fruit fly, and modified the two-hybrid system for high throughput projects. Dr. Evangelista earned his Ph.D. in molecular biology from SUNY Albany.

In March 2006, **Brian Smith-White** converted to NCBI Staff Scientist after having been a Management Systems Design contractor since 2000. Mr. Smith-White earned a M.S. in Biological Chemistry from the Johns Hopkins University School of Medicine for work on reaction mechanism of exonuclease III of Escherichia coli followed by a decade of research in plant biochemical genetics with Jack Preiss. He will continue to develop the plant genomics presence at NCBI.

In March 2006, **Alexandra Soboleva** converted to NCBI Staff Scientist after one year as a contractor. She received her M.S. degree in mathematics and applied mathematics in 1993 from Moscow State University. She will continue to work with Dr. Ron Edgar on the GEO repository and related projects.

In March 2006, **Zhiyun Xue, Ph.D.**, joined LHNCBC as a Postdoctoral Fellow. Prior to arriving at NLM, she completed a doctorate degree in electrical engineering from Lehigh University, Bethlehem, PA. Her thesis topic was in the area of multi-sensor image fusion and object detection and tracking in image sequences. At LHNCBC, Dr. Xue will be working with the Communications Engineering Branch on the creation of a content-based image retrieval (CBIR) system for cervico-graphic images.

In April 2006, **Phillip Osborne** joined NLM as the Director of Office of Acquisitions in the Office of Administration. He received his Bachelor's degree in Business Management from Florida State University. Mr. Osborne was previously the Director of the Division of Grant and Contracts Management at the Food and Drug Administration. Mr. Osborne has also held senior leadership positions in the acquisitions field at the Department of Commerce and the Environmental Protection Agency. He has more than 22 years of acquisition experience and extensive leadership experience.

In April 2006, **Jason Papadopoulos** converted to NCBI Staff Scientist from a contractor position. Mr. Papadopoulos received a Master's degree in electrical engineering from the University of Maryland, College Park. He will continue to work in the BLAST group under Dr. Tom Madden, maintaining and enhancing the BLAST suite

of alignment tools and developing new sequence searching and alignment algorithms.

In May 2006, **Caroline Ahlers, M.D.**, joined the LHNCBC as a Postdoctoral Fellow. She received an M.D. degree from Ross University School of Medicine, West Indies. Prior to training to be a physician, Dr. Ahlers worked as an engineer for four years. Dr. Ahlers joined the LHNCBC after receiving her degree and is working on enhanced natural language processing for pharmacogenomics. She is analyzing text in the domain of pharmacogenomics by extracting and summarizing drug-gene-disease relations.

In May 2006, **John B. Anderson, Ph.D.**, converted to NCBI Staff Scientist after having been a ComputerCraft contractor. Dr. Anderson earned a Ph.D. in Genetics from the University of Hawaii for work on the molecular and population genetics of the Segregation Distortor system in Drosophila melanogaster. He then did a postdoc with Dr. Lawrence Chan at Baylor College of Medicine working on apobec1 and perilipin, two genes important in lipid metabolism. This was followed by a postdoctoral fellowship at NIH. He joined Computercraft in March 2000 as a contractor for the NCBI working as an indexer for GenBank and then switched to the new Conserved Domain Database project with Computational Biology Branch of NCBI in December of 2000. As a Staff Scientist, he will continue to work on the Conserved Domain Database and other projects in the Structure Group of CBB.

In May 2006, **Myra Derbyshire, Ph.D.**, converted to Staff Scientist after having been a ComputerCraft contractor. Dr. Derbyshire earned a Ph.D. in Genetics from the University of Leeds, England for work on nitrogen metabolism in the Moss Physcomitrella patens. She followed this with various research and teaching positions. She will continue to contribute curated protein domain models to the NCBI Conserved Domain Database (CDD).

In May 2006, **John Jackson, Ph.D.**, converted to Staff Scientist after having been a ComputerCraft contractor. Dr. Jackson earned a Ph.D. in Molecular and Cell Biology from the Pennsylvania State University in 1996 for work on the beta-globin Locus Control Region, and then spent several years studying the role of a histone H2A variant in yeast chromatin. He will continue his work as part of Dr. Stephen Bryant's CDD curation team.

In May 2006, **Thomas Lehmann, Ph.D.**, joined LHNCBC as a Visiting Scholar. He has a Ph.D. degree in computer science from Aachen University of Technology (RWTH), Germany where he is currently an associate professor at the Faculty of Electrical Engineering. He heads the Division of Medical Image Processing. Dr. Lehmann's research interests are in discrete realizations of continuous image transforms, medical image processing applied to quantitative measurements for computer-assisted diagnoses, and content-based image retrieval from large medical databases. Currently he is working with researchers in the Communications Engineering Branch on projects in Content-based Image Retrieval.

In May 2006, **Cynthia Liebert, Ph.D.**, converted NCBI Staff Scientist after having been a ComputerCraft contractor. Dr. Liebert earned a Ph.D. in microbiology from the University of Georgia. As a Postdoctoral Fellow, she assisted in a USDA funded drug-resistance study to characterize the integron classes and drug resistance gene cassettes present within the bacterial flora of poultry. At NCBI, she will continue to contribute to the structural and functional annotation of protein families comprising the NCBI CDD, use molecular sequence alignment, three-dimensional structure superposition, and phylogenetic analysis to characterize the evolutionary relationships of conserved protein domains and to identify and annotate sites responsible for their biological functions.

In May 2006, **Fu Lu, Ph.D.**, converted to NCBI Staff Scientist after having been a ComputerCraft contractor since May 2004. Dr. Lu graduated with a Ph.D. degree in Biochemistry from Purdue University, and then worked in Celera Genomics on genome mapping and comparative genomics. He will continue to work on protein domain family classification and annotation as part of the CDD project group.

In May 2006, **Gabriele Marchler, Ph.D.**, converted to NCBI Scientist after working on the CDD project as a ComputerCraft contractor since 2001. Dr. Marchler earned a Ph.D. in biochemistry from the University of Vienna, Austria, for work on stress response in yeast. She then completed a postdoctoral research project at the NCI, working on heat shock response in Drosophila. Dr. Marchler will continue to curate and annotate conserved protein domain hierarchies, and train and support other curators. She will also validate the work of others in the group, as part of the CDD production pipeline.

In May 2006, **Sergey Resenchuk, Ph.D.**, converted to NCBI Staff Scientist after having been an MSD, Inc. contractor since 1999. Dr. Resenchuk earned a Ph.D. in molecular biology in Russia for work on computational analysis of poxviruses. He will continue to work on the creation of software for NCBI resources, such as MapViewer, Genome Project, and others, as well as on data processing for NCBI genome pipeline builds, and complete genomes projects.

In May 2006, **James Song, Ph.D.**, converted to Staff Scientist after having been a ComputerCraft curator since 2001. Dr. Song earned a Ph.D. in molecular pharmacology from the University of Vermont College of Medicine for work on multi-drug resistance, and then did postdoctoral research in the area of signal transduction and cancer biology at NIH. He was an Assistant Professor of Research at Temple University, where he performed structure-

function studies of unique, naturally occurring inhibitors of angiogenesis. At NCBI, he will continue to perform the processing, annotation, curation, and quality control steps of the curation process and review the biological significance and accuracy of each protein family in the CDD.

In May 2006, **Roxanne Yamashita, Ph.D.**, converted to NCBI Staff Scientist after having been a ComputerCraft contractor since 1999. Dr. Yamashita earned a Ph.D. in Genetics and Molecular Biology from the University of Hawaii for work on the expression of foreign proteins in Neurospora crassa, and then spent several years working the unconventional myosins in Aspergillus nidulans at the Baylor College of Medicine and Drosophila melanogaster at NHLBI. She will continue to work at NCBI as a curator for the CDD.

In June 2006, **Jian Zhang, Ph.D.**, converted to NCBI Staff Scientist after having been an MSD contractor since 2002. Dr. Zhang earned a Ph.D. in medicinal chemistry from the Beijing Medical University for work on vaccine development and drug discovery, and then spent several years working on small molecules database design and bioactivity data integration. He will continue to work with the PubChem project on the user interface design, compound and bioactivity data presentation.

In July 2006, **Josh Cherry, Ph.D.**, converted to NCBI Staff Scientist after having been an MSD contractor since 2003. Dr. Cherry obtained a Ph.D. in Human Genetics from the University of Utah. As a Postdoctoral Fellow at Harvard University he did theoretical and computational work in population genetics and molecular evolution. He will continue working on various software systems, including Genome Workbench, scripting language interfaces to the NCBI C++ Toolkit, and algorithms involved in constructing and analyzing genome sequences.

In July 2006, **Sergio Leon, M.D.**, joined LHNCBC as a Postdoctoral Fellow. He has an M.D. degree from Universidad Del Valle in Cali, Colombia and completed an Internal Medicine residency at Prince George's Hospital Center. Dr. Leon is interested in the retrieval of clinical information where patient care is delivered. He is doing research with the Office of High Performance Computing that will look into the use of Smart phones and other wireless handheld devices for real-time Internet access to MEDLINE/PubMed and other clinical references and how these can impact on patient management and improve clinical practice.

In July 2006, **Aleksandr Morgulis** converted to NCBI Staff Scientist after having been an MSD contractor since 2002. Mr. Morgulis earned a Masters Degree in mathematics from the Pennsylvania State University. As a contractor at NCBI, Mr. Morgulis worked on projects related to DNA repeat masking, low complexity filtering,

and BLAST performance improvement. He will continue to work on various BLAST related projects such as database indexing and algorithmic optimization, as well as support and maintain software developed in support of those projects.

In July 2006, **Jewen Xiao, Ph.D.**, converted to NCBI Staff Scientist after having been an MSD contractor since 2002. Dr. Xiao earned a Ph.D. in Chemical Engineering from the University of New Hampshire for work on blood microcirculation, and then spent several years working on Information Technology. He will work on the PCAssay database on PubChem.

In August 2006, **Sheldon Kotzin**, Chief of NLM's Bibliographic Services Division, was appointed Associate Director for Library Operations, the largest of NLM's major components. Mr. Kotzin, who has headed the Library's Bibliographic Services Division since 1981, has a Master of Library Science degree from Indiana University in 1968. Following graduation he came to the NLM as a Library Associate. He served in positions of increasing responsibility in Library Operations before becoming Chief of the Bibliographic Services Division. Since 1998, Mr. Kotzin has also served as Executive Editor of MEDLINE and Administrator of the Literature Selection Technical Review Committee, the body that reviews and recommends journals for indexing in MEDLINE. He is also NLM's representative to the International Committee of Medical Journal Editors, a group of 12 clinical journal editors who establish standards for submission of journal articles.

In September 2006, **Jerry Sheehan** joined the NLM as the Assistant Director for Policy and Legislative Development. Mr. Sheehan has a Master's degree in Technology and Policy and a Bachelor's degree in Electrical Engineering from the Massachusetts Institute of Technology. Prior to joining NLM, Mr. Sheehan was with the Organization for Economic Cooperation and Development in Paris, France and was responsible for the formulation, management and performance of analytical work related to international science, technology, and innovation policy. He has also held positions with the National Research Council and the Office of Technology Assessment, U.S. Congress, working on information technology issues. Mr. Sheehan was the Study Director for two key NRC reports on health information technology initiated and supported by NLM: "For the Record—Protecting Electronic Health Information" and "Networking Health."

NLM Associate Fellows Program for 2006–07

The NLM Associate Fellowship Program is a one-year training fellowship for recent graduates of Masters 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-

month period. Projects relate to key NLM programs areas and are typically of a research, development, or evaluation nature. Seven new Associate Fellows began their year at NLM on September 5, 2006.

Abdrahamane Anne is from Mali and is participating in the program as an International Fellow. Mr. Anne received his M.A. in Librarianship in 1994 from the Bielorussian University of Culture. He received his M.I.S. in 2003 from l'Ecole Nationale Superieure des Sciences de l'Information et des Bibliothèques in Villeurbanne, France. In Mali, he has been a reference librarian at the Faculte de Medecine de Pharmacie & d'Odonto-stomatologie of Bamako since 1998. Mr. Anne's responsibilities include online searching of MEDLINE, Pascal, and other online health information sources; developing Web sites; managing a local database management system; and contributing to publication of "Digest Sante Mali," a quarterly bibliographic review for Malian health workers. Prior to his current position, he was a librarian at the Centre Amadou Hampate Ba in Bamako and also completed two 6-month internships at the Bielorussian National Library. His undergraduate training was in the Humanities.

Marisa Conte received her M.L.I.S. in May 2006 from Wayne State University. She has reference experience as a Graduate Assistant in the Wayne State University Library. She also has varied experience as a volunteer in two hospital libraries. Prior to starting her career in librarianship, Ms. Conte had nine years of customer service experience in the airline industry. Her undergraduate training was in Medieval Studies.

Courtney Crummett received her M.L.I.S. in August 2006 from the University of South Florida. As a Graduate Assistant in library and information science, she provided technical support for electronic instructional environments, developed Web page materials for instruction and research support, and assisted faculty with research and writing. She also has circulation experience at the University of Tampa Library. As a graduate assistant in geology, Ms. Crummett taught physical geology laboratory classes and gained experimental laboratory experience. She holds an M.S. and B.A. degree in Geology.

Robin Featherstone received her M.L.I.S. in May 2006 from Dalhousie University in Nova Scotia. In the health sciences library at Dalhousie, she has experience in reference, circulation, online searching, developing online tutorials, and indexing. She also has experience as a research assistant compiling metadata for a bibliographic database component of the History of the Book in Canada project. Prior to beginning her librarianship training, she was the manager of a book shop. Her undergraduate degree was in English.

Amy McNeely received her M.L.S. in May 2006 from UCLA. She has four years of cataloging experience as a

student and Library Technical Assistant, including experience in original and copy cataloging, authority file creation, and adding local subject headings and notes to catalog records. She also has experience with the California Cultures Digitization Project, where she reviewed and edited digital metadata and previewed digital objects for quality and accuracy. As an IT Lab Assistant at UCLA, she provided hardware and software support for the faculty and students in the Department of Information Studies. Her undergraduate degree is in English.

Alison Rollins received M.L.S. and M.I.S. degrees, with a Chemical Information Specialist Certificate, in May 2006 from Indiana University. She has experience indexing health care services as part of the Indiana GoLocal project at Ruth Lilly Medical Library. She also has reference and Web maintenance experience at the Indiana University Chemistry Library, and special library experience in reference, circulation, and bibliography production. Her undergraduate degree is in Biology and English.

Meredith Solomon received her M.L.S. from Emporia State University in May 2006. She has eight years experience as a Library Technical Assistant at the Tuality Healthcare Health Sciences Library, where she had varied responsibilities in reference, online searching, circulation, and interlibrary loan. She also has circulation experience in a public and academic library. She gained management and staff training experience as team leader for a warehouse distribution department for Hollywood Entertainment. Her undergraduate degree is in Liberal Studies.

Retirements and Separations

On September 30, 2005, **Ronald Stewart** retired from NLM. Mr. Stewart had been with the Library since 1988 after having served in the U.S. Air Force for 26 years. Mr. Stewart started his career at NLM as the Chief of the Office of Administrative Management. In 1990, he was promoted to Supervisory Management Analyst in the Office of Administration, and effective May 7, 2000, he was promoted to the position of Deputy Executive Officer, from which he retired. Prior to his retirement, Mr. Stewart served as Acting Executive Officer, a position he held for several months. Mr. Stewart had advised and assisted all of the Executive Officers he had worked with by sharing his extensive knowledge of administrative policies and regulations and of the NLM itself. Mr. Stewart's insight, understanding, knowledge, and expertise were invaluable in directing the Executive Office through its period of transition from one Executive Officer to another.

On October 31, 2005, **CAPT James Knoben, Pharm.D. M.P.H.,** retired from the Division of Specialized Information Services after serving over 33 years in the Federal government. Dr. Knoben worked at NLM from 1982–83 and 1999–2005, and was appointed Special Assistant to the Associate Director, SIS in July 2000. Most of Dr. Knoben's

career included service at the FDA where he served as a Division Director; he was also co-editor of a drug therapeutics books used throughout the world. He is a graduate of the University of California and Yale University. In July 2006, he was awarded the Surgeon General's Exemplary Service Medal, one of the highest awards in the USPHS Commissioned Corps. Dr. Knoben currently serves at NLM part-time as a contractor.

In October 2005, **Mary Smith** departed from NLM to join the Division of Acquisition Policy and Evaluation, Office of the Director, NIH. Ms. Smith was a Contract Specialist with the NLM Office of Acquisitions Management, Office of Administration since 1987. Previously, she was a Contract Specialist with the Contract Operations Branch, Division of Extramural Affairs, NHLBI. Ms. Smith received an NIH Merit Award in 1998. In 2002, she served as Acting Chief of NLM's Office of Acquisitions Management with responsibility for the planning, oversight, negotiation, award, administration and close of all acquisitions awarded.

On March 31, 2006, **Eve-Marie Lacroix** retired from NLM. Ms. Lacroix had been Chief of the Public Services Division since 1985 after holding several positions with the Canada Institute for Scientific and Technical Information. Among her many accomplishments at NLM, she oversaw the implementation of the preservation program and the DOCLINE interlibrary loan system, the development of NLM's main Web site, and the development of MedlinePlus. Ms. Lacroix is a Fellow of the American College of Medical Informatics and received many awards for her contributions to the Library's programs including the NIH Director's Award, the NLM Director's Award, and the MLA Thomson Scientific/Frank Bradway Rogers Information Advancement Award.

On April 15, 2006, **Marjorie A. Cahn** retired from NLM. Ms. Cahn joined the Library in July 1991 as first Head of the National Information Center on Health Services Research and Health Care Technology (NICHSR). In her 15 years at the Library, Ms. Cahn was instrumental in improving access to health information for health services researchers and the public health community through the development of NLM databases, services and Web sites designed to meet their needs. Her leadership of the Partners in Information Access for the Public Health Workforce initiative brought together diverse public health organizations and libraries in the National Network of Libraries of Medicine to collaboratively produce quality information, training, and outreach resources for the public health workforce and the library and information specialists who serve those communities.

In June 2006, **Carolyn Tilley** retired after a 33-year career at NLM. Her government service spanned 41 years. From 1981 to 2004 she served as Head of the NLM's MEDLARS Management Section with responsibility for support and training of online database users, licensing of NLM databases, and database testing and quality assurance. Throughout her career Carolyn was instrumental in implementing countless database features and services that enhanced access for NLM users throughout the world. During her tenure as Section Head, use of MEDLINE increased from a few thousand searches a year to nearly 700 million. In 2004 Carolyn accepted a new position to lead Library Operations staff into new areas of involvement with the administration, training, and documentation for the Unified Medical Language System. Carolyn is preparing for a new career as a hospital chaplain.

Awards

The 2006 NLM Board of Regents Award for Scholarship or Technical Achievement was awarded to **Jeffrey D. Beck** for leadership in development of the NLM Journal Archiving and Interchange Document Type Definition (DTD).

The Frank B. Rogers Award recognizes employees who have made significant contributions to the Library's fundamental operational programs and services. The recipients of the 2006 awards were **Natalie A. Arluk** in recognition of innovative and substantial contributions to NLM's Medical Subject Headings applications; **Martha R. Fishel** in recognition of substantial contributions to NLM's programs to provide medical information to researchers and the public worldwide; and **Sergey Krasnov, Ph.D.** in recognition of innovative and substantial contributions to NLM's PubMed Central Database.

The NLM Director's Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to **Wei Ma** for exceptional efforts in leading the technical development and operations of MedlinePlus, NIHSeniorHealth, RxNorm, DailyMed, Local Legends, and NLM Outreach and Exhibit databases, and **Marjorie A. Cahn** for exceptional contributions to and leadership of the NLM Public Health Information Program that helps the public health workforce find and use information effectively to improve and protect the public's health.

The NIH Merit Award was presented to three employees: **Deirdre A. Clarkin** for exceptional leadership and management of the CAS Onsite Access Unit, which delivers 250,000 items annually to Reading Room users within an average of 25 minutes; **Judith C. Eannarino** for her exceptional contributions in managing and interpreting the NLM collection development policy to reflect advances in biomedical science and increased access to digital information; and **Janet R. Zipser** for superior management of NLM's MEDLINE Training Program which provides training and education for searching NLM Databases, especially MEDLINE/PubMed, thereby improving patient care and research throughout the world.

The NIH Director's Award was presented to one individual and two groups. The individual award recipient was **Wesley D. Russell** in recognition of his efforts to advance network communications services at the NIH for the past 20 years. The group award recipients were **Lisa Forman, Ph.D., James M. Ostell, Ph.D.**, and **Stephen T Sherry, Ph.D**. as members of the Hurricane Katrina Victim DNA Identification Team: "For extraordinary volunteer efforts in the DNA-based identification of victims following Hurricane Katrina" and **Elliot R. Siegel, Ph.D**. as a member of the Trans-NIH Type I Diabetes Research Strategic Plan Steering Group: "For exemplary service in leading the development of a Trans-NIH Type 1 Diabetes Research Strategic Plan."

The Louis Round Wilson Prize for Lifetime Achievement Award from the Louis Round Wilson Academy at the University of North Carolina was presented to **Donald A.B. Lindberg, M.D.** for a lifetime of knowledge exploration, compilation and stewardship in service to society.

The Association of Academic Health Sciences Libraries Cornerstone Award was presented to **Betsy L. Humphreys.**

Table 17. FY 2005 Full-Time Equivalents (Actual)

Office of the Director	9
Office of Health Information Programs Development	6
Office of Communication and Public Liaison	8
Office of Administration	39
Office of Computer and Communications Systems	50
Extramural Programs	13
Lister Hill National Center for Biomedical Communications	68
National Center for Biotechnology	154
Specialized Information Services	35
Library Operations	273
TOTAL FTEs	**655**

NLM Diversity Council

The NLM Diversity Council welcomed four new members in FY 2006: Robin Hope-Williams, Lynette Rollerson, Crystal Smith and Tim Valin. Each will serve a two-year term. Continuing on the Council are: Carmen Aguirre, Patricia Carson, Donald Jenkins, Sue Levine, Melanie Modlin, Elizabeth Mullen, Helen Ochej, and Bryant Pegram. The Council continues to receive support from its ex-officio members: Kathleen Cravedi, Public Liaison Officer; Todd Danielson, Executive Officer; Mehryar Ebrahimi, Office of Administrative and Management Analysis Services; Pamela Oliver and Blandina Peterson from the NIH Office of Equal Opportunity and Diversity Management; and Nadgy Roey, Program Advisor and Ethics Coordinator for NLM, as well as its talented alumni. Patricia Carson and Melanie Modlin continued as Council Co-Chairs, and Helen Ochej served another year as Council Secretary.

NLM Director's Employee Education Fund: The NLM Diversity Council continued its coordination of the NLM Director's Employee Education Fund. In FY 2006, the Fund enabled 65 staff to take 61 classes. The school with the largest number of NLM enrollees was the University of Maryland (22 attendees). Course disciplines enrolled in included: business, chemistry, computer networking, economics, marketing, mathematics and psychology.

Art 4 Health: In collaboration with the HHS Kansas City Region 7 Office of Public Health, the NLM Diversity Council sponsored the successful "Art 4 Health" exhibition in celebration of the 2006 African American History Month and Women's History Monty reflecting health disparities in a diversified nation.

Language Instruction: The Council continues to support a program to help NLM employees improve their proficiency in speaking and writing English. Following the model employed by local government literacy programs, the NLM programs offers one-to-one tutoring with NLM staff members, who volunteer their time and receive special training to be tutors. The Council also purchased special software, available in the NLM Staff Library, to teach Spanish to staff.

School Supply and Holiday Gift Drives: In 2006, the Diversity Council continued its partnership with Rolling Terrace Elementary School in Takoma Park, Md. The school serves a multi-ethnic population of about 700, including a high percentage of low-income students. In August, NLM staff filled and refilled the collection bins with backpacks and assorted school supplies. In November and December, the staff helped make the holidays special for Rolling Terrace families, purchasing clothes, toys and gift cards to match their specific needs.

Coat and Food Drives: October and November saw the Diversity Council collecting coats and other cold weather gear, to be distributed to Montgomery County, Md. residents in need by The Shepherd's Table, a non-profit agency. Similarly, the Council helped its neighbors by staging a highly successful food drive in March and April, with the items going to the Manna Food Center, in Rockville, Md.

Getting to Know NLM: This popular series for NLM staff continued in 2006, spotlighting programs and different offices within the Library. Among the offerings were "The Joy of Stacks," a subterranean trip into the collection, and a close-up look at the rare book holdings and the conservation lab.

Appendix 1: Regional Medical Libraries

1. **MIDDLE ATLANTIC REGION**
   New York University School of Medicine
   Frederick L. Ehrman Medical Library
   550 First Avenue
   New York, NY 10016
   (212) 263-5394 FAX (212) 263-6534
   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 60612-4330
   (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
   Health Sciences Libraries and Information Center
   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:

BUCHANAN, Holly S., Ed. D. (Chairwoman)
Associate Vice President of Knowledge Management and IT
Health Sciences Library & Informatics Center
University of New Mexico
Albuquerque, NM

CHABRAN, Richard, M.L.S., Chair
California Community Technology Policy Group
3081 Sunrise Court
Chino Hills, CA

COHEN, Jordan J., M.D.
President Emeritus
Association of American Medical Colleges
Washington, D.C.

HARRIS, C. Martin
Chief Information Officer and Chairman
Information Technology Division
The Cleveland Clinic Foundation
Cleveland, OH

ISOM, O. Wayne, M.D.
Terry Allen Kramer Professor of Cardiothoracic Surgery
New York Presbyterian–Weill Cornell Medical School
New York , NY 10021

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

MORTON, Cynthia, Ph.D.
W.L. Richardson Professor
Departments of OB/GYN and Pathology
Brigham and Women's Hospital
Boston, MA 02115

STANLEY, Eileen H., M.A.
1356 Sextant Ave.
Roseville, MN 55113

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:

FRIEDMAN, Carol, Ph.D. (Chair)
Adjunct Professor
Department of Medical Informatics
Columbia University
New York, NY

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

CALIFF, Robert M., M.D., Vice Chancellor
Department of Medicine
Duke University Medical Center
Durham, NC

CARTER, Jerome H., M.D.
Chief Executive Officer
Neck, Time & Money Informatics, Inc.
Atlanta, GA

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

SHNEIDERMAN, Ben, Ph.D., Professor
Department of Computer Science
University of Maryland
College Park, MD

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

SRIHARI, Sargur N., Ph.D.
University 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:

GINSBURG, David, M.D. (Acting Chair)
Distinguished University Professor
Internal Medicine and Human Genetics
University of Michigan
Ann Arbor, MI

FIRE, Andrew Z., Ph.D., Professor
Departments of Pathology and Genetics
Stanford University School of Medicine
Stanford, CA

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

NICKERSON, Deborah A., Ph.D., Professor
Department of Genome Sciences
University of Washington
Seattle, WA

SALEMME, F. Raymond, Ph.D.
President
Imiplex, LLC
Yardley, PA

SALZBERG, Steven L., Ph.D.
Senior Director of Bioinformatics
University of Maryland
College Park, MD

THOMAS, Annette C., Ph.D.
Managing Director & President
Nature Publishing Group
Macmillan Publishers Ltd.
London, United Kingdom

Appendix 5: Biomedical Library and Informatics Review Committee

The Biomedical Library and Informatics Review Committee meets three times a year to review applications for grants under the Medical Library Assistance Act.

Members:

YOKOTE, Gail A. (Chair)
Associate University Librarian for Research Services and Collections
Peter J. Shield Library
University of California, Davis
Davis, CA

ARONSKY, Dominik, M.D., Ph.D.
Assistant Professor
Department of Biomedical Informatics
Eskind Biomedical Library
Vanderbilt University
Nashville, TN

DUNKER, A. Keith, Ph.D., Professor
Biochemistry & Molecular Biology
Indiana University Schools of Informatics & Medicine
Indianapolis, IN

HUNTER, Lawrence E., Ph.D.
Associate Professor, Department of Pharmacology
University of Colorado Health Sciences Center
Aurora, CO

LEHMANN, Harold P., M.D.
Associate Professor, Health Sciences Informatics
Johns Hopkins University
Baltimore, MD

LIDDY, Elizabeth D., Ph.D., Trustee Professor
Center for Natural Language Processing
School of Information Studies
Syracuse University
Syracuse, NY

MARCHIONINI, Gary J., Ph.D.
Cary C. Boshamer Professor
School of Information and Library Science
University of North Carolina at Chapel Hill
Chapel Hill, NC

NADKARNI, Prakash M., M.D.
Associate Professor, Department of Anesthesiology
Center for Medical Informatics
Yale University School of Medicine
New Haven, CT

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

PANI, John R., Ph.D., Associate Professor
Department of Psychological & Brain Sciences
University of Louisville
Louisville, KY

PRATT, Wanda, Ph.D., Associate Professor
Department of Biomedical and Health Informatics
University of Washington, School of Medicine
The Information School
Seattle, WA

SALTZ, Joel H., M.D., Ph.D.
Professor and Chair, Department of Biomedical Informatics
Ohio State University
Columbus, OH

SHEDLOCK, James, A., M.L.S.
Director, Galter Health Sciences Library
Feinberg School of Medicine
Northwestern University
Chicago, IL

SPACKMAN, Kent A., M.D., Ph.D.
Professor, Department of Pathology
Oregon Health and Science University
Portland, OR

TAIRA, Ricky K., Ph.D., Associate Professor
Department of Radiological Sciences and Medical
Informatics
University of California, Los Angeles
Los Angeles, CA

TANJI, Virginia M., Med, M.S.L.S.
Director, Health Sciences Library
John A. Burns School of Medicine
University of Hawaii at Manoa
Honolulu, HI

TEMPLETON, Etheldra, M.L.S.
Executive Director
Library & Educational Information Systems
Philadelphia College of Osteopathic Medicine

Philadelphia, PA

TONELLATO, Peter J., Ph.D.
Chief Scientific Officer
POINTONE Systems, LLC
Wauwatosa, WI

WALKER, James M., M.D.
Chief Medical Information Officer
Geisinger Health System
Danville, PA

WARD, Deborah, M.A., M.L.S.
Director, Health Sciences Libraries
University of Missouri-Columbia
Columbia, MO

ZHOU, Z. Hong, Ph.D., Associate Professor
Departments of Pathology and Laboratory Medicine
University of Texas Health Science Center
Houston, TX

Appendix 6: Literature Selection Technical Review Committee

The Literature Selection Technical Review Committee meets three times a year to select journals for indexing in MEDLINE.

Members:

MCCLURE**,** Lucretia W., M.A. (Chair)
Special Assistant to the Director
Countway Library of Medicine
Harvard Medical School
Boston, MA

BAUCHNER, Howard, M.D., Professor
Professor of Pediatrics and Public Health
Boston University School of Medicine
Boston, MA

DELCLOS, George L., M.D.
Professor of Environmental & Occupational Health
University of Texas Health Science Center
Houston, TX

DOSWELL, Willa M., Ph.D.
Associate Professor, School of Nursing
University of Pittsburgh
Pittsburgh, PA

FLEMING, David A., M.D., Associate Professor
Department of Health Management & Informatics
Director, Center for Health Ethics
University of Missouri
Columbia, MO

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

KAPLAN, Jerry, Ph.D.
Professor of Pathology

University of Utah School of Medicine
Salt Lake City, UT

MANNING, Phil, M.D.
Professor of Medicine Emeritus
(University of Southern California)
Corona del Mar, CA

RACZ, Gabor B., M.D.
Grover Murray Professor & Director of Pain Services
Department of Anesthesiology
Texas Tech University Health Sciences Center
Lubbock, TX

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

SOEHNER, Catherine B., M.L.S.
Director of Science Libraries
University of Michigan
Ann Arbor, MI

SPANN, Melvin, Ph.D.
Retired
Silver Spring, MD

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

VAN PEENEN, Hubert J., M.D.
Professor of Pathology (Retired)
Eugene, OR

Appendix 7:   PubMed Central National Advisory Committee

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

Members:

KILEY, Robert J., M.S. (Chair)
Head, Systems Strategy
Wellcome Library
London, England

ADLER, Prue S., M.S.
Associate Executive Director
Association of Research Libraries
Washington, D.C.

ALIRE, Camila, Ed.D.
Dean Emerita, University Libraries
University of New Mexico & Colorado State U.
Sedalia, CO

BAKER, Shirley K., M.A.
Dean and Vice Chancellor
Libraries and Information Technology
Washington University
St. Louis, MO

GREENSTEIN, Daniel, D.Phil.
University Librarian
California Digital Library
Oakland, CA

HAWLEY, John, B.A.
Executive Director
American Society for Clinical Investigation
Ann Arbor, MI

KOHANE, Isaac S., M.D.
Director, Information Programs
Children's Hospital
Boston, MA

MICHALAK, Sarah, MLS, Professor
School of Information & Library Science
University of North Carolina
Chapel Hill, NC

PARTHASARATHY, Hemai, Ph.D.
Managing Editor, PloS Biology
Public Library of Science
San Francisco, CA

RYAN, Mary L., M.P.H., M.L.S
Director, UAMS Library
University of Arkansas for Medical Sciences
Little Rock, AR

SO, Anthony D., M.D., Director
Program on Global Health and Technology Access
Terry Sanford Institute of Public Policy
Duke University
Durham, NC

SOBEL, Mark E., M.D., Ph.D.
Executive Officer
American Society for Investigative Pathology
Bethesda, MD

VELTEROP, Johannes, Ph.D.
Director of Open Access
Springer Publishing
Guildford, Surrey
United Kingdom

WARD, Gary E., Ph.D., Associate Professor
Department of Microbiology & Molecular Genetics
University of Vermont
Burlington, VT

WILBANKS, John T., B.A.
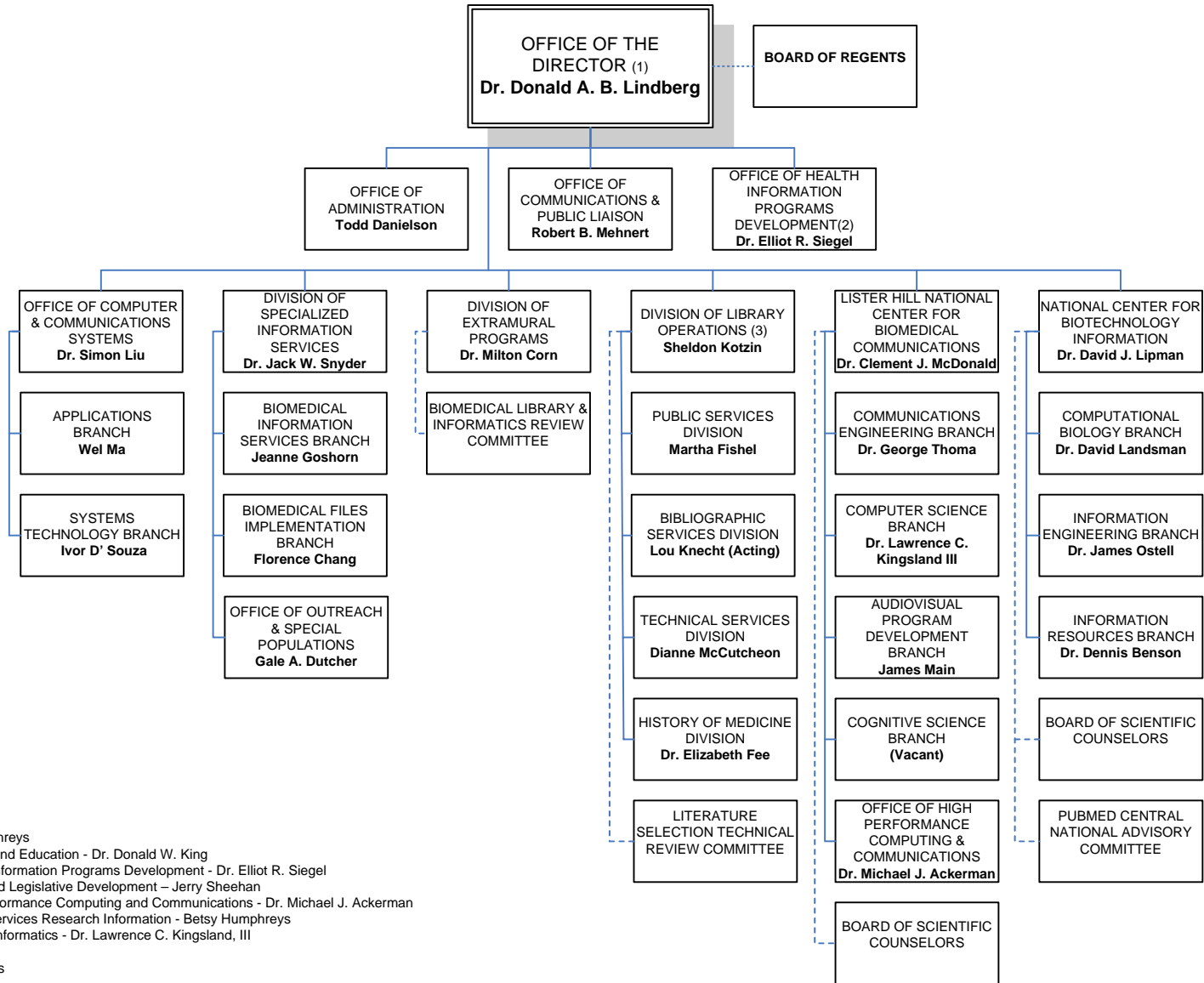Executive Director, Science Commons
South Boston, MA

Appendix 8: Organizational Acronyms And Initialisms Used In This Report

AAHSL   Association of Academic Health Sciences Libraries
ACP   American College of Physicians
ACSI   American Customer Satisfaction Index
AHRQ   Agency for Healthcare Research and Quality
ALTBIB   Alternatives to Animal Testing
AME   Automated Metadata Extraction
AMIA   American Medical Informatics Association
AMPA   American Medical Publishers Association
AMWA   American Medical Women's Association
APDB   Audiovisual Program Development Branch
BISTI   Biomedical Information Science and Technology Initiative
BLAST   Basic Local Alignment Search Tool
BLIRC   Biomedical Library and Informatics Review Committee
BOR   Board of Regents
BSD   Bibliographic Services Division
BSN   Bioinformatics Support Network
CAS   Collection Access Section
CBIR   Content-Based Image Retrieval
CCB   Configuration Control Board
CCDS   Consensus CoDing Sequence
CCRIS   Chemical Carcinogenesis Research Information System
CDD   Conserved Domain Database
CEB   Communications Engineering Branch
CgSB   Cognitive Science Branch
ChemIDplus   Chemical Identification File
CIT   Center for Information Technology
CoreBio   Core Bioinformatics Facility
CPT   Current Procedural Terminology
CRISP   Computer Retrieval of Information on Scientific Projects
CSB   Computer Science Branch
CSI   Commission on Systemic Interoperability
CSR   Center for Scientific Review
CT   Computer Tomography
DART   Developmental and Reproductive Toxicology
dbMHC   Database for the Major Histocompatability Complex
DDBJ   DNA Data Bank of Japan
DCMS   Data Creation and Maintenance System
DEAS   Division of Extramural Administrative Support
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
EMS   Emergency Medical Services
EP   Extramural Programs
EPA   Environmental Protection Agency
eRA   Electronic Research Administration
EST   Expressed Sequence Tag
ETIC   Environmental Teratology Information Center
FDA   Food and Drug Administration
FIC   Fogarty International Center
FNLM   Friends of the National Library of Medicine
FTE   Full Time Employee
Gbps   Gigabits per Second
GDS   GEO DataSet
GEO   Gene Expression Omnibus
GENSAT   Gene Expression Nervous System Atlas
GHR   Genetics Home Reference
GIS   Geographic Information System
GPS   Global Positioning System
GSA   General Services Administration
GSS   Genome Survey Sequences
GUI   Graphic User Interface
HapMap   Haplotype Map
HAVnet   Haptic Audio Video Network for Education Technology
HBCU   Historically Black Colleges and Universities
HHS   Health and Human Services
HIPAA   Health Insurance Portability and Accounting Act
HLA   Human Leukocyte Antigen
HL7   Health Level Seven, Inc.
HMD   History of Medicine Division
HSDB   Hazardous Substances Data Bank
HPCC   High Performance Computing and Communications
HRSA   Health Resources and Services Administration
HSRProj   Health Services Research Projects
HSRInfo   Health Services Research Information
HSRR   Health Services and Sciences Research Resources
HSTAT   Health Services and Technology Assessment Text
IAIMS   Integrated Advanced Information

	Management Systems		NIA	National Institute on Aging
ICMJE	International Committee of Medical Journal Editors		NIAID	National Institute of Allergy and Infectious Diseases
ICs	Institutes and Centers (of NIH)		NIBIB	National Institute of Biomedical Imaging and Bioengineering
IHM	Images from the History of Medicine		NICHSR	National Information Center on Health Services Research and Health Care Technology
ILL	Interlibrary Loan			
ILS	Integrated Library System			
IRIS	Integrated Risk Information System			
IT	Information Technology		NIDCD	National Institute on Deafness and other Communication Disorders
ITER	International Toxicity Estimates for Risk			
ITK	Insight Toolkit		NIDDK	National Institute of Diabetes, Digestive, and Kidney Diseases
JDI	Journal Descriptor Indexing			
LACT	Drugs and Lactation (database)		NIEHS	National Institute of Environmental Health Sciences
LAN	Local Area Network			
LHC	Lister Hill Center		NIGMS	National Institute of General Medical Sciences
LHNCBC	Lister Hill National Center for Biomedical Communications			
LO	Library Operations		NIH	National Institutes of Health
LOINC	Logical Observations: Identifiers, Names, Codes		NINDS	National Institute of Neurological Disorders and Stroke
LSTRC	Literature Selection Technical Review Committee		NIOSH	National Institute for Occupational Safety and Health
MARG	Medical Article Records Groundtruth		NIST	National Institute of Standards and Technology
MARS	Medical Article Records System			
MDoT	MEDLINE Database on Tap		NLM	National Library of Medicine
MDT	Multimedia Database Tool		NLP	Natural Language Processing system
MEDLARS	Medical Literature Analysis and Retrieval System		NN/LM	National Network of Libraries of Medicine
			NNO	National Network Office
MEDLINE	MEDLARS Online		NOSC	Network Operations and Security Center
MeSH	Medical Subject Headings		NOVA	National Online Volumetric Archive
MHC	Major Histocompatability Complex		NRCBL	National Reference Center for Bioethics Literature
MIM	Multilateral Initiative on Malaria			
MIRS	Medical Information Retrieval System		NSF	National Science Foundation
MLA	Medical Library Association		NTCC	National Online Training Center and Clearinghouse
MLAA	Medical Library Assistance Act			
MMDB	Molecular Modeling DataBase		OAM	Office of Administrative Management
MMS	MEDLARS Management Section		OCCS	Office of Computer and Communications Systems
MMTx	MetaMap Technology Transfer			
MTI	Medical Text Indexer		OCHD	Coordinating Committee on Outreach, Consumer Health and Health Disparities
MTMS	MeSH Translation Management System			
NCBC	National Centers for Biomedical Computing		OCPL	Office of Communications and Public Liaison
NCBI	National Center for Biotechnology Information		OCR	Optical Character Recognition
			OD	Office of the Director
NCCS	NIH Consolidated Collocation Site		OHIPD	Office of Health Information Programs Development
NCI	National Cancer Institute			
NCRR	National Center for Research Resources		OMB	Office of Management and Budget
NCVHS	National Committee on Vital and Health Statistics		OMIA	Online Inheritance in Animals (database)
			OMIM	Online Mendelian Inheritance in Man (database)
NEI	National Eye Institute			
NGI	Next Generation Internet		OMSSA	Open Mass Spectrometry Search Algorithm
NHANES	National Heath and Nutrition Examination Surveys		OSIRIS	Open Source Independent Review and Interpretation System
NHGRI	National Human Genome Research Institute		PAHO	Pan American Health Organization
			PCA	Personal Computer Advisory Committee
			PCR	Polymerase Chain Reaction
NHLBI	National Heart, Lung, and Blood Institute		PDA	Personal Digital Assistant

PDR	Publisher Data Review	SPER	System for the Preservation of Electronic Resources
PDB	Protein Data Bank	SII	Scalable Information Infrastructure
PDF	Portable Document Format	STB	Systems Technology Branch
PDR	Publisher Data Review	STTR	Small Business Technology Transfer Research
PHLI	Public Health Law Information Project		
PHS	Public Health Service	STS	Sequence Tagged Site
PLAWARe	Programmable Layered Architecture With Artistic Rendering	TEHIP	Toxicology and Environmental Health Information Program
PMC	PubMedCentral	TIE	Telemedicine Information Exchange
PRS	Protocol Registration System	TILE	Text to Image Linking Engine
PSD	Public Services Division	TIOP	Toxicology Information Outreach Project
RCSB	Research Collaboratory for Structural Bioinformatics	TOXLINE	Toxicology Information Online
		TOXNET	Toxicology Data Network
RefSeq	Reference Sequence (database)	TPA	Third Party Annotation (database)
RFA	Request for Applications	TRI	The Toxics Release Inventory
RFP	Request for Proposals	TSD	Technical Services Division
RML	Regional Medical Library	TTP	Turning the Pages
RNAi	RNA interference	UMLS	Unified Medical Language System
RRF	Rich Release Format	UPS	Uninterrupted Power Supply
RTECS	Registry of Toxic Effects of Chemical Substances	VAST	Vector Alignment Search Tool
		VHP	Visible Human Project
SBIR	Small Business Innovation Research	WebMIRS	Web-based Medical Information Retrieval System
SEF	Serials Extract File		
SIDA	Swedish International Development Agency	Web-STOC	Web-Services Technology Operations Center
SII	Scalable Information Infrastructure	WGS	Whole Genome Shotgun
SIS	Specialized Information Services	WIISARD	Wireless Internet Information System for Medical Response in Disasters
SMART	Scalable Medical Alert and Response Technology		
		WISER	Wireless Information System for Emergency Responders
SNOMEDCT	Systematized Nomenclature of Medicine Clinical Terms		
		XML	Extensible Markup Language
SOAP	Simple Object Oriented Protocol		

National Library of Medicine

OFFICE OF THE DIRECTOR (1)
Dr. Donald A. B. Lindberg

BOARD OF REGENTS

OFFICE OF ADMINISTRATION
Todd Danielson

OFFICE OF COMMUNICATIONS & PUBLIC LIAISON
Robert B. Mehnert

OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT(2)
Dr. Elliot R. Siegel

OFFICE OF COMPUTER & COMMUNICATIONS SYSTEMS
Dr. Simon Liu

- **APPLICATIONS BRANCH**
  Wel Ma
- **SYSTEMS TECHNOLOGY BRANCH**
  Ivor D' Souza

DIVISION OF SPECIALIZED INFORMATION SERVICES
Dr. Jack W. Snyder

- **BIOMEDICAL INFORMATION SERVICES BRANCH**
  Jeanne Goshorn
- **BIOMEDICAL FILES IMPLEMENTATION BRANCH**
  Florence Chang
- **OFFICE OF OUTREACH & SPECIAL POPULATIONS**
  Gale A. Dutcher

DIVISION OF EXTRAMURAL PROGRAMS
Dr. Milton Corn

- **BIOMEDICAL LIBRARY & INFORMATICS REVIEW COMMITTEE**

DIVISION OF LIBRARY OPERATIONS (3)
Sheldon Kotzin

- **PUBLIC SERVICES DIVISION**
  Martha Fishel
- **BIBLIOGRAPHIC SERVICES DIVISION**
  Lou Knecht (Acting)
- **TECHNICAL SERVICES DIVISION**
  Dianne McCutcheon
- **HISTORY OF MEDICINE DIVISION**
  Dr. Elizabeth Fee
- **LITERATURE SELECTION TECHNICAL REVIEW COMMITTEE**

LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS
Dr. Clement J. McDonald

- **COMMUNICATIONS ENGINEERING BRANCH**
  Dr. George Thoma
- **COMPUTER SCIENCE BRANCH**
  Dr. Lawrence C. Kingsland III
- **AUDIOVISUAL PROGRAM DEVELOPMENT BRANCH**
  James Main
- **COGNITIVE SCIENCE BRANCH**
  (Vacant)
- **OFFICE OF HIGH PERFORMANCE COMPUTING & COMMUNICATIONS**
  Dr. Michael J. Ackerman
- **BOARD OF SCIENTIFIC COUNSELORS**

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION
Dr. David J. Lipman

- **COMPUTATIONAL BIOLOGY BRANCH**
  Dr. David Landsman
- **INFORMATION ENGINEERING BRANCH**
  Dr. James Ostell
- **INFORMATION RESOURCES BRANCH**
  Dr. Dennis Benson
- **BOARD OF SCIENTIFIC COUNSELORS**
- **PUBMED CENTRAL NATIONAL ADVISORY COMMITTEE**

---

1. Deputy Director – Betsy Humphreys
   Deputy Director for Research and Education - Dr. Donald W. King
   Associate Director for Health Information Programs Development - Dr. Elliot R. Siegel
   Assistant Director for Policy and Legislative Development – Jerry Sheehan
   Assistant Director for High Performance Computing and Communications - Dr. Michael J. Ackerman
   Assistant Director for Health Services Research Information - Betsy 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 & Health Care Technology, Head – (Vacant)

2006

➖ ➖ ➖ ➖ Dotted lines indicate connections to advisory committees.