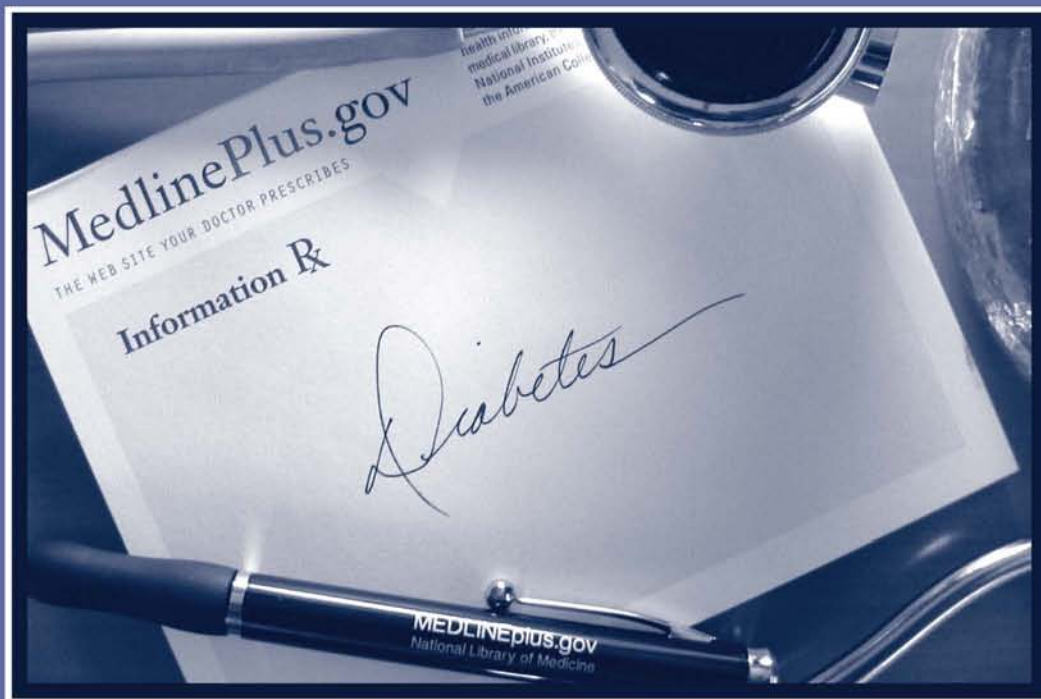# NATIONAL INSTITUTES OF HEALTH
NATIONAL LIBRARY OF MEDICINE
PROGRAMS & SERVICES FY 2004

NIH Publication No. 05-256

U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES

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

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

Cover: Information Rx: "Health Information Prescription" program. A joint project with the American College of Physicians to encourage the use of MedlinePlus by patients.

NATIONAL INSTITUTES OF HEALTH

National Library of Medicine

Programs and Services

Fiscal Year 2004

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

Contents

Tables

Appendixes

Preface

The National Library of Medicine is playing a role of growing importance on the national health information scene. The registration of clinical trials, "open access" for published government-sponsored research information, and creating and providing access to national health data terminology standards are three notable examples in Fiscal Year 2004. Among the advances described by NLM programs in this report:

- ClinicalTrials.gov received Harvard University's prestigious "Oscar" of government awards—the Innovations in American Government Award. ClinicalTrials.gov was created and is maintained by the staff of the Lister Hill Center.
- MedlinePlus® and MedlinePlus en español ranked 1st and 2nd among all U.S. government Web sites in the American Customer Satisfaction Index surveys. The rate of page views of MedlinePlus has more than doubled this year—to 498 million. MedlinePlus is maintained by the Division of Library Operations.
- In August 2004 PubMed reached its 15 millionth citation. That system continues to be a tremendous resource for medical research, logging some two million searches a day. The Library Operations Division and the National Center for Biotechnology Information share the responsibility for maintaining this vital system.
- This year NLM successfully launched NIHSeniorHealth.gov with a demonstration in the Congress. This new information resource is a joint effort of the NLM and the National Institute on Aging.
- The Specialized Information Services Division this year introduced several new services, including the Asian American Health Web site and the Wireless Information System for Emergency Responders (WISER).
- The Office of Computer and Communications Systems established the NIH Consolidated Collocation Site, which provides crucial backup capability for NIH's extensive computer-based operations.

We at the National Library of Medicine are conscious of carrying on a tradition of 169 years of service to the medical community and to the nation. It is a responsibility we are proud to undertake and the accomplishments detailed in this report are the result. I extend my thanks to the Library staff and to the many advisors and consultants we rely on.

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

Office of Health Information Programs Development

Elliot R. Siegel, Ph.D.
Associate Director

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

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

Outreach and Consumer Health

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

NLM Coordinating Committee on Outreach, Consumer Health and Health Disparities
This office has convened and is chairing the NLM Coordinating Committee on Outreach, Consumer Health and Health Disparities. This Committee plans, develops, and coordinates NLM outreach and consumer health activities. A number of the activities described below are conducted under the auspices of the Committee.

American College of Physicians Physician Information Prescription Project
Doctors often prescribe medication after seeing a patient. But what if that doctor also wants to direct the patient to up-to-date, reliable, consumer-friendly information about a health concern? The American College of Physicians (ACP) Foundation teamed with

the NLM in FY2003 to create the "Health Information Prescription" program. Doctors in several pilot states were given customized prescription pads that they can use to point patients to first-rate online health information in NLM's MedlinePlus database.

The Information Rx project was launched nationally on April 22, 2004, the opening day of the American College of Physicians Annual Session in New Orleans. The joint project has been tested in Georgia and Iowa by more than 500 ACP internists and their patients. Among a variety of feedback tools yielding important findings, pre- and post-tests found that 97 percent of the participating internists made information referrals, with 59 percent using the prescription pads for information provided by ACPF and NLM. Twenty percent of participating physicians also reported an increase in patients bringing Internet information to the office visit.

Internists who participated in the pilot programs said that MedlinePlus empowers patients (54 percent), explains difficult concepts and procedures (43 percent), and improves patient-physician communication (42 percent). The project was modified for the third stage of the pilot program in Virginia in March 2004 to partner with Virginia librarians; a toolkit to facilitate participation by libraries was developed.

Diabetes: Consumer Health Diabetes Projects
NLM is exploring the use of new information technologies to enable diabetes patients to manage their disease and avoid or delay the onset of costly and debilitating complications, especially patients from minority and medically underserved populations. In particular, we seek to learn how the use of NLM's MedlinePlus web site, and other computer-based health information resources, can be helpful to patients, their families, and members of the public to learn about and understand the latest research news on diabetes, nutritional requirements, tests, devices, secondary prevention techniques, and for obtaining answers to patient-specific questions.

In the clinical setting, the principal hypothesis is that MedlinePlus can reinforce and supplement the information provided by physicians, nurses and health educators. A related hypothesis is that a combination of individualized training and access to publicly available computer resources at hospital libraries and elsewhere in the community can help reduce the health disparities experienced by minority populations that have less ready access to computer-based health information in the home, school and workplace than the majority population.

The goal is to develop, design, implement and evaluate a comprehensive program of diabetes-focused outreach initiatives in collaboration with academic health science centers and libraries, clinical

centers, community-based organizations and voluntary health organizations.

The latest initiative in collaboration with the Upper Cardozo Health Center in Washington, D.C. and George Washington University, undertakes a controlled field experiment with patients enrolled in the Diabetes Health Disparities Collaborative wherein diabetes patients will receive an individualized information technology-based intervention to complement their regular patient education program. Education and training will be given to (a) physicians on how to incorporate the Information Prescription Project into their daily office practice routine; (b) medical assistants to show patients how to access and use MedlinePlus in English and Spanish; and (c) clerks on how to instruct patients to use MedlinePlus in waiting rooms during clinic hours. A wealth of routinely collected clinical patient data will give NLM a unique opportunity to provide objective evidence of the impact of MedlinePlus use on patient health outcomes.

Web Evaluation
The Internet and World Wide Web now play a dominant role in disseminating NLM information services. And the web environment in which NLM operates is rapidly changing and intensely competitive. These two factors combined suggested the need for a more comprehensive and dynamic NLM Web planning and evaluation process. The continuing Web evaluation priorities of the OCHD include: a) quantitative and qualitative metrics of web usage; and b) measures of customer perception and use of NLM Web sites. During FY2004, the OCHD continued to pursue an integrated approach intended to encourage exchange of information and learning within NLM, and help better inform NLM management decision-making on Web site research, development, and implementation. The year's evaluation activities included: online surveys users of select NLM web sites; several online focus groups; access to a syndicated telephone survey of the 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 FY2004, OHIPD collaborated with other units of NIH to initiate a trans-NIH online user survey project based on the American Customer Satisfaction Index (ACSI), with significant funding support from the NIH Office of Evaluation. The project will extend the ACSI online user survey methodology to about 60 NIH Web sites at about 28 different NIH units, with expected benefits to both. The project will include multi-level evaluation objectives:

- strengthening of each participating component's Web evaluation capability and activity;
- sharing of web evaluation learning and experience on a trans-NIH basis;
- aggregation of ACSI results and learning on a trans-NIH basis; and
- sponsoring an NIH-wide staff workshop that will highlight the contributions and challenges of the ACSI from the NIH perspective, consolidate lessons learned, and identify future directions and opportunities.

The project will be managed by the ACSI Survey Leadership Team consisting of representatives from NLM and several other NIH components who will work closely with representatives from all participating units to be convened under the auspices of the NIH Web Authors Group. The primary ACSI contractor will be ForeseeResults Inc. via an agreement between NLM and the Federal Consulting Group/Department of the Treasury. An evaluation contractor will be selected to provide consultant support to the ACSI Leadership Team and participating NIH units.

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. Also, NLM has funded a Phase 3, in which more intensive community-based outreach and training are being implemented at select Phase 1 and 2 sites to assess if these community-based approaches significantly enhance the project impacts on health information, behavior, and outcomes. The Phase 3 evaluation report is being prepared for publication in 2005. Also NLM has funded a Phase 4 in FY2004 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).

A major new initiative during FY2004 was the planning and implementation of a Native American Listening Circle Project. Listening Circles are a Native American tradition for encouraging dialog and discussion and developing trust among various parties, in this case NLM and representatives of tribes and Native groups. The objectives of the Listening Circles are to promote open dialog between NLM and tribal leaders, share perspectives on each other's capabilities and needs, and identify opportunities for collaborative projects. The idea of Listening Circles was brought to NLM's attention by Dr. Ted Mala, a member of the original Tribal

Connections advisory committee, former President of the Association of American Indian Physicians, currently Director of Traditional Healing for the Southcentral Foundation, and a current member of the NIH Council of Public Representatives. The Listening Circles are consistent with DHHS and White House guidance on Federal agency consultation and coordination with Tribal Governments.

For the Listening Circles, NLM contracted with Cindy Lindquist and the National Indian Women's Health Resource Center to assist with planning and organizing a series of three Listening Circles. Dr. Mala served as a senior advisor. The Resource Center in turn involved local tribal and Native groups in the organizing of each Listening Circle, working in collaboration with the OHIPD. During 2003–2004, three Listening Circles were planned, organized, and implemented: one each in the Dakotas (with American Indians); Hawaii (with Native Hawaiians); and Alaska (with Alaska Natives). The NLM delegation for all three Listening Circles was led by Dr. Donald A.B. Lindberg, accompanied by Dr. Elliot R. Siegel, Dr. Fred B. Wood, Ms. Gale Dutcher, and Dr. Rob Logan. The key facilitators were Ms. Lindquist and Dr. Mala working with local Native organizations.

Also, in 2004 the OHIPD again partnered with NIH and NLM Equal Employment Opportunity offices to participate in the NIH American Indian Pow-Wow Initiative. This included exhibiting at eight pow-wows in the Mid-Atlantic area including one pow-wow each in New Jersey and Pennsylvania and two in North Carolina. Also OHIPD participated with NIH in the Gathering of Nations Powwow in Albuquerque, NM. An estimated 24,000 persons visited the NLM booth over the course of these pow-wows. These activities proved to be another viable way to bring NLM's health information to the attention to segments of the Native American community and the general public.

Outreach to Hispanics
The Lower Rio Grande Valley Hispanic Outreach Project 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. The initial Lower Rio Grande Project is complete. Based on the project results, NLM has funded a series of follow-on projects focusing on outreach to Hispanic populations in South Texas. One follow-on project involves Hispanic residents of the Lower Rio Grande Valley who live in colonias. This project involves

collaboration with Texas A&M University as well as the University of Texas at San Antonio, and its Regional Academic Health Center in Harlingen, TX. Also, OHIPD funded two projects building on the very successful pilot project at MedHigh in the Lower Rio Grande Valley, where high school peer tutors were trained and then in turn taught their peers about health information resources available from NLM. The pilot project received several awards for outstanding performance and service to the students and surrounding communities. The follow-on projects extend the MedHigh concept and collaborations to other high schools in the LRGV and in the San Antonio and Laredo areas.

International Programs

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

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

The web site, http://www.nlm.nih.gov/mimcom, comprises links to Medline®, a variety of free online journals, databases, malaria-related sites, and general information. An NLM reference librarian serves as the webmaster and is expanding the site to include special news releases and articles of interest to researchers.

MIMCom has evolved in order to support the shifting needs of the malaria research sites in Africa. In 1998, the network started with a microwave link to the Internet in Bamako, Mali, and has since assisted 19 other sites in 12 countries. In the intervening years, the telecommunications revolution has moved forward—technology has changed, along with the Internet itself, the latter now bringing us spam and viruses undreamt of in the not too distant past. Conversely, where there was once little or

nothing in terms of telecommunications options, there are now, in some but not all instances, a number of players providing useful services, resulting in competitive pricing. Additionally, some sites have experienced dramatic growth and are no longer properly served by the system as it currently exists.

It has become clear from recent meetings of MIM-TDR and the MIM Funders Forum that the most practical next phase for MIMCom is to focus on medical informatics. We have begun to do this with the Antimalarial Drug Resistance Network; this network has set up a secure server so that researchers can share raw data and post data summaries. It addresses the need for supporting scientists in making innovative use of their new research tools to facilitate new ways of working together.

MIMCom Evaluation
Evaluation continues to be an integral part of MIMCom development. The results of examinations of the project by independent panels and individuals have been critical to assessing the project's effectiveness. Evaluations have concluded that "creation of MIMCom has provided isolated scientists with tools that bring the whole world closer. Reliable communication with collaborators and vastly improved access to the scientific literature have both increased the reach of African scientists and facilitated their participation in the broader scientific world, especially by improving their potential to publish in world-class journals, a key part of being a mainstream scientist."

Another evaluation found that "MIMCom is viewed by many (86%) as one of the most successful and important contributions of the MIM, and it is strongly recommended that MIMCom continues to expand the network to new sites and be involved in the subsequent steps of IT-training and management.... Further expansion and development of associated activities such as more access to online journals and forming web-based research networks was very much desired. The training of on-site IT specialists is identified as very important for the sustainability of a functioning local network."

Another evaluation studied the effects of enhanced connectivity on professional performance of malaria research staff with the use of a Web-based questionnaire. In summary, taking into account the explanations and open answers given in response to the questionnaire, it can be said that enhanced connectivity is generally experienced as a positive contribution to professional performance. It gives access to a world of up-to-date scientific information, facilitates efficient communication, allows effective coordination of research activities and offers improved possibilities for capacity building.

NLM has provided the leadership, resources, and glue for this first phase of MIMCom and has achieved the goals of the initial mandate. With generous support from Swedish SIDA, more sites will be able to benefit from the telecommunications tool. The goal is to promote further capacity building, resulting in strong African research characterized by networks and sharing of information.

African Medical Journal Editors Partnership Project
"I want my journal to be successful, to be a flag bearer for Africa. How do I get there?" James Tumwine, Editor, African Health Sciences

African Medical Journal Editors Partnership Project is a collaboration with the Fogarty International Center in which the International Programs is working closely with Library Operations. The objective is to create four partnerships between four African medical journals and journals from the U.S. (3) and U.K. (2) for the purpose of strengthening the African journals. NLM's specific objective is strengthening African journals so that they are able to get into MEDLINE, and, as a result, make African research available to the world.

Partnerships have been initiated between editors of the following journals: *Ghana Medical Journal* and *The Lancet*; *Malawi Medical Journal* and the *Journal of the American Medical Association*; *African Health Sciences* and *British Medical Journal*; *Mali Medical Journal*, the *American Journal of Public Health* and the *Environmental Health Perspectives*/National Institute of Environmental Health Sciences journal.

Partnerships can encompass sharpening of business, editorial, and technical aspects of: editorial skill development; training for authors, reviewers, and editorial board; sharing manuscripts; joint commissioning of article; exchange of editorial content; staff exchange for skills and experience sharing; increased publication of local research; survey of journal's target audience.

Training
NLM continues to be active in the training of medical librarians, including programs in which the trainers train others. NLM participates in the biannual Association for Health Information and Libraries in Africa conference by offering workshops in PubMed training and sponsoring the travel of African librarians associated with the MIMCom project above. Two librarians from Vietnam came to NLM for training in indexing and MeSH®, so they could begin to make their own collection available to physicians and health workers in that country. NLM's Associate Program has an international fellow who returns home with expertise and resources to carry out projects locally.

International Network Partnerships

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

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

Global Internet Connectivity

End-to-end performance of the Internet, on both national and global scales, continues to be important to NLM in part because the Internet is the primary vehicle for promoting access to and dissemination of health information. This includes the further exploration of the methods and metrics needed to better understand the quality of Internet performance from the end user perspective. NLM is a leader in this field, and several other research and technical organizations now recognize the importance of end-to-end Internet performance. Additionally, NLM has implemented Phase I of its own Internet connectivity performance monitoring network, starting with select U.S. sites (the eight Regional Medical Libraries) but envisioned to extend to other U.S. sites and some international sites in the medium term.

International MEDLARS® Centers

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

AUSTRALIA
National Library of Australia

CANADA
Canada Institute for Scientific and Technical Information (CISTI)

CHINA
Institute of Medical Information Chinese Academy of Medical Sciences

EGYPT
ENSTINET Academy of Scientific Research and Technology

FRANCE
INSERM

GERMANY
German Institute for Medical Documentation and Information

HONG KONG
The Chinese University of Hong Kong

INDIA
National Informatics CenterMinistry of Information Technology

ISRAEL
Hebrew University

ITALY
Istituto Superiore di Sanita

JAPAN
Japan Science and Technology Corporation (JST)

KOREA
Seoul National University

KUWAIT
Kuwait Institute for Medical Specialization

MEXICO
Centro Nacional de Informacion y Documentacion sobre Salud

NORWAY
University of Oslo

RUSSIA
The State Central Scientific Medical Library

SOUTH AFRICA
South African Medical Research Council

SWEDEN
Karolinska Institute Library

UNITED KINGDOM
The British Library

**PAN AMERICAN HEALTH ORGANIZATION
BIREME/PAHO**
Centro Latino Americano e de Caribe
Informcao em Ciencias da Saude

INTERGOVERNMENTAL ORGANIZATION
Science and Technology Information Center
Taipei 10636, Taiwan

International Visitors
In FY2004 the Office of Communications and Public Liaison (and HMD) arranged for 331 tours—107 regular daily (1:30 p.m.) tours and 224 specially arranged tours. There were 6141 visitors in all. They came from the following 67 countries:

Antigua, Australia, Bangladesh, Belgium, Bosnia, Botswana, Brazil, Burundi, Canada, China, Colombia, Côte d'Ivoire, Croatia, Cuba, Ecuador, El Salvador, England, Eritrea, Finland, Georgia, Germany, Ghana, Haiti, Hungary, India, Indonesia, Iran, Ireland, Jamaica, Japan, Kazakhstan, Kenya, Kyrgyzstan, Korea, Malaysia, Mali, Marshall Islands, Mexico, Federated States of Micronesia, Morocco, The Netherlands, New Zealand, Nicaragua, Nigeria, Norway, Pakistan, Palau, Peru, Philippines, Poland, Portugal, Romania, Russia, Serbia, South Africa, Sweden, Switzerland, Tajikistan, Thailand, Trinidad, Turkey, Uganda, United States, Uzbekistan, Vietnam, Zambia, Zimbabwe.

Planning and Analysis

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

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

In September 2004, the NLM Board of Regents approved the initiation of a new effort to develop a Long Range Plan for FY 2005–2010, and appointed a Board Subcommittee on Planning, co-chaired by Hon. Newt Gingrich and Dr. William Stead, to oversee this undertaking.

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. A new Health Disparities Plan for FY2004–2008 was prepared and is available on the NLM web site.

It is important for NLM to be able to describe and analyze its outreach, consumer health, and health disparities projects in order to identify areas of opportunity, report on their progress, and plan for new initiatives. A major activity of the OCHD is the implementation of a database of NLM outreach, consumer health, and health disparities projects. This database, which includes projects from all of the Regional Medical Libraries as well as NLM, is a major source of data for the National Outreach Mapping Center, which is seeking to use mapping as an aid to ensuring the effective distribution of outreach services by the NLM and the National Network of Libraries of Medicine.

In line with its other planning activities, this office worked with senior NLM and Association of Academic Health Science Libraries members to plan a major and well-publicized joint symposium on "The Library as Place: Symposium on Building and Renovating Health Sciences Libraries in the Digital Age," held at NLM November 5–6, 2003. Post-symposium activities in 2004 have included preparation of a DVD "proceedings" and website.

Library Operations

Betsy L. Humphreys
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 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 Center on 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 system features.

Program Planning and Management

LO sets priorities based on the goals and objectives in the NLM Long Range Plan, 2000–2005, and the closely related NLM Strategic Plan to Reduce Racial and Ethnic Disparities. In FY2004, LO contributed to plans for developing a new NLM Long Range Plan for 2006–2016, under the auspices of the NLM Board of Regents. The actual planning sessions, which will involve outside experts and representatives of the Library's many constituent groups, will begin in FY2005.

The current NLM Long Range Plan has a strong focus on the opportunities and challenges arising from electronic publishing and the role of the Web and the Internet in locating and accessing health information. In FY2004, 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. Based on an analysis of work currently performed by the Web Management Team and the Reference and Customer Service Section, LO's Public Services Division (PSD) initiated a reorganization that will merge these two units in FY2005. LO assisted OCCS in developing priorities and schedules for replicating key databases and data creation systems at the new offsite backup computer facility. LO assisted NCBI and the Office of the Director in its work on the development of an NIH Public Access policy involving PubMed Central® by providing a variety of analyses of the volume of articles emanating from NIH-funded research, current journal publication practices and subscription prices, etc. Many other specific projects undertaken to enhance access and handling of electronic information are described throughout this chapter.

In FY2004, LO focused considerable attention on working with other NLM program areas to meet the Library's expanded responsibility for distribution of standard clinical vocabularies within the UMLS® Metathesaurus®. There were major changes and enhancements to UMLS development, distribution, and user support materials that are described in many sections of this chapter and in the LHC chapter. A decision was made to transition responsibility for the Metathesaurus production system from LHC to OCCS, with work to begin in FY2005.

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 to the space and environment for staff, patrons, and physical and electronic collections. In FY2004, LO continued to contribute to plans for a new NLM building and for interim arrangements for housing staff and storing collections until a new building is available. In November 2004, an NLM/AAHSL (Association of Academic Health Sciences Libraries) symposium on "The Library as Place" was held in NLM's Lister Hill Center, which examined trends, issues, and lessons learned in building and renovating libraries and highlighted the continuing need for physical library buildings in an increasingly electronic era. LO co-chaired the Organizing Committee for this highly successful symposium and worked with LHC to produce an interactive DVD version of the proceedings.

In FY2004, LO's Administrative Office continued to assist managers, supervisors and staff with the transition to a range of new administrative systems and the consolidation of human resources functions within the National Institutes of Health. LO continued to encourage staff to take advantage of flexiplace work arrangements as appropriate. Nearly 70 LO employees now 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 FY2004, the NLM collection contained 2,482,585 volumes and 5,469,662 other physical items, including manuscripts, microforms, pictures, audiovisuals, and electronic media.

Selection
In FY2004, NLM completed a total revision of the Collection Development Manual of the National Library of Medicine. Prepared with advice from an external oversight committee chaired by Alison Bunting, former Chair of the NLM Board of Regents, the revised manual is available as an interactive Website, with links to related documents (e.g., the NLM preservation policy, joint NLM/Library of Congress/National Agricultural Library collection statements) and cooperating institutions (e.g., the Kennedy Institute of Ethics Library). The revision of the Collection Development Manual was informed by the deliberations of two working groups established

by the Board of Regents to review NLM's coverage—in its collection, the MeSH vocabulary, and its databases—of the fields of bioethics and of biomedical imaging and bioengineering. Both working groups found that NLM's coverage of these subjects was generally very good, but indicated that improvements were needed to facilitate retrieval of information in these subject areas.

Acquisitions
The Technical Services Division (TSD) received and processed 156,515 contemporary physical items (books, serial issues, audiovisuals, electronic media), which is slightly below last year's total. The increase in electronic publishing has not yet had a significant effect on the number of physical items that NLM acquires. Net totals of 27,101 volumes and 427,921 other items (including nonprint media and manuscripts and pictures acquired by the History of Medicine Division (HMD) were added to the NLM collection. Eighteen libraries offered NLM gifts of retrospective literature; a total of 3,220 journal issues and 1,504 bound volumes were added to the NLM collection as a result.

LO uses subscription agents and book vendors to acquire current literature published around the world. In FY2004, TSD awarded new blanket purchase agreements for monographs to seven U.S. and international vendors. To address the increasing workload and complexity of licensing electronic resources, TSD increased staffing devoted to this activity and worked with NLM's Office of Administration to streamline procedures for reviewing and approving licensing terms.

HMD acquired a wide variety of important printed books, manuscripts and modern archives, images, and historical films during FY2003. Among the books were Fabricius ab Aquapendente, De Respiratione et Eius Intrumentis (Padua, 1615), a work which William Harvey argued against when he later developed his new theory of circulation; Juan Luis Vives, De Anima et Vita Libri Tres (Basel, 1538), a Renaissance work about the relationship of emotions to remembering and forgetting; Relatione dell'Esperienze Fatte In Inghilterra, Francia, ed Italia (Rome, 1668), an extraordinary collection of letters and documents arguing for and against blood transfusion; and Cecilio Follio's Sanguinis a Dextro in Sinistrum Cordis Ventriculum Defluentis Facilis Reperta Via (Venice, 1639), a work contesting William Harvey's new theory of circulation.

Archives and modern manuscript collections acquired included the Food and Drug Administration's "Notice of Judgment Files," a collection 2,679 linear feet documenting fraud prosecuted by the FDA during the first half of the twentieth century. It is one of the largest and most significant additions to the Library in many years. An

index to the collection accompanied it. Other major collections added were the archives of the American Association for the Surgery of Trauma; the papers of Charles Johnson (Dean, Meharry Medical College), and papers of John Watson retired Program Director of the Artificial Heart Program of the NIH's National Heart, Lung, and Blood Institute. Large additions to several existing collections included the papers of James Bosma (audiology), Adrian Kantrowitz (heart-assist devices), and James Harvey Young (history of quackery and regulation of food and drugs). NLM rented space at the University of Maryland Health Sciences Library for temporary storage and processing of several large modern manuscript collections, including the FDA files.

New prints and photograph acquisitions included public health posters and medical ephemera donated by William Helfand, additional photographs by Martha Tabor, and fine art prints by Katherine Du Teil and Rosamond Purcell. New videos and films added included master video tapes of NIH lectures and a large collection of films from the National Hansen's Disease Center, Carville, Louisiana. NICHSR and LHC collaborated to expand the collection of interviews with eminent researchers as part of the effort to document the history of health services research.

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 on more permanent media, conservation of rare and unique items, book repair, maintenance of appropriate environmental and storage conditions, and disaster prevention and response. LO distributes data about NLM's preservation copies to avoid costly duplicate effort by other libraries. LO works with other NLM program areas to develop digital preservation techniques and to promote the use of more permanent media and archival-friendly formats in new biomedical publications.

In FY2004, LO reviewed and revised NLM's preservation priorities. The great majority of the Index Medicus® and Index Catalogue titles that were identified as brittle in a 1985 survey of the Library's print collection have been microfilmed. Although the amount of brittle paper in the NLM collection is still substantial, digitization is emerging as an acceptable alternative to microfilming as appropriate commitments, procedures, and systems for preservation of digital information are established at NLM and elsewhere. In addition, more recent surveys of the condition of NLM's audiovisual and picture collections have highlighted the need to focus more attention on preservation of these materials. Taking these factors into account, LO reallocated

preservation resources to support duplication of historical audiovisuals and increased conservation of prints and photographs. Microfilming will be limited to filling in gaps in filmed Index Medicus and Index Catalogue titles and to other monographs and serials volumes, such as NLM's unusual collection of pre-Revolutionary Russian materials, that are so deteriorated that they are at risk of text loss.

In FY 2004, LO bound 18,311 volumes, microfilmed 2,603 volumes, repaired 1,688 items in the onsite repair and conservation laboratory, made 808 preservation copies of films and audiovisuals, conserved 4,305 prints and photographs and 202 other rare or unique items. Guidelines were developed for selecting post-1970 audiovisuals for duplication, and procedures were established for inspecting newly produced audiovisual copies. A total of 802,069 items were shelved or re-shelved and about 40,000 duplicate unbound journal issues and 7,365 bound volumes were removed from the collection. Stricter inspection procedures were established for bags carried out of the NLM library building.

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, 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 backfiles. LO assists NCBI in soliciting participation of additional journals, particularly in the fields of clinical medicine, health policy, health services research, and public health. In FY2004, LO negotiated the specific terms of an agreement with the Wellcome Trust and the Joint Information Systems Committee in the United Kingdom which will recruit participation of additional journals and fund the digitization of the complete backfiles of these journals. Journals recruited to date are: Annals of Surgery, Biochemical Journal, Journal of Physiology, Medical History, Journal of the Royal Society of Medicine, and the British Journal of General Practice. Negotiations are under way with publishers of other titles.

In FY2004, LO's Public Services Division continued to work closely NCBI to scan and add to

PubMed Central digitized backfiles of journals currently depositing newly published articles in the archive. PSD prepares back issues for scanning, shipped them to the scanning contractor, and manages the human review portion of the quality control of the scanned images, accompanying OCR data, and XML-tagged citations for articles that pre-date current MEDLINE/PubMed coverage. In the initial two years, 25,000 issues have been assembled for scanning, more than 1.8 million pages have been scanned, and 156,000 XML citations created. 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. NLM is particularly grateful to the Marine Biological Laboratory in Woods Hole for donating complete back runs of several titles in FY2004.

NLM is using its own main Web site as a test-bed for procedures and mechanisms for ensuring permanent access to electronic information published by government agencies and private non-profit institutions. With the redesign of NLM's main Web site in FY2004, the Library established an Archives section, which now includes outdated web pages that are important in documenting the history of NLM. Items in the archive are retrievable, but they are segregated and clearly labeled to avoid confusing users about what is currently applicable. In cases where archived items have been replaced by newer versions (e.g., fact sheets), there are "Replaced by" and "Previous version" links between them. When a new NLM Web document is created, a "permanence level" is assigned. Those designated as Permanent with unchanging or stable content will be transferred into the Archive if and when they become outdated. Procedures are currently in place for labeling and archiving html documents on NLM's main web site, using the Teamsite web management software. LO and OCCS are developing mechanisms to handle other types of documents, e.g., PDF, and expect to expand the project to incorporate other NLM web sites in the coming year.

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 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 Standards Committee, provide staff support to the National Committee on Vital and Health Statistics (NCVHS) Standards and Security Subcommittee, and participate in the Public Health Data Standards Consortium. In FY2004, in recognition of the Library's standards activities and expertise in health information technology, the Secretary of Health and Human Services (HHS) acted upon an NCVHS recommendation and designated NLM as the coordinating center for standard clinical terminologies. Funds were transferred to NLM from other HHS agencies to assist with these responsibilities. The Secretary also selected NLM as the operational home of the Commission on Systemic Interoperability, which was established by the Medicare Modernization Act of 2004 to develop a comprehensive strategy for the adoption and implementation of health care information technology standards that includes a timeline and priorities. The Commission is expected to release its final report by the end of calendar 2005.

Medical Subject Headings (MeSH)
The 2005 edition of MeSH contains 22,568 main headings, 83 subheadings or qualifiers, 129 publication types, and more than 146,000 supplementary records for chemicals and other substances. For the 2005 edition, the MeSH Section added 487 new descriptors, replaced 129 descriptors with more up-to-date terminology, deleted 60 descriptors, and added 340 entry terms or "see" references.

The 2005 vocabulary reflects work to reorganize and update the vocabulary for macromolecular substances, including polymers and multiprotein complexes, and intracellular signaling peptides and proteins. Important revisions or additions were made to the terminology for digestive system diseases, cardiomyopathies, endocrine system diseases, morphogenesis, reproduction, and a number of types of organisms. A number of foreign brand names for drugs were added to MeSH supplementary concept records.

MeSH is translated into many other languages by organizations around the world,

including a number of NLM's international MEDLARS partners. In FY2004, LO and OCCS released the first production version of the Web-based MeSH translations database and maintenance system, which can be used by remote translators to improve the currency and accuracy of their translations. The system allows translators to view and translate new terms as they are added by the MeSH Section throughout the year rather than waiting until a complete new edition is released. Three organizations used the system to prepare updated editions of MeSH for 2005. The system is also being used to prepare new translations in additional languages.

In FY2004, the MeSH Section developed and published on the Web files that explicitly document the citation and maintenance procedures that were performed on the MEDLINE database as a result of implementing the new version of MeSH.

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 within the UMLS Metathesaurus. RxNorm was designated as a U.S. government-wide target clinical vocabulary standard by the Secretary of Health and Human Services in 2004. 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 FY2004, RxNorm was linked to additional commercial drug terminology within the UMLS Metathesaurus, and NLM established an agreement with FirstDataBank for regular electronic data feeds to assist in keeping RxNorm up-to-date. LO and OCCS made significant progress on the development of a system that will permit NLM to issue more frequent additions to RxNorm, between editions of the UMLS Metathesaurus. Documentation for RxNorm was published on the NLM Website.

Through LO's NICHSR, NLM supports the continued development and free distribution of LOINC® (Logical Observation Identifiers, Names, Codes) by the Regenstrief Institute. LOINC was designated as a U.S. government-wide target clinical vocabulary standard in 2003. NLM also manages and pays the annual update fees for the U.S.-wide license for Systematized Nomenclature of Medicine. Clinical Terms (SNOMED CT®). In FY2004, work continued on the NLM-commissioned project to examine the overlap between the non-laboratory sections of LOINC and SNOMED CT and to recommend strategies for reducing it.

There is widespread agreement that the existence of authoritative electronic mappings from standard clinical vocabularies to administrative code sets should facilitate automated production of bills and statistical reports as a by-product of the capture of detailed patient data. In FY2004, NLM defined assumptions and parameters for such mappings; enlisted cooperation from relevant federal agencies and private organizations; initiated projects to map LOINC to Current Procedural Terminology (CPT) and SNOMED CT to CPT and to the International Classification of Diseases, 9th edition, Clinical Modification (ICD-9-CM); and began discussions about mappings from SNOMED CT to the Medical Dictionary for Regulatory Affairs and from Medcin to SNOMED CT. To be credible, mapping efforts must: involve both vocabulary producers and intended users, undergo technical review and testing within real clinical systems, and establish effective mechanisms for keeping mappings up-to-date and responding to user feedback.

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 (LHC). In FY2004, Metathesaurus editors accomplished the enormous task of editing the insertion of both the English and Spanish editions of SNOMED CT, the largest single vocabulary ever incorporated into the Metathesaurus. A number of other vocabularies, including several drug vocabularies, were updated in the Metathesaurus. At the close of FY2004, the Metathesaurus contained more than 1 million concepts and 3.8 million concept names from 113 source vocabularies. LO staff assisted LHC in completing the specifications for a new Metathesaurus Rich Release Format, provided in addition to the existing format, that allows completely accurate representation of all relationships present in source vocabularies and supports distribution of purpose-specific mappings between vocabularies. The Bibliographic Services Division coordinated a complete rewrite of the UMLS documentation to reflect the major changes in the Metathesaurus distribution format, a new licensing agreement, use of the Unicode UTF8 character set and associated software tools and assumed full responsibility for publication of the documentation effective with the 2004AC release. (See further information about UMLS activities in the Information Products section of this chapter and the UMLS section of the Lister Hill Center chapter.

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 or selected 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 FY2004, the Cataloging Section cataloged 21,238 contemporary books, serial titles, nonprint items, and cataloging-in-publication galleys, a 6% increase from the previous year. The Section began to provide name authority control for organizations represented in ClinicalTrials.gov, in addition to the similar service already provided for MedlinePlus.gov. With the implementation of the archives of outdated, but important NLM Web documents, the Cataloging Section assumed responsibility for ensuring that all archived documents and those prospectively labeled as permanent have standard and complete metadata and are represented in the NLM catalog.

The Cataloging Section consolidated, expanded, and updated NLM's policies for cataloging subject analysis and classification and published them on the NLM Web site. FY2004 was the first year that the new NLM Classification maintenance system was used to incorporate and validate changes to Classification's MeSH index. The new system allowed the 2004 edition to be released in April, a great improvement from the previous year.

Significant progress was made in providing cataloging records for NLM's historical and special collections. HMD completed a two-year project to catalog a collection of 22,000 unbound pamphlets and also cataloged 152 early monographs, 2,785 pictures, 5,134 historical audiovisuals, and 423 linear feet of manuscripts. New Profiles in Science® Web sites were released for C. Everett Koop, former U.S. Surgeon General, and Wilbur A. Sawyer, a major figure in international public health in the first half of the 20th century. HMD also began a project to create "chapter" cataloging records for medical biographies

present in the American National Biography, the Oxford Dictionary of National Biography, and the Dictionary of Canadian Biography.

Indexing
LO indexes 4,839 biomedical journals for the MEDLINE/PubMed database to assist users in identifying articles on specific biomedical topics. The indexing workload increases steadily, in part due to the selection of additional journals to be indexed, but primarily because of increases in the number of articles published in journals already being indexed. A combination of Index Section staff, contractors, and cooperating U.S. and international institutions indexed 571,000 articles in FY2004, a 9% increase from the previous year. Previously indexed citations were updated to reflect 54 retractions, 5,362 corrections, and 30,678 comments found in subsequently published notices or articles.

In FY2004, indexers created 33,444 annotated links between newly indexed MEDLINE citations for articles describing gene function in selected organisms and corresponding gene records in the NCBI LocusLink database. During the year, additional organisms were incorporated into the gene indexing process, the software supporting gene indexing was improved, and the Index Section participated in testing for the transition from LocusLink to the new Entrez Gene database. The new database will support gene indexing for almost any organism for which information is reported in the published literature.

The Index Section completed installation of dual monitors for all inhouse indexers and began providing dual monitors to contract indexers to speed indexing from the electronic versions of journals. 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. In the case of journals with identical electronic and print versions, indexing from the electronic version frees the print version for immediate use in fulfilling onsite and interlibrary loan document requests.

In FY2004, the Index Section completed the basic data analysis for the indexing consistency study conducted last year and will use the data to establish a baseline for evaluation of continuing efforts to improve automated assistance to the indexing process. Indexer use of the MeSH headings suggested by the Medical Text Indexer system is gradually increasing, and preliminary data indicate that use of the system shortens the time required to train new indexers. Experiments with extracting certain data (e.g., grant numbers) from the full-text of electronic articles indicate that there are other ways to reduce or eliminate certain tasks now performed by human

experts. LO is continuing to work with other NLM program areas to enhance the efficiency and effectiveness of its critical and very high volume indexing operation.

Indexers perform their work after the initial data entry of citations and abstracts has been accomplished. Over the past eight years, great strides have been made in improving the efficiency of data entry. In FY2004, 74% of the citations and abstracts were received from publishers in electronic form (the fastest and most economical method), up from 60% last year; 17% were created by scanning and optical character recognition (OCR); and 9% were double-keyboarded. The combination of increased electronic submissions and enhancements made by LHC to the scanning/OCR system led LO to discontinue the keyboarding contract in June 2004. (Keyboarding was the sole method of indexing data entry from 1967 to 1996.) A total of 315 publishers are now supplying XML-tagged electronic data for 2,966 journals.

NLM selects journals for indexing with the advice of the Literature Selection Technical Review Committee (LSTRC) (Appendix 6), an NIH-chartered committee of outside experts. In FY2004, LSTRC reviewed 473 journals and rated 95 of them highly enough for NLM to begin indexing them immediately. Another 92 titles ranked sufficiently highly to be indexed if their publishers are able to supply electronic citation and abstract data. Following up on the special studies of NLM's coverage of bioethics and of biomedical imaging and bioengineering, the LSTRC reviewed additional journals in these subject areas. NLM implemented a new policy that indexing of electronic-only journals is contingent on their publishers having a credible strategy for ensuring their permanent availability. Deposit in PubMed Central is one way to meet this criterion.

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 African editors in improving the quality of their journals. NLM's role is to improve communications support for African editors so they can use the Internet to recruit authors and reviewers, communicate with editors in other countries, and otherwise become 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 manages the creation, quality assurance, and maintenance of the content of MEDLINE/PubMed, NLM's database of electronic citations; the NLM catalog, which is now 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 ReferenceTM, as well as SIS toxicological, environmental health, and AIDS information services.

In FY2004, LO made significant progress in ongoing efforts to provide online access to NLM's retrospective bibliographic data. Following a multi-year effort, the Library released all five series of the monumental Index-Catalogue of the Library of the Surgeon General's Office in an Encompass database available via the Web. (Encompass is a product of the Endeavor company). Considered an essential resource for the history of medicine and science, Index-Catalogue contains more than 3.7 million entries for books, journal articles, theses, pamphlets, including many not available in other NLM databases. NLM also extended the coverage of PubMed further back in time by adding 243,000 indexed citations from NLM's 1950–1952 printed indexes. Both of these developments improve access to older literature that is newly germane to current health care, including works on smallpox, anthrax, and tuberculosis.

LO and NCBI collaborated to develop a new Entrez database, NLM Catalog, using the new XML catalog distribution format defined and generated by LO and OCCS. The NLM Catalog database was created to provide search capabilities that are not available in LocatorPlusTM, the version of the catalog in the Voyager integrated library system. LocatorPlus will continue to be used for cataloging, onsite circulation, and other library processing functions. The NLM Catalog database does not contain detailed holdings data or provide MARC-formatted output, but it provides links to LocatorPlus for these features.

Use of MEDLINE/PubMed increased to 678 million searches in FY2004, a 35% increase from the previous year, most directly in PubMed and some via the NLM Gateway. Page views totaled 2.5 billion, 39% more than last year. Google is now indexing selected PubMed content, which has contributed to

the growth. MEDLINE/PubMed now includes more than 15 million citations. BSD staff assisted NCBI with design, development, and testing of many enhancements to PubMed and also worked with LHC on the development and testing of many new features in the NLM Gateway. PubMed's Clinical Queries page was updated to reflect refined search strategies developed by Brian Haynes and colleagues at McMaster University. Beta versions of PubMed filters that facilitate retrieval of evidence on cost and outcomes of health services were made available via the NICHSR web site.

Use of MedlinePlus and MedlinePlus en español also continued to increase dramatically. Almost 52 million unique visitors viewed a total of half a billion pages. The number of page views more than doubled and the number of visitors more than tripled in comparison to the previous year. More than 42,000 people subscribe to the weekly announcements of new additions to MedlinePlus content. MedlinePlus and MedlinePlus en español ranked 1st and 2nd among all U.S. government Web sites in the continuous American Customer Satisfaction Index (ACSI) surveys. Last year, PSD worked with SIS and the Office of Health Information Programs Development to obtain NIH evaluation funding for NLM participation in the ACSI program and to implement it as a test for the rest of NIH. Yahoo decided to use an XML file of MedlinePlus health topics, in both English and Spanish, to promote MedlinePlus search results above others due to the quality and authority of the content.

PSD and OCCS continued to expand and improve the content and features of the English and Spanish sites. Forty-seven new health topic pages were added to MedlinePlus to bring the total to 677; 38 were added to MedlinePlus en español for a total of 625. Fifteen new interactive tutorials were added in both languages. Other new features included "Find a Hospital" based on the American Hospital Association database and pages that provide access to all easy-to-read and low vision materials and to English and Spanish materials. "Go Local" was expanded to include a Missouri site that assembles community health service information. NLM released a new Go Local input system for those who wish to build NLM-hosted Go Local sites. A number of groups are actively entering information about local health service Web pages and more Go Local sites are expected to debut by mid-2005. An innovative "talking" version of NIHSeniorHealth was released with additional features and more topics provided by several NIH Institutes.

Under the direction of NICHSR, NLM continues to expand and enhance its databases for health services researchers and public health professionals. In FY2004, NICHSR worked with NCBI to move the entire contents of HSTAT (Health Services and Technology Assessment Text) to the Entrez systems, as part of the Bookshelf. This allows more robust linking between HSTAT documents (including all evidence reports produced by the Agency for Healthcare Research and Quality, CDC's Guide to Preventive Services, etc.), and MEDLINE/PubMed, PubMed Central, and other Entrez databases. NICHSR continued to work through AcademyHealth and the Sheps Center at the University of North Carolina, Chapel Hill to expand the content of HSRProj (Health Services Research Projects) to incorporate work funded by additional foundations and states. Organizations contributing data for the first time in FY2004 included the Idaho Department of Health and Welfare and the states of Kansas and Utah. The HSRR (Health Services and Sciences Research Resources) database also continued to expand to cover additional datasets, surveys, other research instruments, and software packages used with datasets. Among the new resources added were the Health Utilities Index, American Stop Smoking Intervention Study, and the National Children's Study.

HMD is also expanding the Entrez Bookshelf through the "Medicine in the Americas" digital library project, which provides scanned historical American medical books and searchable versions of the texts. In another database effort, HMD has established History of Medicine: Online Syllabus Archive, which already includes 130 syllabi from more than 50 educational institutions in many countries. This new resource has been received with enthusiasm by educators.

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 FY2004, there were 290 MEDLINE licensees, a 32% increase from the previous year. The majority use the data for research and data-mining. LHC and BSD collaborated to produce statistical reports covering the content of the 2002, 2003, and 2004 MEDLINE baseline databases and published them via the Web for use by licensees and other researchers. NLM made its cataloging data available in XML format in FY2004, as an alternative to the MARC format

distribution which has been available since the early 1970s. NLM also redistributed its Chinese cataloging records in MARC format, following completion of the project to add pinyin transliteration to them. A relatively small number of organizations license NLM catalog records or one or more of the SIS toxicological or environmental health files in XML format. Many users execute the online Memorandum of Understanding that permits FTP transfer of the MeSH files in XML, ASCII, or MARC format.

In FY2004, BSD, OCCS, and LHC completely revamped the procedures for licensing UMLS data to allow users to establish licenses via the Web. To obtain the 2004AA release, all UMLS users had to execute a new UMLS license (now applicable to the Metathesaurus only) that incorporates new language covering the terms related to SNOMED CT. As of the end of FY2004, there were 2,115 UMLS Metathesaurus licensees. DVD replaced CD has the hard media distribution mechanism due to the effect of SNOMED CT and the additional distribution format on the size of the Metathesaurus. UMLS users may also obtain the Knowledge Sources and related programs via download, through an application programming interface, or an interactive Web interface, all from the UMLS Knowledge Source Server. During FY2004, BSD staff began assuming an greater role in quality assurance of UMLS releases.

Web and Print Publications
NLM's databases and Web sites are its primary publication media. Demand for the Library's print publications has declined dramatically due to increasing electronic access to NLM data throughout the U.S. and around the world. Reflecting this situation, NLM decided to cease publication of the monthly *Index Medicus*, effective with the December 2004 issue, after 125 years of publication. (The annual *Cumulated Index Medicus* ceased publication in 2000.) Launched by John Shaw Billings in 1879, *Index Medicus* was for many years an indispensable tool for medical librarians, researchers, and practitioners. The desire to publish it in a more timely fashion was the impetus for NLM's pioneering work in automation in the early 1960s, which provided the foundation for the development of MEDLINE in 1971. With the spread of the Internet, the printed *Index Medicus* has outlived its usefulness, but it will survive as a searchable subset within MEDLINE/PubMed. The "Black and White" MeSH, published as a supplement to *Index Medicus* since the 1960s, still receives considerable use as a search tool and will continue to be published in print.

PSD coordinated a complete redesign of NLM's main home page and the secondary pages to which it refers, which debuted in May 2004. The new design is based on feedback from NLM's various customer groups and problems identified in usability testing of the previous version. The new home page has a more flexible three-column format that accommodates news, allows NLM to highlight time-sensitive content, and leads to several different types of portal pages, e.g., for broad subject groupings such as "health services research and public health" and "environmental health and toxicology"; for particular audiences (e.g., public, health professionals, librarians), and for types of NLM services (e.g., training and outreach). In FY2004, NLM's main web site received more than 48 million page hits from users at more than 7.9 million unique Internet addresses. The number of page hits increased 30% from the previous year; the number of unique IP addresses increased 68%.

In conjunction with major changes to the UMLS formats, associated programs, documentation, and licensing procedures, NICHSR consolidated two separate UMLS websites previously maintained by LHC and LO into one revised and expanded set of pages under NLM's main web site. The new UMLS site has an expanded set of resources for UMLS users, including links to information about key UMLS source vocabularies, and prominent links to the UMLS Knowledge Source Server, which is accessible to UMLS licensees only.

Publications available from the main Web site include recurring newsletters and bulletins, fact sheets, technical reports, and documentation for NLM databases. In FY2004, TSD published the List of Serials Indexed for Online Users in XML format for the first time. It was previously available in PDF only. 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.

PSD's Reference and Customer Service Section edits Current Bibliographies in Medicine, a series of special bibliographies on topics of current interest to NLM or other federal agencies. Topics covered this year included health literacy and distance education in public health. In FY2004, PSD reallocated some of the resources previously devoted to Current Bibliographies to other high priority activities, such as periodic systematic review of MedlinePlus topic pages. This change was possible because NIH can now obtain literature search and analysis services for its Consensus Development meetings from the Evidence Centers identified by the Agency for Healthcare Research and Quality (AHRQ). In the past, NLM produced bibliographies for most of these NIH meetings. (The Evidence Reports produced by AHRQ-funded centers are one

of the series that NLM makes available online in the HSTAT collection on the Entrez Bookshelf.).

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 FY2004, PSD's Collection Access Section processed 631,806 requests for contemporary documents. HMD handled 10,031 requests for rare books, manuscripts, pictures, and historical audiovisuals.

The number of onsite users is declining due in part to security measures which make access to NIH facilities more time-consuming and cumbersome, but onsite use of NLM's collection is still significant. Main Reading Room users requested 272,229 contemporary documents from NLM's closed stacks, a 6% decline from last year. Users of the HMD Reading Room requested 8,618 items from the historical and special collections. Paid printing at Main Reading Room workstations increased 31% to 395,915 pages, reflecting significant use of the electronic journals NLM makes available to onsite users. In FY2004, PSD moved the onsite viewing stations for non-print media from the Learning Resource Center to the Main Reading Room. Materials previously shelved in the Learning Resource Center were relocated to the stacks or, in some cases, to the Main Reading Room. Given declining onsite use of non-print materials, the new arrangement provides better service for patrons, is more efficient for staff, and frees up space for other purposes.

The Collection Access Section received 359,577 interlibrary loan requests, a 1% decrease from FY2003, but was able to fill 13,000 more requests than last year. The improvement in fill rate (from 74% to 78%) was due to a collaborative effort with the Index Section and TSD to make more issues of titles indexed for MEDLINE available for document delivery and a major shelf-reading project directed by the Preservation and Collection Management Section. The percentage of requests processed within 12 hours of receipt increased from 80 to 92%. NLM now delivers 92% of interlibrary loan requests electronically. Relais, the system NLM uses to scan and transmit documents, was upgraded to support electronic delivery of documents to libraries behind firewalls. The purchase order for document delivery and first search services was recompeted and a new 5-year procurement awarded.

A total of 3,260 libraries use DOCLINE, NLM's interlibrary loan request and routing system, which received a major interface redesign in FY2004. DOCLINE users entered 2.7 million requests in FY2004, a 5.5% decline from last year; 91% of the requests were filled. Although the absolute number of interlibrary loan requests received by NLM declined slightly in FY2004, the Library's share of all DOCLINE requests continues increase by about half a percent each year—to 13.3% in FY2004. Individuals submitted 809,673 document requests to DOCLINE users via the Loansome Doc® feature in MEDLINE/PubMed and the NLM Gateway, a 6% decline from the previous year. Document request traffic continues to decline in all Regions of the NN/LM due to expanded availability of electronic full-text journals.

In FY2004, NLM expanded and improved the mechanisms for alerting DOCLINE and Loansome Doc users when articles they intend to request are freely available either in PubMed Central or on the Web site of any LinkOutTM provider. This decreased the number of document requests entered by more than 15,000. NCBI and staff at the Regional Medical Libraries continued to promote the use of PubMed's LinkOut for Libraries and "Outside Tool" as 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 31% to 1,091.

DOCLINE requests are routed to libraries automatically based on automated holdings data. At the end of FY2004, DOCLINE's serial holdings database contained 1,401,060 holdings statements for 53,850 serial titles held by 3,049 libraries. In FY2004, LO and OCCS implemented automated transfer of holdings data from OCLC to NLM for NN/LM members who requested this service. Transfer of holdings data from NLM to OCLC was established last year.

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 FY2004, EFTS participation increased 14% to 949 libraries. Participants receive either a single net consolidated bill or a net consolidated payment each month. In FY2004, NLM reviewed the policy for the national maximum charge that Resource Libraries in the NN/LM may levy on network members for filling ILL requests. As a result of this review, resource libraries have the option to conduct a formal study to determine if their actual

costs exceed the national maximum and to charge more if the results justify it. NLM has arranged for resource libraries to make use of the ILL cost study methodology developed by the Association of Research Libraries if they wish to do so.

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 libraries. LO also has primary responsibility for responding to inquiries about NLM's products and services and how to make use of them. With contract assistance, PSD's Reference and Customer 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 107,939 inquiries (excluding spam) were received in FY2004, up 2% from FY2003. The number of onsite inquiries declined 12% to 36,649, reflecting the decline in the number of onsite users. The number of remote inquiries increased 11% to 71,113, with the overwhelming majority arriving via email. NLM uses the Seibel software, integrated with a telephone call system, to track remote inquiries and then applies datamining tools to analyze and characterize customer service inquiries stored (without personal identifiers) in the Seibel database.

PSD also continues to develop the knowledge base of "Cosmo," a virtual customer service representative built with the NativeMinds software designed to answer frequently asked questions about NLM's programs, products, and services. In FY2004, Cosmo responded to 3,678 questions that were within his job description and answered 87% of them correctly, up from 72% last year. Questions that Cosmo can't answer are now transferred, at the user's request, to the Reference staff for response.

In FY2004, PSD conducted customer satisfaction surveys for its telephone and email reference service and revised all Reference and Customer Service fact sheets, FAQs, and Reading Room handouts to reflect "plain language" principles.

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 the secretariat for the Partners in Information Access for the Public Health Workforce; participates in NLM-wide efforts to develop and evaluate outreach programs for under-served 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 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 5666 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, under contract to NLM. (See Appendix 1 for a list of the RMLs.)

In addition to the basic NN/LM contracts and the Electronic Funds Transfer System, NLM funds subcontracts for four national centers that serve the entire network. The activities of one of these centers, the National Online Training Center and Clearinghouse at the New York Academy of Medicine, are described elsewhere in this chapter. The Outreach Evaluation Resource Center at the University of Washington provides training and consulting services throughout the NN/LM and assists in designing methods for measuring overall network programs and individual outreach projects. In FY2004, the Center focused on refining the strategy for measuring progress on network-wide outreach goals for 2001–2006: to bring NLM and NN/LM services to the attention of every public library system and every public health department in the U.S. The National Outreach Mapping Center at Indiana University in Indianapolis assists NLM in displaying the geographic distribution and impact of NN/LM programs and services. In FY2004, work continued on collecting uniform outreach encounter data from all Regions and providing a Web-based tool to the RMLs for use in generating outreach maps. The Web-Services Technology Operations Center (Web-STOC) provides ongoing technical

management of the NN/LM Web sites and also investigates, recommends, and directs the implementation of additional Web technology for teleconferencing, Web broadcasting, distance education, online surveys, etc.

In FY2004, as part of the mid-course evaluation of current NN/LM operations, review teams conducted in-person or audio site visits with the four Centers. Their reports, with recommendations for NLM, the RMLs, and the Centers, will be submitted in early FY2005 in time to be considered in the development for the statement of work for the 2006–2011 NN/LM contracts.

In addition to the work on the public library and public health department outreach goals, the 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. Some projects are identified by individual RMLs through regional solicitations or ongoing interactions with regional institutions; others are identified by periodic national solicitations for outreach proposals issued simultaneously in all NN/LM regions. In FY2004, the NNO initiated a new type of outreach award, the community outreach partnership planning award, to allow health science libraries and community-based organizations to explore opportunities for productive collaboration prior to developing full-fledged outreach project proposals. In all, the NN/LM issued 73 subcontracts for outreach projects in FY2004 as a result of national solicitations. The projects target many rural and inner city communities and special populations in 32 states and the District of Columbia.

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 FY2004, NLM and NN/LM services were exhibited at 150 national, regional, and state meetings across the U.S. These exhibits highlight all NLM services relevant to attendees, not just those to which LO contributes. In FY2004, NLM implemented a new exhibit database to track this activity.

As a result of input from network members at site visits to the 8 RMLs in 2002–2003, NLM and the RMLs established an NN/LM Hospital Internet Access Task Force in FY2003 to identify: barriers to access to the Internet in hospitals; best practices for achieving the twin goals of easy access to the Internet and appropriate security for hospital patient data; and actions the NN/LM and NLM might take to assist hospital libraries in overcoming barriers. In FY2004, the Task Force teamed up with the Hospital Libraries Section of the Medical Library Association (MLA), both to obtain information about barriers and best practices and to disseminate best practices. The Task Force arranged a special open forum on these issues at the MLA Annual Meeting in May 2004.

Also as a result of input from the site visits, NLM and the RMLs established an E-licensing Working Group to identify: state and local group licensing resources available to network members, model licensing language, best practices for negotiating licenses, and methods for disseminating the information to network members. The Working Group, which is also coordinating its efforts with MLA, will submit an initial report in early FY2004.

Partners in Information Access for the Public Health Workforce
The NN/LM is a key member of the Partners in Information Access for the Public Health Workforce, a collaboration 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 Agency for Healthcare Research and Quality and the Medical Library Association are the two newest members, joining 10 other federal agencies and public health associations. 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) provides unified access to public health information resources produced by all members of the Partnership, as well as other reputable organizations. In FY2004, the Web site was migrated from an NN/LM server to one at the NLM. One of the most popular resources on the site is the Healthy People 2010 Information Access project, which includes evidence-based PubMed search strategies and links to MedlinePlus topics for Healthy People 2010 objectives. During FY2004, strategies were completed and tested for objectives in 11 more focus areas, bringing the total number of objectives covered to more than 400, with every focus area represented.

The Partnership also devoted considerable effort to the development of additional training resources. *Public Health Information and Data: A Training Manual* was developed by staff from the New York City Department of Health and Mental Hygiene, the Midcontinental Region of the NNLM®, the University of Michigan, NICHSR, and NNO and made available on the Web site in PDF format. A training course based on its content was conducted at the fall 2004 annual meeting of the American Public Health Association, and a Web-based version of the tutorial is under development.

Special NLM Outreach Initiatives
LO participates actively 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. In FY2004, the Office of the Associate Director and BSD continued to work with other NLM components, the American College of Physicians Foundation, and the NN/LM to launch the Information Rx project nationwide in April at the ACP Annual Session. Information Rx provides physicians with materials to write prescriptions for information from MedlinePlus for their patients. Prior to the national launch, BSD developed an online site for physicians to order their materials and an NLM Library Associate Fellow developed a web-based Information Rx Toolkit for librarians with guidance from NLM staff and input from NN/LM librarians. In FY 2004, a total of 1,450 physicians and librarians requested promotional products for the Information Rx initiative.

The Office of the Associate Director, LO, the NNO, and BSD continued to work with the American Library Association and Public Library Association (PLA) to improve public library awareness of MedlinePlus and MedlinePlus en español. The Office of the Associate Director participated in a panel session at the PLA biennial meeting which focused on web resources available for providing health information to multicultural populations. The Office of the Associate Director also serves on an ALA Advisory Committee for the "Be Well Informed @ Your Library" program which is funding 10 public library systems to conduct seminars on health education issues. BSD staff continued to support a direct mail and library exhibit program to provide all public and health sciences libraries with materials to promote MedlinePlus. The three-year program resulted in 6,318 libraries ordering materials and more than 2.2 million bookmarks were distributed to their readers.

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 FY2004, LO provided summer employment and training opportunities for several students and teachers.

Historical Exhibitions and Programs
HMD directs the development and installation of major historical exhibitions in the NLM rotunda, with assistance from LHC and the Office of the Director. As an important part of NLM's outreach program, the exhibitions are designed to appeal to the interested public, as well as the specialist, and to highlight the Library's rich historical resources. The current exhibition, *Changing the Face of Medicine: Celebrating America's Women Physicians*, debuted on October 14, 2003, with a gala opening program featuring remarks by Dr. Elias Zerhouni, Director of NIH, Dr. Donna Christian-Christensen, delegate from the Virgin Islands, and Dr. Antonia Novello, Commissioner for Health for the State of New York and former U.S. Surgeon-General, and a performance by a string quartet, using instruments made by pediatrics pioneer, Dr. Virginia Apgar.

This well-reviewed exhibition features more than 300 women physicians, living and dead, selected with advice from an advisory committee of eminent physicians (both women and men), chaired by Tenley Albright, M.D., former chair of the NLM Board of Regents. Girls who might be interested in pursuing an M.D. degree are one of the principal audiences for the exhibition, which illustrates the wide range of careers open to women physicians and shows that women from all segments of U.S. society have excelled in the field. The exhibition has a Web site, http://www.nlm.nih.gov/changingthefaceofmedicine, which provides information about the women physicians in the exhibition and educational and professional resources for those considering a career in medicine. The "Share Your Story" section of the Web site encourages people to provide information about outstanding women physicians they have encountered, whether family members, mentors, or their own doctors. To date, more than 6,200 visitors have seen the exhibition at NLM and the accompanying Web site has received 880,000 page hits. The American Library Association and NLM are collaborating on the development of a traveling version, funded by the NIH Office of Research on Women's Health and the Library.

Previous NLM exhibitions live on through heavily used Web sites, printed catalogs, DVD editions, or touring traveling versions. Excluding *Changing the Face of Medicine*, exhibition web sites received more than 4.6 million page hits in FY 2004. The traveling version of *Frankenstein: Penetrating the Secrets of Nature* continued its two-year tour of public, academic, and health sciences libraries across the United States under the auspices of the American

Library Association and garnered favorable publicity at every stop.

In addition to the major exhibitions in the NLM rotunda, HMD installs "mini-exhibits," generally in the cases near the entrance to the HMD Reading Room. Mini-exhibits mounted in FY 2004 included: John Eisenberg: A Life of Service, 1946-2002; C. Everett Koop: From Pediatric Surgeon to Surgeon General; and Time, Tide, and Tonics: The Patent Medicine Almanac in America; and Francisco Goya at the National Library of Medicine, an exhibition of NLM's 13 Goya prints. The Exhibition program also produced a traveling ten-panel exhibit entitled An Odyssey of Knowledge: Medieval Manuscripts and Early Printed Books from the National Library of Medicine. It premiered at the International Congress of Medical History in Bari, Italy in September 2004 and will go on tour.

In November 2003, HMD hosted a major symposium on Visual Culture and Public Health, which featured presentations by invited scholars who drew heavily on NLM's collections. Other historical programs include a monthly series of seminars by historical scholars and several special historical lectures organized by HMD in conjunction with the Diversity Council and the EEO Office. HMD also hosted a number of visiting historical scholars.

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, including the recurring features "Voices from the Past" and "Images of Health" for the American Journal of Public Health, which often feature 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 Online 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 Association of Academic Health Sciences Libraries, and the Association of Research Libraries to increase the diversity of those entering the profession, to provide leadership development opportunities, to promote multi-institution evaluation of library services, and to encourage specialist roles for health sciences librarians.

In FY 2004, the MEDLARS Management Section (MMS) and the NTCC trained 948 students in 76 classes covering PubMed, the NLM Gateway/ClinicalTrials.gov, TOXNET®, and the UMLS. Experiments with remote broadcasts of online training sessions as a means of providing training in more locations were only partially successful so NLM and the NN/LM are investigating other approaches to filling this need. An average of about 31,000 unique users visited the Web-based PubMed Tutorial about 40,000 times each month. Three new animated Viewlet tutorials were created for basic PubMed search features. The PubMed tutorial files were made available on NLM's ftp server in response to a request from the Life Science Library, Academica Sinica, Taipei, Taiwan. The UMLS for Librarians course was revised to reflect the new Metathesaurus release format and greatly enhanced MetaMorphoSys program. LHC and MMS staff presented a revised UMLS tutorial for informaticians at MedInfo in San Francisco in September 2004.

The UMLS Courses are one of a number of NLM training courses useful in preparing librarians for new and expanded roles. LO and the NTCC 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 14 participants in FY 2004: six 2nd year Associates at sites across the country and eight 1st year Fellows, who completed their year at NLM in August 2004. Seven of the latter also chose to participate in the optional 2nd year of the program at sites across the country: the University of Massachusetts, Georgetown University, the University of New Mexico, Johns Hopkins University, the Centers for Disease Control and Prevention, the University of Texas at San Antonio, and the University of Washington. Seven new Fellows began a year at NLM in September, including one International Fellow from the Medical School Library, University of Zambia.

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 5 to 10 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 5 mid-career health sciences librarians. AAHSL contracts with ARL for the leadership training portion of the program. Based on the success of the first two years of the initial three year pilot, LO has decided to fund the program for an additional three.

Table 1

Growth of Collections

Collection	Previous Total (9/30/03)	Added FY 2004	New Total (9/30/04)
Book Materials			
Monographs:			
Before 1500	588	3	591
1501-1600	5,938	25	5,963
1601-1700	10,221	13	10,234
1701-1800	24,637	18	24,655
1801-1870	41,424	36	41,460
Americana	2,341	0	2,341
1871-Present	727,462	14,834	742,296
Theses (historical)	281,794	0	281,794
Pamphlets	172,021	0	172,021
Bound serial volumes	1,269,541	19,301	1,288,842
Volumes withdrawn	(80,483)	(7,129)	(87,612)
Total volumes	2,455,484	27,101	2,482,585
Nonbook Materials			
Microforms:			
Reels of microfilm	137,442	4,625	142,067
Number of microfiche	447,374	3,029	450,403
Total microforms	584,816	7,654	592,470
Audiovisuals	72,965	1,738	74,703
Computer software	2,243	132	2,375
Pictures	58,010	2,422	60,432
Manuscripts	4,323,707	415,975	4,739,682*
Total nonbook	5,041,741	427,921	5,469,662
Total book & nonbook	7,497,225	455,022	7,952,247

* Equivalent to 2,708 linear feet.

Table 2

Acquisition Statistics

Acquisitions	FY 2002	FY 2003	FY 2004
Serial titles received	20,350	20,476	20,769
Publications processed:			
Serial pieces	133,908	134,579	132,192
Other	22,274	24,523	24,323
Total	156,182	159,102	156,515
Obligations for:			
Publications	$5,802,023	$6,217,417	$6,942,747
(For rare books)	($446,039)	($297,894)	($300,831)

Table 3
Cataloging Statistics

	FY 2002	FY 2003	FY 2004
Completed Cataloging	21,419	19,927	21,238

Table 4
Bibliographic Services

Services	FY 2002	FY 2003	FY 2004
Citations published in MEDLINE	502,056	526,338	571,000
For *Index Medicus*	459,558	492,911	537,469
Journals indexed for MEDLINE	4,538	4,697	4,839
Journals indexed for *Index Medicus*	3,834	3,994	4,189
Total items archived in PubMed Central	72,683	109,910	347,680

Table 5
Consumer Web Services

Services	FY 2002	FY 2003	FY 2004
NLM Web Home Page			
Page Views	40,607,752	37,166,023	48,335,875
Unique Visitors	5,300,363	4,792,482	7,934,966
MedlinePlus			
Page Views	116,335,454	214,127,932	498,702,940
Unique Visitors	9,594,429	16,356,444	51,724,958
ClinicalTrials.gov			
Page Views	23,288,683	26,010,359	33,651,851
Unique Visitors	1,422,734	2,387,487	3,190,813
Genetics Home Reference			
Page Views	***	***	8,410,455
Unique Visitors (daily average)			25,617
Household Products Database			
Page Views	***	***	7,096,664
Unique Visitors			1,364,649
Tox Town			
Page Views	***	***	1,732,336
Unique Visitors			365,383

Table 6
Circulation Statistics

Activity	FY 2002	FY 2003	FY 2004
Requests Received	705,069	653,916	631,806
Interlibrary Loan	373,292	363,352	359,577
Onsite	331,777	290,564	272,229
Requests Filled:	539,274	511,032	510,751
Interlibrary Loan	268,816	268,714	281,543
Onsite	270,458	242,318	229,208

Table 7

Online Searches—PubMed and NLM Gateway

	FY 2002	FY 2003	FY 2004
Total online searches	382,000,000	504,000,000*	678,000,000

*Corrected figure

Table 8

Reference and Customer Services

Activity	FY 2002	FY 2003	FY 2004
Offsite requests	49,153	64,010	71,290
Onsite requests	48,395	41,774	36,649
Total	97,548	105,784	107,939

Table 9

Preservation Activities

Activity	FY 2002	FY 2003	FY 2004
Volumes bound	25,609	15,646	18,311
Volumes microfilmed	5,255	2,795	2,603
Volumes repaired onsite	1,542	1,285	1,652
Audiovisuals preserved	283	500	795
Historical volumes conserved	66	111	197

Table 10

History of Medicine Activities

Activity	FY 2002	FY 2003	FY 2004
Acquisitions:			
Books	424	314	498
Modern manuscripts	840,000	498,750	5,516,000*
Prints and photographs	3,176	1,000	1,591
Historical audiovisuals	1,361	97	757
Processing:			
Books cataloged	368	215	13,621
Modern manuscripts cataloged	984,025	203,000	740,250**
Pictures cataloged	0	1,048	2,758
Citations indexed	846	856	5,134
Public Services:			
Reference questions answered	14,898	14,693	18,701
Onsite requests filled	6,870	16,163	8,618

*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 Toxicology and Environmental Health Information Program (TEHIP), known originally as the Toxicology Information Program, was established 35 years ago within the NLM's Division of Specialized Information Services (SIS). Over the years TEHIP has provided for the increasing need for toxicological and environmental health information by taking advantage of new computer and communication technologies to provide more rapid and effective access to a wider audience. We continue to move beyond the bounds of the physical National Library of Medicine, exploring ways to point and link users to relevant sources of toxicological and environmental health information wherever these sources may reside. Resources include chemical and environmental health databases and Web-based information resource collections. The Division's HIV/AIDS information initiative now includes several collaborative efforts in information resource development and deployment, including a focus on the information needs of other special populations.

The SIS Web server provides a central point of access for the varied programs, activities, and services of the Division. Through this server (http://sis.nlm.nih.gov), users can access interactive retrieval services in toxicology and environmental health, HIV/AIDS information, or special population health information; find program descriptions and documentation; or be connected to outside related sources. Continuous refinements and additions to our Web-based systems are made to allow easy access to the wide range of information collected by this Division. Web usage has continued to increase over the past year.

In FY2004 SIS continued to balance efforts to enhance and re-engineer existing information resources with efforts to provide new services in emerging areas. We further developed various prototypes that rely on geographical information systems, innovative access and interfaces for consumers, and graphical display of data from information sources. Highlights for 2004 include:

- WISER, or Wireless Information System for Emergency Responders, a tool designed to provide critical chemical information quickly and conveniently on a Personal Digital Assistant (PDA) for use by emergency responders, especially during the first 24 hours in a "hot-zone";

- ITER, or International Toxicity Estimates for Risk, a resource that presents chemical risk information from authoritative groups worldwide, including the U.S. Environmental Protection Agency, the U.S. Agency for Toxic Substances and Disease Registry, Health Canada, the Dutch National Institute of Public Health and the Environment, and the International Agency for Research on Cancer, as well as independent parties whose risk values have undergone peer review;

- TOXMAP, a prototype system that uses maps of the United States to help users visually view data about chemicals released into the environment and easily connect to related environmental health information;

- ChemIDPlus Lite, a streamlined version of ChemIDPlus, that allows users to retrieve relevant substance records simply by typing chemical names or registry numbers into a single search box;

- A new Special Topic Web resource page that provides information on education, careers, and outreach programs in toxicology and environmental health;

- A new Special Topic information portal devoted to issues affecting the health and well-being of Native Americans;

- Continued support of PAHO/NLM Disaster Preparedness Information Centers in Honduras, Nicaragua, and El Salvador;

- Expanded Native American outreach initiatives; and

- Continuing minority outreach activities with the Historic Black Colleges and Universities, United Negro College Fund Special Projects, and the National Medical Association.

Resource Building

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

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

specific brands and in what percentage; which products contain specified chemicals; who manufactures a specific brand and how can that manufacturer be contacted; what are the potential acute and chronic health effects of the chemical ingredients found in a specific brand; what other information is available about such chemicals in the toxicology-related databases of the National Library of Medicine?

In FY2004, SIS released **TOXMAP**, a prototype system that uses maps of the United States to help users visualize data about chemicals released into the environment. TOXMAP integrates data from the EPA's Toxic Release Inventory (TRI) with information about health effects, research citations, etc. found in TOXNET databases. Users can create nationwide or local area maps that show where chemicals are released into the air, water, and ground. TOXMAP also integrates data from other sources, such as demographic data from Census Bureau. TOXMAP provides region-specific links to chemical and bibliographic information.

In FY2004, SIS also released **WISER** (Wireless Information System for Emergency Responders), designed to provide critical chemical information quickly and conveniently on a Personal Digital Assistant for use by emergency responders (first 24 hours in hot-zone). The application is being developed in partnership with the Agency for Toxic Substances and Disease Registry, using ATSDR Medical Management Guidelines for Acute Chemical Exposures, which were developed to aid emergency department physicians and other emergency health care professionals who manage acute exposure following chemical incidents. The WISER prototype has focused on approximately 400 agents found in the Hazardous Substances Data Bank, and current deployment plans include a user's guide, a tutorial, evaluation methodology, and "in-field" testing.

ITER (International Toxicity Estimates for Risk) is a TOXNET data file that contains data in support of human health risk assessments. It is compiled by Toxicology Excellence for Risk Assessment (TERA) and contains over 600 chemical records with key data from the Agency for Toxic Substances & Disease Registry (ATSDR), Health Canada, National Institute of Public Health & the Environment (The Netherlands), U.S. Environmental Protection Agency, and independent parties whose risk values have undergone peer review. ITER provides a comparison of international risk assessment information in a side-by-side format and explains differences in risk values derived by different organizations. ITER data, focusing on hazard identification and dose-response assessment, is

extracted from each agency's assessment and contains links to the source documentation. Among the key data provided in ITER are ATSDR's minimal risk levels; Health Canada's tolerable intakes/concentrations and tumorigenic doses/concentrations; EPA's carcinogen classifications, unit risks, slope factors, oral reference doses, and inhalation reference concentrations; maximum permissible risk levels; and non-cancer and/or cancer risk values (that have undergone peer review) derived by independent parties.

Haz-Map database, released in 2002 at http://hazmap.nlm.nih.gov, is an occupational toxicology database designed to link jobs and hazardous job tasks to occupational diseases and their symptoms. It is a relational database of chemicals, jobs, and diseases that averaged nearly 20,000 queries per month in 2004. A user may search this occupational database by chemical agent, occupational disease and by job type.

ChemIDplus (Chemical Identification File) is an NLM online chemical dictionary, which contains nearly 370,000 records, primarily describing chemicals of biomedical and regulatory importance, and available to users on the Internet at http://chem.sis.nlm.nih.gov/chemidplus. ChemIDplus features include chemical structure search and display for over 200,000 chemicals, and hyperlinked locator fields that retrieve data for a given chemical from other resources such as TOXLINE®, MEDLINE or HSDB® as well as EPA and ATSDR. Over 15,000 records of regulatory interest collectively known as SUPERLIST are also available and hyperlinked in ChemIDplus. During FY2004 over 75,000 queries per month were made of this database. To assist with spelling errors, a chemical spell checker helps users retrieve substances more efficiently by chemical name. The checker, which can be instantly revised using the SIS DBMaint2 online update system, contains spelling indices for more than 1.3 million chemical names and synonyms. The database was enhanced by the addition of various new locators pointing to international resources. In FY2004, the new ChemIDplus "Lite" and "Heavy" systems were released with new capabilities, including a simpler Web front-end that does not require plug-ins for structure display, and an advanced version that allows numeric searching by acute toxicity data and effect, and chemical/physical properties.

The Hazardous Substances Data Bank (HSDB) continues to be a highly used resource, averaging 60,000–70,000 searches each month (a 5% increase over FY2003). Increased emphasis continues to be placed on providing more data on human toxicology

and clinical medicine within HSDB, in keeping with past recommendations of the Board of Regents' Subcommittee on TEHIP. In 2004, there has also been a continued emphasis on adding to HSDB new chemicals with the potential for high toxicity and high human exposure. Approximately 100 new chemicals were added in 2004, including new pesticides, drugs, and environmental pollutants. The emphasis on the addition of new chemicals will continue in the coming year. Newer sources of relevant data are being examined for incorporation into new and existing data fields within the current 4,757 HSDB records. Special summary information is being prepared to allow easier presentation of information at a health consumer level. The process of developing a new Web-based system for HSDB creation, review, and maintenance is continuing. As part of this effort, a relational HSDB database was created and a new client-server interface was programmed to allow easier updates. The new maintenance system is now poised for integration with other new features, including numeric searching and automatic indexing.

The **Toxicology Data Network** (TOXNET), NLM's information system providing database management for many of its toxicology files, has moved from a networked microprocessor environment to a UNIX-based platform (Solaris Version 2.6) on a SUN Enterprise 3000 computer. SIS continues to integrate this configuration with other database creation systems and Web access to them. Further refinements of the SIS search interface (http://toxnet.nlm.nih.gov) enhance the ability of users to simultaneously search HSDB, TOXLINE, CCRIS, Gene-Tox, DART®/ETIC, IRIS, TRI and ChemIDplus from one input screen. Based on recommendations from the Institute of Medicine, users are presented with a basic search screen with just a single input box for searching, with customized screens for more sophisticated users. These advanced features include Boolean searching and the ability to limit search terms to specific fields. Feedback from TOXNET user online surveys has provided a basis for current and future planning, and as result, SIS will implement a chemical spellchecker, automated indexing, and a virtual meta-search tool during the coming years.

Alternatives to Animal Testing (ALTBIB)—SIS continues to compile and publish references from the MEDLARS files that were identified as relevant to methods or procedures that could be used to reduce, refine, or replace animals in biomedical research and toxicological testing. Staff members search, edit, and categorize citations to create a true value-added resource in this field. The 22 bibliographies issued during the past ten years are available on the Internet

through the SIS Web server, and the primary distribution mechanism for this project is now the Internet, through a new online resource named ALTBIB, which allows search access to all of the 7,595 citations organized from previous bibliographies. ALTBIB uses the TOXNET search engine, and is available at http://toxnet.nlm.nih.gov/altbib.html. A user may search by keyword, author, or one of the 16 subdivisions such as "Quantitative Structure Activity Studies."

TOXLINE (Toxicology Information Online) is a large NLM bibliographic database traditionally produced by merging "toxicology" subsets from secondary sources. By the end of FY2004, the database included over 3 million citations to toxicology literature dating back to 1965. In 2004, users accessed standard journal literature in toxicology and environmental health as part of the enlarging MEDLINE database, while NLM continued to add journals in the area of toxicology and environmental health to MEDLINE to cover some of the literature formerly provided by outside sources. For the non-standard journal literature in this area, SIS further enhanced a Web-based system on TOXNET that allows efficient acquisition and updating of these components. Easy access to this TOXLINE Special database and to TOXLINE Core, the standard journal literature on PubMed, is available from the improved TOXNET user interface.

DIRLINE® (Directory of Information Resources Online) is NLM's online directory of resources including organizations, databases, bulletin boards, as well as projects and programs with special biomedical subject focus. These resources provide information to users which may not be available from one of the other NLM bibliographic or factual databases. DIRLINE continues to receive a high level of use (nearly 7000 searches per month) through an interface that supports direct links to the Web sites of the organizations listed in the database, as well as direct e-mail connections. The quality and utility of the database continue to improve as duplicates have been eliminated through changes in policy and streamlining of maintenance. More than 1000 records were either revised or verified in FY2004. Health Hotlines, the always popular publication of health-related toll-free telephone numbers, has a recently updated Web version which also indicates the availability of Spanish speaking customer service representatives and Spanish language publications from the resources listed.

The **Toxics Release Inventory** (TRI) series of files now includes on-line files TRI86 through TRI2002. These files remain an important resource for

environmental release data and are a useful complement to other SIS databases. Mandated by the Emergency Planning and Community Right-to-Know Act, these EPA databases contain environmental release data for air, water, and soil for over 600 EPA-specified chemicals. These files are used in the new SIS R&D project using a geographical information system, TOXMAP.

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

The **Integrated Risk Information System** (IRIS), EPA's official health risk assessment file, continues to experience high usage and be very popular with the user community. EPA has had a version of IRIS on the agency's Web page since 1996, and we will continue to consider how best to integrate our Web service with what EPA provides. IRIS now contains 540 chemicals.

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

The **Developmental and Reproductive Toxicology** (DART) database now contains over 240,000 citations from literature published since 1989 on agents that may cause birth defects. DART is a continuation of the Environmental Teratology Information Center backfile database. In FY2004, next generation DART consisted of two subsets: DART Core on PubMed, containing over 170,000 citations to the journal literature, and DART Special, containing nearly 70,000 citations to specialized resources (including meeting abstracts, books, technical reports) in this subject area. In FY2004, more than 500 new records were added, and easy access to DART Special and to DART Core was maintained at the new TOXNET interface. DART is funded by NLM, the EPA, the National Institute of Environmental Health Sciences (NIEHS), and the FDA's National Center for Toxicological Research, and is managed by NLM.

The **Environmental Mutagen Information Center** (EMIC) database contains over 24,000 citations to literature on agents that have been tested for genotoxic activity. A backfile for EMIC (EMICBACK) contains over 75,000 citations to the literature published from 1950–1991.

Handheld computer devices known as Personal Digital Assistants (PDAs) are increasingly being used in the fields of toxicology and environmental health. Moreover, software applications covering specialized subject matter in these fields are increasingly being made available to PDA users. In an effort to provide information on the main technical and content features of selected applications, the SIS has undertaken an ongoing Review of PDA Applications in Toxicology and Environmental Health. Individual reports in the review series are usually based on free, downloadable demos. Each individual review typically covers the following topics: general information, intended users, authorship/data source, contents, navigation, requirements, application type/price, availability, useful web links, and updates.

AIDS Information Services

NLM remains as the project manager for the multi-agency AIDS Clinical Trials Information Service (ACTIS) and the HIV/AIDS Treatment Information Service (ATIS), which were merged in December 2002 into a service entitled "AIDSinfo." This service provides access to AIDS-related clinical trials information (through Clinicaltrials.gov) and federally approved treatment guidelines. The contract for this service also provides support services for Clinicaltrials.gov.

Evaluation of the AIDSinfo service (accuracy monitoring) was completed in FY2004 with the goal of assisting federal agencies in determining the future direction of the service, including the web site. The number of Live Help interactions continues to grow; users of this service find it very helpful in learning to navigate and locate information on the AIDSinfo, ClinicalTrials.gov, and SIS web sites. The usage level of the consumer fact sheets also continues to grow; the number of PDF downloads for these documents averages more than 10,000 per month, and project staff continue to evaluate options for optimizing the guidelines documents for PDAs.

Other Interagency Initiatives
In FY2004, SIS personnel continued their leadership of the Interagency Tox-to-Consumer Initiative, which completed an Inventory of Federal Government Consumer Environmental Health Resources.

Evaluation Activities
With funding from the NIH Office of Evaluation, SIS is using the American Customer Satisfaction Index (ACSI) to evaluate user satisfaction with AIDSinfo and TOXNET. Starting in November 2004, the Index

provides continuous results that evaluate all aspects of the user's Web experience based on an online survey. Results are benchmarked against other federal Web sites and against private industry, and the results are published quarterly in the mainstream and trade press. AIDSinfo and TOXNET show very strong results, especially for their primary users: toxicologists, physicians, chemists and scientists. In response to feedback from the survey, TOXNET has made changes to its home page, enhanced ChemIDplus content, and uses the survey data to decide priorities for site improvement. AIDSinfo used the survey results to guide a home page redesign, improve search function, and better address the HIV/AIDS information needs of students and the public. SIS has a lead role in using the ACSI at NIH and was instrumental in expanding its use to an additional 60 NIH web sites through a 2004–6 project funded by the Office of Evaluation.

Outreach / User Support

Special Population Web Sites: The Arctic Health web site (http://arctichealth.nlm.nih.gov), initially developed by SIS staff, is now updated by the University of Alaska at Anchorage; the Asian-American Health web site will now be updated with assistance from the Asian American Pacific Islander Health Forum, and the Native American Health web site has been released. These Web sites include relevant policy, legislative, and organizational information as well as organized links to health and environmental issues of concern to the designated population.

NLM-Tox-Enviro-Health-L listserv was created in June 2003 to send announcements-only about SIS's toxicology and environmental health programs and resources. Messages sent to the nearly 1200 subscribers include lists of new chemicals added to Hazardous Substances Databank, announcements about the new Household Products Database, and new environmental health topics for consumers added to Tox Town or MedlinePlus. The MedlinePlus Environmental Health listserv, created in FY2004, now sends messages to nearly 1400 subscribers.

In FY2003, the Toxicology Information Outreach Panel (TIOP) evolved a new strategic plan and was renamed the Environmental Health Information Outreach Panel (EnHIOP). Dr. Henry Lewis, Dean of the School of Pharmacy at Florida A&M University, became Chair of the new group. The new EnHIOP includes representation from additional Historically Black Colleges and Universities (HBCUs) as well as from Tribal Colleges and Hispanic Serving Educational Institutions. In FY2004, the panel members met

twice, and individual awards of $5000 were made to 15 of the institutions participating in EnHIOP.

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

In FY2004, SIS continued its support of the Regional Disaster Information Center for Latin America and the Caribbean (CRID) to strengthen the capacity to collect, index, manage, store, and disseminate public health and medical information related to disasters. The countries involved are Nicaragua, Honduras, and El Salvador. The main objective of this project is to contribute to disaster reduction by capacity building activities in the area of disaster-related information management. Selected libraries and information centers have been provided with the knowledge, training and technology resources in order to act as reliable information providers to health professionals and others in their countries. Through this initiative, the participating libraries and information centers have been strengthened in several areas:

- Technological Infrastructure (Internet connectivity and computer equipment)
- Information Management (Health science librarian training)
- Information Product Development (Digital Library, Web sites)

This project is also assisting SIS in developing models for collecting and exchanging health information in geographically isolated and disaster-prone environments and for handling non-traditional or unpublished literature, in this case on the health aspects of disasters.

SIS exhibited at over 40 conferences in FY2004. Several of these provided opportunities for presentations or workshops about NLM's information resources. In addition, SIS hosted the UNCFSP e-Health Conference for HBCU's, Empowerment for Health Information, in Bethesda, Maryland.

Research and Development Initiatives

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

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

connections between chemicals, the environment, and the public's health. Tox Town is designed to provide:

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

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

ToxSeek provides a virtual meta-search tool for simultaneous searching of target information systems, displaying search results from targeted systems, and harvesting related concepts. This tool can be configured to define a set of target information/search tools, which for SIS are T&EH databases and searchable resources on the web. Testing of the prototype is underway and a beta version will be ready for public release in FY2005.

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 an ongoing participatory project in which voluntary representatives from participating nations provide crucial input and feedback to assure credible and high-quality sources of information. With support from the Fogarty International Center, this project is scheduled to release fully developed information resources from approximately 15 nations in FY2005.

The Automated Indexing Project for selected HSDB data fields continues to identify appropriate search terms to use in comparing retrieval performance of the MeSH-indexed and non-MeSH-indexed versions of HSDB. Retrieval testing and evaluation has begun, with further work to be completed in FY2005.

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

Lister Hill National Center for Biomedical Communications

Alexa T. McCray, Ph.D.
Director

The Lister Hill National Center for Biomedical Communications, established by a joint resolution of the United States Congress in 1968, is a research and development division of the U.S. National Library of Medicine. Seeking to improve access to high quality biomedical information for individuals around the world, the Center continues its active research and development in support of NLM's mission. The Center conducts and supports research and development in the dissemination of high quality imagery, medical language processing, high-speed access to biomedical information, intelligent database systems development, multimedia visualization, knowledge management, data mining and machine-assisted indexing. An external Board of Scientific Counselors meets biannually to review the Center's research 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. The work of each is described below. An organization chart with the names of Branch and Office Chiefs is on the inside back cover of this report.

Organization

Computer Science Branch
The Computer Science Branch (CSB) applies techniques of computer science and information science to problems in the representation, retrieval and manipulation of biomedical knowledge. CSB projects involve both basic and applied research in such areas as intelligent gateway systems for simultaneous searching in multiple databases, intelligent agent technology, knowledge management, the merging of thesauri and controlled vocabularies, data mining, and machine-assisted indexing for information classification and retrieval. Research issues include knowledge representation, knowledge base structure, knowledge acquisition, and the human-machine interface for complex systems. Important components of the research include embedded intelligence systems that combine local reasoning with access to large-scale online databanks. CSB research staff include the team that has developed NLM's Gateway, the team that annually produces the Unified Medical Language System Metathesaurus, and the staff who coordinate the Center's training programs. The most current information about the Computer Science Branch can be found at http://lhncbc.nlm.nih.gov/csb/.

Cognitive Science Branch
The Cognitive Science Branch (CgSB) conducts research and development in computer and information technologies. Important research areas encompass the investigation of a variety of techniques, including linguistic, statistical, and knowledge-based methods for improving access to biomedical information. Branch members actively participate in the UMLS project and collaborate with other NLM research staff in the Indexing Initiative project, the goal of which is to develop automated and semi-automated techniques for indexing the biomedical literature. The branch also conducts research in digital libraries and collaborates with NLM's History of Medicine Division on Profiles in Science, a project to digitize the archival collections of prominent biomedical scientists. Several branch projects address the challenges involved in providing health information to consumers. ClinicalTrials.gov is an important resource for the public and, additionally, serves as a testbed for conducting consumer health informatics research, and the Genetics Home Reference provides complex information about genes and diseases to the public in easily understood language. The most current information about the Cognitive Science Branch may be found at http://lhncbc.nlm.nih.gov/cgsb/.

Communications Engineering Branch
The Communications Engineering Branch (CEB) is engaged in applied research and development in image engineering and communications engineering motivated by NLM's mission-critical tasks such as document delivery, archiving, automated production of MEDLINE records, Internet access to biomedical multimedia databases, and imaging applications in support of medical educational packages employing digitized radiographic, anatomic, and other imagery. In addition to applied research, the branch also developed and maintains operational systems for

production of bibliographic records for NLM's flagship database, MEDLINE. Research areas include content-based image indexing and retrieval of biomedical images, document image analysis and understanding, image compression, image enhancement, image feature identification and extraction, image segmentation, image retrieval by image content, image transmission and video conferencing over networks implemented via asynchronous transfer mode and satellite technologies, optical character recognition, and man-machine interface design applied to automated data entry. CEB also maintains archives of large numbers of digitized spine x-rays and bit-mapped document images that are used for intramural and outside research purposes. Iinformation about the Communications Engineering Branch can be found at http://lhncbc.nlm.nih.gov/ceb/.

Audiovisual Program Development Branch
The Audiovisual Program Development Branch (APDB) conducts media development activities with several specific objectives. As its most significant effort, the branch participates in the Center's research, development, and demonstration projects with high quality video, audio, imaging, and graphics materials. From initial project concept through project implementation and final evaluation, a variety of forms and formats of visuals are developed, and staff activities include image creation, editing, enhancement, transfer and display. Consultation and materials development are also provided by the branch for NLM's other information programs. From applications of optical media technologies and teleconferencing to support for Web distribution, the requirement for graphics, video, and audio materials continues to increase in quantity and diversification of format. In addition to the development of new techniques and processes, the facilities and hardware infrastructure must reflect state-of-the-art standards in a rapidly changing field. Included within APDB is the Office of the Public Health Service Historian. The office preserves and disseminates information about the history of Federal efforts devoted to public health. The most current information about the Audiovisual Program Development Branch can be found at http://lhncbc.nlm.nih.gov/apdb/.

Office of High Performance Computing and Communications
The Office of High Performance Computing and Communications (OHPCC) serves as the focal point for NLM's High Performance Computing and Communications (HPCC) activities. OHPCC coordinates NLM's HPCC planning, research and development activities with Federal, industrial, academic, and commercial organizations while collaborating with Lister Hill Center research

branches and NLM divisions in the development, operation, evaluation and demonstration of HPCC research programs and projects. In addition, OHPCC plans, coordinates, and administers the interagency HPCC research and development program. Office staff serve as NLM's liaison to scientific organizations at all levels of national, state and international government on planning and implementing research in High Performance Computing and Communications. The major research activities of the Office center on the Visible Human Project®, NLM's Next Generation Internet program, telemedicine, the HPCC Collaboratory, and the 3D informatics research program. The most current information about the Office of High Performance Computing and Communications can be found at http://lhncbc.nlm.nih.gov/ohpcc/.

Training Opportunities at the Lister Hill Center

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 FY 2004 the Center provided training to 53 participants from 17 states and 8 countries. Participants worked on projects in the areas of biomedical knowledge discovery, content-based image retrieval, consumer health systems research, document imaging, image database research, information retrieval research, medical illustration, natural language systems, ontology research, hand-held technology, Web services research, user interface research, telemedicine, ubiquitous computing, Unified Medical Language System research, and visualization research. The Center continues to offer a successful NIH Clinical Elective in Medical Informatics for third and fourth year medical students. The elective provides an overview of the state-of-the-art of medical informatics in a lecture series by nationally and internationally known speakers, and offers an opportunity for independent research under the mentorship of expert NIH research staff. 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. Established in 2001, the NLM Rotation Program continues to

grow. The eight-week rotation program for trainees from NLM funded Medical Informatics programs provides these individuals an opportunity to learn about NLM programs and current Lister Hill Center research. The rotation includes a series of lectures and the opportunity for students to work closely with established scientists and meet fellows from other NLM funded programs.

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

Language and Knowledge Processing

Developing SPECIALIST, an experimental natural language processing system for the biomedical domain, is the focus of the Center's natural language processing work. The SPECIALIST system includes several modules based on the major components of natural language: lexicon, morphology, syntax, and semantics. The lexicon and morphological component are concerned with the structure of words and the rules of word formation. The syntactic component addresses the constituent structure of phrases and sentences, while the semantic component seeks to extract biomedical content from text. All components of the SPECIALIST system rely heavily on the domain knowledge in the Unified Medical Language System Knowledge Sources.

Terminology Research and Services
Lister Hill Center research staff build and maintain the SPECIALIST Lexicon, a large syntactic lexicon of medical and general English terminology released annually with the UMLS Knowledge Sources. New lexical items are continually added to the Lexicon using a lexicon building tool, LexBuild, developed and maintained by the lexical systems research team. LexBuild allows researchers to enter items directly into a central database via a Web browser. A new version of LexBuild featuring internal checks to prevent common data entry mistakes and logical inconsistencies was deployed in FY2004. The FY2005 SPECIALIST Lexicon release tables will be generated entirely using the new LexBuild tool. The SPECIALIST Lexicon increased by over 32% to 242,000 lexical items in the FY2004 release. Lexical access tools are also distributed as open source resources with each UMLS release. During this past year the group also developed several tools to manage diverse vocabularies for a range of language and information processing purposes. The team recently achieved a significant milestone in providing customized UMLS data to several projects, including ClinicalTrials.gov, Profiles in Science, and the

Genetics Home Reference. A number of internal tools were also developed to handle data customization. These incorporate UMLS updates and provide client applications with periodic releases of customized data and the latest terminology enhancements.

Semantic Knowledge Representation
Innovative methods for providing more effective access to biomedical information depend on reliable representation of the knowledge contained in text. The Semantic Knowledge Representation project develops programs that extract usable semantic information from biomedical text by building on existing NLM resources, including the UMLS knowledge sources and the natural language processing tools provided by the SPECIALIST system. Two programs in particular, MetaMap and SemRep, are being used to address a variety of problems in biomedical language and information processing. MetaMap maps noun phrases in free text to concepts in the UMLS Metathesaurus. The MetaMap Technology Transfer program (MMTx) is an exportable, Java-based version of MetaMap that runs under Windows, Mac OS X or Unix/Linux and is provided as a resource to the bioinformatics community. Users are able to create MMTx data files independently of the UMLS. MetaMap Technology Transfer source code is included in the MMTx release, and an error reporting and tracking system ensures that problems reported by users are effectively addressed.

SemRep is a tool that uses the Semantic Network to determine the relationship asserted between concepts developed in MetaMap. SemRep serves as the basis for ongoing research initiatives in biomedical information management, such as projects for extracting medical and molecular biology information from text, processing clinical data in patient records, and research in knowledge summarization and visualization. Recent enhancements to SemRep's linguistic coverage include the addition of a mechanism for interpreting hypernymic propositions. Current work addresses arguments of nominalizations, comparative structures, and coordination of predicates. Semantic predications produced by SemRep serve as the basis for continued work in automatic abstraction summarization of biomedical text, including MEDLINE citations and an online encyclopedia. SemGen, a modification of SemRep, is being developed for identifying and extracting semantic propositions on the causal interaction of genes and diseases from MEDLINE citations. Project staff are also developing methods for automatically suggesting appropriate images as illustrations for anatomically oriented text.

Indexing Initiative

The Indexing Initiative investigates concept-based indexing methods for the automatic selection of subject headings in both semi-automated and fully automated indexing environments at the NLM. The goal of the Indexing Initiative is to obtain retrieval performance equal to or better than performance of systems using manually assigned index terms. A prototype indexing system for testing indexing methods, the Medical Text Indexer (MTI), is being tested by NLM indexers. MTI recommendations are available to all indexers as an additional resource available through NLM's Data Creation and Maintenance System. In addition, results of the MTI system are being used as keywords for AIDS/HIV, health sciences research, and space life sciences collections of meeting abstracts that are not manually indexed.

On-going improvements to MTI continue to be made. Short-term, incremental changes arise from requests made by indexing staff or by a desire to incorporate more of NLM indexing policy into the system. Longer term goals include a word sense disambiguation effort to improve MTI's accuracy. The team has also begun to investigate the use of the full text of articles in addition to their work with MEDLINE titles and abstracts. Additional work investigates an 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) corresponding to biomedical specialties. The JD system associates JDs with words in titles and abstracts in a three-year training set of 1,378,597 MEDLINE records. Each record "inherits" the JDs from the journal in the record. A word in the training set can then be described by a list of JDs ranked according to the number of co-occurrences between the word and the JDs. Text as input to the system can be indexed based on averaging the word-JD co-occurrences for the words in the text that are also in the training set, ranking the JDs in decreasing order of these averages. The journal descriptor approach 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 NLM's participation in the Text Retrieval Conference (TREC 2004).

Unified Medical Language System

Unified Medical Language System research regularly develops and distributes multi-purpose, electronic knowledge sources and associated lexical programs. The Metathesaurus, Semantic Network and SPECIALIST Lexicon are used by system developers to enhance patient data, create digital libraries, retrieve Web and bibliographic data, apply natural language processing, and improve decision support.

The Metathesaurus represents multiple biomedical vocabularies organized as concepts in a common format providing a rich terminology resource in which terms and vocabularies are linked by meaning. The Semantic Network allows users to investigate relationships among semantic types and relations and retrieve a list of Metathesaurus concepts assigned to a particular semantic type. Finally, the data in the SPECIALIST Lexicon provides users with the syntactic and morphologic information about each of its lexical items.

The Metathesaurus continues to grow in size, scope, and mission. As of FY2004, there are more than 1 million concepts with 5 million names from 117 source vocabularies in 15 languages. The Metathesaurus is now released in a new "Rich Release" format that 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. Following the July 2003 announcement by the Secretary, HHS of a government license for nationwide use of SNOMED CT, this widely used standard vocabulary for US clinical medicine has been added to the Metathesaurus. The Metathesaurus installation and configuration program called MetamorphoSys has been enhanced to offer easy extraction of pre-computed subsets, for example all HIPAA (Health Insurance Portability and Accountability Act) vocabularies, or selected natural language processing names. This feature will assist users in many areas including regulatory compliance in electronic medical records. The Metathesaurus team has successfully met several new challenges including meeting increasing demand for frequent updates; developing methodologies for mappings between vocabularies; and the development of tools to meet the changing needs of an expanding community, especially of clinical users.

A significant change in the method of delivery of the UMLS Knowledge Sources to users has occurred along with the increase in size of the Metathesaurus to 18 gigabytes. Approximately one third of all users now access the UMLS through the UMLS Knowledge Source Server, one third request the files on DVD-ROM, and one third download the full Knowledge Sources online. The Metathesaurus group has developed a multi-platform Java program that allows users to decompress, customize, and install the Knowledge Sources on local machines, and has added browsers for users who create local subsets.

Modeling and Learning Methods

The Modeling and Learning Methods project seeks to develop new modeling methods that enable

researchers to rapidly construct effective computational models from large datasets. The objectives of the project are to develop machine learning methods that automate the process of constructing probabilistic models for identifying relevant information among large datasets, mapping identified information to networks of ontologies, accessing queried information accurately, and answering user queries through mining the data located in heterogeneous information sources. Interest in probabilistic models ranges over a wide spectrum of biomedical fields, including computational biology; biomedical, clinical, and healthcare informatics; and epidemiology. The objectives of the project will be evaluated with a set of suitable metrics such as receiver operating characteristic that measure the performance of prospective models in terms of sensitivity and specificity in reaching their target functions. Depending on the domain of the models and the problems of interest, domain subjects and/or experts might be needed to determine the gold standards or the target functions for the performance evaluations of the models if such gold standards or target functions are not readily available. FY2004 research focused on identifying information represented in textual data (e.g., MEDLINE abstracts) using UMLS tools, the SPECIALIST parser, and MetaMap. New computational methods in modeling textual and numerical data are being developed. Staff participated in TREC 2004 and competed in the Physiological Data Modeling Contest at the 21st International Conference on Machine Learning.

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, OpenGA–LEN, the Gene Ontology, and the Foundational Model of Anatomy are being explored as well.

During FY2004, the research team focused on foundational issues and explored the ontological properties of resources such as SNOMED CT and the Foundational Model of Anatomy. Non-lexical approaches to identifying dependence relations in ontologies were studied, with application to acquiring associative relations in the Gene Ontology. A generic framework was also developed for computing semantic similarity from a taxonomy and the

frequency of its nodes in a corpus, applicable to the Gene Ontology, MeSH, and WordNet. Finally, the team pursued work on visualization by developing RxNav, an application for navigating drug information in the RxNorm model. In the future, the research team will investigate a semantic similarity approach to comparing lists of MeSH descriptors assigned to MEDLINE documents and to identifying functionally related gene products annotated with the Gene Ontology.

Image Processing

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

Visible Human Project
The Visible Human Project (VHP) 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. VHP data sets are available through a free license agreement with the NLM. Data sets are distributed to licensees over the Internet at no cost and on DAT tape for a duplication fee. Worldwide use of the data sets continues to grow as they are applied to a wide range of educational, diagnostic, treatment planning, virtual reality, virtual surgeries, artistic, mathematical, and industrial uses by over 2000 licensees in 48 countries. The Visible Human Project has been featured in well over 850 newspaper articles, news and science magazines, and radio and television programs worldwide.

FY2004 saw the continued maintenance of two databases to record information about Visible Human Project use. The first database logs information about VHP license holders and records their plans for using the images. The second database records information about the products that licensees are developing. The Insight Toolkit (ITK), a research and development initiative under the Visible Human Project, completed two official software releases in FY2004. 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. Additional ITK awards have been made to extend the software infrastructure into clinical and research applications through the introduction of database management tools, workbenches for tumor volume measurement for possible use in clinical trials, and

the sponsorship of Web portals for sharing research data and publications. Non-funded researchers are now testing, developing and contributing to ITK in over 30 countries. Research institutions, including the Mayo Clinic, the Imperial College of London, Georgetown University, the University of Utah, Kitware, Harvard University, Cognitica, the University of Pennsylvania, and NLM staff participated in demonstrations and technology exhibits at the 2003 annual Radiological Society of North America conference in Chicago. Tutorials on how to use ITK were presented at the IEEE Vis2003 conference in Seattle, the SPIE Medical Imaging Conference in San Diego, and the MICCAI 2003 conference in Montreal. At the end of FY2004, the NIH Roadmap Initiative for Bioinformatics and Computational Biology awarded a 5-year cooperative agreement to the National Alliance of Medical Image Computing. This $20 million national center for biomedical computing has adopted ITK and its software engineering practices as part of its engineering core.

3D Informatics

During FY2004 the 3D Informatics Program has continued to mature and develop its in-house research efforts around problems encountered in the world of 3-dimensional and higher-dimensional, time-varying imaging. Research is continuing in the areas of image-based implicit rendering, research and systems trials for ITK, and haptic latency analysis for surgical simulation. The team has extended and enhanced its pilot project for creating the framework for an archive of volume image data, the National Online Volumetric Archive. This project includes the physical implementation of the pilot archive for volume image data, as well as a tutorial for data submission, meta-data structure management tools using XML, and Web page structure. The metadata structure management were refined, published and presented at the 2004 SPIE Medical Imaging conference. Research is continuing in an effort to create a software framework for artistic and non-photorealistic rendering of digital models entitled, Programmable Layered Architecture With Artistic Rendering. The framework will consist of a layered software architecture for implementing medical illustration techniques using computer graphics technologies. The framework adopted the infrastructure from the ITK software engineering methodologies in FY2004. Additional work includes: research of implicit surface and its application to surface generation for efficient rendering of anatomic objects, research of finite element modeling and simulation system for human colon straightening and its application in virtual colonography, and research on geometric mapping using the index-check-and-

correct method for human colon polyp detection in helical CT datasets.

In September 2004, the HPCC office together with representatives of the National Institute of Biomedical Imaging and Bioengineering and two directorates at the National Science Foundation, sponsored a workshop on visualization research challenges. The 28-member panel drew national and international participants from industry and academia to begin a discussion on the current grand challenges in visualization and imaging research.

AnatQuest

While the Visible Human images have been used by biomedical scientists and developers worldwide, the goal of this in-house project is to provide widespread access to the Visible Human images for a broad range of users, including the lay public. In line with this goal, a Web-mediated system, AnatQuest (available at anatquest.nlm.nih.gov), was developed. This system is based on a 3-tier architecture in which the first tier consists of Java applets for displaying thumbnails of the cross-section, sagittal and coronal images of the Visible Human Male, from which detailed full-resolution views are accessed. The second tier is a set of servlets that process user requests and compress the requested images prior to shipment back to the user. The third tier is the object-oriented database of high resolution VH images and rendered 3D anatomic objects. Low bandwidth connections are accommodated by a combination of adjustable viewing areas and image compression done on the fly as images are requested. Users may zoom and navigate through the images.

Current work is proceeding in two directions. The first is to increase the number and type of rendered images (beyond the current 300 surface-rendered structures) to make the collection more useful for the public. This would require registering all of the cryosection slice images, segmenting and labeling anatomic structures on each slice, and using these to create surface- and volume-rendered images. The second direction taken in this project addresses a long term NLM goal, that is, 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. This may add value to text resources such as PubMed and MedlinePlus by linking to anatomic images. For this purpose, project staff are developing a modular prototype system (Text to Image Linking Engine, or TILE) to serve as a testbed to investigate the alternatives in the functions needed to accomplish this linkage. These functions involve 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.

WebMIRS

The Web-based Medical Information Retrieval System, a Java application, allows remote users to access data from two surveys conducted by the National Center for Health Statistics. These are the National Heath and Nutrition Examination Surveys II and III (NHANES II and III), carried out 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. Through the WebMIRS graphical user interface, a user may construct a query for the NHANES II or NHANES III data. WebMIRS allows the user to save the returned data to the local disk drive, where it may be analyzed with statistical tools. The WebMIRS NHANES II database also contains vertebral boundary data that was collected by a board-certified radiologist for 550 of the 17,000 x-ray images in WebMIRS. 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.

WebMIRS enhancements include collaborative work with Texas Tech University to develop an advanced compression capability custom tailored to the image characteristics of the x-ray images, to allow delivery of the WebMIRS images in compressed form rather than in the low-resolution form as at present. Software written in Java has been developed for the decompression at four different levels. Work is now under way to improve the performance efficiency of the decompression, before the code is incorporated into the WebMIRS system. Significant progress was made toward the development of the next generation WebMIRS system, the Multimedia Database Tool. This system will provide a software framework for the incorporation of new text/image databases in a much more general way than the current WebMIRS and provide new features for the database end user that extend current WebMIRS capabilities. The specific framework that has been designed has the goal of accommodating new sets of text and images under a flexible database schema and GUI approach that is intended to allow new databases to be incorporated with work done only at the level of the database administrator, and not at the software modification level.

Online X-ray Archive

The complete set of 17,000 NHANES II x-ray images in the full-resolution form in which they were digitized was made publicly available in FY 2000. These images are available by FTP and have been accessed by researchers from both within the U.S. and also from international sites. Staff created the ImViewJ software, a downloadable Java application, which allows users to view images at their full spatial resolutions (e.g., 1463x1755 for the cervical spine images, 2048x2487 for the lumbar spine images). Coordinate data collected under the supervision of a radiologist at Georgetown University are also available on the FTP site for 550 images. This coordinate data defines landmark points for each vertebra in a manner commonly used in the field of vertebral morphometry, and serves as reference data to aid in creating and evaluating the performance of image processing algorithms for segmentation of the vertebrae. Users may access this coordinate data either through the FTP archive or through the WebMIRS system. The number of TIFF 8-bit images publicly available was increased to 1,000 in FY2004.

Content-Based Image Retrieval

The goal of the content-based image retrieval (CBIR) project is to develop methods for effective extraction of biomedical information from biomedical digital images, with the current concentration being on the NHANES II spine x-rays. The focus is both on indexing the image data and search of those data. For example, for the 17,000 NHANES II images, the only indexing data originally available was the collateral alphanumeric data collected in the questionnaires and examinations; no indexing information derived directly from the images was originally available, and the high cost of employing radiological experts to compile such data by physical viewing and interpreting each image makes it unlikely that such information will ever be acquired by purely manual means. These circumstances could be reversed if reliable, biomedically validated software could produce image interpretations automatically. Even in the more likely case that only semi-automated methods should prove feasible, the reduction in labor costs could be sufficient to allow the creation of databases of significant biomedical information where this is not currently economically feasible. This is the motivation for research into computer-assisted image indexing. Computer-assisted image searching is a potential enabler of enhanced information extraction from a database that has already been indexed.

During the current year new and substantially extended CBIR capability was developed with the implementation of the latest version, CBIR3. Highlights of the system are that it can operate in networked or stand-alone modes, uses

XML for reporting, and allows the user to select either a more mature or an experimental version of the system. CBIR3 differs from its predecessors in that all data (text, images, and segmentations) are now stored on a centralized MySQL database. Each user is allocated a unique login that grants certain rights and privileges. The system supports access to multiple data sources that can be selected by the user. CBIR3 also provides a validation sub-mode for expert review, validation, and pathology indication for indexed images. CBIR3 currently allows vertebral shape segmentation using the Modified Active Contour Segmentation and LiveWire segmentation techniques. In addition it has a well defined interface allowing the addition of more techniques. It is now possible to segment images in a database-controlled sequential mode that remembers the user's state when he/she stopped working. The last image and vertebra segmented are saved and automatically brought up the next time the same user segments. Another new feature of CBIR3 is that it allows text searching on the complete NHANES II dataset through the familiar WebMIRS interface. WebMIRS (standalone) has been linked with CBIR3 for allowing hybrid text and image searches. For image queries, CBIR3 supports query by sketch and query by image example. Query shape can be generated by sketch, choosing it from the existing shapes on the database, or by supplying an image and segmenting it to obtain a shape. The query shape can subsequently be edited by moving points, adding points, and removing points.

Digital Archive of Uterine Cervix Images
Work continued in FY2004 towards the creation of an archive database of the 60,000–100,000 digital images of the uterine cervix collected by the National Cancer Institute. This work included analysis of color models, standards and technology for the digital capture of color information from 35 mm slides with high color fidelity, and similar issues related to retaining the color across digital output devices such as monitors and printers. MATLAB programs were created to enable the comparison of images digitized at different scan densities or at different compression levels. A Nikon 4000 slide scanner was acquired, and 200 uterine cervix slides were scanned to generate evaluation data. For each of these slides, a medical expert labeled regions of interest with a MATLAB tool that was developed for that purpose. A compression study was conducted to allow the comparison of uncompressed uterine cervix images with those compressed using the Hybrid Multiscale Vector Quantization method developed by Texas Tech University. Multiple medical experts participated in the study, which used 50 test images compressed at eight different compression levels. A similar study was conducted to determine a suitable

spatial resolution to use for digitizing the 35 mm slide collection.

Engineering Laboratories
The Image Processing Laboratory is equipped with a variety of high end servers, workstations and storage devices connected by a mix of 100 and 1000 Mb/s Ethernet. The laboratory supports the investigation of image processing techniques for both grayscale and color biomedical imagery at high resolution. In addition to computer and communications resources and image processing equipment to capture, process, transmit and display such high-resolution digital images, the laboratory also archives a variety of image content. The equipment includes a Sun Enterprise 4500 server with dual 400 MHz CPUs, and 1.5GB memory, and a SunFire 280R server with dual 1.2 GHz CPUs, 3 GB memory, and two internal 73 GB SCSI disks. Additional computers in the lab include two Sun Ultra 10 workstations, each with a 440 MHz CPU, 512 MB memory, and an external 36 GB SCSI disk; and two Sun Ultra 10s, each with a 300 MHz CPU and 512 MB memory. All of these machines run the Solaris 9 operating system. Large-scale magnetic storage is provided by a Network Appliance FAS960 which is a network-attached storage device connected by redundant Gb/s Ethernet connections and provides 24TB of RAID storage. For the ultra-high-resolution display of x-ray images, two E-systems Megascan monitors provide image display at a spatial resolution of 2048x2560 pixels. The laboratory also contains specialized equipment and software for device calibration and color profile creation. This includes a USB-interfaced MonacoOPTIX colorimeter, capable of color measurement from emissive sources for CRT and LCD monitor color calibration, and used with MonacoOPTIX software; and a USB-interfaced GretagMacbeth Eye-One spectrophotometer, which measures color in the 380–730 nm range, with resolution of 10 nm, from both emissive and reflective sources, used with MonacoProof software, for the creation of standard color profiles which characterize the color I/O of devices such as scanners, monitors, and printers using the International Color Consortium standard.

The Document Imaging laboratory supports DocView, MARS and other research and design projects involving document imaging. Housed in this laboratory are advanced systems to electro-optically capture the digital images of documents and subsystems to perform image enhancement, segmentation, compression, OCR and storage on high density magnetic and optical disk media. The laboratory also includes high-end workstations connected by gigabit Ethernet for performing document image processing. Both in-house developed and commercial systems are integrated

and configured to serve as laboratory testbeds to support research into automated document delivery, document archiving, and techniques for image enhancement, manipulation, portrait vs. landscape mode detection, skew detection, segmentation, compression for high density storage and high speed transmission, omnifont text recognition, and related areas. The laboratory also contains rack-mounted, networked processors running all recent versions of Windows-based operating systems to support the DocView, DocMorph and MyMorph projects. This provides an easily configurable test platform for simulating a variety of potential user environments, including those with firewalls, for testing, modifying and improving software developed in these projects.

The Document Image Analysis Test Facility is an off-campus facility that houses high-end workstations and servers that constitute the MARS production system. While routinely used to produce bibliographic citations for MEDLINE, this facility also serves as a laboratory for research into techniques for the automatic zoning, labeling, and reformatting of bibliographic fields from document images, intelligent spellcheck by pattern recognition techniques, and other key elements of MARS. These techniques are fundamental to the automated extraction of descriptive metadata for the long term preservation of document images. Besides real time performance data, also collected and archived are large numbers of bitmapped document images, zoned images, labeled zones, and corresponding OCR output data. This collection serves as ground truth data for research in document image analysis and understanding

Multimedia Research and Development
Multimedia research and development efforts concentrate on the engineering of technical improvements applied to issues such as image quality and resolution, color fidelity, transportability, storage, and visual communication. In addition to developing new methods and processes, LHC facilities and hardware infrastructure reflect state-of-the-art standards in the rapidly changing field of multimedia research and development. High definition video, for example, represents today's standard for improved electronic, motion imaging quality. Multimedia systems, scientific visualization and networked media are being pursued for their performance, educational, and economic advantages. Three dimensional computer graphics, animation techniques, and photorealistic rendering methods have changed the tools and products of the artists in the branch. Digital video and image compression techniques are central to projects requiring storage of large images and rapid visual file transmission.

CD-ROM, DVD and DVD-ROM technology for capturing media assets including video, audio, Web information, and computer text slides continue to be explored. Web links within these assets are used for updating program content and providing links to additional information tools. A template allowing the simultaneous viewing of multiple interactive windows, including speaker video, slides, and an interactive index was developed to improve access to program content on CD-ROM, DVD and DVD-ROM technology. By selecting any one slide from the index, two other windows immediately synchronize to that point in the presentation. Using the new template technology, project staff developed a symposium DVD-ROM "The Library As Place: A Symposium on Building and Renovating Health Sciences Libraries in the Digital Age" and a Conference DVD "From Double Helix to Human Sequence—And Beyond" featuring over 10 hours of video, Web access, video, and additional information on each disc.

Together with NLM's Office of Communications and Public Liaison and the HMD Exhibition Program, project staff have worked with MacNeil/Lehrer Productions to launch the *Changing the Face of Medicine: Profiles in Achievement* Web-enhanced DVD in FY 2004. The highly interactive DVD features 12 physician profiles, a mentoring program profile, and 200 Web links as an information resource tool for users. The interactive DVD was awarded a 2004 Web DVD Excellence Award by the DVD Association of America. As an element of the *Changing the Face of Medicine* exhibit, the NLM is working on the planning and production phases of video and Web programs featuring the *Local Legends* program, a collaborative project between the NLM and the American Medical Women's Association (AMWA). The Local Legends Web site highlights congressionally nominated women physicians from 50 states. The Web site is designed to include video profiles of one representative from each state, as selected by a committee within the AMWA. The first video interview with the Washington, D.C. local legend, Janelle Goetcheus, M.D., was conducted at the Columbia Road Health Services clinic, and additional video content was produced in the clinic and on the streets of Washington. These materials were edited into an overview video featuring the NLM/AMWA program and presented at the annual AMWA meeting in San Diego, CA in February 2004. The overview video of Dr. Goetcheus won a 2004 Telly Award. Thirty-three on site video interviews with nominees were conducted at the annual meeting to select state representatives of the Local Legends program. All aspects of the Local Legends Web site design have been completed and approved by the NLM Local Legends development team and the AMWA. Future work will include the development of additional Local Legends video profiles.

Additional projects illustrate a variety of technological advancements. *Project 20*, a 15-minute videotape chronicling the last 20 years of the NLM, highlighted the growth of MEDLINE, the development of Grateful Med, Internet Grateful Med, Free MEDLINE, UMLS, the creation of NCBI, and other significant events in the history of NLM. A prototype DVD-ROM based on the NLM Dream Anatomy exhibition was completed in FY2004. The DVD features a video overview, a gallery and timeline, and a virtual tour of the exhibit. The narrated program also features high definition video of the exhibition, video graphics, and an original musical score. Web links to NLM's Dream Anatomy exhibition Web site and a fully functional search tool are also available when the DVD is viewed on a computer. Additional DVDs were prepared in FY2004 including: (1) *LHNCBC Research Projects Video DVD*, (2) *The 2004 Collen Award: Dr. McDonald's Life and Career*, and (3) NLM Board of Regents presentation: *Saving Lives and Saving Money*, by the Honorable Newt Gingrich.

Information Systems

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

Profiles in Science
The Profiles in Science Web site uses innovative digital technology to make available the manuscript collections of prominent biomedical researchers, medical practitioners, and those fostering science and health. Database content is created in collaboration with the History of Medicine Division, which processes and stores the physical collections. Most collections have been donated to the NLM and contain published and unpublished materials, including books, journal volumes, pamphlets, diaries, letters, manuscripts, photographs, audio tapes and other audiovisual resources. The Visual Culture and Health Posters, as well as the collections of C. Everett Koop and Wilbur A. Sawyer were added in FY2004, bringing the total number of archives for prominent biomedical researchers, medical practitioners, and those fostering science and health to 13: Christian B. Anfinsen, Oswald T. Avery, Julius Axelrod, Donald S. Fredrickson, C. Everett Koop, Joshua Lederberg, Barbara McClintock, Marshall W. Nirenberg, Linus Pauling, Martin Rodbell, Florence R. Sabin, Wilbur A. Sawyer and Fred L. Soper. The Reports of the Surgeon General (1964–2000), the history of the Regional Medical Programs (1964–1976), and Visual Culture and Health Posters are also available on Profiles in Science.

In FY2004, project staff continued to enhance the effectiveness of Profiles in Science. The Web site was upgraded to more powerful hardware with up-to-date applications and operating system software. Enhancements to the underlying digital library framework included a new database infrastructure, the creation of additional ways to view information, and faster methods for extracting records in ASCII format. New error detection and correction rules and methods for automatically updating data were also added. Protocols for digitizing collections at other institutions were developed and tested in collaboration with the Wellcome Library staff, United Kingdom. Development began in FY2004 on a Historical Events and Prominent Scientists Timeline to highlight the major historical events (e.g., political, medical, scientific, and social) that occurred at the time of the major achievements of the scientists represented in the collection. Changes to the current Metadata Entry and Editing Program were made in preparation for moving the program to a Web interface. Detailed analysis of workflow in obtaining copyright permissions identified changes needed in the database and user interface for tracking permissions. Finally, the development of an XML-based Web interface and transition to an XML-based search engine, as well as automated testing and verifcation tools, continue to be pursued.

MARS
Document image analysis and understanding research combined with database design, graphical user interface design for workstations, image processing, string pattern matching, lexical analysis, speech recognition and related areas underlie the development of MARS (Medical Article Records System), a system to automate the production of MEDLINE citation records from biomedical journals. MARS has evolved through several generations of increasing capability. Its core engine consists of daemons 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. For some years, its production version has been used to extract bibliographic data to populate MEDLINE. Two other techniques to obtain such data have been manual keyboarding and XML-tagged data directly from publishers.

To meet the NLM's goal of discontinuing the keyboarding contract and thereby realizing savings, MARS design faced the challenge of having to process journals currently handled manually. These journals include ones with page background in color or gray shades which greatly compromises OCR accuracy. Experiments were conducted with

grayscale scanners comparing different approaches to eliminating these atypical backgrounds, and the best approach was found to be by using a library developed with functions in the FineReader OCR toolkit. This library was embedded in the inhouse-developed scan software, and preliminary results from a test set of 101 articles showed that low-confidence characters occur at about the same rate with these grayscale scanners as with the monochrome scanners in production, i.e., effectively eliminating the deleterious effects of gray and color backgrounds. Following the completion of these tests, the grayscale scanners have been placed in production.

The scanning software has also been modified to improve quality control. Images from poorly scanned documents cause OCR errors and compromise downstream processes. Conventional QC relies on the operator viewing the images and deciding on their quality. This is highly subjective and is not always reliable. To make this step more robust, a commercial library from ScanSoft has been incorporated in the scan module to detect low-confidence characters and calculate those as a percentage of the total number of characters on the page. This figure provides the operator a quantitative measure of image quality. Another key element in allowing NLM to eliminate its keyboarding contract is the requirement for MARS to accommodate foreign language journals, which account for 11% of MEDLINE citations. This requirement introduces new rules to extract vernacular titles (in Roman script languages but not in others), and process the second pages of articles (formerly only one page needed to be processed). These have been achieved by the FLEX software suite that incorporates new code in several MARS workstations. Starting with journals in French, German, Italian, and Spanish, MARS enhanced by FLEX now processes five Western European languages using Roman script and three using Cyrillic script.

WebMARS is a system to extract bibliographic data from online journals. A prototype system has been developed to combine downloading and classification of journal articles followed by zoning, labeling and reformatting algorithms to identify and extract the data. The NLM Board of Regents was recently given a talk covering the history of automated bibliographic data extraction from 1996, when NLM's keyboarding contract ran into difficulties, through the evolution and increasing automation in the MARS system, and focusing on the design and functions of WebMARS. A key point in the presentation was a comparison of the relative labor required in producing citations with keyboarding, MARS, XML citations from publishers, and WebMARS (which promises to result in the least amount of labor). WebMARS is undergoing testing with over 60 journal titles. Tests comparing WebMARS output against existing MEDLINE citations for past issues have been useful in refining the labeling and reformatting algorithms.

An additional prototype has been developed to handle meeting abstracts. Testing with four volumes was successful and the prototype is ready for demonstration. "Meeting abstracts" refers to the proceedings of important conferences in HIV, AIDS and other topics of current importance. The contents of these proceedings are not simply "abstracts" as conventionally understood, but include most other bibliographic information: title, author names, affiliations, etc., Most important for automation, meeting abstracts do not follow the familiar layouts of typical biomedical journal articles. The unconventional layout of meeting abstracts requires a modification of the existing zoning, layout and reformatting rules. For instance, since author names are arranged differently from a typical journal (all names in a single line, and separated by semicolons), the existing reformatting rules in MARS required changes to accommodate this format.

Ground Truth Data for Document Image Analysis
In August 2003, the Medical Article Records Groundtruth database was released for research in document image analysis and understanding techniques by the computer science and informatics communities. The data consists of over 1,000 bitmapped images of the first pages of articles from biomedical journals indexed in MEDLINE falling into nine layout types encountered in MARS production. Included are the corresponding segmented and labeled zones all in XML format. Also available from this Web site is Rover, an analytic tool that may be used to compare the results of a researcher's program with the ground truth data. Rover has been enhanced to allow a visual comparison of researchers' algorithmic results with the ground truth data, as well as some statistical metrics.

DocView
DocView facilitates the delivery of library documents directly to the patron via the Internet in multiple ways, but it is most commonly used by library patrons to receive scanned journal articles from libraries that use Ariel software for interlibrary loan services. While Ariel, developed by Research Libraries Group, and now a product of Infotrieve, is used by libraries and document suppliers routinely to send documents via Internet to similar organizations, there are few options for end users to directly receive them. DocView helps fill this void by allowing end users to receive documents sent by Ariel via a modified form of File Transmission Protocol. DocView also enables users to retain the received

documents in electronic form, view the images, organize them into "folders" and "file cabinets," electronically bookmark selected pages, manipulate the images (zoom, pan, scroll), copy and paste images, and print them. In addition, DocView serves as a TIFF viewer for compressed images received through the Internet by other means, such as Web browsers. Users may receive document images either via Ariel FTP or Multipurpose Internet Mail Extensions protocols. With DocView, users may also forward documents to colleagues for collaborative work. 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. The MyMorph Web service consists of Windows-based client software and modifications to DocMorph for accommodating the Simple Object Access Protocol. In-house testing has shown that MyMorph significantly improves user productivity compared to the conventional use of DocMorph through a Web browser, particularly for users who need to convert large numbers of files to PDF.

Document Preservation
Project staff have begun to design a flexible, modular software framework that may be used as a prototype for investigating techniques to preserve NLM's digital resources in a cost-effective manner. A prototype system called SPER (System for the Preservation of Electronic Resources) has been developed. The system allows ingest, metadata extraction and file migration, and the identification of minimum required technical metadata for document files. Developing SPER required careful attention to proposed standards and models for digital preservation and preservation metadata schemas, including the NISO X39.87 proposed standard for digital still images. SPER relies on open source, platform-independent components, as well as current open resources and tools which already provide some functionality required by SPER. A JavaServer Faces-based GUI was chosen to provide the Web interface for SPER users or operators. The SPER prototype was implemented in FY2004, with a first phase model designed to convert TIFF images to PDF documents and/or JPEG2000 images. The Profiles in Science collection and MARS document images will be used as test sets. Additional research and development efforts on metadata extraction and prototype design strategies of SPER will address issues with metadata elements, strengthen tools that automatically learn journal-specific rules using both geometric and contextual features, and strengthen systems that automatically learn the 2D layout models of document page images using Bayesian

learning algorithms. In-house tools (e.g., DocMorph, MyMorph) are being studied as potential tools for electronic preservation. Modifications of DocMorph and MyMorph to produce PDF/A files from image-based files are being explored. This work may lead to a system, accessible from any point on the Internet, that allows users to mass-migrate image-based file collections to a standard archival format. Additional research is being conducted to identify key issues related to the preservation of video.

Turning The Pages Information Systems
Turning the Pages Information Systems research seeks to design more efficient methods to translate paper volumes from the NLM's historic collection to photorealistic electronic form, extend the virtual books into information systems, and to increase the accessibility of historical documents for the public. After the initial development of the Turning the Pages (TTP) format, research began to transform the design into a usable information system (TTP+). Research focused on a "discovery" and a "storyline" model as directions for TTP+. The TTP+ version of Blackwell's Herbal uses the "discovery" model, retaining the photorealism of the original TTP while allowing a patron to "travel" to live sites on the Internet. For example, from highlighted text on the St. John's Wort page, users can go to various search engines (e.g., PubMed, ClinicalTrials.gov, USDA) and obtain citations or general information on St. John's Wort. The TTP+ version of *Vesalius' Anatomy in Photorealistic* uses the "storyline" model and contains images from other sources (e.g., rendered Visible Human images, pictures of Italian cities, etc.). Images are interlinked to present the consumer with several multimedia "stories," including *Man of Padua* and *Modes of Portraying Anatomy*.

Two methods have been investigated in order to combine all existing virtual books for kiosk display. A monolithic approach bundling all software into one file was pursued. Memory limits imposed by the Windows OS rendered this method unscalable. On the other hand, a modular approach where the code for each book is selected by the user, provided a scalable method more suitable for the addition of future books. In addition, while developed under the Windows OS environment, the TTP code has also been successfully tested for operation on a Mac computer running OS X. Future goals are to continue developing efficient, high quality methods for producing and distributing TTP books as more historical books are selected.

NLM Gateway
The NLM offers a number of Internet-based information resources, each with its own user interface. The NLM Gateway provides an easy to

use, "one-stop" search method that allows users to issue simultaneous searches in 15 NLM information resources using 5 retrieval methods from a single interface. The NLM Gateway continued to grow and evolve in FY2004 with several additions and enhancements. NLM Gateway access was added for the MedlinePlus Health Tutorials, MedlinePlus Current Health News, Online Mendelian Inheritance in Man, Hazardous Substances Data Bank, TOXLINE Special, and the Genetics Home Reference. NLM's book, serials, and audiovisual materials were migrated from LocatorPlus to the new NLM Catalog under "Entrez," substantially increasing the searching capabilities in the collection. The NLM Gateway language table was updated with the latest Machine Readable Cataloging language codes. Targeting PubMed, enhancements include the addition of a LinkOut feature for PubMed citations and direct links on the Document Ordering page for PubMed Central articles. A spell checker that automatically searches both British and American spellings of words was also incorporated. Author name truncation for searching was added to the Meeting Abstracts Collection and the Health Services Research Projects database, and approximately 15,000 abstracts were added to the Meeting Abstracts Collection.

A comprehensively redesigned NLM Gateway Version 2.0 entered early testing in FY2004. The new user interface will provide clear, easy to understand, and a cleaner navigation to different areas of the composite result set. At the same time, the new interface will continue to execute simultaneous searches in 15 information resources. The targeted release for the new user interface is early FY 2005.

Consumer Health Informatics Research
Exploring consumer information needs, information seeking behavior, and cognitive strategies, consumer health informatics research projects utilize informatics methods and information technologies to study methods to develop, organize, integrate, and deliver accessible health information to consumers with all levels of health literacy.

ClinicalTrials.gov provides comprehensive, up-to-date information about federally and privately supported clinical trials throughout the US and many other parts of the world. The system grew out of 1997 legislation requiring the HHS, through the NIH, to establish a registry for both federally and privately funded trials "of experimental interventions for serious or life-threatening diseases and conditions," thereby broadening the public's access to information on potential interventions for a wide range of diseases. Launched in February 2000, ClinicalTrials.gov provides patients, families and members of the public easy access to information

about the location of clinical trials, their design and purpose, criteria for participation and, in many cases, further information about the disease and intervention under study. There are also links to individuals responsible for recruiting participants to each study.

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

A Web-based Protocol Registration System allows providers to maintain and validate information about their trials. New views of protocol summaries are supported by geographical location, date added, and by patient recruiting status. A Spanish-language prototype system using Spanish-English cross-language information retrieval technology was developed and is undergoing extensive testing. ClinicalTrials.gov was the recipient of Harvard University's prestigious 2004 Innovations in American Government Award in recognition of its significant achievements. HHS Secretary Tommy G. Thompson noted that ClinicalTrials.gov is a good example of how government can improve access to vital health care information for all Americans.

The Genetics Home Reference is an integrated Web-based information system designed for consumers and others to learn about specific genetic conditions and the genes or chromosomes associated with those conditions. The research results made possible by the Human Genome Project are increasingly being made available in scientific databases on the Internet, but because of the often highly technical nature of these databases, they are not readily accessible to the lay public. The goal is to provide a bridge between the clinical questions of the public and the richness of the data emanating from the Human Genome Project.

The Genetics Home Reference Web site provides basic information in a question and answer format on the nature of genes and how they give rise

to various conditions and diseases. The site currently includes more than 100 condition summaries and more than 160 gene summaries, over half of which were added during FY2004. Additional FY2004 improvements include a new feature that provides information about chromosomes and chromosomal disorders. Several new topics (e.g., pharmacogenomics, multifactorial disorders, and imprinting) were also added to "Help Me Understand Genetics," the site's genetics handbook. Genetics Home Reference achieved significant site navigation improvements in FY2004 with a redesigned home page, as well as newly designed browse, search, and help features. Targeted links were also added throughout the site. The site was integrated with MedlinePlus, Gateway, PubMed Linkout, and the "What's New" series in order to help consumers locate the Genetics Home Reference Web site.

Further Consumer Health Informatics Research focuses on understanding and improving access to online health information. Technologies are being developed that provide measures of text difficulty that help determine the suitability of health-related documents for consumers at different literacy levels. New approaches for providing timely access to consumer health information in order to accommodate the diverse needs of people in the U.S. and abroad are being pursued through cross-language information retrieval research. Finally, the Consumer Health Vocabularies project focuses on mapping words and phrases commonly used by consumers to technical medical terms and concepts.

Research Infrastructure and Support

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.

Communication and Collaborative Technologies
Lister Hill Center staff engages in research to develop technologies that will facilitate easy access to biomedical information through devices such as Personal Digital Assistants (PDAs), wireless portable computers, mobile phones, and other emerging devices.

PubMed on Tap is a research and development project to develop accessible biomedical information at the point of care through handheld devices used by clinicians and other mobile health care providers. User interface, content selection, content organization, and system performance are necessary for effective access to information. Initial research is focused on the design of a user interface for search and retrieval of

MEDLINE bibliographic citations through PubMed. Initial content selection is involved with categorizing citations returned in response to a query, creating multi-document summaries for clusters of highly related documents, and single-document descriptions containing features specific only to a given document in the cluster. System performance research is focused on discovering design factors that ensure the speed and reliability of the hardware and software required for accurate and timely retrieval of data. Areas of investigation include choice of parsers, efficient use of a database to store recent queries and citations, and load testing.

A prototype system, developed for PDAs running the Palm operating system, was built and tested in FY2003. The software uses the PDA's wireless communication interface and HTTP protocol to communicate with a servlet residing on a proxy server. The proxy server communicates with PubMed through the Entrez programming utilities (e.g., Esearch, Efetch and Elink). The proxy server stores queries, results, and citations to provide a quick response to recurring queries and fast delivery of frequently requested citations. The proxy server also monitors performance measures and accumulates aggregate statistics to help in developing clustering and ranking tools. The client program is responsible for the user interface and for storing user-specific information, such as preferred search strategies or recurring queries. FY2004 upgrades, implemented as a result of user feedback, have significantly improved PubMed on Tap usability.

PubMed for Handhelds also explores hand-held technology for use in the clinical setting. During FY2004 several new features were introduced. PICO (Patient/Problem, Intervention, Comparison, and Outcome) is a method used for developing well-formulated clinical queries. This format can also be used for structuring literature searches and may be helpful to those interested in evidence-based medicine. In support of users of newer handheld devices that feature WAP browsers (mobile phones, hybrid PDA-phones) the system has been reformatted. Current services offered are clinical queries, systematic reviews, PICO searching without filters, journal abstracts browser, and access to ClinicalTrials.gov.

Additional projects targeting the use of handheld devices as a portal to information dissemination continued to expand in FY2004. The Biomedical Informatics and Pathology departments at the Uniformed Services University collaborate with Center staff to provide wireless (e.g., infrared, Bluetooth, 802.11b) PDA access to PubMed, MEDLINE, and other NLM databases during small, medical student group discussions. PDAs will allow students to electronically submit reports and case

summaries, which is expected to enhance their interactions with teachers.

The ASKLEPiOS project (Access to Services and Knowledge, multiLingually, Everywhere, Portably, in Open Source) seeks to explore the integration of portable wireless hand-held devices together with non-mobile computer servers and telephones. The integration framework is built with open source tools and includes internet-based telephony, videoconferencing, wireless data services, speech recognition/synthesis services, and a robotic chat service. The framework provides the needed "middleware" layer upon which applications relevant to the mission of the NLM can be built. Portable personalized devices with visual and speech-based interfaces may prove helpful in delivering health care to an increasingly multicultural and multilingual society. Through collaboration with external groups, the project focuses on technologies such as information servers, speech synthesis and recognition software, handheld personal computing devices, wireless networking, and the public-switched telephone network.

The Collaboratory for High Performance Computing and Communication investigates innovative means for assisting health science institutions in their use of online distance learning technologies. The Collaboratory also explores advanced computer and network technologies for distance interactivity, including wireless technology and virtual reality research. Major upgrades to existing videoconferencing codecs were accomplished and new codecs were added in FY2004. Several significant demonstrations were performed using videoconferencing technology, both at NLM and off site at national meetings. Demonstrations of streaming and wireless Webcasting were done and videoconferencing and Webcasting were employed routinely in program activities. One significant upgrade was the purchase of a Click-2-Meet videoconferencing server that allows end points to tunnel through firewalls. The new software required a significant upgrade in computer hardware. Hardware upgrades were also needed for Webcasting and it appears that dual processing machines are increasingly required.

Experiments continued using the conventional h.323 videoconferencing technology with Charles R. Drew University of Medicine and Science and its affiliated medical magnet high school for minorities, the King-Drew Medical Magnet High School. A pilot videoconference featuring NLM librarians was completed. The h.323 videoconferencing technology was also employed in a virtual site visit of the NLM funded medical informatics program at the University of Missouri. As a result of the phase-out of Access Grid version 1.1, another major upgrade was undertaken with the Collaboratory Access Grid node. A commercially developed software application was purchased in order to use the commercial software for standard applications, while also experimenting with open source beta versions. In addition to utilizing the new software, the audio for the current node was upgraded with a state of the art echo cancellation system. The Access Grid node was used in NLM's tutorials on advanced networking at the 2003 Annual Meeting of the Radiological Society of North America, co-sponsored by NLM and Internet2.

The EtherMed database of Web accessible health professions educational materials continued to be expanded through collaborations with colleagues at the University of Utah, UCLA, and the University of Oklahoma. A major FY2004 upgrade allows outside individuals to nominate Web sites and enter information for later review. After initial testing, this improvement is expected to simplify the task of identifying sites to be included in the database. Another major review of EtherMed was completed using an NLM developed set of search queries. Additions to the database are being held until the research is complete.

Scalable Information Infrastructure
The purpose of the Scalable Information Infrastructure (SII) initiative is to encourage the development of health-related applications of scalable, network aware, wireless, geographic information systems, and identification technologies in a networked environment. The initiative focuses on situations that require, or will greatly benefit from the application of these technologies in health care, medical decision-making, public health, large-scale health emergencies, health education, and biomedical, clinical and health services research. Projects must use test-bed networks linking one or more of the following: hospitals, clinics, health practitioners' offices, patients' homes, health professional schools, medical libraries, universities, medical research centers, laboratories, or public health authorities.

FY2004 began the first year of a three year effort for 11 SII research contract awards. Several SII projects have already made notable progress; an early prototype system using wireless networks, GPS, RF tags, and handheld and wearable computers was developed by the University of California, San Diego; an auditorium-scale presentation of 3D anatomy and collaborative surgery with haptics was conducted with Stanford University and collaborators in Australia; a monitoring system was implemented for the Project Sentinel Collaboratory information security program at Georgetown University; a secure XML medical record template for individuals was developed at the Children's Hospital in Boston; and significant progress was made in viewing and

manipulating 4D datasets through the "4D Visible Mouse Project" at the Pittsburgh Supercomputing Center, Carnegie Mellon University.

Telemedicine

The Telemedicine Information Exchange, sponsored by the NLM, is a Web-based resource of telemedicine and telemedicine related activities maintained by the Telemedicine Research Center in Portland, OR. During FY2004, approximately 727 non-NLM bibliographic citations and other records were delivered to the NLM. The University of Pennsylvania Dental School completed its NLM-sponsored project during this past year. Given the declining manpower in dentistry, limited training facilities, and the increasing cost of dental education, the project considered the feasibility of providing a distributed program of dental instruction.

The virtual microscope project has been initiated by in-house staff. The project team has developed a Web-based system that allows users to view an image in an interactive manner, simulating the experience of examining a slide under a microscope. Potential applications of the tool include medical education, quality control and diagnostic proficiency surveys, and telemedicine. Staff continue to participate in the monthly meetings of the multi-agency Joint Telemedicine Working Group. Participating in this group, Lister Hill Center staff made a formal presentation to Congress and the Administration on state-of-the-art Telemedicine and e-Health projects and solutions.

National Center for Biotechnology Information

David Lipman, M.D.
Director

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

NCBI is charged with providing access to public data and analysis tools for studying molecular biology information. Over the past 16 years, the ability to integrate vast amounts of complex and diverse biological information created the scientific discipline of bioinformatics. It is now almost impossible to think of an experimental strategy in biomedicine that does not involve some dependence on bioinformatics. At the core of this shift is the flood of genomic data, most notably gene sequence and mapping information. NCBI will meet the challenge of collection, organization, storage, analysis, and dissemination of scientific data by designing, developing, and distributing the tools, databases and technologies that will enable the gene discoveries of the 21st century.

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

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

NCBI programs are divided into three areas: (1) creation and distribution of databases to support the field of molecular biology; (2) basic research in computational molecular biology; and (3) dissemination and support of molecular biology databases, software, and services. Within each of these areas, NCBI has established a network of national and international collaborations designed to facilitate scientific discovery.

GenBank—The NIH Sequence Database

GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. NCBI is responsible for all phases of GenBank production, support, and distribution, including timely and accurate processing of sequence records and biological review of both new sequence entries and updates to existing entries. Integrated retrieval tools have been built to search the sequence data housed in GenBank and to link the results of a search to other related sequences, bibliographic citations, and other related resources. Such features allow GenBank to serve as a critical research tool in the analysis and discovery of gene function as well as discoveries that lead to identification and cures for a number of diseases.

Important sources of data for GenBank are direct sequence submissions from individual scientists and genome sequencing centers, and substantial staff and resources are devoted to the analysis and assembly of genome 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.

In FY2004, approximately 7 million sequences were added to GenBank, and the base count rose from 33 billion in August 2003 to 40 billion in August 2004. The 34 million sequences in GenBank represent data from over 130,000 organisms.

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." This 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 is maintained as confidential, if requested.

In FY2004, the restriction on sequence length for database records was removed. Previously the International Nucleotide Sequence Database Collaborators (INSD) had agreed to a 350,000 base limit in order to maintain compatibility with various existing biology software packages. Newer software versions are able to analyze long sequences quickly and by removing the length limitation, megabase comparisons can be performed more efficiently.

NCBI is continuously developing new tools, and enhancing existing ones, to improve access to, and the utility of, the enormous amount of data stored in GenBank. Sequence data, both nucleotide and protein, is supplemented by pointers to corresponding PubMed bibliographic information, including abstracts and publishers' full-text documents. 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 available data.

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. NCBI also accepts submissions from Whole Genome Shotgun (WGS) sequencing projects. Annotations are allowed in these assemblies and records are updated as sequencing progresses and new assemblies are computed. Forty-nine WGS sequencing projects were released during FY2004.

Improvement of NCBI's sequence submission software continues to be a high priority. A new version of Sequin, NCBI's stand-alone submission tool, was released in FY2004. In this new version, improvements were made to facilitate ease

of TPA sequence submission, the alignment view in the Record Viewer was improved, and Batch submission features have increased functionality. The GenBank 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 are also available on the GenBank site.

BankIt, another sequence submission software tool, is now in its tenth year of use. Some of the improvements made to BankIt this year include the ability to identify sequences appropriate for the TPA database, options for including strain name for mouse, rat, and *Influenza* virus, and a more explicit example of features that can be added to a record.

GenBank has evolved to contain several types of sequence information, from relatively short Expressed Sequence Tags (ESTs) to assembled genomic sequences that are several hundred kilobases in length. EST data obtained through cDNA sequencing are critical to understanding gene function and therefore continue to be heavily represented in GenBank. 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 STS information can be found in the UniSTS database.

Entrez Genomes contains records representing over 2,000 species including bacteria, archaea, and eukaryotes, complete microbial genomes, a number of viroids, mitochondria, a broad host range of plasmids, and over 1,000 viruses. The genomes represent both completely sequenced organisms and those for which sequencing is in progress. Approximately 20 new complete genomes and over 900 records for viral, microbial, and organellar chromosomes were added to the database in FY2004. Twenty-two organism-specific genome resource pages are now available including chimpanzee, rat, mouse, chicken, cow, dog, pig, sheep, cat, and honey bee.

The Human Genome

NCBI is responsible for collecting, managing, and analyzing human genomic data generated from the sequencing and genome mapping initiatives of the public Human Genome Project. NCBI also plays a key role in assembling and annotating the human genome sequence. This resource is truly an international public sequencing effort due to the cooperation of scientists and sequencing centers from around the world. In FY2004, multiple annotated

builds of the human genome were released to the public. The latest Build 35, version 1, was released in June.

Assembling and Annotating the Human Genome
A team of NCBI scientists is engaged in annotating, or characterizing, the biologically important areas of the genome. In FY2004, annotation for genome builds was based on Gnomon, the new gene prediction program developed by NCBI scientists. Gnomon puts a greater emphasis on coding propensity and matches to existing proteins when predicting genes. To create a gene model, Gnomon finds the best self-consistent set of transcript and protein alignments to a genomic region and uses these alignments as constraints for a Hidden Markov Model (HMM)-based gene prediction. As a result of the Gnomon program, the number of genes for human, mouse, and rat genome builds has decreased significantly, while the number of models identified as pseudogenes has increased. The number of human genes is now predicted as low as 20,000 versus earlier estimates of 35,000.

NCBI Resources Designed to Support Analysis of the Human Genome
NCBI has developed a suite of genomic resources to support comprehensive analysis of the human genome, as well as the complete genomes of several model organisms. Specialized tools and databases have also been designed to facilitate researchers' use of this data. NCBI 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, LocusLink, RefSeq, Map Viewer, Gene, Homology Maps, UniGene, HomoloGene, and GEO. NCBI's Human Genome Sequencing site provides access to information on sequencing efforts and various other types of resources, such as chromosome-specific mapping information, and TaxPlot for genome similarity plotting.

NCBI's Map Viewer provides a graphical display of features on assemblies of genomic sequence data as well as cytogenetic, genetic linkage, physical, and radiation hybrid maps, when available. Map features that can be seen along the sequence include NCBI contigs, the BAC tiling path, the location of genes, exons, STSs, FISH mapped clones, ESTs, GenomeScan models, SAGE tags, and sequence variation. Maps from other sequencing

centers are also available. Genes or markers of interest can be found by submitting a query against a whole genome, or by querying one chromosome at a time. The results table includes links to a chromosome graphical view where the gene or marker can be seen in the context of additional data. The Evidence Viewer is a feature that provides graphical biological evidence supporting a particular gene model and the Model Maker allows users to build a gene model using selected exons.

In FY2004, NCBI continued to improve its Map Viewer. A new Map Viewer home page was released, grouping the organisms for which map information is available. Twelve organisms were added to Map Viewer this year bringing the total number of organisms to 34. An advanced search capability was added which allows restriction of searches to specific chromosomes or searching for objects based on specific attributes. A new comparative maps feature allows users to view maps of different organisms side-by-side for comparison.

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. This collection, part of the NCBI Books site, contains descriptions for over 150 genetic diseases and links to databases and organizations for additional information. For each gene description there is a link to PubMed, the Online Mendelian Inheritance in Man database (OMIM), the Map Viewer, LocusLink, and BLink for related sequences.

OMIM is an electronic version of Dr. Victor McKusick's "Online Mendelian Inheritance in Man," a catalog of human genes and genetic disorders. The database, produced at Johns Hopkins School of Medicine, contains over 15,500 records. 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. During the past year, information connecting diseases to sequence (genes or markers) was used to create a human sequence based-phenotype map. More than 1,600 diseases have been placed in sequence coordinates on the human genome.

The GeneTests database produced at the University of Washington is now being 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. Data produced by this database is now being integrated into NCBI data resources.

LocusLink, NCBI's original single-query interface to curated sequences and descriptive information about genetic loci, continues to grow.

The number of genes represented expanded to 152,000 not counting genes predicted from NCBI's genome annotation pipelines. In FY2004, organisms added to LocusLink include honey bee, chicken, dog, pig, purple sea urchin, African clawed frog, and Western clawed frog, bringing the total number of organisms to 15. LocusLink provides one of the windows into NCBI's annotation of the human genome, with direct links to the Map Viewer, gene annotation, gene ontology terms, and links to other NCBI resources.

The Reference Sequence (RefSeq) database provides a comprehensive, integrated, non-redundant set of sequences, including genomic DNA, transcript (RNA), and protein products for major research organisms. These standards serve as a basis for medical, functional, and diversity studies by providing a stable reference for gene identification and characterization, mutation analysis, expression studies, polymorphism discovery, and comparative analysis. In FY2004, the NCBI RefSeq database grew by 48% and the full release of all NCBI RefSeq records includes over 1.1 million proteins from 2,558 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, which pass quality tests, and are consistently identified by different groups. A comparison between the human genome annotation provided by NCBI and the Ensembl groups determined that approximately 16,000 annotated proteins are identical between the two groups, 99% of which are annotated by sequences provided in the curated RefSeq database.

NCBI has created new infrastructure to support effective communication with research groups that work on genome sequencing, annotation, and biology. During FY2004, pre-existing collaborations continued for seven genomes, in addition to an established viral genomes advisors group, and new collaborations for eleven organisms plus fungi were established. From this effort, eleven sets of supporting web pages were added along with related map and/or genomic sequence information in the NCBI Map Viewer.

The Entrez Gene database debuted early in FY2004 and is a significant step to providing a much larger scope of gene-specific data at NCBI. Entrez Gene integrates information about genes from LocusLink and gene features annotated on RefSeqs from Entrez Genomes. Currently more than 2,275 taxa are represented in the Gene database with a total of about 860,000 genes.

RNA interference (RNAi) is an emerging technology for silencing specific genes that is proving to be of great utility as a research tool and may have important clinical applications. In response to a request from Dr. Zerhouni, the NIH held an RNAi workshop earlier this year and established a cross-institute working group to aggressively pursue RNAi research. To fully realize this investment, the NCBI has established a database to store information on RNAi reagents and experimental results. NCBI scientists are currently working with other NIH scientists to enter the appropriate information and the first public release of the database is planned for early 2005.

As the number of sequenced genomes continues to grow, there is increasing interest in comparative analysis of genes from represented species. The NCBI HomoloGene resource performs such large-scale comparison automatically and presents the results to scientists, obviating the need for individual analyses. Over the past year, the HomoloGene system has been completely revised to use genome-based information, as opposed to the transcript-based information that was available in the pre-genome era. In a recent release, comparisons of over 16 billion pairs of genes were performed, leading to 103,677 gene homology groups. HomoloGene was also added to the Entrez retrieval system in FY2004.

The dbSNP database of genetic variation is a comprehensive catalog of common human polymorphisms for the international research community. dbSNP continued to experience rapid growth in FY2004. New content was driven by ongoing surveys of human sequence variation for the International Haplotype Map Project (HapMap), major submissions by private companies, and variation analysis using whole genome shotgun reads for freshly sequenced organisms. dbSNP expanded support for genotype data in 2004 with a new schema and several rounds of intensive staff curation to uniquely identify the individuals represented by overlapping sets of cell line reagents and pedigree data. dbSNP group members worked in both production and advisory roles for two major NIH projects, the Mammalian Gene Collection (MGC) and HapMap, on issues of SNP mapping, sequence annotation, data interpretation, and final data quality assessment. The topic of haplotype representation in dbSNP and annotation of linkage disequilibrium on reference genome sequence will be a major development issue in FY2005 when HapMap's phase I genotype data and high-resolution haplotype data are released for unrestricted distribution.

Quantitative trait loci (QTLs) have measurable effects on an organism's "phenotype," i.e., an open-ended set of measures of the natural processes of metabolism, growth and reproduction, or the abnormal processes of disease. In 2004, NCBI developed a new repository for phenotype data to include this crucial element of biological data in its public genome data infrastructure. NIH's

commitment to understanding human health and disease through the increasingly powerful lens of genomic biology is producing large systematic datasets relating genotype/haplotype to phenotype for humans and comparative model organisms including yeast, fruit fly, mouse, worm and dog. A retrieval system is being designed to accommodate comparisons across heterogeneous submissions of experimental results; capture the multi-dimensional relationship between phenotype measures and data on sequence composition, haplotype variation or level of expression; and support retrieval across multiple scales of organization.

In 2004, NCBI designed a prototype XML schema for a proof-of-concept implementation. It will include publicly available data for human, mouse, rat, and fly. The model's scope was deliberately confined to those organisms with extensive available genomic sequence and an established ontological representation of anatomy, developmental stage and disease. Phenotype will become a new Entrez database in 2005 to facilitate the association of phenotype data with other Entrez components such as dbSNP, Gene, OMIM, and GEO. In this way, putative risk factors like genetic variants or haplotypes, linked QTLs, drug treatments, diet, and epigenetic status can be associated with measurable traits and compared across studies to generate hypotheses as to the true etiology of diseases. Using the Map Viewer, QTLs can be aligned across organisms to identify syntenic regions of phenotype linkage. The phenotype database can become a new point of entry into the Entrez search space via the ontological concepts reflected in its controlled vocabulary.

Model Organisms for Research

The genomes of model organisms can provide genetic information for human development and gene regulation, genetic disease, and the evolutionary process. NCBI genome resource guides provide information on diverse organism-related resources from multiple centers including sequence, mapping, and clone information, when available. The guides also provide easy navigation to organism-specific BLAST pages, and other NCBI resources. NCBI currently provides genome resource guides for 21 organisms other than human. Resource guides added in FY2004 include Aspergillus (fungus), bee, chicken, cow, Dictyostelium, dog, frog, pig, sea urchin, and sheep.

The mouse genome was the first model organism available on the NCBI website. The mouse genome resource guide has links to mapping and BLAST pages as well as information on sequencing progress, sequencing centers, strain resources, and a monthly newsletter designed for the mouse research

community. In FY2004, NCBI Build 33 was released and represents a third generation composite assembly. Rat genome Build 2 was released in FY2004 with new maps including an assembly map, EST alignment maps for human, rat, and mouse, and an ab initio map.

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. It is the bibliographic component of the NCBI's Entrez retrieval system and provides links to full-text journal articles at web sites of participating publishers, as well as to other related web resources.

Full-text journals with PubMed links have increased from 4,054 in September 2003 to over 4,400 in September 2004. Approximately 60% of all PubMed citations from 1990–2003 now have links to full-text. Usage of PubMed by the scientific and lay communities has also grown considerably since its introduction in 1997, with up to 2.8 million searches and over 300,000 users per day.

In August, the 15 millionth citation was added to PubMed and during this year over 1.7 million OLDMEDLINE citations were added. The OLDMEDLINE citations were originally printed in the hardcopy indexes published from 1951 through 1965. The MeSH database was enhanced with terms that are identified by MeSH as pharmacological actions and a direct link was added to the Clinical Queries page. The Clinical Queries page was also revised and filter strategies were updated. The History page now includes a menu from the search statements number to provide an easier way to combine, delete, and retrieve History statements. The truncation limit in PubMed was increased from 150 variations of a truncated term to 600. The PubMed Batch Citation Matcher was updated to include an email feature and the ability to upload a formatted file.

In September, NCBI released a new Entrez database, NLM Catalog. The NLM Catalog provides access to bibliographic data for over 1.2 million books, journals, audiovisuals, computer software, electronic resources, and other materials in the NLM collection via the Entrez retrieval system. The new database is an alternative search interface to the bibliographic records resident in NLM's online catalog LocatorPlus and supports automated mapping features and MeSH term indexes.

LinkOut is a feature of Entrez designed to provide users with links from PubMed and other Entrez databases to a wide variety of relevant web-accessible online resources, including full-text publications, biological databases, consumer health information, research tools, and more. As of

September 2004, over 1,500 organizations have supplied links to their Web sites, representing a 40% increase from last year. Providers include over 1,000 libraries, 180 full-text providers, and 200 providers of non-bibliographic resources including biological databases. Together they provide links to 29 million Entrez records. LinkOut resources received more than 16 million hits per month, a 35% increase from last year.

The LinkOut for Libraries program continues to provide biomedical libraries the ability to link library patrons from a PubMed citation directly to the full-text of an article. Enhancements to the program include a new upload-holdings function that allows libraries to display print holdings in LinkOut. In addition, a new service, Outside Tool, directs users to a local tool where they can explore information local to their own environment. Approximately 100 institutions have registered to connect their users to internal OpenURL-based link resolvers.

The NCBI Bookshelf provides access to the full text of over 38 textbooks in the clinical and research areas of biomedicine. Books may be searched directly or found through links in PubMed abstracts. An innovative indexing approach developed by NCBI permits readers of electronic books to locate sets of related PubMed articles based on phrase matching. In addition to textbooks from commercial publishers, the Bookshelf also includes monographs authored by NCBI, NLM, and NIH staff. Use of the Books database has increased six-fold in the past year and about two million book pages per month are downloaded by users.

Seven new books were added to the database this year as well as over 100 chapters to continuously published books. One new book, *HSTAT—Health Services/Technology Assessment Text*, was a database transferred from the LO. HSTAT contains 741 entries including AHRQ Evidence Reports, AHCPR Supported Guidelines and Consumer Guides, Guides to Clinical Preventive Services, and NIH Consensus Development Programs. Two NCBI resources, *Genes and Disease and Coffee Break* were also transferred to the Books site. Books added include *Molecular Biology of the Cell*, 6th Edition, *Endocrinology: An Integrated Approach*, and *The Genetic Landscape of Diabetes*. The Bookshelf is developing tools for publishing Microsoft Word documents to XML easily, with no technical knowledge required on behalf of the author. The tools are already being used in collaborative projects with the creators of GeneTests (www.genetests.org), the NIH Roadmap Imaging Agent Database group, and the Fogarty Center/World Bank for their book *Disease Control in Developing Countries*.

PubMedCentral (PMC) is a web-based repository of life sciences journal literature providing free and unrestricted access to full-text life sciences journal literature. This repository is based on a natural integration with the existing PubMed biomedical literature database of abstracts. As of August 2004, PMC included over 160 life science journals. Use of the service has increased by 50 percent relative to last year, reaching 830,000 unique users for the month of September 2004.

PubMedCentral has enhanced its value as a digital archive by scanning back issues of journals for online access. Approximately half of the 350,000 articles in PMC have come from the NLM back issue digitization project in the past year. The complete run of the *Bulletin of the Medical Library Association* (1911 forward) was released online in November 2003, making it the first of what will be many journals providing archival access.

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. BLAST also provides users the option of retrieving results with a request ID within 24 hours of searching.

The BLAST suite of programs is continuously enhanced for easier use. Many versions of the database were released this year with BLAST 2.2.9 the last released build in FY2004. BLAST genome pages were added for chicken, cow, pig, dog, sheep and cat as well as an environmental samples data page. The BLAST sequence searching server is one of NCBI's most heavily used services and its usage continues to grow at a pace reflecting the growth of GenBank. Each day more than 200,000 BLAST searches are performed, with users submitting their requests through server/client programs and the Web. 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.

Several programming changes to BLAST queuing and calculation of final alignments have improved the turnaround time for answering a user's query and reduced the peak load on the formatting machines, allowing more searches with fewer resources. A new BLAST report formatter was made available to the public, improving the presentation and value of results. The improvements include a

new alignment style for closely related sequences as well as applying existing alignment styles to searches of translated nucleotides that were not previously supported. A new graphical viewer added the option of retrieving results in HTML format. This option makes it easier for users to store or even produce the results on their own computer and simplifies NCBI processing of formatting requests.

Improvements are routinely made in order to allow easier access to the tools and database by users. Standalone BLAST software is distributed to allow users to run BLAST searches within their own institution. FASTA BLAST database files were migrated from ZIP to GZIP compression format for improved efficiency and storage. Algorithmic improvements have been added to the MegaBLAST program, allowing it to run three times faster for some searches.

Other Specialized Databases and Tools

Documenting the interaction of human immunodeficiency virus type 1 (HIV-1) proteins with those of the host cell is crucial to our understanding of the processes of HIV-1 replication and pathogenesis. To meet this need, the Division of Acquired Immunodeficiency Syndrome of the National Institute of Allergy and Infectious Diseases, in collaboration with the Southern Research Institute and NCBI, has begun to compile a comprehensive "HIV Protein-Interaction Database" to provide a concise summary of documented interactions between HIV-1 proteins and host cell proteins, other HIV-1 proteins, or proteins from disease organisms associated with HIV/AIDS. The database, introduced in April of this year, has been designed to track information for each protein-protein interaction identified in the literature.

The new Cancer Chromosomes database, made public in March, integrates three databases, the NCI/NCBI SKY (Spectral Karyotyping)/M-FISH (Multiplex-FISH) and CGH (Comparative Genomic Hybridization) Database, the NCI Mitelman Database of Chromosome Aberrations in Cancer, and the NCI Recurrent Chromosome Aberrations in Cancer, into NCBI's Entrez retrieval system. Cancer Chromosomes supports searches for cytogenetic, clinical, or reference information using the flexible Entrez search and retrieval system.

Searches in Cancer Chromosomes are based on case information and underlying cytogenetic features. From the results list, users can access the pull-down menu and display a variety of features, including the corresponding literature from PubMed, or the "Similarity Report" showing common elements relating to diagnosis, site, and other cytogenetic abnormalities. As in PubMed, there are pre-computed related links for each case related cytogenetically, diagnostically and/or textually.

The new Entrez Genome Project database is organized around 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 designed around a hub-and-spoke model, with an organism comprising the hub, and individual projects the spokes. This allows the collection of disparate data that all refer to a single organism, conveniently displayed for easy access with references to all subprojects. Currently the database contains 1,408 eukaryotic and 517 microbial genome projects. 37 complete microbial genomes were processed this year which brings the total number to 192 complete genomes of important plant and human pathogens.

The protein clusters database (Proteus) is designed for reference sequence re-annotation by applying consistent and up-to-date annotation to every protein in a protein family across complete microbial and viral genomes. The intent is to increase the speed of re-annotation, increase accuracy and consistency by applying the same annotation across genomes, and to automatically re-annotate proteins that enter the database from new genomes automatically. Proteus is currently under development, it contains approximately 550,000 protein sequences in about 40,000 low-level clusters.

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 plant genomic effort has one technical hurdle relative to other genomic efforts: the range of plant genome size is very large extending from approximately the same size as the genome of many small animals to more than five times as large as the human genome. In September 2004, there were over 80 organisms included the Plant Genomes database, many of which appear in the NCBI Map Viewer.

The Viral Genomes website provides a convenient way to retrieve, view and analyze complete genomes of viruses and phages. This site now contains over 1,600 records for more than 1,200 different species.

The Influenza Virus Resource was created at NCBI with data obtained from GenBank and the National Institute of Allergy and Infectious Diseases Influenza Genome Sequencing Project. This project aims to produce "real time" sequence information during flu season to provide assistance in flu vaccination decisions. This resource will prove to be valuable due to the rapid evolution of flu viruses and will include sequence analysis tools for flu sequences

as well as links to other resources on flu viruses. The database currently contains over 12,000 sequences.

The Gene Expression Omnibus, 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. Currently, GEO contains over 30,000 user-submitted microarrays. GEO Profiles, previously Entrez GEO, contains seven million profiles accounting for hundreds of millions of individual expression points. GEO DataSet (GDS) contains dataset definitions to facilitate identification of experiments of interest. At this time there are 640 curated experiments in the database. Over 100 organisms are represented in these two databases. Graphical and text query tools for gene profiles and datasets have been developed. Multiple clustering methods are available and link to other resources such as HomoloGene and Entrez Gene.

Serial Analysis of Gene Expression, or SAGE, is an experimental technique designed to quantitatively measure gene expression. The SAGEmap tool compares computed gene expression profiles between SAGE libraries generated by the Cancer Genome Anatomy Project (CGAP) and submitted by others through GEO. SAGEmap also includes a comprehensive analysis of SAGE tags in human GenBank records. Data can be retrieved by tag, sequence, UniGene cluster ID, and library name. Links to genomic sequence via the Map Viewer are also available. SAGE includes a total of over six million tags from 12 organisms and 389 experiments.

The NCBI Taxonomy project provides a standard classification system used by the international nucleotide and protein sequence databases. The Taxonomy database contains the names and lineages of more than 130,000 organisms, both living and extinct, represented by at least one nucleotide or protein sequence in GenBank. 40,326 taxa from newly submitted sequences were added to the taxonomy database over the past year, representing a 22% increase from the previous year.

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 also provided. The Taxonomy system also provides a 'Common Tree' function that builds a tree for a selection of organisms or taxa.

A major redesign in FY2004 includes the addition of genomic data and richer links to internal resources, the trace archive, and select external resources through the LinkOut program. A particularly productive collaboration has developed over the last year between the taxonomy group, the NCBI viral genomes project, and the International Committee on Taxonomy of Viruses (ICTV). The PubMedCentral archive is now being scanned and indexed with links to organisms in the Taxonomy database.

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. New organisms added to UniGene this year include: Maxus x domestica (cultivated apple), Ovis aries (sheep), Apis mellifera (honey bee), Bombyx mori (domestic silkworm), Canis familiaris (dog), Helianthus annuus (sunflower), Lactuca sativa (garden lettuce), and Salmo salar (Atlantic salmon).

The PubChem project, a key component of the NIH Roadmap project in Molecular Libraries and Imaging, was initiated in FY2004. 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. Following a rapid development cycle, a public search service became available in September. The PubChem database contains some 900,000 small molecules including their structures, properties, and activities. PubChem marks the first time that this type of comprehensive information on the chemical structures and biological activities of thousands of small molecules will be freely available to the public sector. PubChem in its first version contains legacy data from the NCI's Developmental Therapeutics Program (250,000 compounds), NIST's physical properties database (300,000), NLM's ChemIDplus (100,000), and NIAID's anti-HIV screening program (50,000). PubChem, part of the Entrez database, contains an extensive set of links to PubMed literature citations as well as links to the proteins and/or genes representing a protein they bind to. Compounds are searchable by chemical structures, by chemical properties, and by bioactivity.

NCBI's Molecular Modeling DataBase (MMDB) is Entrez's 'Structure' database, a compilation of all the structures in the Protein Data Bank (PDB). PDB is a collection of all publicly available three-dimensional protein structures, nucleic acids, carbohydrates and a variety of other complexes experimentally determined by X-ray crystallography and NMR and is maintained by the Research Collaboratory for Structural Bioinformatics and the European Bioinformatics Institute.

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 also features custom labeling options, coloring by alignment conservation, and a variety of file export formats that together make Cn3D a powerful tool for structural analysis.

The Conserved Domain Database (CDD) is an 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 now run by default for each protein BLAST search.

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

An interagency agreement with the National Institute of Justice in 2003 commissioned NCBI to develop high-throughput forensic interpretation software for use in state crime labs and for mass fatalities. In 2004, the NCBI development team created and released two public domain software packages: BatchExtract, a software utility to convert DNA electropherogram instrument files into ASCII text for independent analysis, and a beta test version of OSIRIS, an Open Source Independent Review and Interpretation System. NCBI and the Florida Department of Law Enforcement are currently validating OSIRIS for state crime lab use through a series of concordance studies wherein the forensic genotype calls (i.e., DNA "fingerprint") for 20,000 samples from the instrument's genotyping software and the independent OSIRIS genotypes are compared. A new technology for compressing the fingerprint image will permit the immediate interpretation of database "hit" quality when suspect or convicted felon profiles are compared to crime scene samples, and thus reduce the wait time to act on potential leads from days to minutes.

Database Access

Entrez Retrieval System
The major database retrieval system at NCBI, Entrez, was originally developed for searching nucleotide and protein sequence databases and related MEDLINE citations. With Entrez, users can search gigabytes of sequence and literature data with techniques that are fast and easy to use. A key feature of the system is the concept of "neighboring," which permits a user to locate references or sequences which are related to a given citation or sequence. The ability to traverse the literature and molecular

sequences via neighbors and links provides a very powerful and intuitive way of accessing the data.

At this time, Entrez consists of 27 integrated databases providing information on sequences, taxonomy, genes, and literature. Databases added in FY2004 include: HomoloGene, Cancer Chromosomes, NLM Catalog, PubChem Compound, PubChem BioAssay, and PubChem Substance. Entrez Global Query was expanded to include 27 databases for simultaneous searching.

Other Network Services
Usage of NCBI's Web services continues to expand as more information and services are added. NCBI staff continued to make access and usage easier with improved documentation and tutorials. A web usability group was established this year to address issues such as improving awareness of underutilized services, implementing a better and more consistent means to navigate the NCBI site, establishing a content management system, and evaluating user experience of all services.

The NCBI web provides an integrated approach to accessing all of NCBI's database and services as well as general information about NCBI, its research, data submissions, and updates. At the end of FY2004, NCBI's site was averaging over 40 million hits daily. Because of the mission-critical nature of NCBI's computing platforms for PubMed, Entrez, BLAST, and other services, extensive system monitoring is performed. Based on measurements taken every 15 minutes from 50 ISP monitoring sites across the U.S. and overseas, the average time to load the entire NCBI home page is 0.82 seconds, an average PubMed search takes less than 2.5 seconds and availability has been better than 99.5 percent.

NCBI has a number of network services that provide programmatic access to several important NCBI databases. A monitoring program was developed to make sure all of these services are responsive and producing correct information. The program quickly notifies relevant staff members if any service for which they are responsible becomes unavailable or starts producing unexpected or incorrect results. The detailed diagnostic information provided by the program allowed coding bugs, configuration errors, and hidden dependencies to be understood and fixed rapidly, greatly increasing the reliability and utility of the services being monitored.

Software development in NCBI has largely shifted from "C" to "C++" programming language as the relatively new NCBI C++ Toolkit has matured and stabilized enough to replace the older NCBI Toolkit written in "C."

NCBI started providing sequence data in an XML format known as INSDSeq in FY2004. This format is an XML structured mapping of the GenBank flatfile fields that annotate DNA and

protein sequence records. It is designed to be used by academic groups and biotechnology companies, who have always parsed the data from the GenBank flatfile into their own analysis programs. One significant advantage to INSDSeq is that biological feature intervals are presented in an expanded format that is much easier to parse than the condensed form required in GenBank format.

Software used to extract the actual nucleotide and protein sequence letters from within a GenBank or RefSeq sequence record was also redesigned this year. The new code was placed in programs that produce FASTA files, write GenBank or INSDSeq format, and validate sequence records and decreased processing time up to 40%. Changes to the GenBank flatfile generator greatly sped up the performance of the Entrez web site when producing GenBank reports on genomic sequences. By eliminating the need to reload components of a large genomic record, the overall speed of transfer improved by a factor of two.

The new NCBI computer room in the B2 level of Building 38A now houses a major part of NCBI's computing infrastructure. This room is connected to NCBI's portion of the NLM computer room by multiple gigabit Ethernet connections. FY2004 saw a major expansion and upgrade to the "NCBI Compute Farm," a batch queue processing system that functions as a virtual supercomputer and supports many CPU-intensive production and research activities. NCBI's plan to centralize storage using Network Attached Storage was substantially advanced with a total increase in network storage capacity to approximately 100 TB. Also, NCBI's Computers and Networks Section established the basic network infrastructure necessary to support the NIH Consolidated Collocation Site and began to provision public services at the site. This site, located in Sterling, Virginia, is in addition to NCBI's existing facility in NLM. It will also provide continuity of critical services such as PubMed, BLAST, and FTP in the event that the IT infrastructure in Building 38A is unavailable for any prolonged period.

Research

Research is at the core of NCBI's mission. The Computational Biology and Information Engineering Branches are the main research branches of NCBI, with the latter branch concentrating on applied Research and Development. Each Branch comprises a multidisciplinary team of scientists that carries out research on a broad range of fundamental problems in molecular biology by developing and applying mathematical, statistical, and other computational methods. Research conducted by NCBI investigators has strengthened applications and database work and has led to the development of many new theoretical

and practical models that have opened doors to new areas of research.

NCBI's basic research group is within the Computational Biology Branch and consists of 70 senior scientists, staff scientists, research fellows, and postdoctoral fellows. Projects focus on new computer methods to accommodate the analysis of genome sequences and molecular sequence databases due to the rapid growth in large-scale sequencing efforts. Other projects focus on such techniques as the analysis of particular human disease genes and the genomes of several pathogenic bacteria, viruses and other parasitic organisms, as well as collaborations with experimental laboratories. New areas of research include: development of novel amino acid substitution matrices for improved sensitivity of sequence alignment programs, evolutionary genetics, analysis of gene regulatory pathways, the development of new modeling tools for tumor DNA data, single nucleotide polymorphism data analysis, analysis of malaria genomes for vaccine development, evolutionary analysis of protein domains and comparative genomics, and development of mathematical models of genome evolution. New databases are also being designed for data on conserved protein domains and mRNA expression experimental results. Staff continued collaboration with other NIH institutes for sequence analysis, gene identification, and the analysis of experiments on gene expression. Collaboration was also continued with several institutes worldwide on genetic linkage analysis problems.

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 this group is evidenced by the number of peer-reviewed publications, over 100 this year with more in press. The staff participated in numerous oral presentations and mounted posters at various scientific meetings, and at universities worldwide. Presentations were also made to visiting delegations, oversight groups, and steering committees. NCBI also hosted numerous 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. The NCBI uses the NIH Intramural Research Training Award Program and the Fogarty Visiting Fellow mechanisms to recruit for this program.

Outreach and Education

NCBI continues to expand its outreach and education programs to increase awareness of its myriad public databases and specialized tools and services. Over the

past year NCBI staff maintained a general web site on NCBI resources; 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 increasing diversity of NCBI's products and services, the course, "A Field Guide to GenBank and NCBI Resources," was expanded and is taught at NIH and throughout the United States as requested. The course consists of a three-hour lecture, a two-hour hands-on practicum, and one-on-one sessions if requested. In FY2004, additional modules were developed that focus on specific tools and databases including structures and gene expression. An extended two-day course was also presented that combined the main course and separate modules. The 11-member teaching staff presented 64 courses to over 5,000 people in FY2004.

Education: Mini-Courses and Lecture Presentations
NCBI offers 10 mini-courses to provide a practical introduction to various programs. Three new mini-courses introduced in FY2004 include, "GenBank Quick Start," "Identification of Genes and Disease," and "Correlating Disease Genes and Phenotypes." This year, 28 mini-courses were offered to over 1,200 participants.

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 an intramural Core Bioinformatics Facility (CoreBio)— a network of bioinformatics specialists serving individual institutes within the NIH. Individual CoreBio Members are trained over a nine-week period in the use of bioinformatics tools provided to the research community by NCBI. 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, conducting eight 9-week training programs, two in the past year since the program began in 2001. Twenty-five update sessions and two special topic sessions for the institute representatives have also been held. One-on-one consultations are available on an ongoing basis for NIH scientists with NCBI faculty in the NCBI

Learning Center, the NIH Library, and the NCI-Frederick Cancer Research and Development Center.

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

The third "NCBI Advanced Workshop for Bioinformatics Information Specialists" was held in FY2004. The collaborators and 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 original materials. Together, the collaborators and course alumni form the growing Bioinformatics Support Network, a group supported by NCBI which has been established for the purpose of communication and continuing education among members.

A regional training program throughout the country for the three-day introductory course was launched to complement the introductory course taught by NLM by increasing its accessibility nationwide. Four courses were offered with a total of 63 participants in addition to the 16 at the NLM course. Participants support users of NCBI resources at their institutions through instruction integrated into their training curriculums, introductory workshops, and direct clientele assistance. Regional courses were taught by Educollab members who also work with NCBI to teach the five-day Advanced Workshop. 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.

Outreach: User Guides for NCBI Resources
NCBI has continued to develop a comprehensive list of fact sheets that outline the services and databases offered by NCBI. These fact sheets and guides are available for printing via the "About NCBI" site. In addition, a number of other informational and educational resources are available on the NCBI Web site. Links are available that discuss the fundamental principles of biomolecular research and underlying sequence similarity search tools. Interactive tutorials may be found for a number of databases and search and retrieval tools such as Entrez, PubMed, Structure, and BLAST.

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 FY2004, over 18,000 printed copies of the NCBI News were distributed quarterly. Access to the newsletter via the NCBI Web site has increased dramatically as more people have become aware of its availability online.

Biotechnology Information in the Future

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

Extramural Programs

Milton Corn, M.D.
Associate Director

The Extramural Programs Division (EP) of NLM continues to receive its budget under two authorizing acts: the Medical Library Assistance Act (unique to NLM), and Public Health Law 301 (covers all of NIH). The funds are expended mainly as grants-in-aid, and in some instances as contracts, to the extramural community in support of NLM 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 database.

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

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

Highlights of FY 2004:

- The number of applications assigned to NLM and reviewed by BLIRC continued to increase. Statistics show a drop in success rates for most NLM grant programs due to the increase in applications and a leveling of the budget after several years of substantial increases.
- A new support staffing model for extramural grants offices has been under development at NIH, following the A-76 competition which was won by NIH. EP participated in early training for several of these people.
- EP continues to refine and sharpen its statements about priorities and interests for grant projects in biomedical informatics and bioinformatics.
- The new EP web site and new NLM web site are providing improved access to grants

information for prospective applicants as well as simplified instructions for those new to preparing applications.

Resource Grants (MLAA)

Resource Grants, authorized by the Medical Library Assistance Act, support access to information, connecting computer and communications systems and promote collaboration in networking, integrating, and managing health-related information. The four Resource Grant programs range in complexity as well as in dollar amounts and duration. Internet Access to Digital Libraries (IADL), Information systems, and Integrated Advanced Information Management Systems (IAIMS) grants are considered "seed" grants designed to initiate and deploy elements of the information environment that are expected to become self-sustaining after grant funding ends. Publication grants support the development of scholarly works in selected areas relevant to health and biomedical sciences. All Resource Grants are open to public and private, nonprofit health institutions engaged in health education, research, patient care, and administration. Many include health sciences or public librarians as active participants.

Internet Access to Digital Libraries Grants.
IADL grants enable organizations to offer access to health-related information provided by NLM and others, to transfer files and images, and to interact by e-mail and videoconferencing with colleagues throughout the world. IADL grants provide up to $45,000 for a single institution and up to $8,000 each for up to 15 additional performance sites. The applicant may propose two years as the project period, but a longer project period does not increase the total size of the award. Forty-four applications for IADL grants were reviewed in FY2004, from which 7 new grants were funded. The average priority score of new IADL grants funded was 155. Nineteen of the applications were received from community organizations or health centers and seventeen from independent hospitals. Of those awarded, 6 went to community organizations or health centers. In addition, 2 grants were awarded that were approved for funding in FY2003 but not funded until FY2004.

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

resources and services. Eighty four information system grant applications were reviewed in FY2004, and 15 new grants were funded. The average priority score for new information system grants funded was 163. Forty-three of the applications came from academic centers, 22 from community or health centers and 13 from independent hospitals. Of those awarded, 8 went to community or health centers and 6 to academic centers. In addition, 4 grants were awarded that were approved for funding in FY2003 but not funded until FY2004.

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

The IAIMS program contains five options, of which two are funded with MLAA funds. IAIMS Planning Grants provide up to $150,000 per year for one or two years, with an optional infrastructure supplement of $100,000 in the second year; IAIMS Operations Grants provide up to $400,000 per year for up to four years. Twenty three IAIMS grant applications were reviewed in FY2004, of which 14 were planning grants. Two awards were made for IAIMS Operations grants, and 3 new IAIMS planning grants were awarded. The average score of successful IAIMS planning grants was 169.

Publication Grant Program
The Publication Grant Program provides short-term financial support for scholarly research that will lead to a publication. Studies prepared or published under this NLM program include critical reviews or research monographs in the history of medicine and life sciences; special areas of biomedical research and practice; medical informatics, health information science and biotechnology information. Unique at NIH, the publication grant is also unusual among NLM's grant programs in that it accepts applications from individuals without an organizational affiliation. Seventy publication grant applications were reviewed in FY2004, and 18 new grants were awarded. The average priority score of new publication grants funded was 167. In addition, one new grant was awarded that was approved for funding in FY2003 but not funded until FY2004.

Training and Fellowships (MLAA)

Overview
Exploiting the potential of computers and telecommunication for health care information

requires investigators who understand biomedicine as well as fundamental problems of knowledge representation, decision support, and human-computer interface. NLM remains the principal support nationally for research training in the fields of biomedical informatics as applied to clinical medicine and to basic research. NLM provides both institutional and individual training support.

NLM-Supported Training Programs
Five-year institutional training grants support over 250 trainees at pre-doctoral and postdoctoral levels. Eighteen training programs were funded for a new five-year period beginning July 1, 2002. Eleven of the previous twelve were again funded, and seven new programs were added to the set. NLM is expanding its support for such programs in response to the marked recent interest in biomedical computing and the consequent need for trained informaticians. Among our programs, training for bioinformatics is now receiving significantly more attention and opportunity than in previous years, and, for the first time, a program dedicated to imaging informatics is included. For the latter, NLM receives some co-funding from NIBIB, the new NIH Institute for bioengineering and imaging. NIDR continues to contribute funds to NLM to help support slots at these training sites for applicants interested in dental informatics. The 18 programs currently funded are at the following universities: California (Irvine), California (Los Angeles), Columbia, Harvard, Indiana, Johns Hopkins, Minnesota, Missouri, Oregon Health Science, Pittsburgh, South Carolina, Stanford, Rice, Utah, Vanderbilt, Washington, Wisconsin, and Yale.

This program is scheduled to be re-competed in FY2006. To provide EP with a timely overview of what the programs are doing, EP embarked on a cycle of evaluative site visits in FY2004 Each site visited is asked to provide a pre-visit report, giving statistics and other background on the curriculum, students and faculty. One-day visits were made by a team of three EP staff members plus an outside consultant to the following locations: California (Irvine), California (Los Angeles), Missouri, South Carolina, Rice, and Wisconsin. Following each visit, the principal investigator receives a letter which summarizes the findings of the team. This letter becomes part of the official grant file.

Individual Fellowships

Informatics research training
NLM offers two 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 10 or

more years of professional experience in an appropriate field (F38). In FY2004 21 applications were received for the F37 program, of which four were awarded. The average score of successful proposals was 159 for the F37 program. Seven applications were reviewed for the F38 program, and one award was made.

Training for Informationists
In October 2003, NLM issued program announcements for two new fellowships, both aimed at supporting the training of in-context information specialists. These programs use the F37 and F38 mechanisms, but emphasize training for professional careers, not research training. One F37 application was received and it was funded.

IAIMS Fellowships
No applications were received for this program.

Early Career Development Awards
This program provides transition assistance for biomedical informaticians who are establishing their initial independent research programs. Applicants may apply without yet having identified their home institution; once a position is secured, the award process is completed. Fourteen applications were reviewed in FY2004 for this program, and 3 awards made. The average score of a successful application was 165. Two K22 awards were issued that received approval in late FY2003 but could not be funded until FY2004.

Loan Repayment Program
NLM participates in NLM's loan repayment program by identifying applications it is willing to sponsor. These applications are reviewed for merit by a Special Emphasis panel. A central NIH office checks the suitability and substance of the applicant's debt and employment status. For FY2004, NLM funded 7 LRP awards of the 14 received.

Biomedical Ethics
Ethical issues in health care and research produce an enormous literature. This literature comes from law, medicine, public health, philosophy, and government publications. The National Reference Center for Bioethics Literature at Georgetown University continues to offer invaluable resources and guidance for workers in this area. A contract from NLM's Library Operations program area now supports the Center as well as the indexing and cataloging of materials cited in MEDLINE and LocatorPlus. Arrangements were completed in FY2004 to consolidate the two separate contracts previously managed by Extramural Programs and Library Operations into one. Transition will take place in early FY2005.

Research Support (PHS 301)

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

Biomedical Informatics and Bioinformatics

Research Grant Program
In the early years of the R01 grant program, the majority of NLM's research support in informatics focused on the informatics of health care delivery with support both to applied projects (e.g., the electronic medical record, telemedicine) and related basic problems (e.g., natural language processing, data-mining, knowledge representation). In recent years there has been marked expansion in research support for informatics issues related to biological and medical research. Thus, the research grant program now has two "branches," both of which are funded from PHS 301 funds. In FY2004, a new program announcement was issued, updating the language and clarifying NLM's research interests in biomedical informatics and bioinformatics. Forty eight applications were reviewed for this program, and 7 awards made. The average score of awarded applications was 174. In addition, 7 grants were funded that were approved in FY2003 but could not be funded until FY2004. All but three research grant applications came from academic centers.

Small Grant Program
To complement its traditional R01 grants, in 2003 NLM issued a program announcement for small project research grants, a mechanism used by most of the NIH Institutes. These grants provide $50,000 per year for one or two years, and are designed to help researchers who are just starting out in an area of inquiry. Feasibility and proof of concept studies, and the gathering of preliminary data that might support a subsequent R01 study are typical uses of the R03 grant. Thirty nine R03 grants were reviewed in FY2004, and 4 new grants were funded. The average priority score of funded R03 grants was 148. Like R01grants, most R03 applications came from academic organizations. In addition, 5 R03 grants were funded that were approved in FY 2003 but could not be awarded until FY 2004.

Informatics for Disaster Management
NLM's program of research grants exploring the application of informatics approaches in natural and man-made disasters. initially an R01 mechanism, is

now an R21 mechanism to better accommodate projects that are more akin to engineering research & development than to hypothesis-testing experimental research. During the formal change process for this program, two other institutes (National Institutes of Mental Health and National Institute of Biomedical Imaging and Bioengineering) signed onto NLM's program announcement. Nineteen new applications in this program were reviewed in FY2004, and no new awards were made. The average score of Informatics for Disaster Management applications reviewed was 308. Eleven of the new applications were from academic centers, four from for profit firms.

NLM Exploratory/Developmental Grants
NLM's new Exploratory/Developmental grant fills a niche between Resource and Research grants and was issued in concert with development of such a mechanism by NIH. Announced in April 2003, the R21 grant supports high risk/high yield projects, proof of concept, and work in new interdisciplinary areas. Preliminary data are not required for these grants, and emphasis in review is shifted from hypothesis testing to achievement of milestones during R&D. Eight applications were reviewed in FY2004, and 1 new award was made. The average score of new applications reviewed was 237.

Resource Grants for Biomedical Informatics/Bioinformatics
In August 2004, NLM issued a program announcement for an earlier, expired program of support for scientific research resources. This program, which uses the P41 grant mechanism, is similar to an R01 grant but contains a service component and support for maintenance of a resource or service. The applicant must demonstrate that the proposed resource or service is already actively used by researchers or clinicians across the US or the world. Seven new applications were reviewed in FY2004, and 3 were funded. The average score of the awards was 139.

Conference Grants
Support for conferences and workshops is intended to help scientific communities in focused areas of informatics and bioinformatics to identify research needs, share results, and prepare for productive new work. The average conference grant is about $10,000. The program allows multi-year awards. EP generally caps conference awards at $20,000 per year. To expedite processing of these grants, NIH permits a two-level review to be done by NLM staff. Of three applications received in FY2004, one was funded.

Pan-NIH Projects

NLM and Roadmap Activities
A major pan-NIH enterprise initiated by the Director, NIH, is resulting in requests related to three themes: New Pathways to Discovery, Research Teams of the Future, and Reengineering Clinical Research. NLM is a participant in all of the Roadmap initiatives, and EP staff was actively involved in NIH Roadmap teams for the National Centers for Biomedical Computing and a number of interdisciplinary research initiatives. Although NIH Roadmap grants are considered pan-NIH grants, and awards will be managed by teams of program officers, each Roadmap grant has a "home" Institute.

NCBC and BISTI
Initially, following the award of several planning grants for BISTI (Biomedical Information Science and Technology Initiative) Centers, NIH intended to issue an RFA to support a selection of those centers. Instead, the NIH Roadmap issued an RFA for National Centers for Biomedical Computing NCBC). P.I.s on the existing Planning Grants were eligible to apply but were not accorded preference in the review. Forty-one grant proposals were received in response to the RFA, and 4 were awarded. NLM is administrative home for one NCBC grant. Because the Roadmap initiative provided only $12 million dollars for Centers, NLM, NIGMS, and NCRR combined to contribute an additional $4 million so that 4 Centers could be funded, each at a total cost per year of $4 million. These cooperative agreements last for five years and can be renewed for another five years.

Special Multi-institute Projects

Multi-institute Program Announcements
In addition to its involvement in the NIH Roadmap, NLM also participates with other NIH and federal organizations in a number of multi-agency projects, including the Human Brain Project, the Pharmacogenetics Research Network, and a number of individual program announcements that focus on tool development, innovation in computational sciences for biomedicine, and other informatics-related topics. The applications for these programs are reviewed by the NIH Center for Scientific Review, and then participating institutes select grants for full or shared funding. NLM participation has been steady but is rarely more than one new grant each year, and in some years none is funded. The statistics for these programs are folded into regular grant program counts; most are R01 grants. An updated listing of the multi-institute initiatives in which NLM participates is available on the EP Web site.

Informatics for the National Heart Attack Alert Program (Research Contracts)
Although some small supplements were added to several of these projects in FY2004, funding for the National Heart Attack Alert Informatics Program is essentially complete. A contractors' conference was held for spring of FY 2004. NLM's involvement in the program has now ended.

Shared Funding for Research Grants
The NLM provides funding for Bioinformatics and Biomedical Informatics by its continuing support of collaborative extramural funding with other agencies. The NLM continues its support of the Protein Sequence Databank at Rutgers University jointly with the NSF and has increased its fiscal commitment to the project. This databank serves as the single worldwide repository for the processing and distribution of 3-D biological macromolecular structure data.

The NLM has collaborated with the Fogarty International Center (FIC) in support of their International Training in Informatics by funding the "AMAUTA Health Informatics Research and Training Program" that involves the collaborative efforts of the University of Washington and the Unversidad Peruana Cayetano Heredia (UPCH) in Lima, Peru. This training addresses informatics for global health and will help the UPCH to establish a health informatics research program within Peru. The FIC also received complementary co-funding support from the NLM for their International Training Program in Informatics for the "Informatics Training for Public Health in Tanzania." This program includes the collaborative efforts of the Harvard School of Public Health and the Muhimbili University College of Health Sciences. This ten-year alliance has been focused on addressing major public health problems in Tanzania through multidisciplinary teams of investigators in research and training. This program will support advanced degree programs in public health and work towards a sustainable training program in Tanzania.

As part of the NLM's support of its Training Program for Bioinformatics, the NLM receives ongoing co-funding support from the National Institute of Dental and Craniofacial Research for support of Dental Informatics trainees. The NLM also receives co-funding support from the National Institute of Biomedical Imaging and Bioengineering for trainees in bioinformatics.

The NLM has also provided joint funding to the National Institute of General Medical Sciences providing continued support of a cooperative agreement for the Stanford Phamacogenetic Knowledge Base. In addition, NLM provided co-funding support for NIH Roadmap directed to the National Center of Biocomputing at the Brigham and Women's Hospital, "Informatics for Integrating Biology and the Bedside." This multifaceted cooperative agreement supports seven core projects and four research projects. The informatics domains include: clinical informatics, functional genomics and genetics of complex traits.

The National Human Genome Research Institute (NHGRI) is continuing to co-fund the United Protein Databases (UniProt) of the European Bioinformatics Institute (EBI) under a cooperative agreement. This support provides a single database for protein sequence and function linking existing information from other databases supporting protein structure information. The NLM has also provided co-funding support for a NHGRI grant entitled the "Oral History of Human Genetics: The Intelligent Archive" that will include a collection of over 100 oral histories from clinicians, scientists, theorists, organizational leaders and others covering the ethical, legal and social issues surrounding the field of human genetics. The NLM and the NHGRI jointly funded a new research grant entitled, "BioMediator: Biologic Data Integration & Analysis System" for searching across various genomic databases for the purpose of curating the GeneClinics Database formerly supported by a NLM grant. The GeneClinics project, under NLM contract support, has now been integrated into the NCBI as one of their genetics resource databases available to clinicians and research scientists.

The NLM provided co-funding to the National Center for Research Resources (NCRR) in support of a Neuroimaging Analysis Center that would support further development and extensions of the Insight Toolkit that will be used in working in a grid computing environment. This grid computing application will allow for computation of very large datasets that otherwise could not be analyzed effectively. The NLM also provided NCRR with co-funding support for the pan-NIH initiative on Electronic Research Administration. This cooperative agreement, "Electronic Submission of Grant Applications," also supports the overall Federal Government's e-Grants and e-Government initiatives.

The NLM provided support for the 7th Annual International Protégé Workshop hosted by Stanford University and jointly funded by the National Cancer Institute and the NLM through an intra-agency agreement.

In support of the Small Business Technology Transfer Research (STTR) Program at the NIH, the NLM provides co-funding support of a grant at the National Institute of Nursing Research for the continued development of a home caregiver device for people that are cognitively impaired preventing personal injury at night. The "Night Alert Prompting System" is a collaboration of the Amron Corporation and the University of Florida a part of

the STTR program joining private companies with universities in support of a business-research partnership.

The NLM provided funding support as part of a co-funding agreement to the National Institute of Neurological Disorders and Stroke for a project entitled, "A Mature Brain Architecture Knowledge Management System (BAMS)." The objective of this project is to develop a user-friendly neuroinformatics workbench for the Web allowing the neuroscience community to access, evaluate and visualize neuroanatomical literature. BAMS will facilitate basic research into the cause and treatment of all diseases that affect the brain.

The NLM provided approximately $1.4 million in collaborative co-funding agreements in FY2004.

Joseph and Rose Kennedy Institute of Ethics/Georgetown University
The Division of Extramural Programs has continued its support of the National Reference Center for Bioethics Literature (NRCBL) in 2004. Early in the year the NRCBL staff worked with its counterpart in Bonn, Germany, Deutsche Referenzzentrum für Ethik in Den Biowissenschaften, and extended its ethics classification scheme to include French and German to the already existing English. The expanded classification table currently allows those searchers more familiar with French or German access to the "ETHX on the Web" database. The National Information Resource on Ethics and Human Genetics, NCRBL staff has evaluated its existing computer platform and concluded that a more modern and flexible platform (not provided by NLM contract) was required and this would be incorporated with improvements during 2004. The NRCBL continues to publish volumes of its ongoing collection, *New Titles in Bioethics*. The primary measure of success of this ethics resource has been its use as a library, and as such, has provided personal responses to requesters from around the world. There are approximately 1 million Web queries per year. NRCBL staff has also provided training sessions on the effective use and access to the library databases and resources for a graduate seminar on nursing ethics taught at the George Mason University. There are a number of new bioethics publications for 2004 including *Digital Library Projects: Beyond the Beltway and Bioethics Searchers Guides: Using Databases of the National Library of Medicine*. The NCRBL will continue to be supported by NLM in 2005 as one of the premier national and international bioethics literature and information resources.

GeneTests
The NLM has in the past awarded several grants to the University of Washington in support of clinical genetics databases titled Helix, GeneClinics and GeneTests, primarily in support of clinical health professionals. The competitive application for renewed funding support for the GeneTests project proposed the consolidation of all three databases. The renewal proposal was reviewed, positively scored, and considered for funding. The NCBI expressed interest in incorporating this GeneTests resource into the NLM database resources by converting the grant proposal to a contract proposal. The new sole-source contract provides public access though an NLM web interface.

The new website for GeneTests provides four categorical breakdowns for information. The GeneReviews portion provides online publications of expert-authored disease reviews. A laboratory directory provides an international directory of genetic testing laboratories and the clinic directory an international directory of genetics and prenatal diagnosis clinics. In addition, a repository for educational materials that includes an illustrated glossary, information about genetic services and PowerPoint slide presentations are available. Approximately 30,000 entries are viewed per day.

Shared Funding for Training
In June 2003, the Fogarty International Center issued an RFA for Informatics Training for Global Health. The review of applications was handled administratively by EP's Scientific Review Unit. NLM is providing full funding ($250,000 per year) for one of the programs, and is co-funding a second.

There are discussions with the Robert Woods Johnson Foundation exploring possible RWJ support of training slots for Public Health Informatics Research at some of NLM's existing Informatics Research Training Programs.

SBIR/STTR (PHS 301)
All NIH research grant programs, including NLM's, by Congressional mandate allocate a fixed percentage of available funds every year to Small Business Innovation Research (SBIR) grants. These projects may involve a Phase I grant for product design and a Phase II grant for testing and prototyping. SBIR and STTR applications are reviewed by CSR. Sixty three SBIR/STTR applications were assigned to NLM tin FY and reviewed by CSR. Two awards were made. Of these applications 35 were 'unscored,' indicating reviewer assessment that they were not in the top ½ of applications received. The average score for SBIR/STTR grant awarded was 199.

EP Operating Units - Highlights

Grants Management Office
The Grants Management staff reviews NLM grant applications for compliance with guidelines and

directives; prepares and disseminates grant awards; maintains official grant files for NLM; provides consultation and assistance to grantees on appropriate business management concepts; and advises NLM officials on grants management policy and procedures.

The Grants Management staff, which consists of four employees, issued a total of 322 awards for FY2004, including grants, administrative supplemental awards, fellowships and administrative actions. Details of the grants are provided in Appendix 1, Table 2. Of these, 212 were for new and non-competing awards in NLM's own grant programs. Grants Management staff continues to provide budget oversight for all awards, and prepares reports for NLM staff and Congress as requested.

Committee Management Activities
Board of Regents: The Board of Regents met three times in FY 2004 on February 10-11, May 19-20, and September 21-22. The Extramural Programs Subcommittee was held prior to each of these meetings. The Board approved 373 grant applications, including any special reviews made by the EP Subcommittee. These special reviews are conducted when the recommended amount of financial support is larger than some predetermined amount; when at least two members of the scientific merit review group dissented from the majority; when a policy issue is identified; or when an application is from a foreign institution. The EP Subcommittee makes recommendations to the full Board, which votes on the applications. The Board Operating Procedures were reviewed and approved without change at the February 10-11, 2004 meeting.

Presentations of Programs to the Board of Regents in FY 2004
EP programs presented to the BOR for Concept Review and Approval:
- NLM participation in a collaborative effort, Interagency Opportunities in Multi-Scale Modeling in Biomedical, Biological, and Behavioral Systems. Partners and purpose of the initiative were described.
- NLM possible participation in a broad variety of NIH Roadmap activities, including (1) Re-engineering the Clinical Research Enterprise: Feasibility of Integrating and Expanding Clinical Research Networks, (2) Training for a New Interdisciplinary Research Workforce, (3) Interdisciplinary Health Research Training: Behavior, Environment and Biology, (4) Short Programs for Interdisciplinary Research Training Exploratory Centers (P20) for Interdisciplinary Research, and (5)

National Centers for Biomedical Computing.
- NLM's Informatics Research Program—approval of announcement describing the program.

EP programs presented to BOR as updates included the Publication Grant Program, the Loan Repayment Program, the NIH Roadmap Initiative, the Small Project Grant Program, National Centers for Biomedical Computing, and NLM's Informatics Training Conference.

Scientific Review Office
NLM's initial review group, the Biomedical Library and Informatics Review Committee (BLIRC), evaluates grant applications for scientific merit. BLIRC met three times in FY2004 and reviewed 204 applications. The Committee (see Appendix for roster of members) operates as a "flexible" review group. BLIRC reviews applications for medical informatics and biotechnology research projects, information systems, and publications. BLIRC has two standing subcommittees: the Networked Information Access Subcommittee and the Medical Informatics Subcommittee. The subcommittees consider applications for informationist fellowships, and training awards in medical informatics and biotechnology information, respectively.

The Amended Charter of the Biomedical Library and Informatics Review Committee reflects 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: 18 Special Emphasis Panels were held during FY2004. These panels are convened on a one-time basis to review applications for which the regularly constituted review group lacks appropriate expertise, or when a conflict of interest exists between the applicant and a member of the BLIRC. Lately, due to the increase in number of applications received, the panels have also been convened to review applications that simply cannot be reviewed in the BLIRC. The panels reviewed a total of 178 applications during FY2004. One site visit to evaluate an IAIMS Operations application was also carried out by an ad hoc panel.

A Special Emphasis Panel was convened in February 2004, at the request of the Fogarty International Center, to review applications responding to their RFA for training grants, "Informatics Training for Global Health."

A second level peer review of applications is performed by the Board of Regents as described above. One of the Board's subcommittees, the Extramural Programs Subcommittee, meets the day before the full Board for the review of "special" grant

applications. Examples include applications for which the recommended amount of financial support is larger than some predetermined amount; when at least two members of the scientific merit review group dissented from the majority; when a policy issue is identified; and when an application is from a foreign institution. The Extramural Programs Subcommittee makes recommendations to the full Board, which votes on the applications.

Program Office

Program activities in FY2004 were focused on clarifying NLM's research interests, evaluating the university-based informatics training programs, building new collaborations, and publicizing NLM's grant programs.

Program referral guidelines
The creation of the National Institute on Biomedical Imaging and Bioengineering (NIBIB) brought a new set of research interests to NIH that overlap in several areas with those of NLM. New referral guidelines for EP were prepared this year and sent to the Center for Scientific Review. These will continue to be refined as the other "non-categorical" Institutes increase their funding in informatics. The primary overlaps are with NCRR, NIGMS, NHGRI and NIBIB.

Program announcements
NIH requires all standing programs to be re-issued every three years. Program staff identified all current and expired programs and developed a timetable for re-issuing expired programs. The expired announcements replaced in FY2004 were for the R01 Research Grants and the P41 Research Resource Grants. Draft text was completed for updates to the Information Systems grant and Publications Grant, which will be issued early in FY2005.

Collaborations
EP was a co-sponsor and active participant in organizing the 2004 BECON Symposium, entitled "Biomedical Informatics for Clinical Decision Support: a Vision for the 21st Century." The meeting, held June 21–22, was well-attended. In addition to participation in NIH Roadmap workgroups for the National Centers for Biomedical Computing and

Interdisciplinary Research, EP program were involved in several new multi-agency grant program announcements.

Training-related Initiatives
The Annual Training Conference was held June 9 and 10 in Indianapolis. Poster sessions and break-out sessions were included in the meeting for the first time and with great success. At that meeting, Training Directors were briefed on the interest of Robert Wood Johnson Foundation in partnering with NLM to provide training in public health informatics. Parameters of such a program were developed, to be presented to the RWJF Board for approval in November 2004.

Administration and Operations

Personnel Activities
EP has had several personnel changes over the past year and the Division experienced five losses of permanent staff. The NIH A-76 competition for grants management, review and program support staff was completed and officially implemented on October 4, 2004. Only one position was transferred to the NIH's Division of Extramural Administrative Support. Three additional staffers from the NIH Division were also assigned to fill the remaining 3 support vacancies in EP.

Some Issues That Impact NLM Extramural Budget and Programs
- Increasing numbers of applications while the budget is flat may require some narrowing of NLM's funding interests if a reasonable payline is to be maintained
- NLM's participation in Roadmap, BISTI and other multi-Institute computing initiatives inevitably decreases available funds for NLM's own grant programs.
- Because some of the informatics areas NLM has supported for many years are now also being funded generously by other Institutes, the proper future focus for NLM's grant programs in biomedical computing would benefit from reevaluation. The upcoming long-range plan meetings could provide a useful forum for such analysis.

EXTRAMURAL PROGRAMS
FY 2004 Final
($ in 000)

EXTRAMURAL PROGRAM BUDGET

	NON COMPETING		COMPETING		TOTAL	
	NO	AMT	NO	AMT	NO	AMT
MLAA		($)		($)		($)
IAIMS	2	$904	6	$1,237	8	$2,141
TRAINING						
TRAINING PROGRAMS (T15)	18	$14,210	0	$0	18	$14,210
FELLOWSHIP(F37/F38)	5	$270	8	$598	13	$868
CAREER(K22)	0	$0	5	$750	5	$750
TOTAL TRAINING	23	$14,480	13	$1,348	36	$15,828
PUBS(G13)	17	$1,144	20	$1,298	37	$2,442
RESOURCE						
IADL(G07)	3	$108	7	$452	10	$560
INFO. SYS.(G08)	20	$2,765	16	$2,033	36	$4,798
TOTAL RESOURCE	23	$2,873	23	$2,485	46	$5,358
BIOETHICS(N01)*	0	$0	0	$0	0	$0
GENETESTS(N01)	0	$0	0	$0	0	$0
LOAN REPAYMENT(L30)	0	$0	7	$419	7	$419
NN/LM CONTRACTS(N01)	8	$12,884	0	$0	8	$12,884
TOTAL MLAA:	73	$32,285	69	$6,787	142	$39,072
PHS 301						
BIOMED-INFORM. RESEARCH (R01/R03/R13/R21/R24/P41)	36	$9,334	11	$1,923	47	$11,257
PROTEIN SEQ. DATABANK(IAG)	0	$0	1	$200	1	$200
CHAIRMAN'S GRANT(U09)	1	$200	0	$0	1	$200
BIOMED-INFORM. RESEARCH TOTAL	37	$9,534	12	$2,123	49	$11,657
BIOINFORM. RESEARCH(R01/R03/R21)	14	$4,240	7	$2,473	21	$6,713
BIOINFORM. RESOURCE(P41)	2	$950	1	$2,293	3	$3,243
BISTI(R21/R33/P20/P41/U54)**	7	$3,476	2	$967	9	$4,443
BIOINFORM. RESEARCH TOTAL	23	$8,666	10	$5,733	33	$14,399
SBIR/STTR(R43/R44/R41/R42)*	2	$955	2	$238	4	$1,193
TOTAL PHS 301:	62	$19,155	24	$8,094	86	$27,249
TOTAL EP:	135	$51,440	93	$14,881	228	$66,321

TABLE 11

Office of Computer and Communications Systems

Simon Y. Liu, Ph.D.
Director

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

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

OCCS helps to coordinate, integrate, and standardize the vast array of computer services available throughout all of the organizations comprising NLM. The Division also serves as a technological resource for other parts of the NLM and for other Federal organizations with biomedical, statistical, and administrative computing needs.

Executive Summary

Enhanced MedlinePlus: OCCS continued an aggressive campaign of major MedlinePlus releases this year including Release 16 of the Go Local Input System and Release 15 of a public directory of health services. Major upgrades and enhancements to MedlinePlus included:
- Database software upgrade to Oracle 9i.
- Modifications and testing of the MedlinePlus input system to run in NLM's disaster recovery/failover (NCCS) site in Sterling, Virginia.
- Modification and testing of MedlinePlus public pages to run in active-active (load-balancing) mode at the NCCS.
- Added Spanish-language news, a Spanish email listserver, and a Spanish language translation of ASHP drug information.

NIH Consolidated Collocations Site (NCCS): OCCS continued to lead the effort on the NIH Consolidated Collocation Site Project. The NCCS became operational in November 2003 in Sterling, Virginia. The facility provides disaster recovery and continuity of operations by reducing the risk of service interruptions due to a variety of unpredicted threats and disasters. Communications systems were planned and deployed to support the NCCS disaster recovery capabilities. These systems include access to Internet 1, a 622 Mbps link between NIH and the NCCS, load-balancing systems between NLM and the NCCS, and internal communications systems within the NCCS-NLM space.

High Speed Communication Network: OCCS improved the redundancy of equipment and network paths to eliminate single points of failure in the network. NLM's network perimeter connections to external networks provide an aggregate of 2 gigabits per second (Gbps), while the interconnection between NLM and the NIH/CIT campus backbone operates at 1 Gbps. Also in FY2004, OCCS implemented the remote-access Citrix terminal (and cable modems) as an effective solution for NLM flexi-place workers. OCCS is also expanding secure wireless access to the Internet and internal applications.

A-76 Competitive Sourcing Review: The OCCS computer center was one of three NIH centers subject to an A-76 competitive sourcing review in FY 2004. After a streamline review, the NIH Most Efficient Organization (MEO) won the competition with a cost savings of over 4 million dollars. The MEO is expected to increase the effectiveness and efficiency of computer center operations.

Multi-faceted IT Security Program: OCCS continued its multi-faceted and multi-layered IT security program that successfully prevented over 2.7 million virus attacks this year and detected more than 26,000 probes, scans, denial of service (DOS) attacks and other security events on a monthly basis. OCCS also performed a monthly cycle of vulnerability scanning, detection, and remediation; implemented automatic virus scanning and signature update mechanisms; implemented a perimeter firewall cluster; and implemented an automatic patch management system.

Enhanced DOCLINE: Expanded the functionality and improved the usability of DOCLINE, the NLM interlibrary loan system, to support 3,200 domestic and international libraries in processing approximately 3 million interlibrary loan transactions a year. Version 2.1 was released in April and Version 2.2 was released in August and included a total of 25 enhancements in response to user and Library Operations requests.

RxNorm Project: Designed and developed a prototype to prove the concept of RxNorm nomenclature management. This will standardize the labeling data mandated for clinical drugs by the FDA.

By the end of FY2004, development and planning were in an advanced stage.

Enhanced Medical Subject Headings (MeSH): Final development of the MeSH Translation Management System (MTMS) was completed in the first quarter of FY2004. Various foreign-language data sets, including Japanese, Spanish, Portuguese, and Dutch, were loaded. MTMS is an interlingual database of translations that permits automatic updating of the MeSH terminology tree in all languages.

NLM Main Page Redesign: OCCS played a significant role in the redesign of NLM's Main Web site. The redesign of the site improved its look and feel. In addition, using an audience of the general public, health care providers/professionals and librarians, usability testing was conducted to improve site navigation.

NIH E-mail Consolidation: OCCS participated in the planning and transition of NLM e-mail accounts to the NIH Central Email System (CES), an NIH IT consolidation initiative. OCCS maintained responsibility for local e-mail clients, assuming the task of configuring, deploying, and managing the MS-Outlook 2002 client. This consolidation resulted in the retirement of the Novell GroupWise e-mail system that had been in use for over seven years at NLM.

Active Directory Consolidation Project: OCCS contributed to the smooth transition of user accounts from the NLM Active Directory to the NIH Active Directory, a project that affected virtually all NLM users, and required changes to login credentials and local machine settings. Through planning in June, testing in July, and cutover in August, this project was accomplished on schedule.

Enhanced Relais: Relais was modified to accommodate innovations in DOCLINE, including color copy service and the ability to enter alternate Ariel delivery addresses. Ariel is a scanner-based document transmission system from Infotrieve that uses Internet protocols and Adobe's Portable Document Format (PDF) instead of telephone fax services for faster delivery.

Enhanced Voyager: OCCS completed initial development and testing of the new XML distribution of Voyager bibliographic data, which allows data sharing between Library Operations and NCBI's PubMed Entrez search and retrieval system.

Enhanced Data Creation and Maintenance Systems (DCMS): Reengineered DCMS to improve functionality and maintainability. This included a new Java version of the XML Loader and Extractor that will support Meeting Abstracts and OLDMEDLINE data as well as the redesign of Gene Indexing to work with NCBI's Gene Entrez database rather than LocusLink.

MEEC License Savings: OCCS renewed NLM's participation in the Maryland Education Enterprise Consortium (MEEC) licensing agreement that provides a bundle of Microsoft products at the lowest cost available in the U.S. MEEC seat renewal, priced this year at $16, provide licenses and product updates for the current Windows operating system, Microsoft Office Professional, Visual Studio.Net, and BackOffice clients. By contrast, GSA prices for these same products total $1,712.

Computer Facility Reengineering: The NLM computer room has tripled its use of electrical power, cooling and data transmission capacity over the last three years due to the rapid growth in IT systems. Recognizing this growth will continue in the years ahead, OCCS began a detailed process for evaluating the safety, reliability and performance requirements of the computer room. Reengineering activities include:

- Expanded the Uninterrupted Power Supply (UPS) capacity to support the growing needs for electrical power protection and redundancy of systems housed in the NLM Computer Room.
- Developed plans to bring additional power to the computer room, and to streamline the delivery of electrical power to the IT systems.
- Initiated plans for a pre-action sprinkler system to improve the reliability and safety of the fire suppressing system.
- Developed plans for an overhead ladder rack in the computer room, as a separate pathway for running data networking cables to improve the reliability, availability, and maintainability of data communication services.

The following describes in more detail OCCS accomplishments in FY2004:

Customer Services

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 10,700 NLM ticket requests for IT support were entered and tracked. NLM IT staff resolved over 72% of the calls (7,700 tickets)

with 28% of support calls being completed by NIH staff.

OCCS conducted over 80 desktop training courses this year, in topics such as "SPAMology," "Outlook 2002 FUNdamentals," and "Office XP Differences." Additionally, public briefings were conducted in support of the Active Directory migration project and many one-on-one sessions were held in relation to Outlook PST file reduction.

Desktop Support

OCCS worked with the NIH Center for Information Technology (CIT) this year to transition to the CIT-managed Microsoft Exchange 5.5 services known as the NIH Central E-Mail System (CES). OCCS participated in the planning and transition of NLM e-mail accounts to the NIH Central Email System (CES), an NIH IT Consolidation initiative. In addition, OCCS maintained responsibility for local e-mail clients, assuming the task of configuring, deploying, and managing the MS-Outlook 2002 client, and developing a deployment model that kept disruption at NLM's desktops to a minimum. In order to adopt the latest and most secure Microsoft e-mail client, Outlook 2002, OCCS developed and deployed an upgrade package for the MS-Office XP suite. This consolidation resulted in the retirement of the Novell GroupWise e-mail system that had been in use at NLM for over seven years.

OCCS staff contributed to the smooth transition of user accounts from the NLM Active Directory to the NIH Active Directory, a project that impacted virtually all NLM users and required changes to login credentials and local machine settings. Through planning in June, testing in July, and cutover in August, this project was accomplished on schedule. OCCS support contractors coordinated the project, created the detailed mapping of each NLM account to an NIH account, and developed the techniques for effectively delivering the revised credentials to each user's PC. The re-mapping effort enabled the preservation of users' network permissions and personality settings.

The Software Update Server security hotfix deployment solution introduced by OCCS makes possible the expeditious deployment of critical security updates, keeping 1,200 NLM systems better insulated from attack. The system enforces the application of previously released patches, ensuring continuous oversight and active management of security on the desktop. Vulnerability assessments now trend much more favorably. Two hundred nine (209) security patches are now consistently applied as needed to OCCS-supported desktops running the Windows operating system.

Network Support

OCCS continued to fulfill its mission of providing 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, services and communications. OCCS/STB/NES took steps to increase the capabilities and reliability of network services and storage, by providing for the following:

- NCCS data communications services
- Enhanced network and service monitoring and management
- Increased IT security
- New networked services to support the NLM user community
- Increased performance and throughput for networks
- Additional redundancy to eliminate single points of failure
- Enhanced backup for use in disaster recovery scenarios
- Expanded, centralized and efficient storage

Public Internet connectivity services continued to be provided through a contract with Level3/Genuity. Internet connectivity was provided via an OC3 (155Mbps) circuit to the Level3/Genuity network node in McLean, VA. The contract also provides an OC3 link for CIT/NIH to the Level3/Genuity network. NLM and NIH collaborate in using these links 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 addition to supporting the indexing system, the remote access Citrix terminal server solution has been implemented as an effective solution for NLM flexi-place workers. The terminal server system provides authentication into the NLM network, access to office and NLM business applications, network-based files, and the Internet. Network support continues to provide 56K dial-in access and cable modem access for a wide range of NLM staff and contractors. High-speed access is provided mainly through cable modems provided by COMCAST.

NLM consolidated wireless LAN networks into the support services of CIT. The initial wireless capabilities were implemented, and further expansion of the wireless systems will continue in selected areas. Wireless access to the Internet and public services of NLM and NIH is provided for guests and typical users. Through a Virtual Private Network,

authorized users can access internal applications in a secure manner.

Steps were taken to consolidate dial-up remote access services to the NIH Parachute system, whenever possible.

OCCS continued the use of iTRACS for documenting the LAN cabling and infrastructure. The data entry process (Phase I) was continued. Phase II, which includes layering the iTRACS information on AutoCAD drawings of NLM building plans, has also begun, and is expected to be completed during FY2005.

Systems Support

In order to protect NLM's mission-critical systems, CIT and NLM have implemented an NIH Consolidated Collocation Site (NCCS) in Sterling, Virginia. Since November of 2003, NCCS has operated to reduce the risks of service interruptions due to a variety of unpredicted threats and disasters. At present, all systems under MEDLARS and TOXNET are either under Active/active, active/passive or active/cold-backup mode depending on their business requirements. In addition, NLM has established plans for tape backups. The Disaster Recovery and Business Continuity Plan for MEDLARS and TOXNET 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.

OCCS deployed various applications at the NCCS this year, including the NLM Home Page, NIH Senior Health, the MeSH Browser, the Intranet, PHP Partners, and MedlinePlus Directories. Numerous servers and storage systems were deployed at the site to support these applications. In the coming months, OCCS staff will work with the Library Operations Division to deploy additional applications. The success achieved thus far shows that the rewards are well worth the efforts.

OCCS continued to make improvements to the UNIX architecture. Various upgrades in additional servers, increased memory, and subnet communication capacity were performed.

The ILS Oracle server was moved from direct attached storage to a high-speed network attached storage system thereby increasing the capacity of the production ILS Oracle server. The Web servers for this application were also moved into a high-speed gigabit environment this year.

The production Oracle database server for DOCLINE and DCMS was replaced this year with a new system having much faster CPUs and more memory capacity. This server will accommodate additional growth in these applications as well as host added applications.

Several new servers and large storage systems were procured and deployed for the RxNorm Support project. Deployment and testing is currently underway.

Several new servers were deployed for the Siebel development system. This provides for increased capacity and better software life cycle management.

IT Security

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

The perimeter firewall cluster was implemented to enhance NLM perimeter defense capability. Testing the alternate firewalls continued in preparation for their eventual deployment at the perimeter. Basic multicast testing was completed. OCCS servers that host NLM public applications such as MedlinePlus, DCMS, DOCLINE, and NIH SeniorHealth were migrated behind the firewall on the OCCS public firewall boundary in order to improve access control. Strong consideration was given to implementing a defense-in-depth (DiD) "best practices" architecture that provides varied forms of defenses at the different layers of the NLM network. OCCS will continue to emphasize the DiD architecture concept in the ensuing years in order to proactively maintain a strong security posture at NLM.

OCCS performs a monthly cycle of vulnerability scanning, detection, and remediation to improve NLM security posture. The Internet Security System's Internet Scanner provides network vulnerability assessment across servers, desktops, and infrastructure devices. Internet Scanner performs distributed probes of network services, operating systems, routers/switches, servers, firewalls, and application routers to identify potential risks.

OCCS implemented automatic virus scanning and signature update mechanisms to combat ever increasing cyber-attacks. OCCS utilizes antivirus software at the client level with McAfee Virus Scan where signature updates and scans and other various settings are set individually on the client.

Since the majority of all security breaches are caused by a missing patch, OCCS implemented an automatic patch management system to eliminate security breaches. Patch management of the Windows operating system is handled by a Windows SUS server. Settings for automatic update are controlled via group policy for all members of the NLM domain. Application patches are delivered by using industry standard delivery methods, such as

scripting and pushing through Active Directory and Novell Zenworks.

OCCS continued Web URL filtering this year in accordance with NIH Policy 2806. All NIH Institutes have been mandated to filter out access to inappropriate Web sites while simultaneously not affecting NIH business activity.

OCCS responded to IT security incidents that were observed on the intrusion detection system (IDS) console. These incidents were pursued by contacting the appropriate systems administrators and requesting them to take the necessary corrective action. The IDS rules were fine-tuned to reduce false positives. In addition, OCCS continued to run regularly scheduled monthly vulnerability assessments for OCCS, SIS, LHC, and the NCCS.

This year, OCCS successfully completed Inspector General reviews for both MEDLARS and TOXNET. Originally, MEDLARS included five major systems: the Voyager Integrated Library System (ILS), the Data Creation and Maintenance System (DCMS), Medical Subject Headings (MeSH), the Serials Extract File, and DOCLINE. In 2004, the MEDLARS umbrella grew to cover MedlinePlus and NIH Senior Health from OCCS; PubMed and Basic Local Alignment Search Tool (BLAST) from NCBI; and Clinical Trials from LHC.

The Office of Management and Budget requires that 100% of HHS computer users complete annual IT security awareness training. NLM has completed 100% of the mandatory FY04 Security Awareness Training for employee and contractors.

Policies and Product Standards

OCCS promoted the review and consideration by the Personal Computer Advisory (PCA) committee of the document OCCS (NLM) IT Support Policy for Remote Access from Non-NLM-Owned Computers. The committee reviewed the document, and has formed a subcommittee to discuss other issues relating to IT support.

OCCS, participating with the PCA, developed technical standards and product selections for two classes of notebook systems to join the PC Desktop selection in PCA consolidated orders. One system can reliably be connected to AC power sources and high performance is paramount. For the other system, weight, size, and battery life are important, but moderate performance and functionality is also needed These laptop systems will be available as offerings on the recurring consolidated PC purchases conducted by OCCS.

Quality Management (QM) and Configuration Control (CC)

OCCS convened a Configuration Control Board (CCB) to provide oversight of configuration changes made to production IT systems managed by the Systems Technology Branch. Implementation of the CCB concept across all of OCCS is anticipated in the next fiscal year. Quality management is a top priority of OCCS and quality management improvements are expected to lead to significantly greater maturity and repeatability in the day-to-day operations of the Division.

Computer Room Facilities

NLM systems continue to be supported in a safe environment in NLM's computer facility, which is available 24x7x365. The Network Operations and Security Center (NOSC), which was established in 2002, continues to serve as a central point in IT system and service monitoring, IT system administration, IT security event monitoring, and after-hours Help Desk support.

The NOSC display system consists of four 32-inch 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/MedlinePlus, and NLM services as seen by remote users around the world. Near real-time utilization data for NLM's Internet-1 and Internet-2 data communications links are also displayed.

The NLM computer room has tripled its use of electrical power, cooling and data transmission capacity over the last three 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 room. Those reengineering efforts include the following:

- Expanded the Uninterrupted Power Supply capacity to support the growing needs for electrical power protection and redundancy of systems housed in the NLM Computer Room. The Computer Room currently can maintain electrical power for up to 39 minutes after losing commercial power.

- Developed plans for an overhead ladder rack in the computer room, as a separate pathway for running data networking cables to improve the reliability, availability, and maintainability of data communication services.

- Initiated plans for the implementation of a pre-action sprinkler system to improve the reliability and safety of the fire suppressing system. The current system is a traditional wet pipe system. The proposed pre-action sprinkler system would require two actions before water will be released onto the fire: First, the smoke detection system must identify a developing fire and then open the pre-action valve. Second, the sprinkler head must release to permit water onto the fire. Modern day computer rooms are using this approach with upwards of 80% of all computer rooms already in this category.

- OCCS staff worked closely with NIH to develop plans to bring additional power to the computer room, and to efficiently streamline the delivery of electrical power to the IT systems. This will be a multi-year effort.

Consumer Health

MedlinePlus: In 2004, OCCS continued an aggressive campaign of major MedlinePlus releases including the Go Local Input System (Release 16) and a public directory of health services (Release 15). The Go Local initiative debuted in December 2002, allowing users in North Carolina to search for local medical service providers while viewing descriptive material in MedlinePlus. Additional functionality and a site for Missouri were subsequently implemented. In FY 2004, the Go Local Input System, MedlinePlus Release 16, was released. This system can be used by organizations to input site records for links to local services for their areas. Users can associate site records with local service terms that are mapped to MedlinePlus health topics. NLM makes this service available remotely via a Web interface. Ultimately, localities lacking the resources to maintain local sites will be able to create them on the NLM MedlinePlus site.

The public directory of health care resources (Release 15) allows the public to search for and geographically locate hospitals. Setbacks in acquiring a mapping service from MapQuest and hospital data from the American Hospital Directory caused Release 15 to be delayed for seven months, but the first of several planned medical resource directory releases occurred in July.

Upgrades, fixes, and enhancements to MedlinePlus included a database software upgrade to Oracle 9i; modifications and testing of the MedlinePlus input system to run in NLM's disaster recovery/failover site; modification and testing of MedlinePlus public pages to run in active-active (load-balancing) mode at the remote site; and a number of production changes requested by Library staff.

The MedlinePlus team also added Spanish-language news, a Spanish email listserver, and a Spanish language translation of ASHP drug information.

Senior Health Project: NIHSeniorHealth.gov is a joint NLM and National Institute on Aging project that provides health information on the Web using modes of delivery—video and narration—appropriate for older Americans with access limitations (low vision and low hearing, etc.). The system uses the Accent "Talking Web" module developed by OCCS to provide the accessibility enhancements. The TeamSite workflow application was integrated with Accent. Content originators can now preview new material by listening to it. When the originator is satisfied with the new or revised material, he or she can release it with a mouse-click. TeamSite automatically routes the new material for review and further revision. Finally, the new pages are permanently Access-encoded and moved into production on the Senior Health site.

Virtual Customer Service (Native Minds/Cosmo): NLM adopted Frequently Asked Questions software from Native Minds to provide first-level automated customer assistance. Dubbed Cosmo, the system uses artificial intelligence to answer customer questions in a conversational mode. Cosmo can answer hundreds of common questions, freeing reference librarians and other staff for more complex and demanding queries. The look and feel of Native Minds was redesigned to conform to the enhanced NLM Main Web site.

Professional Health Information

NLM Classification System: OCCS completed development of the NLM Classification System. The system allows public and institutional access to the NLM Classification and related services and includes a Classification Editor. he NLM Classification is updated annually in tandem with MeSH.

MeSH Browser: A DCMS connection to the MeSH Browser was made that allows DCMS users to enter MeSH terms directly into DCMS.

DOCLINE: DOCLINE, the NLM interlibrary loan system, supports 3,260 domestic and international libraries in processing approximately 3 million

interlibrary loan transactions a year. In FY2004, the DOCLINE team worked with WestLake Services to redesign the Loansome Doc component of DOCLINE. WestLake produced HTML in August, and OCCS developer coding and database redesign were well under way by the end of the fourth quarter. At the end of FY2004, DOCLINE version 2.3 (including an implementation of ISO's ILL protocol) was in beta testing. Twenty-five enhancements were implemented during the fiscal year

Relais: Relais was modified to accommodate innovations in DOCLINE, including color copy service and the ability to enter alternate Ariel delivery addresses. Ariel is a scanner-based document transmission system from Infotrieve that uses Internet protocols and Adobe's Portable Document Format instead of telephone fax services for faster delivery. A new Relais version (4.1) was implemented during the year, enabling delivery of Ariel requests to servers behind firewalls and email security for PDF documents. OCCS provided new server hardware for Relais email functions and provided new scanner hardware to facilitate the Ariel transmission process. An Access application for Relais Express was created which allows staff to query any request they process.

UMLS Licensing System: NLM's Unified Medical Language System provides UMLS Knowledge Sources (databases) and associated software tools (programs) for the development of computer systems that behave as if they "understand" the languages of biomedicine and health. In 2004, OCCS implemented an online system to license UMLS components.

Voyager Integrated Library System (ILS): OCCS completed initial development and testing of the new XML distribution of Voyager bibliographic data, which allows data sharing between Library Operations and NCBI's PubMed Entrez search and retrieval system. The team also loaded the Voyager LocatorPlus database into NCBI's Entrez NlmCatalog system. A Unicode test version of the next release of Voyager (2003.1) entered testing at the end of the quarter.

Literature Selection Technical Review Committee (LSTRC): The OCCS support team tuned and modified the application to improve performance and enhance functionality. LSTRC was successfully converted to ColdFusion MX.

OLDMEDLINE: In FY2004, the December 2003 OLDMEDLINE data set was made available to NLM licensees for the first time. During the fourth quarter of 2004, after final data modifications were made, over 1.7 million OLDMEDLINE citations were

prepared for incorporation into the DCMS. In mid September, 55,850 citations were sent to NCBI for inclusion into PubMed.

Medical Subject Headings: The Mesh Translation Management System (MTMS) is an interlingual database of translations that permits automatic updating of the MeSH terminology tree in all languages. Final development of the MTMS was completed in the first quarter of FY2004. Various foreign-language data sets, including Japanese, Spanish, Portuguese, and Dutch, were loaded. The Global Change Maintenance System (GCMS) allows propagation on demand. The MeSH component of GCMS (called MHGCMS), which entered production in FY 2003, is in use for propagating MeSH term changes. In FY 2004, development was completed on a Keyword maintenance system (called KWGCMS) which provides a similar capability for citations in SPACELINE and other specialty areas managed by NLM's collaborating partners. Numerous MeSH improvements, including changes related to YEP, were implemented in FY 2004.

Data Creation and Maintenance System (DCMS): The major year-end event for DCMS is the baseline extract, a re-release of all DCMS citations with the new MeSH headings. The baseline extract was separated for the first time into three groups based on publication year. Also in FY 2004, the DCMS team accomplished the following:

- Completed the DCMS-to-MeSH Browser connection. This allows the indexers to search and save terms from the MeSH Browser back to DCMS.
- Completed most testing of the new Java version of the XML Loader and Extractor for DCMS. This new version will support Meeting Abstracts and OLDMEDLINE data as well as the 2005 suite of DTDs. Changes are included to support invalid authors and new publishing models.
- Completed a new process to import suggested terms from the Lister Hill Medical Text Indexer (MTI) into the DCMS database. This allows the current DCMS function to be modified by an HTTP database-lookup call to Lister Hill Center's server.
- Completed the redesign of Gene Indexing to work with NCBI's Gene Entrez database rather than LocusLink.
- Completed twelve monthly issues of Index Medicus.

Serials Extract File (SEF): The SEF team made the Serials Viewer compatible with the modified XML DTDs. The team worked with NCBI to perform its

journal database update using the modified DTDs employed by the Serials Viewer. The team also continued to process LSTRC History data into the SSM table and created a List of Serials Indexed for Online Users and a List of Journals Indexed in Index Medicus for the year 2004. Programs were created to correct errors in DCMS MeSH.

Research and Development Efforts

Advanced Search Engine: During 2004, the RecomMind MindServer™ Retrieval System, an advanced search engine, became a standard component of NLM's Web services. The RecomMind search engine analyzes search terms entered by the user to infer meaning from the context. For example, a user researching work-related lung diseases might type in the terms "occupational lung diseases." The string contains no exact matches in the database, but the search engine recognizes an associated concept and locates articles with information on occupational asthma, occupational cancer, etc. Such an expert or intelligent search greatly expands the information available to a researcher, either in the professional or the public health arena, and assists in the critical task of filtering through the ever-increasing amount of information available.

Load Testing: The SilkPerformer server load-testing product from Segue was extensively evaluated, tuned, and moved into production as a standard part of NLM's application development and quality assurance toolkit. Load testing refers to the practice of modeling the performance of a software program through simulation of access by multiple concurrent users. SilkPerformer allows application developers and testers to predict the breaking points in applications and application infrastructures accurately before deployment.

Accent (Accessibility Enhancement) Project: Accent enables a Web server to provide content to vision-impaired users through text magnification and a machine-generated spoken version of Web-site text. The enhancements are produced at the server so the user requires no additional hardware/software. Accent was integrated with NIHSeniorHealth.gov during the first quarter of FY 2004. The application was integrated with TeamSite workflow management so authors and editors can preview text changes before they are final. Research continues into expanding the range of languages available via Accent, particularly Spanish. Also under study is further enhancement of enunciation and grammatical detail. Site navigation by speech has been prototyped and is being refined as FY2005 begins.

Digital Archive: A project is ongoing to ensure that "permanent" Web-based material remains accessible without adversely impacting searches for more current material. Development has been completed and implementation is expected to conclude early in FY2005. In FY2004, the Advanced Search Engine acquired Phase 1 ability to search the Digital Archive.

RxNorm Project: OCCS designed and developed a prototype to prove the concept of RxNorm nomenclature management. This will standardize the labeling data mandated for clinical drugs by the FDA. By the end of FY2004, development and planning were in an advanced stage. Final setup and testing of the RxNorm Testing (QA) J2EE servers, QA application server and QA Oracle database server and final setup and testing of the RxNorm Production J2EE servers, production application server, and production Oracle database server were well under way. Requirements and click-through mockups of the new RxNorm Editing System were also in an advanced stage of development. (The RxNorm Editing System allows NLM RxNorm editors to create Semantic Normal Forms for clinical drugs.)

NLM Web Support

Web Content Management: The TeamSite workflow feature was applied to previewing Accent "Talking Web" content. Numerous enhancements were made to TeamSite and additional TeamSite features were evaluated. Testing of TeamSite 6.1 began and continues into FY2005.

Web Statistics: The Web Support team installed the WebTrends Web site data analysis software during the first quarter and followed through with training, fixes, and enhancements as needed throughout the year. Library staff can now analyze Web statistics for leading NLM sites such as MedlinePlus, the NLM Main Web, Senior Health, NativeMinds, the NLM Intranet, and a variety of Web applications. An interface to export geographically oriented Web statistics to the MedMap Project was also developed.

NLM Main Web Redesign: The OCCS Web Support team worked with Westlake Solutions to carry out a redesign of the NLM Main Web. The Web Team significantly reworked Westlake's code to improve accessibility and enhance site maintenance. The project was completed in approximately 60% of the scheduled time.

NLM Link Checker: A custom NLM Link Checker was launched to replace the ailing Mom Spider. During the course of the year, the team modified the

Link Checker code to optimize its running time and fix a number of outstanding problems.

Technical Bulletin: A template capable of generating a printable bulletin was provided and multiple changes to the template were implemented in response to ongoing requests.

Miscellaneous: The OCCS Web Support team provided detailed and intensive technical research, support, and development related to all NLM Web and Intranet pages, sites, configurations, and functions.

Outreach

Consumer Outreach and Health System: OCCS developed this system to support NLM's Consumer Outreach and Health System (COHS). The system entered production in the spring of 2003. During FY2004, the team developed an interface to export Outreach Project data via HTTP post requests made by local Outreach systems. The data set is extracted on demand and converted to XML before being transferred to the requester. Additionally, the team upgraded the XML data service to provide project funding information, implemented rules updates for local partner organizations, and analyzed and fixed a local Output data discrepancy.

Web-based Exhibits: This system tracks OCCS exhibit activity and presence at national meetings. The database includes activities initiated both by internal NLM staff and the staffs of the Regional Medical Libraries. During FY2004, numerous enhancements were implemented, including a reformatted navigation menu with clusters of related functions, functionality for state and local exhibits, a method for identifying past national exhibits for which reports cannot be found, Add/Create User and Delete User functionality for administrators, and a number of database enhancements. New reports were also added.

HSRProj: In response to user requests, OCCS created a new interface for the HSRProj database and moved it into production at the end of May. The new interface included options that allow searchers to find Projects, Investigators, and Supporting Agencies

quickly. A total of 324 new project records were added to the database in early May.

Health Services Research Resources (HSRR): HSRR is used by the National Information Center on Health Services Research & Health Care Technology to post information on datasets, instruments, and software frequently used in health services research and in the behavioral and social sciences. In 2004, OCCS implemented advanced search functionality for the HSRR Web site. By request of the customer, work on HSRR was then suspended pending the completion of the HSRProj and Public Health Partners sites.

Public Health Partners: A result of collaboration between U.S. government agencies, public health organizations, and health sciences libraries, Public Health Partners provides the public health workforce with timely, convenient access to information resources. During FY2004, the Public Health Partners site was converted from static HTML to a dynamic ColdFusion application.

Administrative Support Systems

Customer Service Support System: The latest compiled version of Customer Service was implemented during the first quarter of FY2004. A new, productivity-enhancing Smartscript agent now provides first-tier staff with a form for quick capture of caller information. A hierarchical view of service requests allows each manager to see requests for all agents in his or her managed department. A Firewall Service Request Management System, rolled out in September of 2004, greatly enhances the efficiency and manageability of network security operations.

Cataloging Statistics Management System (CSMS): This system entered production in the first quarter of FY2004. It comprises ColdFusion, Oracle, HTML, and JavaScript-based functionality to create individual and section-level statistical reports for monthly and yearly production and similar activities. Employee management functions and other enhancements were added during the course of the year.

Small Purchase Management System: This system received report modifications and other maintenance and enhancements during FY2004.

Administration

Jon G. Retzlaff
Executive Officer

Table 12

Financial Resources and Allocations, FY 2004
(Dollars in Thousands)

Budget Allocation:
Extramural Programs $69,597
Intramural Programs 237,119
    Library Operations .... (89,895)
    Lister Hill National Center for
    Biomedical Communications .... (60,744)
    National Center for Biotechnology Information
    (72,324)
    Toxicology Information .... (14,156)
Research Management and Support 11,218
Total Appropriation 317,997
Plus: Reimbursements 11,101

Total Resources $329,089

Table 13

FY 2003 Full-Time Equivalents

Office of the Director 10
Office of Health Information
    Programs Development 7
Office of Communication and
    Public Liaison 8
Office of Administration 41
Office of Computer and
    Communications Systems 50
Extramural Programs 15
Lister Hill National Center
    for Biomedical Communications 81
National Center for Biotechnology
    Information 143
Specialized Information Services 35
Library Operations 285
TOTAL FTEs 675

Personnel

In August 2003, **Kin Wah Fung, M.D.,** joined the Lister Hill Center staff as a postdoctoral fellow. He received his medical degree from the University of Hong Kong and a master's degree in Medical Informatics from Columbia University. Dr. Fung has over 15 years of clinical practice in surgery. At the Hong Kong Hospital Authority, his informatics work was in user communication. At the Lister Hill Center, Dr. Fung is working on the UMLS project.

In November 2003, **Hua Florence Chang** was appointed Chief, Biomedical Files Implementation Branch, Division of Specialized Information Services. Ms. Chang earned a M.S. in Computer Science from Johns Hopkins, and a B.S. in Biology from the University of Maryland. She joined NLM in 2001 as a computer specialist and has played a key role in the design and implementation of several SIS products.

In December 2003, **Malay Kumar Basu, Ph.D.,** joined the staff of the NCBI Computational Biology Branch as a Visiting Fellow. Dr. Basu received his B.Sc. and M.Sc. in zoology from the University of Calcutta. In 2003, he received his Ph.D. from the Center for Cellular and Molecular Biology, Hyderabad, India. Dr. Basu will conduct research on phylogenetic classification of genes and proteins and develop tools to advance the database of Cluster of Orthologous Groups of protein and other systems.

In January 2004, **Liran Carmel, Ph.D.,** joined the staff of the Computational Biology Branch, NCBI as a Visiting Fellow. Dr. Carmel received his Master's degree in physics from the Israel Institute of Technology, Israel and his Ph.D. in mathematics and computer science from the Weizmann Institute of Science, Israel in 2003. Dr. Carmel will conduct research on genome evolution.

In January 2004, **Barend Johannes Mans, Ph.D.,** joined the staff of the Computational Biology Branch, NCBI as a Visiting Fellow. Dr. Mans received both his M.Sc. and his Ph.D. degrees in biochemistry from the University of Pretoria, South Africa. At NCBI, Dr. Mans will conduct research on the evolution of protein families on genome scale.

In February 2004, **Rampriya Ramarathnam, Ph.D.,** Dr. Ramarathnam joined the staff of the Computational Biology Branch, NCBI as a Visiting Fellow. She obtained her Ph.D. in bioengineering from the University of California, San Diego in 2003. Dr. Ramarathnam will conduct research on the classification of protein sequences and structures.

In April 2004, **Alice E. Jacobs** was appointed Acting Head of the Cataloging Section, Technical Services Division. Ms. Jacobs graduated Phi Beta Kappa from Washington University with a B.A. in French in 1972 and received her M.S. in library science from Simmons College in 1974. She came to the Cataloging Section of NLM in 1975, first working as an audiovisuals cataloger on the development of the AVLINE® database and later becoming Head of Unit

I in the section in 1986. Since 1991, Ms. Jacobs has served as Assistant Head of the Cataloging Section.

In May 2004, **Sunghwan Sohn, Ph.D.,** joined the staff of the NCBI Computational Biology Branch as a Visiting Fellow. Dr. Sohn received his Master's degree in computer engineering from the University of Missouri-Columbia and his Ph.D. in engineering management from the University of Missouri-Rolla. Missouri. Dr. Sohn will focus his research on the problem of mining data from the literature relevant to classes of genes such as arise from gene expression arrays.

In June 2004, **Patricia L. Gibbons** joined NLM as Chief, Office of Acquisitions Management. Ms. Gibbons comes to NLM from the National Institute of Mental Health (NIMH). She brings with her 17 years of contracting experience and unlimited Contracting Officer authority. Ms. Gibbons received her BA in political science from Pennsylvania State University and she is currently an MBA student at the University of Maryland.

In June 2004 **Alvin L. Harris** was appointed Acting Chief of the Office of Administrative and Management Analysis Services. Mr. Harris joined the NLM in 1971 and has served as the Deputy of the Office of Administrative and Management Analysis Services since 1988. Mr. Harris will allow NLM to have continuity of service and provide firm leadership until a new permanent Chief of the Office of Administrative and Management Analysis Services is selected.

In June 2004, **Janaki Ananth Mahadevan, Ph.D.,** joined the staff of the Computational Biology Branch, NCBI as an IRTA Fellow. Dr. Mahadevan obtained her M.Sc. degree in chemistry from the Indian Institute of Technology, Madras, India and her Ph.D. in computational chemistry, with honors, from the University of Kansas in 2001. Dr. Mahadevan will conduct research on new strategies based on genome-wide analysis of sequence and structural data.

In June 2004, **Balaji Santhanam, Ph.D.,** joined the staff of the Computational Biology Branch, NCBI as a Visiting Fellow. Dr. Santhanam received both his M.Sc. in physics and Ph.D. degree in biophysics from the Indian Institute of Science, Bangalore, India. Dr. Santhanam will conduct research on the large scale analysis of protein structures and sequences using computational methods.

In July 2004, **Joyce Backus** assumed the position of Head, Reference and Customer Services, Public Services Division. Ms. Backus received her A.B.

from Duke University and M.S. in library science from Catholic University of America. Ms. Backus has been with NLM since her 1985/86 NLM Associate Fellow year. She has served as a Reference Librarian and a Systems Librarian. Joyce has made significant contributions to many NLM products and programs including Grateful Med, Locatorplus, NIHSeniorHealth, the Intranet, and MedlinePlus.

In July 2004, **Saikat Chakrabarti, Ph.D.,** joined the staff of the Computational Biology Branch, NCBI as a Visiting Fellow. Dr. Chakrabarti received his Master's degree in biophysics, molecular biology and genetics as well as his Ph.D. in computational approaches to protein sciences from the National Centre for Biological Sciences, Bangalore, India. Dr. Chakrabarti will conduct research on development of automated structure-based multiple alignment techniques.

In July 2004, **Haixia Du, Ph.D.,** joined the staff of the Lister Hill Center as a Postdoctoral Fellow. Dr. Du received her doctorate degree from the Department of Computer Science at Stony Brook University. At NLM Dr. Du will work with the Office of High Performance Computing and Communications as part of the 3D Informatics research program under the Visible Human Project.

In July 2004, **Incheol Kim, Ph.D.,** joined the staff of the Lister Hill Center as a Postdoctoral Fellow. He received his doctorate degree in information processing engineering from Kyungpook National University, Taegu, Korea. At LHNCBC, Dr. Kim will conduct research in document metadata extraction using image processing and Web document analysis techniques.

In August 2004, **Raja Jothi, Ph.D.,** joined the staff of the Computational Biology Branch, NCBI as a Visiting Fellow. Dr. Jothi obtained both his M.Sc. and his Ph.D. in computer science from the University of Texas at Dallas, in 2004. Dr. Jothi will do research on models and algorithms for studying protein networks and co-evolution of interacting proteins.

In August 2004, **Jane Bortnick Griffith**, was appointed Acting Deputy Director, NLM. Ms. Bortnick Griffith joined NLM in 2000 as Assistant Director for Policy and Legislative Development. Ms. Bortnick Griffith holds a BA in American history from the University of Wisconsin and a MA in American history from Rutgers University. Prior to joining NLM, she worked as a senior specialist at the Library of Congress and served as director of a task force (under the aegis of National Academy of Sciences, National Academy of Engineering, and the

Institute of Medicine) that examined the goals, organization, and operational effectiveness of the National Research Council.

NLM Associate Fellowship Program

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 5-month curriculum phase, and conduct individual projects over the remaining 7-month period. Projects relate to key NLM programs areas and are typically of a research, development, or evaluation nature. Six new Associate Fellows began their year at NLM on September 1, 2004.

Margaret A. Basket received her MSI in May 2004 from the University of Michigan. She has library intern experience at the Minnesota State Archives and the 3M Company, as well as experience as head librarian for a university residence hall library. Prior to beginning her career in librarianship, she had 10 years' experience as a project and technical service engineer at the 3M Company. She also spent four years from 1998-2002 as a Peace Corps volunteer in Senegal. Her undergraduate training was in Mechanical Engineering.

Stephanie N. Dennis received her MLS in May 2004 from the University of Maryland. As a Graduate Assistant, she gained experience in the digital conversion of paper-based records. She also has experience developing Web sites, creating Web tutorials, and categorizing online health information for a search engine development project. Prior to beginning her career in librarianship, she worked on a variety of projects within the Grants Resource Center of the American Associate of State Colleges and Universities. Her undergraduate degree is in English Language and Literature.

Loren R. Frant received her MLIS in June 2004 from the University of California, Los Angeles. She has varied experience in libraries and museums, including cataloging visual history material, providing reference assistance, and conducting training sessions for library users. She also served as a volunteer librarian in South Africa during the summer of 2003. Prior to beginning her career in librarianship, she provided client support for companies delivering information management systems. Her undergraduate degree is in American Studies.

Rachel A. Gyore received her MLS in May 2004 from the University at Buffalo, the State University of New York. She has four years' experience as a library assistant in the Edward G. Miner Library at the University of Rochester, working in reference, circulation, archives, and Web management. She also has three years' experience as an assistant manager at Borders Books & Music. Her undergraduate degree is in History.

Lidia Y. Hutcherson received her MLIS in May 2004 from the University of Illinois at Urbana-Champaign. She has experience as a Graduate Assistant in the Library of the Health Sciences and in the University Library's Office of Planning and Budgeting. She also has four years' experience as a library technician, working in public services at Thomas Jefferson University and in technical services at Washington University in St. Louis. Her undergraduate degree is in History.

Sandy D. Tao received her MLIS degree in May 2004 from San Jose State University in California. She has experience in library automation, serving as a metadata support specialist at the Stanford University Library. Prior to beginning her career in librarianship, she had 5 years' experience in information systems development, including database, Web site, and Web applications development. She has laboratory experience as a research technician on a human genome research project. Her undergraduate training is in Biology.

Retirements and Separations

In November 2003, **Merlyn Rodrigues, M.D.**, departed NLM to join the National Center for Minority Health and Health Disparities, NIH. Dr. Rodrigues joined the Division of Extramural Programs, as NLM's Scientific Review Administrator in February 2001. At EP, Dr. Rodrigues expertly and efficiently arranged the timely review of all grant applications.

In December 2003, **Maria Korab-Laskowska, Ph. D.**, resigned her Staff Scientist position with the NCBI. Dr. Korab-Laskowska joined NCBI's Information Engineering Branch in December 1999. She was responsible for developing and maintaining the locusXref database.

In January 2004, **John Parascandola, Ph.D.,** retired from the Federal government and his most recent position as Public Health Service Historian, Audiovisual Program Development Branch, LHNCBC. Prior to accepting his current post, Dr. Parascandola served as Chief of NLM's History of Medicine Division from 1983 to 1992. Dr. Parascandola's contributions have been recognized by the PHS through such honors as the Surgeon

General's Exemplary Service Award (1989 and 1996), the Assistant Secretary for Health's Superior Service Award (1999), and the NIH Merit Award (1988). He is also the recipient of several awards in the history of science and medicine. His book on The Development of American Pharmacology: John J. Abel and the Shaping of a Discipline was awarded the George Urdang Medal for distinguished pharmaco-historical writing by the American Institute of the History of Pharmacy in 1994.

In February 2004, **Christa F.B. Hoffmann**, retiree from the Federal government and her position Head of the Cataloging Section, Technical Services Division, LO since October 1980. She came to NLM from the University of Nebraska–Lincoln Libraries where she was an Associate Professor of Library Science and head of the catalog department. During her career at NLM, Ms. Hoffmann led the Cataloging Section into a fully automated environment and played a key role in NLM's participation in national bibliographic programs. In September 2003, Ms. Hoffmann received the Frank B. Rogers Award.

In April 2004, **Duane W. Arenales** retired from her position as Chief, Technical Services Division, Library Operations, after 34 years of service with the Federal government, 32 of them at NLM. She came to the Library in 1971 after receiving an MLIS from the University of Maryland. As Chief, Technical Services Division, Ms. Arenales was responsible for NLM collection development policy; for the selection, acquisition and cataloging of material for the NLM's general collection; for overseeing the development and implementation of related processing systems; and for representing NLM in national bibliographic programs. She received the NIH Director's Award in 1998.

In April 2004, **Theodore E. Youwer** retired from his position as Chief, Office of Administrative Management and Analysis Services, Office of Administration. This was Mr. Youwer's second retirement from Federal service as he initially retired from the U. S. Air Force. In 1990 he joined the NLM staff as Chief, OAMAS. During his tenure at NLM, Mr. Youwer managed a host of major projects that significantly improved both the functional and aesthetic appearance of the library buildings and grounds and he directed major improvements that enhanced the quality of the workplace environment. Mr. Youwer received many accolades for his numerous contributions including the NIH Award of Merit and the prestigious NLM Director's Award.

In July 2004, **Kent Smith** retired from his position as NLM Deputy Director after 42 years of service with the Federal government. He received his B.A. degree

from Hobart College in mathematics and economics and his M.A. degree from the Johnson School of Management at Cornell University. During his tenure, Mr. Smith also served in various committees holding leadership positions. These include: President of the International Council of Scientific and Technical Information (ICSTI), Chair of the Policy Group of the Federal Library and Information Center Committee (FLICC), and Vice President of UNESCO General Information Program. He received numerous Senior Executive Service Achievement Awards, the Assistant Secretary for Health Exceptional Achievement Award, NLM Director's Award, the HHS Superior Service Medal, 1997 Medical Library Association President's Award, and the 1998 NFAIS Miles Conrad Lecture.

In July 2004, **Robert H. Cross** retired from his position as Education Specialist, Audiovisual Program Development Branch, LHNCBC after 40 years of service with the Federal government, 26 of which were with NLM. Between 1970 and 1980, he served as the Personnel Officer for various NIH Institutes including NLM and the U.S. Department of Agriculture. He returned to the NLM in 1980 as a Program Analyst for the Office of the Director, LHNCBC and in 1986 became a Staff Assistant in the Audiovisual Program Development Branch.

In September 2004, **Jon G. Retzlaff** resigned from his position as NLM Executive Officer. Mr Retzlaff came to NLM in 2002 from the National Institute of Neurological Disorders and Stroke. While at NLM he provided advice to the Director and other senior staff on administrative management matters and directed the administrative programs and services of the NLM. Mr. Retzlaff accepted a position as Director of Legislative Relations with the Federation of American Societies for Experimental Biology.

In Memoriam

In July 2004, **William Leonard**, the NLM Audiovisual Information Officer, passed away suddenly. Mr. Leonard worked for years in the field of broadcast journalism, most notably for NBC, where he won four Emmy Awards. Mr. Leonard came to the NLM in the mid 1970s, where he worked on programs designed to connect poorly served rural communities with the latest in medical information. For the last two decades, Mr. Leonard acted as Producer and Director on scores of audiovisual programs highlighting important project. Last year Mr. Leonard was the recipient of the NLM Director's Award, the Library's highest honor. He will be truly missed.

Awards

The 2004 Secretary's Award for Distinguished Service was awarded to **David J. Lipman, M.D.** for exceptional leadership in establishing NIH as the major resource in the filed of computational molecular biology.

The NIH Director's Award was awarded to **Martha R. Szczur** for developing consumer information resources to assist in identifying harmful chemical and environmental hazards.

The NLM Board of Regents Award for Scholarship or Technical Achievement was awarded to **Dr. Stuart J. Nelson** for initiating, designing, and directing the development of RxNorm, a clinical drug nomenclature designated as a U.S. Government-wide interoperability standard.

The Frank B. Rogers Award recognizes employees who have made significant contributions to the Library's fundamental operational programs and services. The recipient of the 2004 award was Ms. **Gail A. Dutcher** in recognition of significant contributions to many NLM programs including outreach to minority communities, consumer health and HIV/AIDS activities, and development of related health information resources.

The NLM Director's Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to three employees: **Yuen-Yin Kathy Kwan** (NCBI) for creating and managing the LinkOut project; **Julia C. Royall** (OHIPD) for unique contributions to strengthening NLM's international outreach to developing countries through the Multilateral Initiative on Malaria; and **Patricia Tuohy** (LO) for outstanding management of the design, development, and installation of NLM's major exhibitions and associated educational programs.

The NIH Merit Award was presented to four individuals and a group: **Dr. Valerie Florance** (EP) for her sustained and diligent excellence in administering, improving, and publicizing NLM's grant programs; Ms. **Judy C. Jordan** (LO) for highly successful management of NLM's ILL Serials Processing Group; **Dr. Craig Locatis** for his leadership and continuing support to NLM's ongoing partnership with the Radiological Society of North America; Ms. **Jane L. Rosov** for superior management of NLM's Licensing and Data Distribution Program which extends access to MEDLINE data; and the Extramural Program Special Meeting Team (**Ms. Christine C. Ireland, Ms. Michelle D. Krever, Ms. Susan Wilcox**) for

excellence in the organization, planning, and coordination of special meetings of critical importance to the Division of Extramural Programs.

The Philip C. Coleman Award recognizes significant contributions to the NLM by individuals who demonstrate outstanding ability to motivate colleagues. The recipient of the 2004 award was Ms. **Deirdre A. Clarkin** (LO) for the successful management and motivation of student employees in NLM's Collection Access Section Onsite Unit.

The NLM EEO Special Achievement Award was presented to **Dr. James E. Knoben** for his work with the Diversity Council and for spearheading the implementation of the "English Language Program," an initiative aimed at helping improve the language proficiency of employees whose first language is not English.

The Pehr Edman Award was presented by the International Association for Protein Structure Analysis and Proteomics to **Dr. Stephen F. Altschul** (NCBI) for outstanding contributions to protein and nucleic acid bioinformatics.

The 2004 Senior Scientist Accomplishment Award was presented by the International Society for Computational Biology to **Dr. David J. Lipman** for his contributions to the field of computational biology through research.

The 2004 Medical Library Association President's Award was presented to **Ms. Martha Fishel** and Ms. Betsy Humphreys in recognition of their leadership and contributions to the professional development programs of the Association.

The Thomson Scientific/Frank Bradway Rogers Information Advancement Award was presented by the Medical Library Association to the following individuals: **Ms. Joyce Backus, Ms. Paula Kitendaugh, Ms. Lori Klein, Ms. Eve Marie Lacroix, Ms. Wei Ma, Ms. Jennifer Marill**, and **Ms. Naomi Miller** in recognition of distinguished professional contributions to the application of technology in the delivery of health care information in the development of MedlinePlus.

NLM Committee Activities

NLM Diversity Council

The NLM Diversity Council began 2004 by welcoming four new members: Patricia Carson, Melanie Modlin, Helen Ochej, and Bryant Pegram. Each will serve a two-year term from January 2004 through December 2005. Continuing on the Council

are: Kathleen Cravedi, Felicia Derricott, James Knoben, Tameka Gore, Renee McLean-Banks, Donald Jenkins, and Linda Tang. The Council continues to receive support from its ex-officio members: Ronald Stewart, Acting Executive Officer, David Nash from the Equal Employment Opportunity Office, and Nadgy Roey from the Office of Human Resources, as well as its distinguished alumni. Kathleen Cravedi accepted the responsibilities of Council Chair and James Knoben became Council Vice-Chair.

FY2004 Accomplishments:

- *NLM Director's Employee Education Fund:* The NLM Diversity Council continued its coordination of the NLM Director's Employee Education Fund. In FY2004, the Fund enabled 77 NLM staff to take 85 classes from 18 area schools. This is up from 46 staff taking 65 classes from 13 area schools in FY2003. Undergraduate classes made up the majority of classes supported. The school with the largest number of NLM enrollees was the University of Maryland (21 attendees) with Montgomery College coming in second (17 attendees). Course disciplines enrolled in included psychology, business, marketing, computer networking, chemistry, economics, and biology. In addition to traditional classroom instruction, some courses were taken on the Internet. The Diversity Council continues its effort to publicize the availability of the fund. In fact, the Director's Employee Education Fund is featured under "Benefits" in a new NLM brochure entitled "Working at the NLM."

- *Facility Accessibility and Reasonable Accommodations:* The Council continued efforts to upgrade access at NLM for people with disabilities. Accessibility features in many of the bathrooms in NLM have now been added to accommodate the disabled community and Conference Room B has had LED Caption Display installed to provide a scrolling LED display of CART and real-time captioning to be seen by everyone in the room. The Diversity Council has approved and is working with the NLM's Office of Acquisitions Management to acquire an electric wheelchair for use by requesting patrons.

- *Communication of NLM Diversity*: The Diversity Council again collaborated with the Office of Communications and Public Liaison to promote various activities on the NLM Staff Bulletin Board located outside the cafeteria. This display has provided an excellent setting for celebrating the diversity found at the NLM. The Council voted to have OCPL staffer Fran Sandridge attend meetings on an ex-officio basis to assist in the design of needed bulleting displays.

- *English Language Courses:* The Council is continuing to support an English language program to enable NLM employees to improve their linguistic proficiency in speaking and writing English. Following the model used by local literacy programs, the NLM program offers one-on-one tutoring. NLM staff who volunteer to serve as tutors receive specialized training from the Literacy Council of Montgomery County. Four English language instructors and four students were selected in 2004, with one of those tutors currently on standby pending completion of tutor training. In 2005, the Diversity Council may consider whether to expand the program to include Spanish language instruction for those employees whose work involves that language.

- *NLM Health Education Expo:* In 2004, the Diversity Council sponsored its first annual Health Education Expo for Employees. The Expo, organized by Linda Eisenstadt and the Council, was titled *"Keep a Check on Your Health."* While women were the primary audience, men were encouraged to attend. The program was held at Lister Hill Auditorium with a one-hour presentation by Dr. Patricia Davidson, followed by questions on "Hypertension and Heart Disease in Women and Preventive Measures." The keynote presentation was followed by exhibits in the Lister Hill Lobby where numerous groups including the American Diabetes Association, the National Women's Health Information Center, the Washington Health Center, among others, provided valuable health information to NLM employees. In addition, the American Heart Association was on hand to answer employee questions. Blood pressure and cholesterol screening were provided by NIH. There were raffles for door prizes, including a membership to the NIH Fitness Center. The Expo was a great success and the Council decided to make it an annual event.

- *Diversity Council Honors NLMers with Awards and Ice Cream Social.* The Diversity Council sponsored a "Laborless" Moment to honor NLMers whose volunteer activities helped to promote diversity and improve employment opportunities at the NLM in 2003–04. Ben and Jerry's supplied

the ice cream on the patio adjacent to the Lister Hill Auditorium following a brief awards ceremony in the auditorium. About 400 NLMers attended. It was agreed that the Diversity Council awards ceremony and ice cream social should remain an annual event.

- *Reading Club:* The Diversity sponsors a reading club that meets regularly for interested employees.
- *Looking Under Your Hood:* In 2004, the Diversity Council sponsored a series of monthly lectures by Dr. Donald Jenkins, a member of the Council. The lectures, as the title suggests, provided a valuable and fascinating overview of regions of the body, as shown in the David Bassett archive of images of human cadaver anatomy. This lecture series was based on the belief that personal knowledge about the intricate structure of the human body is beneficial to health and well-being.
- *Diversity Council Coat and Clothing Drive:* The Diversity Council sponsored a coat and clothing drive during the 2004 Thanksgiving and Christmas holidays. Over three carloads of clothing and more than 400 coats were collected by the Diversity Council and delivered to the Shepherds Table in Silver Spring, Maryland. This Center provides food and clothing to approximately 150 needy people daily.

Board of Regents

The Board of Regents (BOR) met three times in FY 2004 on February 10–11, May 19–20, and September 21–22. The Extramural Programs Subcommittee and the Subcommittee on Outreach and Public Information were held during each of these meetings.

During the FY2004 meetings, the Board of Regents reviewed several new and ongoing projects:

- In February, the Board was given presentations on Entrez databases and features, NLM website user studies, Hawaii Access to Computerized Health Information grant program, a Native American internship outreach project, and outreach activities in Africa. The Regents approved a draft Board

statement on the "Library of the 21st Century," and approved the Board Operating Procedures for 2004.

- In May, the Board was given updates on the Information Rx Project, Just in Time Information for Clinicians Grant Project, a report from the Midcontinental Regional Medical Library, International Toxicity Estimates of Risk database, the Wireless Information System for Emergency Responders project, facilitating foreign language versions of MeSH, WebMARS data extraction from online journals, and an outreach project on high school students connecting with MedlinePlus. The Board gave concept approval for Roadmap Initiatives, a Multi-Agency Modeling Project, and a Specialized Information Services Public Health Law Information Project.
- In September, the Board reviewed several projects including the NCBI Bookshelf, the NCBI PubChem system (part of the Roadmap Initiative), a Hospital Elder Life grant project, the CINID-Model Disability Information Network grant project, consumer health information services, a revised version of the Collection Development Manual, and the Visible Human Project. The Board provided concept approval for NLM's support of informatics research, and approval of the revised Collection Development Manual.

During all Board meetings, the committee performed the secondary peer review process for the NLM grant program. Other grant-related activities are listed under the Extramural Programs section of this annual report.

A new Chair was elected to the Board of Regents, Dr. William W. Stead, Professor of Biomedical Informatics at Vanderbilt University in Nashville, Tennessee. Two new members joined the Board in September: The Honorable Newt Gingrich, Chief Executive Officer of The Gingrich Group, in Washington, D.C., and Mr. Richard Chabran, Chair of the California Community Technology Policy Group, in Chino Hills, California.

APPENDIX 1:   REGIONAL MEDICAL LIBRARIES

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

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

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

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

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

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

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

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

APPENDIX 2:   BOARD OF REGENTS

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

Appointed Members:

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

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

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

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

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

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

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

GINGRICH, Newt, Ph.D.
Chief Executive Officer
The Gingrich Group
Washington, DC

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

Ex Officio Members:

Librarian of Congress

Surgeon General
Public Health Service

Surgeon General
Department of the Air Force

Surgeon General
Department of the Navy

Surgeon General
Department of the Army

Under Secretary for Health
Department of Veterans Affairs

Assistant Director for Biological Sciences
National Science Foundation

Director
National Agricultural Library

Dean
Uniformed Services University of the Health Sciences

APPENDIX 3: BOARD OF SCIENTIFIC COUNSELORS/ LISTER HILL CENTER

The Board of Scientific Counselors meets periodically to review and make recommendations on the Library's intramural research and development programs.

Members:

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

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

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

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

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

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

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

APPENDIX 4: BOARD OF SCIENTIFIC COUNSELORS/ NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

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

Members:

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

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

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

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

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

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

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

APPENDIX 5:   BIOMEDICAL LIBRARY AND INFORMATICS REVIEW COMMITTEE

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

Members:

HRIPCSAK, George, M.D. (chair)
Associate Professor
Department of Medical Informatics
Columbia University
New York, NY

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

BALAS, Andrew, M.D., Ph.D.
Dean and Professor
College of Health Sciences
Old Dominion University
Norfolk, VA

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

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

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

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

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

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

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

McKNIGHT, Michelynn, Ph.D.
Assistant Professor
School of Library and Information Science
Louisiana State University
Baton Rouge, LA

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

PRATT, Wanda, Ph.D.
Assistant Professor
Department of Biomedical & Health Informatics
University of Washington School of Medicine
Seattle, WA

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

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

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

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

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

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

YOKOTE, Gail A.
Associate University Librarian
Peter J. Shield Library
University of California
Davis, CA

ZHOU, Z. Hong, Ph.D.
Associate Professor of Pathology
University of Texas Health Science Center – Medical
School
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 *Index Medicus* and MEDLINE.

Members:

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

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

CHEN, Jinkun, DDS, Ph.D.
Professor of General Dentistry
Director, Oral Biology Division
Tufts University School of Dental Medicine
Boston, MA

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

DOUGLAS, Janice E., M.D.
Professor of Medicine, Physiology & Physics
Case Western Reserve University
Cleveland, OH

FREY, John J., 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

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

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

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

SOEHNER, Catherine B., M.L.S.
Head, Science & Engineering Library
University of California
Santa Cruz, CA

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

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

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

APPENDIX 7:   PUBMED CENTRAL NATIONAL ADVISORY COMMITTEE

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

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

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

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

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

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

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

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

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

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

LAPPIN, Debra R., J.D.
Consultant
Princeton Partners Ltd.
Englewood, CO

ROEHR, Bob, B.A.
Writer
Washington, D.C.

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

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

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

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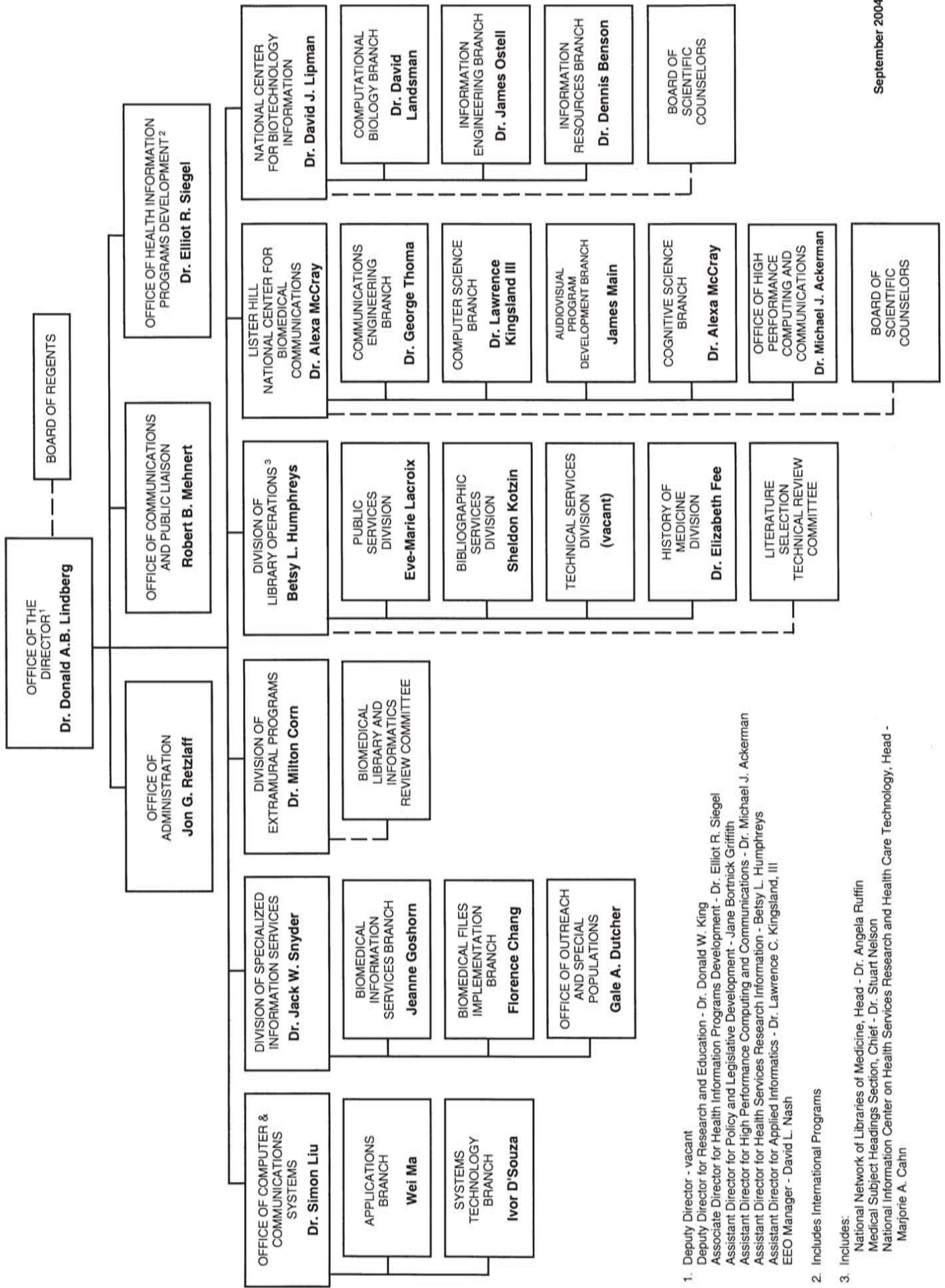University of Virginia
Charlottesville, VA

APPENDIX 8: ORGANIZATIONAL ACRONYMS AND INITIALISMS USED IN THIS REPORT

AAHSL	Association of Academic Health Sciences Libraries
ACP	American College of Physicians
ACSI	American Consumer Satisfaction Index
ACTIS	AIDS Clinical Trials Information Service
AHCPR	Agency for Health Care Policy and Research
AHRQ	Agency for Healthcare Research and Quality
ALTBIB	Alternatives to Animal Testing
AMPA	American Medical Publishers Association
AMWA	American Medical Women's Association
APDB	Audiovisual Program Development Branch
ARL	Association for Research Libraries
ATIS	HIV/AIDS Treatment Information Service
ATSDR	Agency for Toxic Substances and Disease Registry
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
CBIR	Content-Based Image Retrieval
CCB	Configuration Control Board
CCRIS	Chemical Carcinogenesis Research Information System
CDC	Centers for Disease Control and Prevention
CDD	Conserved Domain Database
CEB	Communications Engineering Branch
CES	(NIH) Central Email System
CGAP	Cancer Genome Anatomy Project
CgSB	Cognitive Science Branch (CgSB
ChemIDplus	Chemical Identification File
CIT	Center for Information Technology
CPT	Current Procedural Terminology
CRID	Regional Disaster Information Center for Latin America and the Caribbean
CSB	Computer Science Branch
DART	Developmental and Reproductive Toxicology
DDBJ	DNA Data Bank of Japan
DCMS	Data Creation and Maintenance Systems
DHHS	Department of Health and Human

	Services
DiD	Defense-in-Depth
DIRLINE	Directory of Information Resources Online
DTD	Document Type Definition
EBI	European Bioinformatics Institute
EEO	Equal Employment Opportunity
EFTS	Electronic Funds Transfer Service
EMBL	European Molecular Biology Laboratory
EMIC	Environmental Mutagen Information Center
EnHIOP	Environmental Health Information Outreach Panel
EP	Extramural Programs
EPA	Environmental Protection Agency
EST	Expressed Sequence Tag
ETICBACK	Environmental Teratology Information Center backfile
FDA	Food and Drug Administration
FIC	Fogarty International Center
FNLM	Friends of the National Library of Medicine
GEO	Gene Expression Omnibus
GPRA	Government Performance and Results Act
GSA	General Services Administration
GSS	Genome Survey Sequences
GUI	Graphical User Interface
HapMap	International Haplotype Map Project
HBCU	Historically Black Colleges and Universities
HHS	Health and Human Services
HIPAA	Health Insurance Portability and Accounting Act
HMD	History of Medicine Division
HSDB	Hazardous Substances Data Bank
HPCC	High Performance Computing and Communications
HSRProj	Health Services Research Projects
HSRR	Health Services and Sciences Research Resources
HSTAT	Health Services and Technology Assessment Text
IADL	Internet Access to Digital Libraries
IAIMS	Integrated Advanced Information Management Systems
ICs	Institutes and Centers (of NIH)
ICTV	International Committee on Taxonomy of Viruses
ILL	Interlibrary Loan
ILS	Integrated Library System

INSD	International Nucleotide Sequence Database Collaborators
IRIS	Integrated Risk Information System
IT	Information Technology
ITER	International Toxicity Estimates for Risk
ITK	Insight Toolkit
ITP	Informatics Training Program
JD	Journal Descriptor
LAN	Local Area Network
LHC	Lister Hill Center
LHNCBC	Lister Hill National Center for Biomedical Communications
LO	Library Operations
LOINC	Logical Observations: Identifiers, Names, Codes
LSTRC	Literature Selection Technical Review Committee
MEDLARS	Medical Literature Analysis and Retrieval System
MEEC	Maryland Education Enterprise Consortium
MeSH	Medical Subject Headings
MGC	Mammalian Gene Collection
MIM	Multilateral Initiative on Malaria
MIRS	Medical Information Retrieval System
MLA	Medical Library Association
MLAA	Medical Library Assistance Act
MMDB	Molecular Modeling DataBase
MMS	MEDLARS Management Section
MMTx	MetaMap Technology Transfer
MTI	Medical Text Indexer
MTMS	MeSH Translation Management System
NCBC	National Centers for Biomedical Computing
NCBI	National Center for Biotechnology Information
NCCS	NIH Consolidated Collocation Site
NCI	National Cancer Institute
NCRR	National Center for Research Resources
NCVHS	National Committee on Vital and Health Statistics
NHANES	National Heath and Nutrition Examination Surveys
NHGRI	National Human Genome Research Institute
NHII	National Health Information Infrastructure
NHLBI	National Heart, Lung, and Blood Institute
NIA	National Institute on Aging
NIAID	National Institute of Allergy and Infectious Diseases
NIBIB	National Institute of Biomedical Imaging and Bioengineering
NICHSR	National Information Center on Health Services Research and Health Care Technology

NIEHS	National Institute of Environmental Health Sciences
NIGMS	National Institute of General Medical Sciences
NIH	National Institutes of Health
NIOSH	National Institute for Occupational Safety and Health
NIST	National Institute of Standards and Technology
NLM	National Library of Medicine
NN/LM	National Network of Libraries of Medicine
NNO	National Network Office
NOSC	Network Operations and Security Center
NRCBL	National Reference Center for Bioethics Literature
NSF	National Science Foundation
NTCC	National Online Training Center and Clearinghouse
OAM	Office of Administrative Management
OCCS	Office of Computer and Communications Systems
OCPL	Office of Communications and Public Liaison
OCR	Optical Character Recognition
OD	Office of the Director
OHIPD	Office of Health Information Programs Development
OMB	Office of Management and Budget
OMIM	Online Mendelian Inheritance in Man (database)
OSIRIS	Open Source Independent Review and Interpretation System
PAHO	Pan American Health Organization
PCA	Personal Computer Advisory Committee
PDA	Personal Digital Assistant
PDB	Protein Data Bank
PDF	Portable Document Format
PHS	Public Health Service
PICO	Patient/Problem, Intervention, Comparison, and Outcome
PLA	Public Library Association
PMC	PubMedCentral
PRS	Protocol Registration System
PSD	Public Services Division
QTL	Quantitative Trait Loci
RefSeq	Reference Sequence (database)
RML	Regional Medical Library
RNAi	RNA interference
RTECS	Registry of Toxic Effects of Chemical Substances
SAGE	Serial Analysis of Gene Expression
SBIR	Small Business Innovation Research
SEF	Serials Extract File
SEP	Special Emphasis Panel
SIS	Specialized Information Services
SNOMED CT	Systematized Nomenclature of Medicine Clinical Terms

SPER	System for the Preservation of Electronic Resources
SSEUS	SIS SQL Entry Update System
SSI	Scalable Information Infrastructure
STB	Systems Technology Branch
STTR	Small Business Technology Transfer Research
STS	Sequence Tagged Site
TEHIP	Toxicology and Environmental Health Information Program
TERA	Toxicology Excellence for Risk Assessment
TILE	Text to Image Linking Engine
TIOP	Toxicology Information Outreach Project
TOXLINE	Toxicology Information Online
TOXNET	Toxicology Data Network
TPA	Third Party Annotation (database)
TRI	Toxics Release Inventory
TSD	Technical Services Division
TTP	Turning the Pages
UMLS	Unified Medical Language System
UPS	Uninterrupted Power Supply
VAST	Vector Alignment Search Tool
VHP	Visible Human Project
Web-STOC	Web-Services Technology Operations Center
WGS	Whole Genome Shotgun
WISER	Wireless Information System for Emergency Responders

National Library of Medicine

September 2004

BOARD OF REGENTS

OFFICE OF THE DIRECTOR[1]
Dr. Donald A.B. Lindberg

OFFICE OF ADMINISTRATION
Jon G. Retzlaff

OFFICE OF COMMUNICATIONS AND 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
  Wei 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 AND SPECIAL POPULATIONS
  Gale A. Dutcher

DIVISION OF EXTRAMURAL PROGRAMS
Dr. Milton Corn

- BIOMEDICAL LIBRARY AND INFORMATICS REVIEW COMMITTEE

DIVISION OF LIBRARY OPERATIONS[3]
Betsy L. Humphreys

- PUBLIC SERVICES DIVISION
  Eve-Marie Lacroix
- BIBLIOGRAPHIC SERVICES DIVISION
  Sheldon Kotzin
- TECHNICAL SERVICES DIVISION
  (vacant)
- HISTORY OF MEDICINE DIVISION
  Dr. Elizabeth Fee
- LITERATURE SELECTION TECHNICAL REVIEW COMMITTEE

LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS
Dr. Alexa McCray

- COMMUNICATIONS ENGINEERING BRANCH
  Dr. George Thoma
- COMPUTER SCIENCE BRANCH
  Dr. Lawrence Kingsland III
- AUDIOVISUAL PROGRAM DEVELOPMENT BRANCH
  James Main
- COGNITIVE SCIENCE BRANCH
  Dr. Alexa McCray
- OFFICE OF HIGH PERFORMANCE COMPUTING AND 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

1. Deputy Director - vacant
   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 - Jane Bortnick Griffith
   Assistant Director for High Performance Computing and Communications - Dr. Michael J. Ackerman
   Assistant Director for Health Services Research Information - Betsy L. Humphreys
   Assistant Director for Applied Informatics - Dr. Lawrence C. Kingsland, III
   EEO Manager - David L. Nash

2. Includes International Programs

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