

NATIONAL INSTITUTES OF HEALTH

National Library of Medicine

Programs and Services

Fiscal Year 2007

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

National Library of Medicine Catalog in Publication

Z
675.M4
U56an

National Library of Medicine (U.S.)
National Library of Medicine programs and services.—
1977- .—Bethesda, Md. : The Library, [1978-
v.: ill., ports.
Report covers fiscal year.
Continues: National Library of Medicine (U.S.). Programs and Services. Vols. For
1977-78 issued as DHEW publication; no. (NIH)
78-256, etc.; for 1979-80 as NIH publication; no. 80-256, etc.
Vols. 1981 – present Available from the National Technical Information Service,
Springfield, Va.
ISSN 0163-4569 = National Library of Medicine programs and services.

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

DISCRIMINATION PROHIBITED: Under provisions of applicable public laws enacted by Congress since 1964, no person in the United States shall, on the ground of race, color, national origin, sex, or handicap, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance. In addition, Executive Order 11141 prohibits discrimination on the basis of age by contractors and subcontractors in the performance of Federal contracts. Therefore, the National Library of Medicine must be operated in compliance with these laws and executive order.

CONTENTS

Preface.....	v
Office of Health Information Programs and Development	1
Planning and Analysis	1
Outreach and Consumer Health.....	1
International Programs	3
Library Operations	5
Program Planning and Management.....	5
Collection Development and Management.....	6
Vocabulary Development and Standards.....	8
Bibliographic Control	9
Information Products	10
Direct User Services	12
Outreach	12
Specialized Information Services	22
Toxicology and Environmental Health Resources	22
AIDS Information Services	24
Evaluation Activities	24
Outreach Initiatives	24
Research and Development Initiatives	25
Lister Hill Center.....	27
Next Generation Electronic Health Records to Facilitate Patient-centric Care, Clinical Research, and Public Health.....	27
Biomedical Imaging and Multimedia	28
Automated Concept Extraction from Documents.....	33
Information Resource Delivery for Care Providers and the Public	35
Communication Infrastructure Research and Tools	36
Language and Knowledge Processing	39
UMLS and Focused Clinical Vocabularies	40
Training Opportunities	41
National Center for Biotechnology Information.....	42
Molecular Biology Information Resources.....	42
Integration of Clinical, Genetic, and Environmental Databases	48
Entrez Retrieval System	48
Discovery Initiative	48
Research	50
Bioinformatics Training and Support	50
Extramural Programs.....	53
Success Rates of Grant Applicants	53
Research Support for Biomedical Informatics and Bioinformatics	54
Resource Grants.....	55
Training and Fellowships	55
Pan-NIH Projects.....	58
EP Operating Units.....	60
Office of Computer and Communications Systems	73

Executive Summary.....	73
Business Continuity and Disaster Recovery.....	75
Consumer Health.....	76
IT Security.....	76
Professional Health Information.....	77
Network and Systems Support.....	78
Desktop Support.....	79
Outreach.....	79
Research and Development Initiatives.....	80
NLM Web Support.....	80
Customer Services.....	81
Computer Facilities Operations.....	81
Administrative Support Systems.....	81
Administration.....	82
Personnel.....	82
NLM Diversity Council.....	90
NLM Organizational Chart.....(Inside back cover)	

Appendixes

Appendix 1. Regional Medical Libraries.....	93
Appendix 2. Board of Regents.....	94
Appendix 3. Board of Scientific Counselors/LHC.....	95
Appendix 4. Board of Scientific Counselors/NCBI.....	96
Appendix 5. Biomedical Library and Informatics Review Committee.....	97
Appendix 6. Literature Selection Technical Review Committee.....	100
Appendix 7. PubMed Central National Advisory Committee.....	101
Appendix 8. Organizational Acronyms and Initialisms Used In This Report.....	102

Tables

Table 1. Growth of Collections.....	17
Table 2. Acquisition Statistics.....	18
Table 3. Cataloging Statistics.....	18
Table 4. Bibliographic Services.....	18
Table 5. Consumer Web Services.....	19
Table 6. Circulation Statistics.....	19
Table 7. Online Searches—PubMed and NLM Gateway.....	20
Table 8. Reference and Customer Services.....	20
Table 9. Preservation Activities.....	20
Table 10. History of Medicine Activities.....	21
Table 11. Success Rate, Core NLM Grant Programs.....	53
Table 12. Applications and Awards, FY2003-2007.....	54
Table 13. T15 Trainees for FY2007.....	56
Table 14. Map of EP-Supported Training Programs.....	57
Table 15. Extramural Programs Budget FY2006–2008.....	64
Table 16. Extramural Programs Budget FY2003–FY2007.....	65
Table 17. EP Budget FY2007.....	66
Table 18. RFA/PA Actions in FY2006 and 2007.....	67
Table 19. NLM Grants Awarded in FY2007.....	68
Table 20. Financial Resources and Allocations FY2007.....	82
Table 21. FY2007 Full-Time Equivalents (Actual).....	90

Preface

Fiscal Year 2007 saw many important advances on several fronts. The National Library of Medicine is providing more and varied services to a wider audience than ever before.

This year, NLM launched two new important databases. First, NCBI significantly expanded its breadth of public resources by developing the database of Genotypes and Phenotypes, or dbGaP. This database will be NIH's public repository for linking up genotype data with phenotype information in order to identify the genetic factors that influence health, disease, and speed dramatically the development of diagnostic methods and treatment. Second, this past September, NLM introduced an expanded ClinicalTrials.gov database that can accept "basic results" information. The addition of the results information was responsive to recently enacted legislation and growing public interest in improving access to clinical trial information and increasing transparency in clinical research.

Among other highlights of this year's *NLM Programs and Services*:

- FY2007 included the 35th anniversary of MEDLINE, the Library's flagship database (October 27, 1971). In 1971, MEDLINE served 25 users. This year, nearly 80 million unique visitors viewed 4.4 billion pages in MEDLINE/PubMed.
- Use of MedlinePlus and MedlinePlus en español continues to increase significantly. Nearly 120 million unique visitors viewed over 907 million pages. Four new Go Local sites were released, bringing the total to 22 projects in 20 states and providing coverage for 34% of the U.S. population. Hispanic star Don Francisco helped spread the word about MedlinePlus en español with a series of TV and radio public service announcements.
- SIS teamed with HHS to launch the Radiation Event Medical Management (REMM) Web site. This downloadable resource provides medical information to health care providers during mass-casualty radiation events.
- PubMed Central, the Library's free digital archive of full-text journal articles, reached the one million-article mark. Now in its seventh year, PMC is expanded each week with new articles from over 350 important life sciences journals, and with digitized versions of publications that were previously available only in printed form.
- "Profiles in Science," a popular educational Web site of the History of Medicine Division and Lister Hill Center, was enriched this year with the papers of Harold Varmus, Rosalind Franklin, Mary Lasker and Sol Spiegelman.
- The Library even extended its reach to children with a special exhibit, "Do Mandrakes Really Scream? Magic and Medicine in Harry Potter," featuring NLM's historical holdings.

While making progress on these and other priorities, NLM continues to pursue and develop innovative programs to support clinical and public health systems, improve public understanding of health research findings, enhance resources and infrastructure, and promote discoveries through genomics in the 21st century. I extend my thanks to the Library staff and to the many advisors and consultants we rely on to achieve the accomplishments detailed in this report.

Donald A.B. Lindberg, /S/
Director

OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT

Elliot R. Siegel, Ph.D.
Associate Director

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

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

Planning and Analysis

The NLM Long Range Plan remains at the heart of NLM's planning and budget activities. Its goals form the basis for NLM operating budgets each year. *Charting a Course for the 21st Century: NLM's Long Range Plan 2006–2016* was published in print and on the NLM web site during FY2007. Print copies are available from the NLM Office of Communications and Public Liaison. The report includes the following chapters:

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

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

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

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

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

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

efforts to understand and eliminate health disparities between minority and majority populations. NLM's Health Disparities Plan is available on the NLM Web site.

Outreach and Consumer Health

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

One activity, the Physician Information Prescription Project ("Information Rx"), initiated with the American College of Physicians and reported in previous years, has been expanded to include the National Medical Association, the American Osteopathic Association, hospital librarian members of the Medical Library Association, and disease-focused organizations such as the Fisher Center for Alzheimer's Research. An article reporting an evaluation of the Information Prescription project was published in *Information Services & Use* 26(2006) 1–10. Another outreach activity, reported in past years, focuses on clinical research in which Information Rx is included as a health provider intervention in the routine management of patients with Type 2 Diabetes and HIV/AIDS. Data from two controlled field experiments with patients enrolled in a diabetes program in a Washington, DC clinic serving primarily Hispanic patients, and also a general medicine clinic in Columbia, MO. are now being evaluated.

OHIPD staff is also pursuing several outreach initiatives intended to encourage underrepresented minority high school students to pursue careers in medicine and the health sciences; these have been carried out in collaboration with other divisions of NLM and have taken the form of special symposia in New York City and Houston, TX commemorating the work of Dr. Michael DeBakey. Another collaboration with the Student National Medical Association supports that organization's efforts to encourage African-American medical students to pursue careers as physician researchers; NLM's support is focused on promoting consideration of research careers in biomedical and public health informatics. A pipeline for a strong and diverse workforce is a high priority goal.

Web Evaluation

The Internet and World Wide Web play a dominant role in dissemination of NLM information services. And the Web environment in which NLM operates is rapidly changing and intensely competitive. These two factors combined suggested the need for a more comprehensive and dynamic NLM Web planning and evaluation process. The Web evaluation priorities of the OCHD include both quantitative and qualitative metrics of Web usage and measures of customer perception and use of NLM Web sites. During FY2007, the OCHD continued to pursue an integrated approach intended to encourage exchange of information and learning within NLM, and help better inform NLM management decision-making on Web site research, development, and implementation. The year's evaluation activities included: access to a syndicated telephone survey of the US public's online and offline health information seeking behavior; analysis of NLM Web site log data; and access to Internet audience measurement estimates based on Web usage by user panels organized by private sector companies.

Also during FY2007, OHIPD worked with other units of NIH to complete a trans-NIH online user survey project based on the American Customer Satisfaction Index (ACSI), with significant initial and supplemental funding support from the NIH/OD Evaluation Branch. The project extended the ACSI online user survey methodology to about 60 NIH Web sites at about 28 different ICs/OD units. The project contributed to: 1) strengthening participating IC/OD Web evaluation capability; 2) sharing of Web evaluation learning and experience on a trans-NIH basis; 3) aggregating ACSI results and learning on a trans-NIH basis; and 4) sponsoring several NIH-wide meetings and a final workshop that highlighted the contributions and challenges of the ACSI from the NIH perspective. The project was managed by a multi-institute ACSI Survey Leadership Team. The primary ACSI contractor was ForeseeResults Inc. via an arrangement brokered by NLM. The primary evaluation contractor was Westat Inc. A paper based on the project will be published in the *Journal of Medical Internet Research*. Also the project was a recipient of the NIH Office of the Director Group Merit Award.

Tribal Connections

NLM has continued to focus on improving Internet connectivity and access to health information services in American Indian and Alaskan Native communities. Phase I (Pacific Northwest) and Phase 2 (Pacific Southwest) of tribal connections are complete, and a final project evaluation was published. Also, Phase 3, which involved more intensive community-based outreach and training at select Phase 1 and 2 sites, is complete. A Phase 3 evaluation report is available from the Pacific Northwest

Regional Medical Library, University of Washington, Seattle.

NLM has funded Phase 4, in collaboration with the University of Utah (Midcontinental Regional Medical Library), emphasizing a needs assessment of Web-based tribal health information needs in the Four Corners Region (Arizona, Colorado, New Mexico, Utah). Phase 4 was completed in FY2005, and an evaluation report is available from the Midcontinental RML. NLM has also funded a Phase 5 that includes outreach to Four Corners public libraries serving American Indians in that region, the convening of an NN/LM meeting to consolidate lessons learned from Native American outreach as of July 2006, several pilot projects now underway, and a Web site to share tribal outreach best practices and facilitate collaboration.

Other Native American Outreach

Also, in 2007 OHIPD again participated in the NIH American Indian Pow-Wow Initiative. This included exhibiting at eight pow-wows mostly in the Mid-Atlantic area. An estimated 4,000 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. In other parts of the country, OHIPD supported several projects in the Dakotas, Hawaii, and Alaska. These projects resulted largely from the Native American Listening Circles conducted in FY2003–2004, and have continued into FY2007:

- North Dakota—Cankdeska Cikana Community College (via the Greater Midwest RML), Spirit Lake Nation, Ft. Totten, ND, continuing project to develop a health-related educational program at the Community College, and improvements at the tribal library;
- North Dakota—MHA Systems Inc., a tribal enterprise of the MHA Nation, continuing economic development outreach project to provide outreach assistance to a tribal information technology company that would ultimately result in jobs creation on the reservation (in this case, the Ft. Berthold Indian Reservation); the project is intended to improve the competitive capabilities of MHA Systems Inc., and also to refine, test, and strengthen the company's core scanning services.
- Hawaii—Papa Ola Lokahi (via the Pacific Southwest RML), two continuing Native Hawaiian Community Health Education Projects:
 - Community of Miloli'i, Hawaii (The Big Island)—increased the knowledge of community members about health information and health resources by providing computer hardware and software to the community's library, training for the librarian and other community members, and by increasing multi-

media resources at the Miloli'i Community Library; and supporting community-based initiatives founded on Hawaiian concepts of health (involving a balance between body, mind and spirit).

- Waimanalo Health Center, Oahu (Windward side)—increased the knowledge of community members about health information resources in order to better understand their own health conditions or health conditions of family members and to enable more effective self-management and more informed communication with health service providers; to be achieved by providing training and access to Web-based sources of health and medical information.

Tribal Consultations on Exhibition Concepts

In FY2007, OHIPD helped coordinate and facilitate two major consultations with tribal leaders on the planned NLM exhibition on "Native Concepts of Health and Illness." The consultations were held in Anchorage, AK and Santa Fe, NM. A broad spectrum of Native Americans participated in the consultations, and discussed a variety of ideas, topics, and perspectives for possible use in the exhibition. In conjunction with the consultative meetings, NLM staff conducted site visits of nearby Indian reservations, Native villages, and historical museums. OHIPD staff participated in both consultations and related outreach activities.

Outreach to Hispanics

The Lower Rio Grande Valley Hispanic Outreach Project was a collaboration with the University of Texas at San Antonio Health Sciences Center to conduct a needs assessment and various health information outreach projects with Hispanic-serving community, health, and educational institutions. This was the beginning of an intensified NLM effort to meet the health information needs of the Hispanic population in Texas and elsewhere. In FY2007, NLM continued its support for MedHigh, a magnet health high school in the Lower Rio Grand Valley of Texas. The MedHigh Peer Tutors Program is an award-winning effort to involved high schools students in teaching their peers about online health information.

International Programs

The focus of the office of International Programs is on outreach to researchers, physicians, and librarians in developing countries with an additional more recent emphasis on health workers and end users. This office continues to develop pilot programs, dissemination strategies, and training opportunities as well as evaluation, presentation, and publication of results. In addition, NLM has a Web site, Resources for

International Librarians, Health Professionals and Researchers in Developing Countries at <http://www.nlm.nih.gov/psd/ref/international.html>.

Ms. Julia Royall, Chief, International Programs, NLM, is spending the year in Uganda as a Fulbright Scholar. She is based at Makerere University, Kampala, Uganda, where she is working closely with medical students, professors, and librarians to identify information needs, strengthen medical libraries in Africa, introduce students to the use of MEDLINE, and test MedlinePlus tutorials developed specifically for African use.

MIMCom and Beyond

MIMCom, a project of the Multilateral Initiative on Malaria and the National Library of Medicine, was conceived by African malaria researchers in 1997 and designed and implemented by NLM in collaboration with partners in Africa, US, UK, and Europe. The mandate for Internet access to medical literature came from African scientists: "Access to e-mail and the Internet will promote rapid communication between investigators working at different sites as well as access to online literature and data available to scientists outside Africa." Having established or enhanced connectivity at 21 research sites in 12 countries, NLM's current focus is on products and databases to aid the efforts of malaria research.

MedlinePlus Tutorials for Africa

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

MIMCom News and Web Sites

MIMCom News, a weekly newsletter reaching more than 1,500 malariologists around the world, has recently been spun off and is now a privately run publication known as Malaria World. It still offers an effective medium to the entire MIM community to communicate messages among a large group of professionals. The newsletter aims to be the most complete electronic malaria information resource, covering announcements, contributions from subscribers, scientific publications, reports, events, jobs, grants, training and research opportunities, and news.

From surveys, it is clear that this publication has become an invaluable resource to malaria researchers and managers in Africa and the rest of the world. In addition, NLM hosts a Web site for MIMCom at <http://www.nlm.nih.gov/mimcom>.

African Medical Journal Editors Partnership Program

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

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

Network of African Librarians

NLM organized a three-day workshop, the first of its kind, in FY2006 at the Ethiopian Civil Service College/Global Distance Learning Center in Addis Ababa. Librarians from five Ethiopian universities attended: Black Lion/Addis Ababa University, Debub, Mekelle, Jimma, and Gondor universities. The workshop was designed for librarians, working in health institutions or medical schools in Ethiopia, who are not yet trained in searching medical databases. A follow-up videoconference between Addis Ababa and Washington, D.C. was held on October 10, 2006. One outcome from videoconference was a survey of Ethiopian medical faculty and students about their knowledge of NLM products and services, being conducted along with site visits by a librarian from the Ethiopian Civil Service College/Global Distance Learning Center. Additionally, both undergraduate and graduate students at Ethiopian universities are being trained in the use of NLM resources.

On May 24, 2007, NLM hosted a meeting of African librarians, including NLM Associates African alumni and former Cunningham Fellows from Africa and NLM senior staff. The librarians from Africa included Grace Ajuwon/Nigeria, Abda Anne/Mali, Agnes Chikonzo/Zimbabwe, Cristina Horta/Mozambique (by phone), Nancy Kamau/Kenya, Christine Kanyengo/Zambia (by phone), Alison Kinengyere/Uganda (by phone).

All of these librarians have been busy training faculty and students. They have carried out workshops for librarians and researchers from around the country, produced regular newsletters, presented at faculty board meetings, and conducted lunchtime training sessions for staff. Several have developed institutional repositories which can be accessed online from anywhere.

The objectives of this meeting were to assist African librarians who already have links to NLM in creating an approach for strengthening libraries through outreach and training in Africa, and to explore how this librarian corps can be brought together with the African Medical Journal Editor Partnership Project (Mali, Uganda, Malawi, Ghana) and African research and clinical communities.

Visitors

In FY2007, the Office of Communications and Public Liaison and the History of Medicine Division gave 95 regular daily (1:30 pm) tours and 320 specially arranged tours. For the first time in NLM history, the number of tour guests exceeded 10,000. There were 10,829 visitors in all, from the following 70 countries:

Albania, Argentina, Brazil, Cameroon, Canada, Chile, China, Colombia, Republic of the Congo, Denmark, Dominican Republic, England, Ethiopia, France, Georgia, Germany, Ghana, Guatemala, Guyana, Haiti, Iceland, India, Iran, Iraq, Republic of Ireland, Italy, Jamaica, Japan, Kenya, Republic of Korea, Kyrgyzstan, Liberia, Lithuania, Madagascar, Malawi, Malaysia, Mali, Mongolia, Morocco, Nepal, the Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Northern Ireland, Papua New Guinea, Paraguay, Peru, Philippines, Russia, Serbia, Singapore, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Taiwan, Tanzania, Thailand, Trinidad and Tobago, Tunisia, Turkey, Uganda, United States, Venezuela, Vietnam, Zambia, and Zimbabwe.

International MEDLARS Centers

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

LIBRARY OPERATIONS

Becky J. Lyon
Deputy Associate Director

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

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

LO employs a multidisciplinary staff of librarians, technical information specialists, subject experts, health professionals, historians, museum professionals, and technical and administrative support personnel and relies on the services of a wide range of contractors. LO is organized into four major Divisions: Bibliographic Services, Public Services, Technical Services, and History of Medicine; three units: the Medical Subject Headings (MeSH) Section, the National Network of Libraries of Medicine Office, and the National Center on Health Services Research and Health Care Technology (NICHSR); and a small administrative office. LO staff members participate actively in efforts to improve the quality of work life at NLM, including the Diversity Council and the NLM Intranet.

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.

LO also participates with national information standards organizations in the development of standards related to preservation, bibliographic control, collection holdings, vocabulary control, and data exchange. In FY2007, more than 30 standards actions were reviewed.

Program Planning and Management

Priorities for LO programs are based upon the goals and objectives identified in the *NLM Long Range Plan 2006-2016* and, where appropriate, in the NLM Strategic Plan to Reduce Racial and Ethnic Disparities. The current NLM Long Range Plan emphasizes connecting and making the results of research from scientific data to published literature to patient and consumer health information readily available. In the next ten years, NLM's programs and services will become even more central to scientific discovery, treatment, and prevention. In FY2007, work began in LO on the development of a three-to-five-year LO plan which will align closely with the NLM Long Range Plan.

In FY2007, LO continued to review and revise policies, procedures, and services 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. LO also 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. Transitioning responsibility for the Metathesaurus production system from LHC to LO and OCCS was completed in FY2007.

In FY2007, LO continued to review and revise policies, procedures, services, and organizational lines to reflect shifting workloads, to use electronic information to enhance basic operations and services; and to work with other NLM program areas to ensure permanent access to electronic information. An LO-wide group which includes members from OCCS developed functional requirements for an NLM Digital Repository as an initial step in moving forward with a plan for permanent access to digital information. Work also continued on the Indexing 2015 initiative, an NLM-wide research and development effort to improve indexing performance and productivity which is being led by LO. In FY2007, the Cataloging Section began using some of the technology designed for indexing to improve its productivity.

Though many of its efforts are directed toward creating and promoting use of electronic information resources and to supporting NLM's high priority outreach initiatives, LO also devotes substantial resources and attention to the care and handling of NLM's extensive collections of physical library materials and to the space and environment in which staff and patrons work and collections are stored. In FY2007, LO also continued to make improvements to conditions in the existing NLM library building and develop strategies for handling the

projected growth of the collections until a new facility becomes available. These improvements and strategies are discussed elsewhere in this report.

In FY2007, LO's Administrative Office continued to assist managers, supervisors and staff with the transition to a range of new administrative systems and the challenges of changes in the functioning of human resources within NIH. LO continues to encourage its staff to take advantage of flexiplace work arrangements as appropriate. More than 108 employees work at home at least one day per week.

Collection Development and Management

NLM's comprehensive collection of biomedical literature is the foundation for many of the Library's services. LO ensures that this collection meets the needs of current and future users by updating NLM's literature selection policy; acquiring and processing relevant literature in all languages and formats; organizing and maintaining the collection to facilitate current use; and preserving it for subsequent generations. At the end of FY2007, the NLM collection contained 2,558,973 volumes and 8,919,851 other physical items, including manuscripts, microforms, pictures, audiovisuals, and electronic media.

Selection

In FY2007, selectors worked on a number of projects to enhance the NLM collections. Approval plans were established to cover most of Eastern Europe, including Hungary and the former Yugoslav states. Review was completed of over 1100 Cyrillic and Slavic medical publications donated by the Countway Library of Medicine at Harvard University. This resulted in the addition of 341 items to NLM's collections. Electronic Data Interchange (EDI) ordering of monographs was expanded to four additional U.S. and foreign vendors. Due to the retirement of four selectors in FY2007, most selection activities experienced significant decreases; selection of new print serials decreased by 31% and selection of new electronic journals by 56%.

Acquisitions

TSD received and processed 155,935 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. A net total of 26,915 volumes and 2,256,164 other items (including nonprint media, manuscripts and pictures acquired by HMD) were added to the NLM collection. In FY2007, 3,018 licensed and 1,813 free electronic journals were available to NLM users, many through the NIH Library licenses to Elsevier's Science Direct and Wiley Interscience for NIH employees. NLM installed the Meridian electronic

resource management system, which will be used to track the acquisition and licensing processes for electronic journals and databases.

A new five-year contract for subscription services for U.S. and foreign periodicals for 2008-2012 was awarded to Swets Information Services. Australia and New Zealand and optional services for electronic journal licensing support were added to the contract which now covers over 9,000 orders and accounts for over 80% of serials expenditures. A new five-year contract covering 2008-2012 was also awarded to JPTC for acquiring journals from Japan.

HMD acquired many splendid early printed books, manuscripts, images, and historical films for the NLM collection in FY2007. Important books acquired included: Joannes de Kethan, *Fasciculus Medicine* (Venice, 1513)-- this anthology went through many editions, this one containing ten full-page woodcuts, including the famous dissection scene which begins the *Anatomy* of Mudinus; Girolamo Mercurio's *La Comare Oricoglitrice di Scipione Mercuri* (Venice, 1596) an early vernacular work on obstetrics illustrated and written for midwives; and Christophe Landre's *Haussartzney* (Augsburg, 1578), a collection of household remedies.

Additions to the archival and manuscript collections included the Earl and Thressa Stadtman papers, a husband and wife team at the National Heart, Lung, and Blood Institute; papers of Samuel S. Epstein, a cancer researcher at the University of Illinois; the papers of Henry Swan, a cardiologist who induced hypothermia in open-heart surgery; and the papers of Emery Johnson, Director of the Indian Health Service from 1969-1981. The library also acquired the entire historical collection of the National Center for Homeopathy located in Alexandria, VA. Of the 1,870 volumes included, most were published before 1915; some archival and manuscript materials were also included.

Prints, photographs, and audiovisuals acquired included photo archives from the National Institute of Mental Health, more than 400 contemporary posters (including a dozen from the Army Bicentennial Magazine and many more on public health topics such as AIDS, SARS, teen pregnancy, and drug abuse), oral histories of past directors of the Indian Health Service, a film on Indian health, and a tape of the open hearings on servicemen and veteran's health donated by the National Naval Health Center. A new "Guide to Donating Materials to the Images and Archives Collections of the National Library of Medicine" was completed.

Preservation and Collection Management

LO undertakes a broad range of activities to preserve NLM's archival collection and keep it readily accessible for use. These activities include: binding, reformatting, conservation of rare and unique materials, book repair, maintenance of appropriate storage and environmental conditions, and disaster prevention and response. LO

distributes data about what NLM has preserved to avoid duplicate effort by other libraries. LO works with other NLM program areas to conduct experiments with new preservation techniques as warranted and to promote the use of more permanent media and archival-friendly formats in new biomedical publications.

In FY2007, LO bound 19,670 volumes, repaired 2,313 items in NLM's onsite repair and conservation laboratory, made 317 preservation copies of motion pictures and videos, and conserved 163 rare items.

A multi-year inventory of the serials collection, which began in FY2006, continued with the completion of *Index Medicus* titles and the contractor began inventorying *Index Catalogue* and most heavily requested titles. Detailed procedures were developed for inventorying analyzed titles. During the year, 8,367 titles were inventoried and 6,898 missing items were identified for IM/MEDLINE titles. To date, 16,337 titles have been inventoried. In FY2007, PSD staff developed a plan to obtain from other libraries print copies of journal issues or volumes that are missing from the collection. More than 10,000 missing issues or volumes have been identified from binding preparation and the inventory. The program will be launched in Spring 2008.

Dwindling space for growth of the NLM collections continues to require significant resources for shifting of collections and contingency planning in the event that NLM does not receive funding for a new facility. Over the course of 167 years, the NLM has repeatedly outgrown its physical facilities as its collection grows. This unparalleled resource of books, journals, and other materials contains much that does not exist elsewhere and thus, as hospital and research libraries face increasing budget and space constraints, the NLM collection will assume even greater importance to the nation. Unfortunately, shelf space will be filled in 2010, and some special collections are already out of space.

In FY2007, work began on the first phase of a multi-year project to improve building systems in the stacks and expand existing collection space by strengthening the B2 floor and installing compact shelving. This will be accomplished by applying a fiber reinforced polymer (FRP) to the B2 floor and the ceiling of the floor below (B3). A small area on B3 was vacated, FRP was applied to the ceiling, the HVAC, lighting, flooring and sprinklers were upgraded, and rails for compact shelving were installed. The prints and photos collection was moved into the space in July. Preservation and Collection Management continues to search for offsite storage space which will be needed while the renovations and floor strengthening are in process and to refine the space plans.

Permanent Access to Electronic Information

The preservation of electronic information presents

unique challenges that are not yet fully understood. NLM's general approach to addressing these challenges is to use its own electronic services and publications as test-beds and to work with other national libraries, the National Archives and Records Administration, and other interested organizations to develop, test, and implement strategies and standards for ensuring permanent access to electronic information. LO works closely with other NLM program areas on activities related to the preservation of digital materials. As part of an emerging digital preservation program, an NLM Digitization Selection Criteria document was issued to provide guidelines for decisions about what materials should be preserved using digitization.

PubMed Central, a digital archive of medical and life sciences journal literature developed by the National Center for Biotechnology Information (NCBI), is NLM's primary test-bed for the development of procedures and methods for ensuring permanent access to electronic journals. LO continues to assist NCBI in expanding current deposits to PubMed Central by soliciting the participation of additional journals, primarily in the fields of clinical medicine, health policy, health services research, and public health. In addition, the Public Services Division also works closely with NCBI to scan and add the backfiles of journals depositing current issues in the digital archive. The back issue scanning project proceeded smoothly throughout the year and is vastly increasing the material in PubMed Central. At the end of FY2007, more than 761,433 articles have been scanned and made available. In June, PMC reached the one-million article milestone. Complete runs of 85 journals have been processed. Each journal is made available in its entirety, including front and back covers, tables of contents, administrative material such as masthead and editorial boards, and advertisements. Views include a full-text summary, HTML view, separate views of TIFF images, and the full PDF. Thirteen historically significant titles were added to PMC during the year: *American Journal of Human Genetics*, *American Journal of Public Health*, *Annals of Surgery*, *Biophysical Journal*, *Canadian Journal of Comparative Medicine and Veterinary Science*, *Canadian Medical Association Journal*, *Canadian Veterinary Journal*, *Canadian Journal of Veterinary Research*, *Immunology*, *Journal of Anatomy*, *Journal of Anatomy and Physiology*, *Skull Base: Official Journal of North American Skull Base Society*, and *Skull Base Surgery*.

The first phase of a project to digitize historical pamphlets and monographs on the subject of cholera was completed. The project included implementation of specifications and procedures for the creation, quality assurance, and storage of preservation level master images and their accompanying metadata. This project is providing valuable experience as NLM moves into digitization as its primary means of preservation.

Vocabulary Development and Standards

LO produces and maintains the Medical Subject Headings (MeSH), a thesaurus used by NLM and many other institutions to describe the subject content of the biomedical literature and other types of biomedical information; develops, supports, or licenses the U.S. vocabularies designed for use in patient records and clinical decision support systems; and works with the Lister Hill Center and OCCS to produce the Unified Medical Language System (UMLS), a large database that incorporates many vocabularies, including MeSH and other vocabularies produced or supported by NLM. A multi-purpose knowledge source used in operational systems and informatics research, the Metathesaurus also serves as a common distribution vehicle for classification, 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 1966 (HIPAA). In this capacity, LO staff members serve on the Department of Health and Human Services Data Council, provide staff support to the National Committee on Vital and Health Statistics (NCVHS) Standards and Security Subcommittee, participate in the Public Health Data Standards Consortium, and contribute to the Federal Health Architecture (FHA) initiative, and participate in Healthcare Information Technical Standards Panel technical committees. 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. In FY2007, NLM worked with the HHS Office of the Secretary and ONC to encourage establishment of a secure funding mechanism for standards development activities for FY2009 and beyond. Establishing this funding mechanism would enable NLM to expand its current activities to fully support our role as the coordinating center for standard clinical terminologies.

Medical Subject Headings (MeSH)

The 2008 edition of MeSH contains 24,767 main headings, 83 subheadings or qualifiers, and more than 172,000 supplementary records for chemicals and other substances. For the 2008 edition, the MeSH Section added 456 new descriptors, replaced 288 descriptors with more up-to-date terminology, and deleted 46 descriptors.

The 2008 vocabulary reflects a major project to completely revise the Leukemia and Lymphoma trees. A large number of new and merged descriptors were created. ICD-O, 3rd edition, supplemented by recent papers and texts, were the authorities used in this revision. The Publication Types were also revised along with

associated subject descriptors. The project goal was to make clear the distinction between the descriptors that are "about" a form or type of format and those which label the form of information presented. For example, an article may discuss biography as a type of creative presentation or an article may be a biographical treatment of a person. The Publication Types and other types of descriptive headings have been inserted into MeSH as part of a new tree category, Publication Characteristics.

It was announced that the printed MeSH will be discontinued and the last year printed will be 2007. This decision followed a comment period by members of the NN/LM and analysis of sales data provided by the Government Printing Office.

Clinical Vocabularies

The MeSH Section and its contractors also produce RxNorm, a clinical drug vocabulary that provides standardized names for use in prescribing medications. RxNorm was designated as a U.S. government-wide target clinical vocabulary standard by the Secretary of Health and Human Services. 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. RxNorm provides a mechanism for connecting information from different commercial drug information services. In FY2007, RxNorm was linked to additional drug sources within the UMLS Metathesaurus: Gold Standard Alchemy and Metathesaurus FDA Structured Product Labels (MTHSPL). The latter drug names are created from the structured labels submitted for DailyMed; these are inserted as Metathesaurus forms of the FDA Structured Product Labels.

Through LO's NICHSR, NLM supports the continued development and free distribution of LOINC[®] (Logical Observation Identifiers Names and Codes) by the Regenstrief Institute. LOINC was designated as a U.S. government-wide target clinical vocabulary standard in 2003, and proposed for adoption as standard under the administrative simplification provision of the Health Insurance Portability and Accountability Act (HIPAA) in a notice of proposed rulemaking issued in September 2005.

NLM continued to support and pay the annual fees for the US-wide license for the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) in FY2007. In April 2007, the International Health Terminology Standards Development Organisation (IHTSDO) was established and assumed ownership and responsibility for maintenance and distribution of SNOMED CT; its mission is to significantly promote global standardization of health information. NLM, on behalf of HHS, participated in the negotiations and is now the U.S. member of the IHTSDO. This new organization will allow NLM to establish a robust process for input to SNOMED CT development that fully represents the needs

of the U.S. healthcare industry. In addition, NLM is working with the IHTSDO to facilitate negotiations for the alignment and harmonization between SNOMED CT and key health terminologies including LOINC and RxNorm.

The existence of authoritative electronic mappings from standard clinical vocabularies to administrative code sets is likely to facilitate automated production of bills and statistical reports as a by-product of the capture of detailed patient data. 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. NLM has enlisted cooperation from relevant federal agencies and private organizations and initiated and launched projects to map LOINC to Current Procedural Terminology (CPT) and SNOMED CT to CPT and to the International Classification of Diseases, 9th edition. The first draft LOINC to CPT map was distributed for review via the UMLS Knowledge Sources Server in 2006. An expanded and updated version is currently under review and will be published in early 2008.

In September 2004, NLM entered into a three-year contract arrangement with Health Level Seven, Inc. (HL7) to a) align HL7 message standards with CHI standard vocabularies, specifying which subsets of standard vocabularies are valid for particular message segments and replacing HL7-maintained lists of coded values with subsets of standard vocabularies where feasible (NLM-initiated) and b) create implementation guide(s) for transmitting an entire Electronic Health Record (EHR) between disparate systems (on behalf of HHS). As part of this effort, code sets defined for use in version 2 and 3 of HL7 were added to the UMLS Metathesaurus. The EHR work on behalf of HHS was completed in FY2007. A one-year no cost extension was put in place to enable HL7 to complete the NLM-initiated portion of the contract.

UMLS Metathesaurus

The MeSH Section manages the content editing of the UMLS Metathesaurus, using systems developed by the Lister Hill Center (LHC) and maintained by OCCS. In FY2007 NLM completed transitioning responsibility for UMLS Metathesaurus production, from the LHC to OCCS and LO. At the close of FY2007, the Metathesaurus contained more than 1.5 million concepts and 7.2 million concept names from 141 source vocabularies in 17 different languages. The following new sources and updates were included in the UMLS: Gene Ontology; Nursing Outcomes Classification, 3rd edition; HUGO Gene Nomenclature; SNOMED Clinical Terms; Physician Data Query; Online Mendelian Inheritance in Man; Medical Dictionary for Regulatory Activities Terminology (MedDRA); and Medical Subject

Headings. (See further information about UMLS activities in the Information Products section of this chapter).

Bibliographic Control

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

Cataloging

In general, LO adheres to the *Anglo-American Cataloging Rules, 2nd edition (AACR2)*, when creating cataloging and name authority records that reduce the level of cataloging effort required in other libraries. In FY2007, LO continued to review and provide feedback on drafts of *Resource Description and Access*, the anticipated successor to AACR2.

In FY2007, TSD's Cataloging Section cataloged 22,520 contemporary books, serial titles, non-print items, and cataloging-in-publication galleys, exceeding last year's production totals. Cataloging implemented the CONSER standard record guidelines for serials cataloging, which should reduce the cataloging time for serials by approximately 20%. The Section also initiated other activities to maximize resources including: testing the use of the Medical Text Indexer (MTI) to aid in subject heading assignment; implementing exclusive use of electronic galleys and data for CIP cataloging; and reallocating resources from the English language cataloging contract toward contracts for scripts and languages that cannot be handled by staff.

The Cataloging Section released the 2007 edition of the *NLM Classification* with a revised QuickTour tutorial. A PDF version was made available for the first time.

HMD cataloged 3,339 monographs, 1,276 linear feet of manuscripts, 2,928 historical audiovisuals, and 1,274 pictures. A project to catalog Russian pamphlets and dissertations completed records for 333 pamphlets and 862 dissertations.

Indexing

LO indexed 5,194 biomedical journals for the MEDLINE/PubMed database to assist users in identifying articles on specific biomedical topics. The indexing workload continues to rise, partly due to the selection of new journals for MEDLINE/PubMed, but primarily due

to increases in the number of articles published in journals already being indexed. In FY2007, a combination of in-house staff, contractors, and cooperating U.S. and international institutions indexed 671,000 articles, an 8% increase from last year. Previously indexed citations were updated to reflect 124 retractions, and 31,956 comments found in subsequently published notices or articles.

In FY2007, indexers created 51,823 annotated links between newly indexed MEDLINE citations for articles describing gene function in selected organisms and corresponding gene records in the NCBI Entrez Gene database. This represents a 10% increase over the previous year. Index Section assumed indexing of journals from the History of Medicine Division and the Kennedy Institute of Ethics, implemented a new policy for journal supplements requiring disclosure statements indicating any financial relationship that an author or editor has for the supplement sponsor and with any product discussed in the article text, and implemented a new procedure to process/index journals recommended by the Literature Selection Technical Review Committee from the electronic version if NLM has online access and the publisher can submit XML citation data. Of the 117 journals recommended by LSTRC in FY2007, a total of 62 were implemented under this new procedure to process from the electronic version. Index Section switched another 123 journals to process from the electronic version rather than the print version, resulting in a grand total of 185 titles for FY2007 being handled from the online version.

Increases in the number of articles and journals that are indexed for MEDLINE require new approaches to indexing to improve effectiveness and efficiency of the critical and high volume indexing process. In FY2007 the Index Section continued installing dual monitors for contract indexers to speed indexing from electronic versions of journals. Approximately 90% of contract indexers now have dual monitors. The highly successful new Web-based training program for indexing was made available to the public in June. Index Section now uses the Web-based program as the only means of training new MEDLINE indexers, which enables training to occur as staff and contractors are brought on board. LO continues to work with other NLM program areas on the Indexing 2015 initiative to identify, test, and implement ways to reduce or eliminate tasks now performed by human indexers. Progress has been made with the Medical Text Indexer (MTI) to generate subheading suggestions and the feature will now be incorporated into the indexing Data Creation and Maintenance System.

Indexers perform their work after the initial data entry of citations and abstracts is accomplished by one of two means: electronic submission from publishers (the fastest and most economical method), or scanning and optical character recognition (OCR). Work continues to improve the efficiency of data entry. In FY2007, 85% of the citations and abstracts were received electronically from publishers and the remaining 15% were scanned. A

total of 380 publishers are now supplying XML-tagged electronic data for 4,144 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 FY2007, the Committee reviewed 425 journals and 117 were rated highly enough to be accepted for indexing.

Information Products

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

Databases

LO manages the creation and maintenance of the content of MEDLINE/PubMed, NLM's database of indexed citations; the NLM catalog which is available to the public in two different databases; MedlinePlus and MedlinePlus en español, NLM's primary information resources for patients, 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. DailyMed, a database of Food and Drug Administration (FDA) approved labels, provides access to the prescribing information approved by the FDA. These databases are richly interlinked with each other and with other important NLM resources, including PubMed Central, other Entrez databases, ClinicalTrials.gov, Genetics Home Reference, as well as Specialized Information Services (SIS) toxicological, environmental health, and AIDS information services.

Use of MEDLINE/PubMed was approximately 845 million searches. Page views in PubMed totaled 4.4 billion, a 2% increase from FY2006. The June implementation of Entrez 2.0 resulted in changes to the statistical gathering process. MEDLINE/PubMed now includes more than 17 million citations. MEDLINE increased by 1.2 million citations when OLDMEDLINE citations were mapped to current MeSH and gained MEDLINE status.

BSD staff assisted NCBI with the implementation of Entrez 2.0 "portals" which offers more flexibility for creating the next generation of features and capabilities for PubMed. The implementation resulted in a 100% increase in the number of customer service calls in the second half of the fiscal year. BSD also assisted with the implementation, based on library feedback, to move the library icon to a more prominent position above the Related Articles list on AbstractPlus display.

Use of MedlinePlus and MedlinePlus en español continues to increase significantly. Nearly 120 million unique visitors viewed over 907 million pages. The

number of page views increased by more than 10% and the number of visitors increased by almost 25%. There are more than 98,670 subscribers to the weekly announcements of new content. Both sites remain in the top five of government sites based on the American Customer Satisfaction Index (ACSI). PSD and OCCS continued to expand and improve the basic content and features of MedlinePlus. Both the English and Spanish health topic pages were restructured. The redesigned pages include: health topic summaries written exclusively for MedlinePlus; images of a high quality medical illustration or photograph; synonyms or lay terms for medical conditions; and *Start Here* to get a user started on the topic. Seven new English health topic pages were added to MedlinePlus to bring the total to 745; 13 were added to the Spanish site for a total of 712. Public Service Announcements featuring Hispanic star Don Francisco began airing in July. The campaign features five 30-second video and audio spots that were sent to 100 Spanish language television stations and over 450 radio stations nationwide. Initial reports indicate that the video PSAs aired more than 9,400 times during the first six weeks of the campaign.

Four new Go Local sites were released during FY2007, bringing the total to 22 projects in 20 states. These sites provide coverage for 34% of the U.S. population. Sites released were Illinois, Minnesota, Michigan and Nebraska. Nine new proposals for Go Local sites were approved.

NIHSeniorHealth grew by four topics in FY2007 bringing the total to 33 topics. The National Cancer Institute added Skin Cancer; the National Institute for Mental Health added Depression, and the National Institute on Aging added Falls and Older Adults and Talking with your Doctor.

Under the direction of NICHSR, NLM continues to expand and enhance its Web-based resources and databases for health services researchers and public health professionals. In FY2007, NICHSR continued to work with NCBI to expand coverage of health services research and public health information by adding to the Entrez system, as part of the Bookshelf, 33 documents in the HSTAT (Health Services and Technology Assessment) database; these documents were mostly evidence reports and technology assessments from the Agency on Healthcare Quality and Research (AHRQ). NICHSR continued to work through AcademyHealth and the Sheps Center at the University of North Carolina, Chapel Hill to expand Health Services Research Projects in Progress (HSRProj) 1112 new records were added. The HSRR (Health Services and Sciences Research Resources) database also continued to expand with 96 new records covering additional datasets, surveys, other research instruments, and software packages used with datasets from both health services and public health systems research projects.

Machine-Readable Data

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

Demand for MEDLINE/PubMed data in XML format continues to increase. There are currently 472 MEDLINE licensees, an increase of 36 from last year. The new research license program continues to be a success, as 176 non-U.S. organizations use this option. A relatively small number of organizations lease NLM catalog records or one or more of the SIS toxicological or environmental health files in XML format. At the end of FY2007, there were 3,049 UMLS Metathesaurus licensees, down 14% from last year. Most of these non-renewals occurred when the licensee did not enter into the new agreement required by the change in ownership of SNOMED CT.

Web and Print Publications

NLM's databases and Web sites are its primary method of publication. Demand for the Library's few remaining print publications continues to decline due to increasing electronic access to NLM information and data throughout the world. LO will cease publishing the printed Medical Subject Headings (MeSH) after 2007.

NLM's main Web site showed increased use in FY2007 with 9.3 million unique visitors viewing 63.6 million pages. This was a 15% increase in visitors and a 4% increase in the number of pages viewed in FY2006.

Publications available from the main Web site include recurring newsletters and bulletins, fact sheets, technical reports, and documentation for NLM databases. TSD published the 2007 edition of the *NLM Classification* with a revised Quick Tour and made a PDF version available for the first time. New web resources included: a Bioethics Information Resources page providing links and search boxes to a variety of NLM and NIH bioethics resources; a Veterinary Medicine and Animal Health Special Query page linked from PubMed; a guide linking users to a wide range of MEDLINE/PubMed resources on the NLM Web site; a Journals Recently Accepted by NLM for Inclusion in MEDLINE page that announces the results of the LSTRC recommendations; and new A-Z navigation on the NLM Databases and Electronic Resources page.

Dream Anatomy (Washington, DC: Government Printing Office, 2006) by HMD staff member Michael Sappol was published. It was based on an NLM exhibition of the same name in 2002-2003 and displays and interprets the anatomical imagination from the fourteenth century to the present.

Direct User Services

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

Document Delivery

LO retrieves materials 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 FY2007, PSD's Collection Access Section processed 530,473 requests for contemporary documents. HMD handled 12,345 requests for rare books, manuscripts, pictures, and historical audiovisuals.

The number of onsite requests in the NLM Main Reading Room continues to decline due in part to increased security measures, but primarily to online access to more journals in the Reading Room and free access to journals in PubMed Central. Onsite users requested 232,040 contemporary items from the stacks, a 5% decline from last year. However, users of the HMD Reading Room requested 11,316 items from historical and special collections, a 22% increase from last year attributed to increases in requests for prints and photographs and modern manuscripts. The Reading Room photocopy contract was extended for a few months to monitor the impact that downloading to memory sticks had on photocopying and printing volume. The impact was found to be minimal and patrons are now allowed to download to personal memory stick.

In FY2007, PSD's Collection Access Section processed a total of 298,433 interlibrary loan requests, a 9% decline from the previous year. A rise in the fill rate, to a new high of 83%, occurred for the fourth consecutive year. The one-day processing rate for requests rose from 94% to 98%. NLM delivers 96% of requests electronically.

A total of 3,152 libraries use DOCLINE, NLM's interlibrary loan request and routing system. NLM released three DOCLINE upgrades with system enhancements: to display the borrower's LinkOut holdings during the process to inform the requester about local availability of the item in question; the option of creating a transaction file for electronic billing via EFTS; and, increased system security. DOCLINE users entered

2.1 million requests in FY2007, a 9% decrease from the previous year; 92% of the requests were filled. The Library's share of all DOCLINE requests remained steady at 14% of the total. Individuals submitted 484,320 requests to DOCLINE libraries via the Loansome Doc feature in MEDLINE/PubMed and the NLM Gateway, an 11% 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.

NCBI and staff at the Regional Medical Libraries continued to promote the use of PubMed's LinkOut for Libraries and "Outside Tool" as a means for libraries to customize PubMed to display their electronic and print holdings to their users. The number of libraries participating in LinkOut Outside Tool increased to 456, 54% more than in the previous year; 1855 libraries participate in LinkOut, a 19% increase over FY2006.

DOCLINE requests are routed to libraries automatically based on holdings data in its serial holdings database. At the end of FY2007, the holdings database contained 1.5 million holdings statements for 55,870 serial titles held by 3,015 libraries.

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. The EFTS staff continued to work toward making EFTS self sufficient.

Reference and Customer Service

LO provides reference and research assistance to onsite and remote users as a backup to services available from other health sciences libraries. LO also has primary responsibility for responding to inquiries from those seeking information about NLM's products or services or assistance in using these services. PSD's Reference and Web Services Section handles all initial inquiries with contract assistance and many of those requiring second-level attention. Staff throughout LO and NLM assist with second-level service when their specialized expertise is required.

In FY2007, the Reference and Web Services Section handled a total of 89,421 customer inquiries from onsite and remote patrons, a 5% decrease from the previous year. Onsite requests declined 7%, but remote requests increased 4%.

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 of medicine and science, researchers, educators, health professionals, and the general public. LO coordinates the National Network

of Libraries of Medicine (NN/LM) which works to equalize access to health information services and information technology for health professionals and the public throughout the United States; serves as the secretariat for the Partners in Information Access for the Public Health Workforce; participates in NLM-wide efforts to develop and evaluate outreach programs designed to improve information access for underserved minorities and the general public; produces major exhibitions and other special programs in the history of medicine; and conducts a range of training programs for health sciences librarians and other professionals. LO staff members give presentations, demonstrations, and classes at professional meetings and publish articles to highlight NLM programs and services.

National Network of Libraries of Medicine

The NN/LM works to provide timely, convenient access to biomedical and health information resources for U.S. health professionals, researchers, educators, and the general public, irrespective of their geographical location. It is the core component of NLM's outreach program and its efforts to reduce health disparities and improve health information literacy. The network includes 5,801 full and affiliate members. The full members are libraries with health sciences collections, primarily in hospitals and academic medical centers. Affiliate members include some small hospitals, public libraries, and community organizations that provide health information services, but may have little or no collection of health sciences literature. In an effort to maximize funding available for the NN/LM, the Midcontinental and Pacific Southwest Regions are conducting a pilot to share their Technology Coordinator. The pilot will be conducted from July 2007 to April 2008 to determine how effective a staff sharing arrangement will be in meeting regional needs and in optimizing the time and talents of NN/LM personnel between regions.

In addition to the basic NN/LM contracts, NLM funds subcontracts for three national centers that serve the entire network. The activities of one of these centers, the National Training Center and Clearinghouse at the New York Academy of Medicine, are described elsewhere in this chapter. The Outreach Evaluation Resource Center (OERC) at the University of Washington provides training and consulting services throughout the NN/LM and assists NLM, the RMLs, and other network members in designing methods for measuring the effectiveness of overall network programs and individual outreach projects. In FY2007, the Center focused on assisting with the development of an NN/LM Emergency Preparedness Plan. In September, a Coordinator was hired on a part-time basis to facilitate development of the plan and ensure its implementation in the NN/LM.

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. A major activity for the Web-STOC in FY2007 was ongoing usability testing of NN/LM Web sites. The contractor, American Institutes for Research, completed final testing of the regional Web sites, the OERC, and the membership directory and reported findings. Many of the recommended improvements were made in FY2007.

The RMLs and other network members conduct many special projects to reach underserved health professionals and to improve the public's access to high quality health information. Most of these projects involve partnerships between health sciences libraries and other organizations, including public libraries, public health departments, professional associations, schools, churches, and other community based organizations. 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 all, the NN/LM issued 54 subcontracts for outreach projects in FY2007. Many of these projects focused on improving access to information for health professionals, consumers and the public health workforce.

With the assistance of other NN/LM members, the RMLs conduct most of the exhibits and demonstrations of NLM products and services at health professional, consumer health, and general library association meetings around the country. LO organizes the exhibits at the Medical Library Association annual meeting, the American Library Association annual meeting, some of the health professional and library meetings in the Washington, DC area, and some distant meetings focused on health services research, public health, and history of medicine. In FY2007, NLM and NN/LM services were exhibited at 429 national, regional, and state meetings across the U.S. These exhibits highlight all NLM services relevant to attendees.

Partners in Information Access for the Public Health Workforce

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

The Partners Web site (PHPartners.org), managed by NLM with assistance from the New England RML, provides unified access to public health information resources produced by all members of the Partnership, as well as other reputable organizations. In FY2007, the Web site was expanded significantly, with more than 530 new links added. Unique visitors to the Web site increased more than 27 percent in FY2007, indicating a growing audience. Work continues on six new topic pages that will be added to the site, complementing the existing pages on Bioterrorism, Environmental Health, and HIV/AIDS that were added in FY2007. One of the most popular resources already on the site is the Healthy People 2010 Information Access Project (HP2010 IAP). For every focus area of Healthy People 2010, the IAP resource includes four or more objective-specific evidence-based PubMed search strategies and links to MedlinePlus topics.

In FY2007, NICHSR received the final report of the purchase order with the Association of State and Territorial Health Officials (ASTHO) for *Public Health Workforce Enumeration Strategy for Environmental Health and One Additional Public Health Field*. The report provides a specific enumeration strategy for two distinct fields within public health, environmental health and public health nursing, and will make a more general enumeration possible. NICHSR staff also participated in the MLA/NLM Task Force on electronic Personal Health Records.

Special NLM Outreach Initiatives

LO participates in many NLM-wide efforts to expand outreach and services to the general public and to address racial and ethnic disparities and participates actively in the Library's Committee on Outreach, Consumer Health, and Health Disparities. For the fourth consecutive year, LO has worked in collaboration with NLM's Director of International Programs to improve health information capacity in sub-Saharan Africa by devoting one position in the NLM Associate Fellowship Program to an African librarian. In addition to supporting this training position, a meeting was held in May at NLM to discuss with these and other health librarians in Africa, an approach for strengthening this outreach program. Former AFP participants from Kenya, Mali, Malawi and Mozambique were joined by librarians from Uganda and Zimbabwe for a lively discussion of ideas including establishing a listserv, creating a Web space for sharing online training materials, and information about institutional repositories. For the past several years, LO has also been overseeing a project to build journal capacity and enhance the quality of African medical journals by establishing partnerships between the editor of an established medical journal and the editor of an African medical journal. The following partnerships have been established: *African Health Sciences* with *British Medical Journal*; *Ghana Medical Journal* with *Lancet*; *Malawi Medical Journal* with

JAMA; and *Mali Medical with Environmental Health Perspectives* and the *American Journal of Public Health*.

LO worked with other NLM components and the American Osteopathic Association to begin another Information Rx project to test the use of "information prescriptions" for MedlinePlus in offices of osteopathic physicians in the states of Florida, Pennsylvania, Michigan and Ohio. Presentations were made to the AOA House of Delegates meeting in January and at the meeting of the American College of Osteopathic Family Physicians in March.

A new NLM collaboration with the National Agricultural Library, the US Agricultural Information Network (USAIN), the American Veterinary Medical Association, and the Medical Library Association is developing a Web site to improve information services for veterinarians in clinical practice who may not have access to a veterinary library.

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 FY2007, LO provided summer employment and training opportunities for several students.

Historical Exhibitions and Programs

HMD directs the development and installation of major historical exhibitions in the NLM rotunda, with assistance from LHC and the Office of the Director. Designed to appeal to the interested public as well as the specialist, these exhibitions highlight the Library's historical resources and are an important part of NLM's outreach program. The current exhibition, *Visible Proofs: Forensic Views of the Body*, opened on February 16, 2006. Exploring developments in scientific methods that translate views of bodies and body part into visible proofs, it tells stories of the people, sciences, and technologies that make visible the cause and manner of a death. Visitors to the exhibition observe, analyze, and decipher different forensic views of the body and examine important historical and contemporary cases and forensic techniques through the use of objects, graphics, and multimedia presentations. They also encounter experts whose contributions and discoveries have changed the field of forensic medicine.

This widely reviewed and highly praised exhibition won an NIH Plain Language Award and interactive media in the exhibition won a 2007 MUSE award from the American Association of Museums Media and Technology Committee. In FY2007 approximately 8,730 visitors have taken tours of the exhibition at NLM and the accompanying Web site has received 446,270 page views.

The very successful exhibition, *Changing the Face of Medicine: Celebrating America's Women Physicians*, continues to reach the public in a traveling version, funded by the NIH Office of Research on Women's Health and NLM. It began touring libraries in

the U.S. through collaboration with the American Library Association in FY2006. It will visit 60 libraries over a five-year period.

Previous NLM exhibitions live on through heavily used Web sites, printed catalogs, DVDs, or touring traveling versions. Exhibition Web sites received more than 4,180,271 million page hits in FY2007.

In addition to the major exhibitions in the rotunda, HMD installs "mini-exhibits" generally in cases near the entrance to the HMD Reading Room. In collaboration with the Reginald F. Lewis Museum of Maryland African American History and Culture, an exhibit entitled *Opening Doors: Contemporary African American Academic Surgeons* opened simultaneously in Bethesda and Baltimore on February 1. A travelling version went on the road later in the year. The travelling version of the mini-exhibit *The Horse, a Mirror of Man* also visited several sites during the year. *Do Mandrakes Really Scream? Magic and Medicine in Harry Potter* was popular with both children and adults and another focused on *Hooke's Books: Books that Influenced or were Influenced by Robert Hooke's Micrographia*. In addition, a mini-exhibit in tribute to the late John B. Blake, first chief of NLM's History of Medicine Division, and an accompanying Web site were also created.

New Profiles in Science Web sites were released for Nobel Prize winner and former NIH Director Harold Varmus (1989), Mary Lasker, Rosalind Franklin, and Lasker Award winner Sol Spiegelman (1974). Varmus and J. Michael Bishop shared the Nobel Prize in Physiology or Medicine "for their discovery of the cellular origin of retroviral oncogenes."; Lasker was a tireless supporter of medical research and the National Institutes of Health; and Franklin was a British chemist and crystallographer best known for her role in the discovery of the structure of DNA. It was her x-ray diffraction photos of DNA and her analysis of that data--provided to Francis Crick and James Watson without her knowledge--that gave them clues crucial to building their correct theoretical model of the molecule in 1953. Spiegelman was an American molecular biologist whose pioneering discoveries accelerated the study of gene mechanisms and laid the foundations of recombinant DNA technology. Later he developed RNA-DNA hybridization, one of the most important techniques of molecular biology, and in 1965 became the first to synthesize biologically competent and infective virus RNA in test tubes.

In November, HMD sponsored a two-day workshop on Meat, Medicine, and Human Health in the Twentieth Century. The workshop brought together historians from the United States and Europe whose work dealt with the relationship of meat to human health during the twentieth century. The workshop was also sponsored by the Institut de Recherche sur les Sciences et la Technologie at the Louis Pasteur University, the Medical Faculty at Louis Pasteur University, and the Maison InterUniversitaire des Sciences de l'Homme-Alsace.

HMD also continued to sponsor a number of public lectures and film screenings throughout the year.

HMD staff members present historical papers at professional meetings and to publish the results of their scholarship in books, chapters, articles, and reviews. HMD continued to play a lead role in preparing the recurring features "Voices from the Past" and "Images of Health" for the *American Journal of Public Health*, which often features materials from the NLM collection.

Training and Recruitment of Health Sciences Librarians

LO develops online training programs in the use of MEDLINE/PubMed and other databases for health sciences librarians and other search intermediaries; oversees the activities of the National Online Training Center and Clearinghouse (NTCC) at the New York Academy of Medicine; directs the NLM Associate Fellowship program for post-Masters librarians; and develops and presents continuing education programs for librarians and others in health services research, public health, the UMLS resources, and other topics. LO also collaborates with the Medical Library Association, the Association of Academic Health Sciences Libraries, and the Association of Research Libraries to increase the diversity of those entering the profession, to provide leadership development opportunities, to promote multi-institution evaluation of library services, and to encourage specialist roles for health sciences librarians.

In FY2007, the MEDLARS Management Section (MMS) and the NTCC trained 1,017 students in 74 classes covering PubMed, the Gateway, ClinicalTrials.gov, TOXNET, and the UMLS. A new Quick Tour of MyNCBI, "Editing Collections in My NCBI" was introduced. An average 14,729 unique visitors used the PubMed Tutorial for nearly 124,000 page views each month.

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 locations, 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 make available its suite of courses on health services research, public health, and health policy.

The NLM Associate Fellowship program had eleven participants in FY2007, with seven 1st year Fellows on site at NLM and four 2nd year Fellows placed at libraries outside NLM. Field placements for the 2nd year fellows were at Indiana University, George Washington University, Johns Hopkins University, and the University of North Carolina-Chapel Hill. In September 2007, a new group of seven 1st year Fellows

began their year at NLM, including one International Fellow from the E. Latunde Odeku Medical Library in the College of Medicine at the University of Ibadan, Nigeria. Four of the 2006-07 Fellows chose to continue for a 2nd year fellowship and started their program in September at the following sites: Yale University, MIT, University of Michigan, and Uniformed Services University of the Health Sciences.

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, Medical Library Association, and the Association for Research Libraries (ARL). LO also supports the NLM/AAHSL Leadership Development Program, which provides leadership training, mentorship, and site visits to the mentor's institution for an annual cohort of five mid-career health sciences librarians.

Table 1

Growth of Collections

Table 1 A: Book Materials

<i>Collection</i>	<i>Previous Total (9/30/06)</i>	<i>Added FY2007</i>	<i>New Total (9/30/07)</i>
Monographs: Before 1500	591	3	594
Monographs: 1501-1600	5,982	23	6,005
Monographs: 1601-1700	10,267	17	10,284
Monographs: 1701-1800	24,736	39	24,775
Monographs: 1801-1870	41,685	75	41,760
Monographs: Americana	2,341	0	2,341
Monographs: 1871-Present	773,522	15,774	789,296
Theses (historical)	288,091	0	288,091
Pamphlets	172,021	0	172,021
Bound serial volumes	1,325,128	20,794	1,345,922
Volumes withdrawn	-112,306	-9,810	-122,116
Total volumes	2,532,058	26,915	2,558,973

Table 1 B: Nonbook Materials

<i>Collection</i>	<i>Previous Total (9/30/06)</i>	<i>Added FY2007</i>	<i>New Total (9/30/07)</i>
Microforms: Reels of microfilm	148,194	574	148,768
Microforms: Number of microfiche	452,709	4,072	456,781
Microforms: Total microforms	600,903	4,646	605,549
Audiovisuals	80,452	1,432	81,974
Computer software	2,549	10	2,559
Pictures	68,994	291	69,285
Manuscripts	5,911,132	2,250,325	8,161,457*
Non-book items added	6,664,120	2,256,704	8,920,824
Non-book items withdrawn	-433	-540	-973
Total non-book items	6,663,687	2,256,164	8,919,851

*Total manuscripts equivalent to 4,664 linear feet

Table 1 C: Total Book and Nonbook Materials

<i>Collection</i>	<i>Previous Total (9/30/06)</i>	<i>Added FY2007</i>	<i>New Total (9/30/07)</i>
Total book & nonbook items	9,195,745	2,283,079	11,478,824

Table 2

Acquisition Statistics

<i>Acquisitions</i>	<i>FY2005</i>	<i>FY2006</i>	<i>FY2007</i>
Serial titles received	20,989	20,815	20,165
Publications processed: Serial pieces	132,347	134,020	136,543
Publications processed: Other	24,659	22,368	21,905
Publications processed: Total	157,006	156,388	158,448
Obligations for: Publications	\$8,255,443	\$8,715,869	\$10,329,902
Obligations for: (For rare books)	-\$324,398	-\$337,386	-\$402,153

Table 3

Cataloging Statistics

	<i>FY2005</i>	<i>FY2006</i>	<i>FY2007</i>
Completed Cataloging	21,238	21,662	22,520

Table 4

Bibliographic Services

	<i>FY2005</i>	<i>FY2006</i>	<i>FY2007</i>
Citations published in MEDLINE	606,000	623,000	670,900
Journals indexed for MEDLINE	4,928	5,020	5,194
Journals indexed for <i>Index Medicus</i>	4,279	4,416	4,530
Total items archived in PubMed Central	432,587	755,860	1,115,778

Note: For Index Medicus (ceased publication in 2004)

Table 5**Consumer Web Services**

<i>Services</i>	<i>FY2005</i>	<i>FY2006</i>	<i>FY2007</i>
NLM Web Home Page: Page Views	56,400,000	61,000,000	63,600,000
NLM Web Home Page: Unique Visitors	8,000,000	9,000,000	9,300,000
MedlinePlus: Page Views	661,000,000	820,000,000	906,000,000
MedlinePlus: Unique Visitors	74,600,000	95,000,000	120,000,000
ClinicalTrials.gov: Page Views	61,303,796	129,673,811	479,926,975
ClinicalTrials.gov : Unique Visitors	3,499,091	6,105,822	6,188,562
DailyMed: Page Views		842,977	9,187,248

DailyMed: Unique Visitors	***	96,522	1,298,775
Genetics Home Reference: Page Views	13,633,547	20,181,044	39,829,732
Genetics Home Reference: Unique Visitors (estimated)	827,590	1,394,706	1,906,055
Household Products Database: Page Views	10,547,690	11,622,189	14,045,949
Household Products Database: Unique Visitors	1,908,604	1,219,764	2,375,254
Tox Town: Page Views	3,232,466	7,271,692	10,993,836
Tox Town: Unique Visitors	228,220	261,938	248,349

***Data not available for 2005

Table 6**Circulation Statistics**

<i>Activity</i>	<i>FY2005</i>	<i>FY2006</i>	<i>FY2007</i>
Requests Received: (Total)	580,072	573,828	530,473
Requests Received: Interlibrary Loan	341,239	328,661	298,433
Requests Received: Onsite	238,833	245,996	232,040
Requests Filled: (Total)	475,623	467,143	442,731
Requests Filled: Interlibrary Loan	273,870	265,562	246,902
Requests Filled: Onsite	201,753	201,581	195,829

Table 7

Online Searches—PubMed and NLM Gateway

	<i>FY2005</i>	<i>FY2006</i>	<i>FY2007</i>
Total online searches	754,000,000	896,000,000	845,000,000*

* The Entrez infrastructure was changed in June 2007, altering the statistical gathering method as well.

Table 8

Reference and Customer Services

<i>Activity</i>	<i>FY2005</i>	<i>FY2006</i>	<i>FY2007</i>
Offsite requests	73,493	76,582	79,502
Onsite requests	22,298	15,202	9,919
Total	95,791	91,784	89,421

Table 9

Preservation Activities

<i>Activity</i>	<i>FY2005</i>	<i>FY2006</i>	<i>FY2007</i>
Volumes bound	18,417	19,317	18,684
Volumes microfilmed	1,564	1,509	Ceased in 2006
Volumes repaired onsite	2,095	1,814	2,313
Audiovisuals preserved	936	863	318
Historical volumes conserved	140	154	163

Note: PSD corrected the FY2006 statistics for Audiovisuals preserved.

Note: Microfilming ceased during FY2006.

Note: In FY2007, PSD replaced microfilming statistics with statistics on titles, items and pages scanned (digitally preserved):

<i>Activity</i>	<i>FY2007</i>
Titles Scanned	222
Items Scanned	222
Pages Scanned	6,796

Table 10

History of Medicine Activities

Table 10 A : Acquisitions

<i>Activity</i>	<i>FY2005</i>	<i>FY2006</i>	<i>FY2007</i>
Books	1,070	1,401	938
Modern manuscripts	1,107 (in ft)	783 (in ft)	372 (in ft)
Prints and photographs	11,252	9,945	839
Historical audiovisuals	176	1,864	476

Table 10 B : Processing

<i>Activity</i>	<i>FY2005</i>	<i>FY2006</i>	<i>FY2007</i>
Books cataloged	2,668	3,952	3,339
Modern manuscripts cataloged	453 (in ft)	374 (in ft)	1,276 (in ft)
Pictures cataloged	2,974	6,342	1,274

Note: Citations now being done by Index Section

Table 10 C : Public Services

<i>Activity</i>	<i>FY2005</i>	<i>FY2006</i>	<i>FY2007</i>
Reference questions answered	20,655	23,818	22,237
Onsite requests filled	12,738	14,140	12,588

SPECIALIZED INFORMATION SERVICES

Steven Phillips, MD
Associate Director

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

The Toxicology and Environmental Health Information Program (TEHIP), known originally as the Toxicology Information Program, was established 40 years ago within the National Library of Medicine in the Division of Specialized Information Services (SIS). Over the years, TEHIP has provided for the increasing need for toxicological and environmental health information by taking advantage of new computer and communication technologies to provide more rapid and effective access for 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. Development of HIV/AIDS information resources has been a focus of the Division for many years, and now includes several collaborative efforts in information resource development and deployment, including a focus on the information needs of other special populations. Our outreach program has continued to evolve and reach out to under-served communities through implementation of innovative information access-enabling approaches and dissemination of NLM's resources.

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

In FY2007 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 2007 include the following:

Toxicology and Environmental Health Resources

The **TOXNET** (TOXicology Data NETwork) is a cluster of databases covering toxicology, hazardous chemicals, environmental health and related areas. These databases continue to be highly used resources, and in FY2007 customer surveys, 86% of the respondents reported that they would "return to this site" and "recommend it to others." In FY2007, enhancements to TOXNET were based on user feedback/requests and routine upgrades/additions of data and capabilities. Databases in TOXNET include:

- **LactMed** (Drugs and Lactation Database) provides information on drugs and other chemicals to which breastfeeding mothers may be exposed. It includes information on the levels of such substances in breast milk and infant blood, and the possible adverse effects in the nursing infant and includes links to other NLM databases.
- **HSDB®** (Hazardous Substances Data Bank), a peer-reviewed database focusing on the toxicology of over 5,000 potentially hazardous chemicals. This flagship database was enhanced with records on radiological compounds during FY2007.
- **IRIS** (Integrated Risk Information System), a database from the U.S. Environmental Protection Agency (EPA) containing carcinogenic and non-carcinogenic health risk information on over 540 chemicals.
- **ITER** (International Toxicity Estimates for Risk), a database containing data in support of human health risk assessments. It is compiled by Toxicology Excellence for Risk Assessment (TERA) and contains over 560 chemical records.
- **CCRIS** (Chemical Carcinogenesis Research Information System), a scientifically evaluated and fully referenced data bank, developed by the National Cancer Institute (NCI) and now maintained by SIS, with over 9,000 chemical records with carcinogenicity, mutagenicity, tumor promotion, and tumor inhibition test results.
- **GENE-TOX** (Genetic Toxicology), a toxicology database created by the U.S. Environmental Protection Agency (EPA) containing genetic toxicology test results on over 3,000 chemicals.
- **TOXLINE**, a bibliographic database providing comprehensive coverage of the biochemical, pharmacological, physiological, and toxicological effects of drugs and other chemicals from 1965 to the present. TOXLINE contains over 3.5 million citations, almost all with abstracts and/or index terms and CAS Registry Numbers.
- **DART/ETIC** (Development and Reproductive Toxicology/Environmental Teratology Information

Center), a bibliographic database covering literature on reproductive and developmental toxicology.

- **Toxics Release Inventory (TRI)**, a series of databases that describe the releases of toxic chemicals into the environment annually for the 1987-2005 reporting years.
- **ChemIDplus**, a database providing access to structure and nomenclature authority databases used for the identification of chemical substances cited in NLM databases. ChemIDplus contains over 380,000 chemical records, of which over 276,000 include chemical structures.
- **Household Products Database**, which provides information on the potential health effects of chemicals contained in more than 7,000 common household products used inside and around the home.
- **Haz-Map**, an occupational toxicology database designed primarily for health and safety professionals, but also for consumers seeking information about the health effects of exposure to chemicals and biologicals at work. It links jobs and hazardous tasks with occupational diseases and their symptoms. Added this year in collaboration with the Department of Labor were tasks and chemicals associated with work at the Department of Energy hazardous sites.
- **ALTBIB**, a bibliographic database on alternatives to the use of live vertebrates in biomedical research and testing, developed as part of NLM's participation in the Interagency Coordinating Committee to Validate Alternate Methods.

WISER (Wireless Information System for Emergency Responders) is a tool developed for use by emergency responders during hazardous materials incidents, as well as during training sessions/exercises in preparation for such events. Version 3.1 was released this year and included the addition of radiation information about radiological compounds, treatment algorithms, a dose estimator, and reference guides. Also, in collaboration with the Department of Transportation, the DOT Emergency Response Guidelines 2004 was added as an optional PDA stand-alone resource as well as integrated within WISER. Usage among first responders continued to grow with over 46,000 downloads of WISER onto PDAs (Palm and Pocket PC) and Windows-based desktop/laptops over the FY2007. Total number of WISER downloads is over 90,000. Positive accounts from users about their application of WISER continue to be received, and staff has participated in emergency drills in which WISER was used.

REMM (Radiation Event Medical Management) is a tool developed in an interagency collaboration between SIS and the DHHS Office of the Assistant Secretary for

Preparedness and Response (ASPR) and released in December 2007. REMM is a system intended for use by emergency physicians and related emergency health care providers that includes algorithm-based guidelines for evaluation and management of individuals exposed to radiation during accidental releases, use of radiological dispersion devices, and use of improvised nuclear devices. In FY2007, the REMM system was released, following peer review by experts in the radiation domain, and changes based on user feedback.

Tox Town was enhanced with new content, in English and Spanish, in the Tox Town 'neighborhoods' of Tox Town, Tox City, Tox Farm and a U.S. Mexico Border scene. A new Port scene was added in FY2007. To promote the use of Tox Town by teachers, a teacher page was developed with sections on activities and discussion questions, interactive and illustrated resources, checklists and quizzes, career information and general resources for teachers, and the resource was demonstrated at several educational conferences.

TOXMAP, a Geographic Information System (GIS) 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, released Version 4.0 in FY2007, featuring advanced searching and the ability to build and save custom-made regional views.

Enviro-Health Link pages continue to be useful to our users, especially the new *Dietary Supplements* page with links to many sources of relevant information and the *Pesticide Exposure* page with links to websites about the acute and chronic exposure to pesticides.

ToxSeek is a meta-search tool that enables simultaneous searching of many different information resources and databases on the World Wide Web. The ToxSeek user interface allows selection of resources from a wide range of authoritative sources in environmental health and toxicology. It provides integrated search results from the selected resources and displays related concepts to use in refining searches. Based on user feedback and focus group evaluations, work has continued on enhancements for future releases.

ToxMystery, an interactive Web site for children between the ages of 7-10, was released at the end of FY2006. It provides an animated game-like interface, which includes finding potential chemical hazards in a home and includes fun sound effects. Focus groups and feedback from the targeted user community have indicated that this innovative Web site is a fun and educational experience for kids. A Spanish version was developed and released in FY2007.

AIDS Information Services

NLM is the project manager for the multi-agency AIDSinfo service (aidsinfo.nih.gov). This service provides access to federal HIV/AIDS treatment guidelines, AIDS-related clinical trials information (through Clinicaltrials.gov), and prevention and research information. In April 2007, a Spanish language site called InfoSIDA was released.

The American Customer Satisfaction Index (ACSI) continues to be used to evaluate AIDSinfo. The 2007 score for AIDSinfo is 82, which places its ranking among the top government Web sites.

The National Library of Medicine (NLM) has continued its HIV/AIDS-related outreach efforts to community-based organizations, patient advocacy groups, faith-based organizations, departments of health, and libraries. This program provides support to design local programs for improving information access for AIDS patients and the affected community as well as their caregivers. Emphasis is on providing information or access in a way meaningful to the target community. Projects must involve one or more of the following information access categories: information retrieval, skills development, Internet access, resource development, and document access. In FY2007 NLM made 13 awards.

Evaluation Activities

In FY2007, several SIS web products were professionally assessed via on-line surveys, focus groups, heuristic review, or usability studies. User satisfaction with TOXNET and AIDSinfo continues to be measured by the American Customer Satisfaction Index (ACSI). The new Radiation Event Medical Management (REMM) Web site conducted usability testing with 16 physicians. ToxMystery held a focus group with teachers. An expert heuristic review on usability was conducted for TOXMAP. Over the past three years, other SIS web products have also been professionally evaluated: World Library of Toxicology, ToxSeek, LactMed, Toxicology and Environmental Health home page, WISER, Tox Town, Asian American Health, Arctic Health, American Indian Health, and the Household Products Database. Feedback from these evaluation activities is used to identify improvements, new capabilities, and expanded content that would better serve SIS web users.

Outreach Initiatives

SIS outreach programs reach health professionals, public health workers and the general public especially about health issues that disproportionately impact minorities such as environmental exposures and AIDS. Highlights from FY2007 include:

- United Negro College Fund Special Programs/ NLM – HBCU Access Project, one of NLM's major outreach projects with Historically Black Colleges and Universities, continued during the FY2007 and awarded four HBCUs small grants to develop and implement projects that help to increase the awareness and utilization of NLM resources on campuses and in their communities. The annual June workshop featured a keynote address by Dr. Robert M. Kolodner, Director, Office of the National Coordinator for Health Information Technology Department of Health and Human Services (DHHS).
- Adopt-a-School program with Woodrow Wilson Senior High School, Washington, D.C., encourages students to take an active interest in consumer health and promotes interest in science. Five Wilson students had summer internships at NLM, working in a variety of program areas. NLM also provided tours and library materials for the school.
- Consumer Health Resource Information Service (CHRIS) Project is a faith-based pilot initiative designed by the Medical Education and Outreach group (MEO) of the Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee. The project, which initially involved six African American Churches in Knoxville, has been expanded to a state-wide program due to its success locally. A "Consumer Health Resource Information Service (CHRIS) Program: Guide and Tool Kit" has been developed for faith-based organizations that wish to implement such a program in their communities.
- The mission of the Environmental Health Information Outreach Program (EnHIOP) is to enhance the capacity of minority serving academic institutions to reduce health disparities through the access, use and delivery of environmental health information on their campuses and in their communities. Two successful meetings were held in FY2007 – *Advances in Science: Environmental Cardiology* held at Morehouse School of Medicine in January 2007 and the second, *Advances in Science: Indian Health and Health Disparities*, hosted by the Oglala Tribal College in South Dakota. EnHIOP meetings included representation from 14 HBCUs, three tribal colleges and three Hispanic-serving institutions.
- The Scared Root Tribal Information Fellowship provides an opportunity for representatives from American Indian tribes, Native Alaskan villages, and the Native Hawaiian community to learn about the National Library of Medicine, the National Network of Libraries of Medicine, and to use that knowledge to improve access to health information and technology for their communities. This year's Scared Root Fellows are from the Chickasaw Nation in Ada, Oklahoma. They started their

fellowship in April 2006 and in addition to their own work, will be supporting an ongoing NLM-funded Chickasaw Health Information Center (CHIC) being implemented at the Carl Albert Indian Health Facility in Ada. The two fellows from the Navajo Nation are continuing their training and project planning.

- SIS is a partner in the Refugee Health Information Network (RHIN), which is a national collaborative partnership of several state Refugee Health offices, NLM, and the Center for Public Service Communication (CPSC). RHIN is committed to providing quality multilingual, multi-cultural health information resources for patients and those who provide care to resettled refugees and asylees. A new Web site was released this fiscal year with improved access to a database of approximately 1,800 documents with almost 70 languages represented. In addition, the search system simultaneously retrieves information from several trusted sources including Ethnomed at the University of Washington, and Health Roads Media.
- Minority Health Professional Outreach includes a variety of training and outreach projects with several minority health professional organizations. NLM provides focused online training, demonstrations, presentations at the National Hispanic Medical Association, Black Nurses Association, Association of Hispanic Nurses, Student National Medical Association, and at the National and regional meetings of the National Medical Association. In addition NLM has been collaborating with the Student National Medical Association in areas such as mentoring and developing the pipeline of minority students going to biomedical careers.
- The Central American Network for Disaster and Health Information (CANDHI) is a group of health science libraries and information centers working together to enhance local health and disaster information management capacities with a goal of contributing to disaster preparedness in the region. It is a partnership between the U.S. National Library of Medicine, the Pan American Health Organization, and the United Nations International Strategy for Disaster Reduction. CANDHI consists of centers in Honduras, Nicaragua, El Salvador, Guatemala (with support from U.K. Department of International Development), and Panama and Costa Rica (with financial support from the European Community Humanitarian Office (ECHO)). The CANDHI centers enable health professionals, government agencies, and others in their countries to access vital information previously unavailable. These libraries have acquired the knowledge, skills, and resources that promote delivery of reliable information including. There are now over 12,000

full-text documents available online. During FY2007, a Web site for the overall program was developed (<http://www.candhi.org>). This Web site includes a search engine to retrieve information from all of the CANDHI center web sites and the CRID web site, including the full-text documents. In addition, the tools for digitization of the documents were improved. The CANDHI centers continue to provide critical information services to disaster managers, health care professionals and the public on disasters in their countries.

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.

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. The World Library has been populated with information resource sets from more than forty countries and collaborations with many other countries are in progress. With support from the Fogarty International Center, this project is scheduled to release fully developed information resources in FY2008.

Another resource under development in FY2007 was the **Dietary Supplements Database**, a resource of comprehensive information on supplements used by U.S. consumers. Information on more than 2,000 dietary supplement brands will be available and searchable by brand name, active ingredient, or manufacturer, with links to TOXNET and PubMed searches and other authoritative government information. Following review during 2007 by other agencies with responsibilities in the area of dietary supplements, SIS will release the database in early October of 2007.

The goal of the **Public Health Law Information Project (PHLIP)** is to create in the public domain a searchable database of public health law information that will be not only a guide for non specialists (e.g., concerned citizens, attorneys, public health practitioners, academics, legislators), but also an excellent technical resource for those who are specialists in the field. In FY2007, the pilot project was continued with the state of Delaware, the Widener University School of Law, the Delaware Academy of Medicine and SIS to produce a searchable database containing statutes, regulations, and other information from Delaware that pertain to public health.

SIS is leading a NLM-wide collaborative initiative to produce a **Drug Information Portal** that will make it easier for consumers and health professionals to find drug information in the NLM and other governmental resources. The prototype uses ChemIdplus drug records for searching and resource locator selection, but operates with a user-friendly simple interface. Release is expected in early FY2008.

During FY2007 SIS participated actively in the evaluation of disaster-related information resources across

NLM in preparation for the creation of a Disaster Information Management Research Center in response to recommendations in the most recent NLM Long Range Plan. It is anticipated that this Center will operate from a new Office within SIS.

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, disaster information resources, and other specialized information resources.

LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS

Clement J. McDonald, MD
Director

The Lister Hill National Center for Biomedical Communications (LHNCBC), established by a joint resolution of the United States Congress in 1968, is a research and development division of NLM. Seeking to improve access to high quality biomedical information for individuals around the world, the Center continues its active research and development. It leads a research and development program aimed at creating and improving biomedical communications systems, methods, technologies and networks, and enhancing information dissemination and utilization among health professionals, patients, and the general public. An important new focus of the LHNCBC is the development of Next Generation electronic health records, to facilitate patient-centric care, clinical research and public health, an area of emphasis in the *NLM Long Range Plan 2006-2016*.

The Lister Hill Center research staff is 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 NLM, other Institutes at NIH, other organizations within the Department of Health and Human Services, and academic and industry partners. Staff regularly publish their research results in the medical informatics, computer and information science, and engineering communities. The Center is often visited by researchers from around the world.

The Lister Hill Center is organized into five major components: Cognitive Science Branch (CgSB); Communications Engineering Branch (CEB); Computer Science Branch (CSB); Audiovisual Program Development Branch (APDB) (which currently includes the Office of the Public Health Historian); and the Office of High Performance Computing and Communications (OHPCC).

An external Board of Scientific Counselors meets biannually to review the Center's research projects and priorities. The most current information about the Lister Hill Center research activities can be found at <http://lhncbc.nlm.nih.gov/>. The Center's principal research activities and accomplishments are described in the remainder of this chapter.

Next Generation Electronic Health Records to Facilitate Patient-centric Care, Clinical Research, and Public Health

These projects are early efforts to target the overall recommendations of the NLM Long Range Plan (LRP) Goal 3: "Integrated Biomedical, Clinical, and Public Health Information Systems that Promote Scientific Discovery and Speed the Translation of Research into Practice."

NLM Personal Health Record (PHR)—The early version of this new project aims to help individuals who are caring for their elderly parent(s) and/or young children. This initial version of the NLM PHR supports entry and tracking of key measurements and test results, prescriptions, problems and immunizations. It also will produce digital and paper copies of its contents in various formats. The current version gives access to MedlinePlus information resources about prescriptions and (soon) ailments through one click on the name of recorded medication or ailment. The PHR has logic that can remind the care giver about preventive care interventions that are due (e.g. it is time for the annual flu shot, or ask your doctor about use of cholesterol lowering drugs in patients with high cholesterol). The current system attaches codes to the medications, observations, and problems it carries. These codes come from terminologies supported by NLM and designated as national standards by HHS. The automatic inclusion of these codes within the NLM PHR will facilitate automatic downloading of clinical information to the PHR in future versions of the system.

De-identification Tools—De-identification can unlock the research potential of long-term clinical records, and no well-supported and freely available de-identification tools exist. Taking advantage of experience with de-identified procedures within the NCI Shared Pathology Informatics Network (SPIN) grant in my previous position, and existing Lister Hill Center tools that recognize sensitive information such as dates, person names and locations, text, numbers, and speech, LHNCBC initiated an effort to develop an open source text de-identification tool. This project is in its early stages and has been using a scrubbed 70-patient database to test the initial tools. We are negotiating with a number of sources for narrative clinical data for test purposes. The investigators have completed their IRB training and will be taking HIPAA (Health Insurance Portability and Accountability Act) training as well.

Clinical Data Entry Tools—The initial goal of this project is to develop a tool that can generate data entry forms dynamically based on specifications stored in a database. Currently the development platform chosen is Ruby on Rails, an open-source web application framework. One potential area of use is in the data capture function of personal health records. Several terminology resources from the Unified Medical Language System (UMLS) (e.g. RxNORM, ICD9-CM) are used in some data entry fields that require a set of controlled terms. Further development will involve work with very large databases of de-identified patient data. Additional reusable software tools

will be created, including those involving biostatistical analysis with the “R” package.

Collaboration with Centers for Medicare and Medicaid Services (CMS)—LHNCBC has assisted CMS in the development of many aspects of Medicare’s Post Acute Care data collection project that will be demonstrated in the spring of 2008. One of Medicare’s goals in this project is to standardize and meld different data collection forms from four different post-acute care settings. LHNCBC proposed and demonstrated a Web approach to auto-completion of entered text, much of which has been adopted by Medicare for this demonstration project. LHNCBC has proposed and delivered the full content of many of the look-up tables that Medicare will use in this project (including a revised subset of RxNorm) and shaped their data conceptualization to fit a LOINC/HL7 model. A memorandum of understanding (MOU) to formalize this collaboration is now in final review stages within CMS.

Concept Recognition in Narrative Clinical Reports—The LHNCBC Natural Language Processing group has embarked on an effort to examine all of the words and phrases in a narrative clinical report, identifying the clinical concepts and converting them into UMLS Concept Unique Identifiers (CUIs). Tools already exist for converting strings found in published literature into concepts. Clinical text presents the additional challenge of dealing with telegraphic phrasings (the absence of grammatically complete sentences) and distinguishing positive from negative statements about a finding, disease or symptom. The goal here is not full understanding but the ability to comb through and find important clinical concepts in narrative records. Such tools would have widespread use in quality assurance, clinical research, and decision support, with the right level of sensitivity and specificity. The NLM has existing tools that can facilitate this effort.

Biomedical Imaging and Multimedia

The overall goal of this major research area is to address fundamental questions that arise in the handling, organization, storage, access and transmission of very large electronic files in general and digitized biomedical images in particular. A special focus is research into these topics as applied to heterogeneous multimedia databases consisting of both images and text. Projects in this area have benefited from collaborators in several universities as well as at agencies such as the National Center for Health Statistics (NCHS) and the National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS), and a continuing partnership with the National Cancer Institute (NCI) in their research in cervical cancer caused by the Human Papillomavirus (HPV). These projects address the recommendations of the NLM Long Range Plan Goal 1: “Seamless, Uninterrupted Access to Expanding Collections of Biomedical Data, Medical Knowledge, and Health Information.”

Interactive Publications Research—This effort is a major priority under NLM LRP Recommendation 1.4, to “Evaluate interactive publications as a possible means to enhance learning, comprehension, and sharing of research results.” LHNCBC investigators have developed prototype interactive publications (IP) using Adobe tools-based procedures created by the IP Research team. One is an article by CEB authors (published in the *SPIE Proceedings*) augmented by dynamic tables and graphs, a microscopy video of cell evolution, an animated spine in Flash, digital x-rays, and clinical DICOM images (CT, MRI, ultrasound). In a separate development, IPs were created from two published articles and the raw results for the underlying studies acquired from the American Psychiatric Institute for Research and Education (APIRE), and the raw data were linked to every (tabular) result presented in the articles. Starting with SAS scripts provided by the APIRE authors, we generated executable files and SAS scripts to reproduce every table and statistical test. In discussions of the interactive capabilities with the authors, we identified a set of core data analysis capabilities needed by readers who do not intend to export data for rigorous statistical analysis. The descriptive exploratory analysis tools (means, confidence intervals, linear and non-linear regression analysis) are being developed for ITAG+, a data visualization tool being created in-house.

Steps were taken to recruit a publisher to host IPs for dissemination. Our IP prototypes were demonstrated to the publisher, including the functions of converting tables to graphs, zooming into graphs, creating subsets of the tabular data, zooming into images and changing contrast in DICOM images.

In light of the large sizes of such publications - up to hundreds of megabytes, we are studying techniques and protocols to download such publication rapidly and progressively. We are also developing a Download Manager on the basis of these efforts.

Multimedia Database R&D—The goals of this project are: to research latest technological approaches for information retrieval and delivery for biomedical databases that include non-text data, with an emphasis on biomedical images; and to develop prototype systems for the retrieval and delivery of such information for use by the research and, potentially, the clinical communities. This project encompasses the following systems and capabilities:

- **WebMIRS (Web-based Medical Information Retrieval System)**

Developed some years ago and still in active use, WebMIRS continues to provide access to images and text from nationwide surveys conducted by the National Center for Health Statistics. At present, there are 492 users of WebMIRS in 56 countries. This Java application allows remote users to access data from the National Health and Nutrition Examination Surveys II and III (NHANES II and

III). The NHANES II database 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, 550 of the 17,000 x-ray images collected in NHANES II contain vertebral boundary data collected by a board-certified radiologist. Users may do queries for both radiological and/or health survey data.

- **Digital Atlas of the Cervical and Lumbar Spine**

The Digital Atlas remains available for the public from the CEB Web site, either as a Java applet or a downloaded Java application or as a CD version of the Java application. The Java application version allows the user to add images (either grayscale or color) in a special "My Images" section, and to annotate and title those images for later use.

In addition, the FTP x-ray archive of 17,000 digitized spinal x-rays continues to be very active, with over 500 users worldwide. This archive allows access to the x-rays, available in full 12-bit flat file format and also in TIFF 8-bit format which is easier for many researchers to use.

A suite of newer systems motivated by, but not restricted to, our joint research with National Cancer Institute (NCI), are at various stages of development.

- **Multimedia Database Tool (MDT)**

The MDT extends current WebMIRS capabilities. This new system is intended to accommodate the existing WebMIRS databases and the new text/image database currently being created from the collection of uterine cervix images from NCI. New features allow for both data dissemination and distributed data collection. A working implementation of the MDT is regularly used for demos by NLM and NCI. It currently operates on a database of 60,000 JPEG cervigrams and associated clinical text data from the Guanacaste Project. The latest version includes capability to query and display images segmented with the Boundary Marking Tool.

- **Boundary Marking Tool (BMT)**

This tool provides Web capability to manually mark boundaries on cervicography images, and to manage collected data with a MySQL database. The BMT has supported several studies of the uterine cervix which were carried out by NCI. Dr. Mark Schiffman of NCI presented results from the 20-observer, 919-patient, BMT study done earlier this year, to the 23rd International Papillomavirus Conference and Clinical Workshop in Prague, Czech Republic. The BMT was also used in a pilot study to determine appropriate methods for selecting biopsy sites in colposcopic images. This collaborative work using the BMT was used in a paper, "Colposcopy at a crossroads," which was

published in the August 2006 edition of the *American Journal of Obstetrics and Gynecology*. BMT results were also used in the July 2007 paper, "Visual appearance of the uterine cervix: correlation with human papillomavirus detection and type," published in the same journal. In addition to this cervix-related work, the BMT has been used to mark and label dermatological lesions for the NCI Viral Epidemiology Branch, headed by Dr. Jim Goedert for a study which was completed in 2007 to assess the effects of nicotine patch treatment on the regression of Kaposi sarcoma lesions, and the study results have been summarized in the paper, "Treatment of classic Kaposi sarcoma with a nicotine dermal patch: a phase II clinical trial."

- **Virtual Microscope (VM)**

The VM, currently at an intermediate level of development, will provide Web capability to view and collect information on histology images from expert observers. There are both a simple demo system of basic histology image and data collection capability and a fully-functional system, currently being used to support a multiple-observer study of lung histology images, in collaboration with the NCI Genetic Epidemiology Branch, the NCI Cell and Cancer Biology Branch, and their medical collaborators in Italy.

- **Teaching Tool (TT)**

This system is for training medical personnel in cervix anatomy/pathology. It displays uterine cervix images and quizzes an observer in the categories of medical knowledge, pattern recognition, and patient management, and enables a medical expert to tailor exams by specifying images and questions to use on an examination. A prototype system is available for experimentation by NCI and American Society for Cervical Pathology and Colposcopy (ASCCP) experts, which includes capability to administer and score the ASCCP "Resident's Online Exam."

- **Visual Triage Study (VTS) Software**

The study goal is to determine whether non-oncologists can answer screening questions as reliably as oncologists. Two phases of the Visual Triage Study have now been completed, the latest in Peru. The VTS software is now supporting a third, follow-on phase of this study, which has observers in Costa Rica, Ghana, Peru, and Thailand.

For our work with NCI, these systems are interrelated through the data that is used. The MDT distributes images and text data from the NCI Guanacaste and ALTS projects; the BMT allows the collection of additional graphical and text data that is added to the MDT database for distribution; similarly, for the VM; the TT uses data collected by the BMT to create the content for the examinations

it supports, and the VTS software uses images and text data from Guanacaste and ALTS.

- **Content-Based Image Retrieval (CBIR)**—This project investigates approaches for query and retrieval of biomedical images by direct use of image data, possibly in association with text related to the biomedical images. The emphasis is primarily on the NHANES spine images, using shape methods on vertebrae in the images, and on NCI cervigrams, using color and texture methods to differentially identify tissue regions and tissue characteristics within these images. One goal is to develop effective CBIR methods that may be incorporated into our multimedia database programs (such as the MDT) or into separate, prototype systems for use and evaluation by the biomedical research and/or clinical communities. A third goal is to develop geospatially distributed computing approaches to multiscale CBIR. As envisioned, the results from this effort would allow export and Web-based reuse of open methods that may be particular to local pathologies in various image collections to interact with others that are designed for gross image analysis and classification.

Currently there are two CBIR systems and an interface for a geospatially distributed computing approach to multiscale CBIR. Our earlier CBIR system, CBIR3, has been entirely redeveloped to have a Web-based focus. It has been renamed SPIRS (Spine Pathology Image Retrieval System) and it supports two shape retrieval methods as well as retrieval by descriptive text, using a database of several thousand pre-segmented vertebral shapes and text data from the NHANES II database. The new design is multi-platform and allows the system to run in any Web browser, as long as the host platform has the Java Runtime Environment (JRE) installed.

Using a similar philosophy, the Web-based CBIR system has been developed for NCI uterine cervix images. This system supports retrieval from a database of over 900 images on which medical experts have outlined the boundaries of regions of interest. To create a query, the user specifies weighted combinations of color, texture, size, and locations of marked region of interest. This system is designed as a CBIR prototype suitable for medical experts to use, experiment with, and provide feedback to the engineering community. In addition, we have created an internal engineering version of the Web-based system that allows more fine-grained, technical query specification over a broader range of image features and query methods. SPIRS also exports its algorithms through a query interface to its core algorithms (dubbed cSPIRS) using structured XML queries. A test environment has been setup with the IRMA (Image Retrieval for

Medical Applications) project at Aachen University in Germany. The IRMA project aims at classification of images using overall appearance, but is insensitive to local pathologies in these individual image groups. In contrast, SPIRS supports queries of local pathology in digitized x-ray images of the spine.

In data collection, we have segmented an additional 600 images, bringing the total number of vertebral outlines to 13,000. We are currently using 7,000 shapes in our SPIRS system. We have obtained a three-way expert data collection of 500 images using our Path-Va system. In addition, we have over 1,500 shapes that have expert-marked boundaries.

The Visible Human Project—The Visible Human Project image data sets are designed to serve as a common reference for the study of human anatomy, as a set of common public domain data for testing medical imaging algorithms, and as a test bed and model for the construction of image libraries that can be accessed through networks. The Visible Human data sets are available through a free license agreement with the NLM. They are distributed to licensees over the Internet at no cost; and on DAT tape for a duplication fee. The data sets are being applied to a wide range of educational, diagnostic, treatment planning, virtual reality, and virtual surgeries, in addition to artistic, mathematical, legal, and industrial uses by over 2,450 licensees in 49 countries. The Visible Human Project has been featured in more than 900 newspaper articles, news and science magazines, and radio and television programs worldwide.

FY2007 saw the continued maintenance of two databases to record information about Visible Human Project use. The first, to log information about the license holders and record statements of their intended use of the images; and the second, to record information about the products the licensees are providing NLM in compliance with the Visible Human Dataset License Agreement.

A new edition of the *NLM Current Bibliographies in Medicine, Visible Human Project* was made available on the Visible Human Project Web site. This publication contains over 900 citations and covers the period from January 1987 through March 2007. The Visible Human Project bibliography is an attempt to identify all publications in the scientific and technical literature which discuss the Visible Human Project and its derivative products.

In FY2007, a planning workshop, “VHP: Scope and Scale for the Future,” assembled an expert panel of radiologists, anatomists, pathologists, computer scientists, and engineers from across the country to advise NLM on future directions for the Visible Human Project. Topics such as human variation, community data annotation, algorithm validation, and multiscale anatomy emerged as leading areas of interest.

3D Informatics—The 3D Informatics Program has expanded research efforts concerning problems encountered in the world of 3-dimensional and higher-dimensional, time-varying imaging. Among its many projects, the 3D Informatics (TDI) Group has continued work on image databases, including ongoing support for the National Online Volumetric Archive (NOVA), an archive of volume image data. This collection contains 3D data from across medicine. Contributors to the collection include the Mayo Clinic Biomedical Imaging Resource and the Walter Reed Army Medical Center Radiology Department. Integrated and multimodal data such as virtual colonoscopy matched with recorded video from endoscopic interventions, time-varying 3D cardiac motion, and 4D MRI of a human hand appear in the archive.

The 3D Informatics group continues its partnership with the NLM Specialized Information Systems Division and the U.S. Veterans Administration to study content-based retrieval methods for medical image databases. In the pharmaceutical identification project, we are assisting in the acquisition of imagery through digital macro-photography of the thousands of prescription pharmaceuticals dispensed routinely by the VA Centralized Mail-Order Pharmacies. Together we are creating a new, updated, visual database of all these products and developing techniques for automatically identifying any product in the inventory from a representative photograph. New OHPCC research has developed computer vision approaches for the automatic segmentation, measurement, and analysis of solid-dose medications. In particular, recent focus has been on robust color classification tools to help identify prescription drugs.

3D Telepresence for Medical Consultation—This project tests the efficacy of 2D versus 3D representations of video data transmitted in real time in remote clinical consultations. The technology infrastructure is being developed at the University of North Carolina and its efficacy is being researched there with help from colleagues at other institutions. The research team continues to make substantial progress in implementing the technology infrastructure. A prototype portable camera unit was added to the stationary one and calibrated. The PDA application was completed and all the basic components of the system proposed are in place. The current focus continues on optimizing camera and sensor placement, refining calibration and rendering algorithms, and dealing with problems when perspective changes from different points of view, such as occlusion when an intervening object obstructs the view of interest.

Advanced Network Infrastructure for Distributed Learning and Collaborative Research—This project builds on previous work with HAVnet (Haptic Audio Visual Network for Educational Technology) and is collaboration between Stanford University and the University of Wisconsin at La Cross. The project's focus is on developing visual and haptic applications for

anatomy and surgical training and includes aspects of self scaling technology, self-optimizing end-to-end, network aware, real time middleware, wireless technology, and GIS. The technology is being developed and refined in the context of teaching anatomy and surgical skills and addresses issues concerning network bandwidth and latency and the integration of 3D visualization, haptic, and real time online collaboration tools.

The project proposes to deliver: enhancement and integration of two existing middleware applications, Information Channels and Weather Stations, allowing correlations to be made between network metrics and actual application performance; addition of self-optimizing features to the six applications using the core middleware; development of a new application, Anatomy Window, that uses a handheld computer to map a cadaver and present corresponding images derived from the Visible Human data set; development of a Remote Tactile Sensor, capable of capture and transmission of tactile dermatology information over a network; implementation of the anatomy teaching suite over local, national and global networks for use in early, laboratory based and actual field teaching; and implementation of the clinical skills test bed, primarily in early phase and laboratory testing.

Work on the remote stereo viewer and haptic probe was completed and preliminary trials suggest videoconferencing is essential for dermatologists to see and communicate with patients while using the haptic device. Research was conducted on sense of touch and ability to detect thickness and resistance of membranes as part of the effort and to compare this feedback to haptic feedback generated by computer. The SPRING surgical simulator engine and its Remote Tactile Sensor component were made open source. The engine allows building of software modules providing haptic feedback for simulated surgical tools. Work continued on the iAnatomy collaboration with the Northern Ontario School of Medicine involving the use of the stereo viewer for anatomy teaching and distance learning and a similar collaboration with the University of California, Davis is being explored.

Insight Tool Kit (ITK)—The Insight Toolkit, a research and development initiative under the Visible Human Project, is now in its sixth year with a recent official software release of ITK 3.4. ITK makes available a variety of open source image processing algorithms for computing segmentation and registration of high dimensional medical data on a variety of hardware platforms. Platforms currently supported are PCs running Visual C++, Sun Workstations running the GNU C++ compiler, SGI workstations, Linux-based systems and Mac OS-X. Support, development, and maintenance of the software are managed by a community of university and commercial groups, including OHPCC intramural research staff. The ITK continues to have an impact on the medical imaging research community. Researchers are testing, developing, and contributing to ITK in more than

40 countries, with more than 1,500 active subscribers to the global mailing list for the project.

Across NIH, ITK is providing a foundation for new imaging investigations. The National Alliance of Medical Image Computing (NA-MIC), an NIH Roadmap National Center for Biomedical Computing (NCBC), has adopted ITK and its software engineering practices as part of its engineering infrastructure. NA-MIC is currently using medical imaging techniques to study the physiological sources of schizophrenia and other mental disorders. Staff members participate as science officers and lead science officer for the NIH-Roadmap for the NA-MIC consortium.

ITK also serves as the software foundation for the Image Guided Surgery Toolkit (IGSTK), a research and development program sponsored by the NIH National Institute for Biomedical Imaging and Bioengineering (NIBIB) and executed by Georgetown University's Imaging Science and Information Systems (ISIS) Center. IGSTK is pioneering an open API for integrating robotics, image-guidance, image analysis, and surgical intervention. The external advisory board for IGSTK includes members of the Lister Hill staff.

From 2002 to 2007, approximately 20 purchase orders were awarded for reference data sets and enhanced algorithms to support the further development of ITK. This effort supported the integration of ITK into research platforms such as the Analyze from the Mayo Clinic, SCIRun from the University of Utah's Scientific Computing and Imaging Institute, and the development of a new release of VolView, free software for medical volume image viewing and analysis. Among the data acquisitions for NLM, the Mayo Clinic Biomedical Imaging Resource has provided over 100 datasets collected across dozens of animals and clinical cases representing a wide cross section of anatomy, pathology, modality, and pre- and post-operative clinical conditions.

Image and Text Indexing for Clinical Decision Support—The title of a publication is not always sufficient in determining the Evidence-Based Practice (EBP) relevance of a publication. Given that medical illustrations often convey essential information in compact form, this project seeks to automatically identify illustrations from the articles that could help clinicians evaluate the potential usefulness of a publication in a clinical situation. We explored feasibility of automatic image annotation by utility for EBP, and if such images can be reliably extracted from the original articles.

Our study showed that images presented in clinical journals can be successfully annotated by their usefulness in finding evidence to assist a clinical decision. The feasibility of automatic image classification with respect to its utility in finding clinical decision support demonstrated in this study provides several venues for further exploration. We plan to study the influence of augmenting bibliographic references retrieved from a database search with images; new ways of organizing and presenting retrieval results using annotated images; and

further improvement in the automatic single and multi-panel image extraction, annotation, and complementary text extraction.

"Turning the Pages" Information Systems—Continuing to bring the magnificent rare books at the NLM to public view, a sixth book has been added to the "TTP" collection: Robert Hooke's *Micrographia*, the first book written about microscopes and in which reportedly the word "cell" was first used. New technical challenges in converting this book included the handling of fold-out pages and the inclusion of images of historic and present day microscopes. Library visitors may touch and turn the pages of these books on onsite kiosks, and online users may use the Web version of "TTP," "TTP Online."

A 3D wireframe model of a scroll document has recently been completed. The objective is to create a "touch and scroll" function for an ancient Egyptian medical document, the Edwin Smith papyrus. It is the world's earliest known medical document, written in hieratic around the 17th century BCE, but thought to be based on material from a thousand years earlier. It is a textbook on trauma surgery, and describes anatomical observations and the examination, diagnosis, treatment, and prognosis of numerous injuries in exquisite detail. The Edwin Smith papyrus shows that the heart, vessels, liver, spleen, kidneys, and bladder were recognized, and that the blood vessels were known to be connected to the heart. Other vessels are described, some carrying air, some mucus, while two to the right ear are said to carry the breath of life, and two to the left ear the breath of death. The physiological functions of organs and vessels remained a complete mystery to the ancient Egyptians.

"TTP Online in French" ("Tournez Les Pages") was launched on Bastille Day, July 14, 2007, thanks to translations done by the director of the Bibliothèque Interuniversitaire de Médecine et d'Odontologie (BIUM) in Paris. "Tournez Les Pages" offers explanations of the text, curators' notes and instructions for a francophone audience spread over many countries around the globe. On "TTP Online"'s main introductory page, a user may click on an icon of the French tricolor to go to the equivalent page in French. Translations for five books are planned.

Video Retrieval and Reuse Project—In response to the NLM Director's request for an automated system for storing and retrieving an historic collection of videos, APDB developed the Personal Digital Library (PDL) application which operates on a personal computer (Mac or PC). Videos are stored on an external 160 GB drive connected to the PC. The system software was an outgrowth of an application originally designed for the Movement Disorders Video Database Project, developed by APDB, which dynamically linked patient metadata to patient videos. Upon launching the PDL application, a multi-windowed interface reveals all the titles in the video library. There are currently 122 titles, listed alphabetically, occupying 143 GB of disc space. Titles can also be viewed by category. The entire library can be

searched for any word contained in the audio track of any video or in the video's metadata. The results of a search immediately show all instances across the video library where that search term appears. The user can then play any of those search results within the multi-windowed interface or select the full screen option. QuickTime videos are encoded using the H.264 codec also known as MPEG 4 AVC. The system allows the user to bookmark video, create QuickTime video clips, and create, manage and merge multiple video libraries. The PDL application is fairly compact, occupying 23 MB and when operating the executable uses less than 3 MB of system memory. The program is written with open source code using the Eclipse Platform and Java development tools. The PDL was installed on the NLM Director's laptop in December 2006. The PDL was demonstrated to the LHNCBC Board of Scientific Counselors in May 2007.

In addition, APDB completed production of the "Visible Proofs" exhibition DVD program, based on the NLM History of Medicine's exhibition scripts. The final program contains highly interactive rich media content, including additional video and animation materials to enhance the content of the exhibition. New production elements include high-definition (HD) video of the entire exhibition structure and case materials, and 17 animated segments, based on the exhibition booklet, *The New Forensic Science*, which features scientific theories and tools currently at the forefront of forensic science. Additional video interviews were conducted with: Mike Sappol, PhD, Exhibition Curator; Clyde Snow, PhD (Norman, OK); Sir Alec Jeffries (London, England); Brian Andresen (Livermore, CA); and Kirk Bloodsworth from the Washington, DC-based Justice Project. The graphical user interface design represents the exhibition themes and incorporates the existing color palette of the exhibition. Also, APDB provided project management support and HD video recording of several events including the production of the Collen Award video, the conversion of the Barbara McClintock "Profiles in Science" module to HD DVD, and a DVD version of the Paul Ortega interview.

Automated Concept Extraction from Documents

Research in this area is directed toward developing techniques and algorithms to extract bibliographic data from biomedical journal articles, both digitized and Web documents, to build MEDLINE citations. The projects in this category are MARS and its various spin-offs and the Indexing Initiative.

Medical Article Records System (MARS)—The MARS production system currently generates bibliographic data for 500 articles per day, the remaining citations coming in as XML-tagged data directly from publishers. 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.

Center researchers continue to make changes to the MARS production system to accommodate new requirements from the NLM Indexing section. We modified three MARS software modules (Edit, Reconcile, and Upload) to support Unicode. Two conversion libraries (ASCII-to-Unicode and Unicode-to-ASCII) were created to read and write zone information to and from the MARS database. The ASCII data in zones is converted to Unicode for the operators conducting the Reconciling and Editing stages so they can see the diacritics as they appear in the article. The operators' output in Unicode is then converted back to ASCII for storage in the MARS database prior to eventual upload to DCMS, the database used by indexers to complete the indexing before citations are included in MEDLINE.

WebMARS—The goals of the Indexing 2015 Initiative are being addressed by CEB's development of two systems relying on WebMARS to assist both operators and indexers. The initial versions of both systems, WebMARS Assisted Indexing (WAI) and Publisher Data Review (PDR) are currently under test. PDR is designed to provide operators data missing from the XML citations sent in directly by publishers (such as databank accession numbers, NIH grant numbers, funding sources, check tags and PubMed IDs of commented articles) thereby reducing the burden on operators. In addition, incorrect data sent in by the publishers can be corrected by PDR. Correcting or augmenting the publisher data is a labor-intensive process, since the operators currently perform these functions manually by looking through an entire article to find these items, and then keying them in.

The second system, WAI, is for the indexers; it will help them search for terms in an article that correspond to biomedical terms in a predefined list. Again, indexers currently have to read through the entire article to confirm the occurrence of these terms, a labor-intensive process. WAI will automatically sift through the text and highlight these terms for the indexer to simply confirm and select, thereby reducing manual effort. In all, the systems consist of seven software modules.

ACORN—The ACORN (Automatically Creating OldMedline Records for NLM) system is intended to extract bibliographic information from 60 volumes of the printed *Quarterly Cumulative Index Medicus (QCIM)* from 1927 to 1956, to populate the OLDMEDLINE database. The design of the system is rooted in research in document image analysis and pattern matching techniques. A Web-based Reconcile module is being designed to allow operators to access, verify, and create QCIM citations through the Internet.

Validated Test Set for Document Image Analysis—By the end of May 2007, the Medical Article Records Groundtruth (MARG) database had 12,224 unique IP

visits from 96 countries. That is an increase of 2,536 visits since September, 2006. MARG provides TIFF images of biomedical journal articles and corresponding operator-verified OCR data, page segmentation and labeling results. This data set, obtained from the normal operation of the MARS production system, is used by researchers in the computer science and informatics communities conducting document image analysis to validate their own zoning and labeling algorithms.

Indexing Initiative—The Indexing Initiative project investigates language-based and machine learning methods for the automatic selection of subject headings for use in both semi-automated and fully automated indexing environments at NLM. Its major goal is to facilitate the retrieval of biomedical information from textual databases such as MEDLINE. Team members have developed an indexing system, Medical Text Indexer (MTI), based on two fundamental indexing methodologies. The first of these calls on the MetaMap program to map citation text to concepts in the UMLS Metathesaurus which are then restricted to MeSH headings. The second approach, a variant of the PubMed “Related Articles” algorithm, statistically locates previously indexed MEDLINE articles that are textually related to the input and then recommends MeSH headings used to index those related articles. Results from the two basic methods are combined into a ranked list of recommended indexing terms, incorporating aspects of MEDLINE indexing policy in the process.

The MTI system is in regular, increasing use by NLM indexers to index MEDLINE. MTI recommendations are available to them as an additional resource through the Data Creation and Maintenance System (DCMS). This year MTI recommendations are being augmented by the attachment of subheadings to some of the MeSH headings it recommends. Indexers will now have the option of accepting MTI heading/subheading pairs in addition to unadorned headings. In addition, indexing terms automatically produced by stricter version of MTI are being used as keywords to access collections of meeting abstracts via the NLM Gateway. These collections include abstracts in the areas of AIDS/HIV, health sciences research, and space life sciences.

Automatic Extraction of Outcomes from Published Documents—Originally part of the MDoT project, research was conducted toward automatically finding patient outcomes (e.g., the population under study) from MEDLINE citations using knowledge extractors that rely upon NLM Unified Medical Language System (UMLS) and tools. Our Extractor system identifies an outcome and determines whether a found outcome pertains to the topic of interest, the type of treatment studied, and the quality of the study. We evaluated the ability of the Extractor both to find outcomes in general, and to find high quality outcomes that answer specific clinical questions. Possible application areas might include clinical trials design,

EMR, and a patient-oriented service. Developed to provide access to the repository, a server accepts requests containing information about a patient (at present, current problems, age and medications) and searches MEDLINE via any of three search engines (Essie, PubMed, or the RIDeM database). The extracted information is sent to the client. The repository will be evaluated in a planned pilot study of supporting Evidence-Based Nursing Practice at the NIH Clinical Center.

Digital Preservation Research—In line with the NLM mandate to preserve the medical literature, the goal of this project is to investigate key issues related to the long term preservation of digital material, both for digitized documents and video. Our work in document preservation is more mature, and has resulted in a prototype System for Preservation of Electronic Resources (SPER). SPER is a flexible, modular system that demonstrates key functions such as ingest, automated metadata extraction (AME) and bulk file migration. AME is implemented for the extraction of descriptive metadata from scanned and online journal articles as well as obsolete NLM Web pages. Bulk file migration is implemented through an existing system, DocMorph. While these functions are developed in-house, for the necessary infrastructure capabilities in SPER we have incorporated into the system, and customized, the latest version (1.4) of MIT’s open source DSpace software. Our Java client GUI has the capability to do batch metadata extraction and ingest for journal article TIFF pages, online journal articles and NLM Web pages (HTML).

SPER, in an abbreviated form, is being used in the preservation of a new collection at NLM consisting of over 65,000 historical Food and Drug Administration court records (Notices of Judgment) from the early 20th century. Since the manual identification and entry of descriptive metadata from these records is labor-intensive, our focus is on their automated extraction. In collaboration with the curator for this collection, we identified more than a dozen metadata items which could be extracted automatically. After scanning the paper documents and performing OCR, the system then auto-zones the TIFF files, performs feature extraction and optimal feature selection, classifies text lines in the documents by a Support Vector Machine (SVM) classifier, and extracts the specific metadata by text pattern matching. Currently, the system extracts all metadata for a notice of judgment in less than 10 seconds on an ordinary Pentium machine.

Research was also conducted into video preservation. This effort centered on identifying an open file format such as Motion JPEG 2000 (MJ2) for archiving digitized video on disk media. This effort was guided by the findings of a one-day invitational meeting (at NLM) in 2005 with about 50 archivists and technologists involved in the long term preservation of video and film.

Information Resource Delivery for Care Providers and the Public

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

Clinical Research Information Systems—ClinicalTrials.gov provides the public with comprehensive information about all types of clinical research studies, both interventional and observational. The site has over 43,000 protocol records sponsored by the U.S. Federal government, pharmaceutical industry, academic and international organizations from all 50 states and in over 140 countries. Some 44% of the trials listed are open to recruitment, and the remaining 56% are closed to recruitment or completed. ClinicalTrials.gov receives over 19 million page views per month and hosts approximately 29,000 visitors daily. Data are submitted by over 3,400 study sponsors through a Web-based Protocol Registration System, which allows providers to maintain and validate information about their trials. ClinicalTrials.gov was actively involved in promoting the standards of transparency in clinical research through trial registration. These standards were communicated to a broad range of U.S. and international stakeholders via presentations and peer-reviewed publications. As a result of increasing awareness of the importance of trial registration, more than 9,000 new registrations were received over the last year. The results of an online evaluation aimed at identifying the needs of various groups of ClinicalTrials.gov users were applied in the design of a new user interface.

ClinicalTrials.gov continues to collaborate with other registries and professional organizations, working towards developing global standards of trial registration. In anticipation of a new mandate to develop a clinical trial results database, ClinicalTrials.gov convened two expert meetings to discuss scientific issues related to submitting and displaying published and unpublished clinical trial results, including the design of a feasibility study.

Genetics Home Reference—Genetics Home Reference (GHR) provides basic information about genetic conditions and the genes and chromosomes related to those conditions. This online resource provides a bridge between the public's questions about human genetics and the rich technical data emerging from the Human Genome Project. Created for the general public, particularly individuals with genetic conditions and their families, the site currently includes summaries for more than 230 genetic conditions, more than 395 genes, all the human chromosomes, and information about disorders caused by mutations in mitochondrial DNA. On average, nine new summaries are added per month. Additionally, new tutorial materials have been developed for the *Help Me Understand Genetics Handbook*. Topics in the latest *Handbook* include direct-to-consumer genetic testing, the

International HapMap Project, and an introduction to genome-wide association studies. GHR's usage, as measured by the number of hits per day, increased more than 60% in the past year, and the site is continually recognized as an important health resource.

"Profiles in Science" Digital Library—The Digital Library Research project investigates all aspects of creating and disseminating digital collections, including standards, emerging technologies and formats, copyright and legal issues, effects on previously established processes, protection of original materials, and permanent archiving of digital surrogates. Research focuses on long-term preservation of digital archives and new techniques for creating and accessing collections. The investigators are also exploring ways to achieve interoperability among digital library systems, using well-structured metadata, and varying "points of view" on the same data sets.

"Profiles in Science" showcases digital reproductions of items selected from the personal manuscript collections of prominent biomedical researchers, medical practitioners, and those fostering science and health. The content of "Profiles in Science" is created in collaboration with the History of Medicine Division of NLM, which processes and stores the physical collections. Most collections have been donated to NLM and contain published and unpublished materials, including manuscripts, diaries, laboratory notebooks, correspondence, photographs, journal volumes, poems, drawings, audio tapes and other audiovisual resources. The collections of Harold Varmus, Rosalind Franklin, Mary Lasker, and Sol Spiegelman were added this year. An additional 147 digital items composed of 599 image pages were also added to the 22 existing Profiles in Science collections. Presently the Web site features the archives of 23 prominent individuals:

*Christian B. Anfinsen *Virginia Apgar

*Oswald T. Avery *Julius Axelrod

*Francis Crick *Rosalind Franklin *Donald S. Fredrickson

*Edward D. Freis *Michael Heidelberger *C. Everett Koop

*Mary Lasker *Joshua Lederberg *Salvador E. Luria

*Barbara McClintock *Marshall W. Nirenberg *Linus Pauling

*Martin Rodbell *Florence R. Sabin *Wilbur A. Sawyer

*Fred L. Soper *Sol Spiegelman *Albert Szent-Györgyi

*Harold Varmus

The 1964-2000 *Reports of the Surgeon General*, the history of the Regional Medical Programs, and Visual Culture and Health Posters are also available on "Profiles in Science." The Society of American Archivists awarded "Profiles in Science" the Philip M. Hamer and Elizabeth Hamer Kegan award "in recognition of successful efforts to increase public awareness through the use of archival or manuscript materials." New additions to "Profiles in Science" were highlighted in *Science* magazine's "NetWatch" and *JAMA's* "Health Agencies Update."

Developers made several optimizations to software algorithms underlying the next generation

“Profiles in Science” Web server, resulting in significant performance improvements affecting the ingest of data, copying of images, and display of Web pages. They also developed methods to identify handwritten items and target them for creation of transcripts to improve searching.

User-Focused Portal

- **NLM Gateway**

The NLM Gateway is an ongoing production system that communicates with multiple NLM information resources. Since these resources are frequently updated, improved, and otherwise modified, the Gateway must change with them. The Gateway’s XML parser was replaced to fix a particular issue and upgrade to the latest technology. All TOXLINE data was consolidated on the SIS Web portal, combining toxicology citations from PubMed and TOXLINE Special. Gateway search translations were changed so the user would continue to retrieve only the TOXLINE special citations. Similarly, the DART (Developmental and Reproductive Toxicology) data was consolidated on the SIS Web portal, combining toxicology citations from PubMed and DART Special. Users continue to retrieve only the DART special citations.

FY2007 Activities included DTD changes to MEDLINE/PubMed and the NLM Catalog and automated indexing of the 100,000 meeting abstracts using 2007 MeSH, periodic updates of the UMLS and the MeSH mapping file and the HSRProj (Health Services Research) database. The Gateway system now communicates with the latest version of NCBI’s Esummary Eutility API. Developers upgraded server hardware and the Java and Visibroker software and are optimizing the system to support the MedlinePlus use of the Vivisimo search engine. Researchers continue to conduct usability testing.

Communication Infrastructure Research and Tools

The Lister Hill Center performs and supports research to develop and advance infrastructure capabilities such as high-speed networks, nomadic computing, network management, and wireless access. Other aspects that are also investigated include security and privacy.

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

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

In FY2007, the research team completed the substantial infrastructure required to test the scenarios, including the implementation of color algorithms to real time volume rendering of CT and MRI data and stereo display in the AG environment as well as the use of the technology in surgical education and planning. Virtual reality methods have been employed in a haptic environment allowing surgeons to rehearse liver operations. Several additional successful wide area wireless demonstrations of transmitting video and other patient data from ambulances using 3G and mesh cellular technology have been completed.

Scalable Information Infrastructure Initiative—The Scalable Information Infrastructure (SII) Project encourages the development of relevant health applications that are network aware and able to automatically adjust to changing network conditions and resources. Public next generation networks with SII capabilities hold the promise of adding advanced networking capacity to the tools available to healthcare professionals. Virtual reality and home health care may become realizable at reasonable costs based on next generation networking technology. Applications include wireless and geographic information system (GIS) techniques.

In order to carry out its program in Next Generation Networking Research, OHPCC has established the Collaboratory For Interactive Technology to serve as a laboratory, testbed, and demonstration area for NGN, SII, medical imaging, and high performance medical communications and networking applications. This facility is used to develop and test interactive imaging and communications protocols as applicable to telemedicine and distance collaboration under conventional, internet, and NGN/SII conditions. The facility also is configured to act as a hands-on demonstration site for remote interactive imaging, telemedicine, collaboration, and distance learning paradigms. Communications infrastructure connected to the Collaboratory, including an Internet2 Access Node capability, gives research staff the ability to collaborate with distant research colleagues, and at the same time, demonstrate much of the work being sponsored by the NLM Visible Human and NGN/SII programs.

The Collaboratory is used for a variety of research projects: to develop and test image analysis; to test manipulation and segmentation algorithms for use with the Visible Human data set in particular and in medical imaging in general; and to test advanced communication and collaboration technologies that complement those employed in OHPCC sponsored research. Research staff also use Collaboratory resources to demonstrate advances in telemedicine, imaging, and collaboratory techniques, including haptics, and to support ongoing NLM programs. Three-dimensional projection capability with access to remote 3D data sets is available.

The Collaboratory includes a fully functioning immersive Access Grid Room Node, with multiple projection capability, camera sources, and audio inputs. Several demonstrations were conducted in FY2007, including one for the Federal Communications Commission Panel on Rural Telemedicine. Experimental distant learning programs were conducted, working with SIS, NCBI, and remote collaborators at the Charles R. Drew University of Medicine and Science, the University of Puerto Rico Medical Campus and the University of Michigan. Resources were acquired to begin experimentation with IPHDTV, working with the University of Washington and the Internet2 Research Channel Working Group.

Videoconferencing and Collaboration—A new initiative was undertaken to experiment with uncompressed video over IP as well as high definition television. Compressed HD videoconferencing codecs were investigated using the H.264 technology that is compatible with and part of the revised H.323 standard. Digital Video Transport System (DVTS) technology was implemented, both as a standalone technology and as a component of the Access Grid. DVTS was developed by the WIDE consortium in Japan and is used by various Internet2 members to send uncompressed digital video at 30 megabits per second over IP. In addition, the Collaboratory became part of the Research Channel Working Group within the Internet2 and started acquiring components to implement uncompressed HD video at 1.5 gigabits per second. Collaboratory staff developed major enhancements for the Access Grid's (AG) shared browser and presentation tools. The use of open source browsers and presentation software as the basis for making the enhancements is being considered.

A distance learning program in collaboration with SIS, coordinator of the NLM Adopt-a-School Program, continued to provide on-site and distance education about varied health science topics and information sources to students at the King Drew Medical Magnet High School, affiliated with the Charles R. Drew University of Medicine and Science in Los Angeles. The NIH Office of Science Education participated again in the program and conducted several sessions on health science careers. Initial exploratory work was completed in trying to link a second school serving Native Americans in

Alaska. Each session was assessed as in previous years. As in the past, a statistical analysis of student ratings of teaching showed students rated the distant presentations higher than those on-site.

Methods for providing application sharing and image manipulation with low latency were identified and methods developed enabling the instructor at NLM to view each remote student's desktop. Successful pilot training sessions have been done with the University of Puerto Rico using the application sharing methods in conjunction with H.323 videoconferencing and with the University of Michigan with Access Grid (AG) technology.

A study of collocation as a factor in synchronous learning was completed with the University of Alabama at Birmingham. Students were tested on lectures delivered by videoconference and asked to collaborate on search tasks before being tested. They also were asked to rate teaching effectiveness of the lectures. Students were either physically collocated in a computer lab or meeting virtually in a multipoint videoconference. The data collected is currently undergoing analysis to determine if physical presence or absence affects performance, patterns of interaction, and perceptions of learning experiences.

The Center for Public Service Communication (CPSC) completed a successful pilot test of the use of video over IP to provide remote medical interpretation services at public health clinics in Duval County Florida. Valuable information about how the technology was and should be used was obtained, and the CPSC will move the technology to another public health environment in Florida.

Both the Web casts of the bi-monthly Washington Area Computer Assisted Surgery (WashCAS) Special Interest Group (and videoconferencing, added last year) continued. There is now two-way interaction between those attending the meeting in the Lister Hill Auditorium, where the presentations are made, and those in an auditorium at the Allegheny Hospital System in Pittsburgh. Attendees are able to obtain continuing medical education credits because of this linkage.

Access for Evidence-Based Medicine: PubMed for Handhelds—PubMed for Handhelds was publicly released in FY2005. Developed to facilitate evidenced-based medical practice with Medline access at the point of need via smartphones, wireless PDA's or portable laptops, PubMed for Handhelds requires no proprietary software and reformats the screen display as appropriate for the wireless handheld device being used. In support of evidence-based clinical practice, clinical filters feature easy access to relevant clinical literature. Newly developed resources allow searching Medline through text-messaging.

BabelMeSH—BabelMeSH is a multi-language and cross-language search tool for healthcare personnel who prefer to search MEDLINE/PubMed in their native languages.

Through international collaborations, including WHO Eastern Mediterranean Regional Office in Cairo, users can now search in Arabic, French, German, Italian, Japanese, Portuguese, Russian, Spanish, and English.

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

Advanced Networking—Center staff are exploring advanced optical networking techniques involving dynamic allocation of network resources using Quality of Service (QoS) characteristics in protocols such as RSVP. The purpose of the project is to gain experience with means of creating end-to-end channels with defined or reserved bandwidth and other qualities of service. This effort builds on the work of the DRAGON (Dynamic Resource Allocation in GMPLS Optical Networks) project, a collaboration of the University of Southern California Information Sciences Institute, the University of Maryland/Mid-Atlantic Crossroads (MAX), and George Mason University.

DocView Project: Tools for Using and Exchanging Library Information—This research area applies communications engineering and digital imaging techniques to document delivery and management, thereby addressing the NLM mission of providing document and information delivery to end users and libraries. An additional focus is to contribute to the bulk migration of documents for purposes of digital preservation, also part of the NLM mission. The active projects in this area are DocView, MyDelivery, DocMorph, and MyMorph.

- **DocView**

Originally released in 1998, this Windows-based client software is widely used to facilitate delivery of TIFF documents for interlibrary loan services. More than 18,600 users in 195 countries have downloaded it since it was released, with 600 new users in the last year. Many of these users are patrons of biomedical libraries, who receive electronic journals sent by the libraries via the Internet to DocView running on the patron's desktop computer. The DocView software has seen declining usage in recent years because the PDF file format, rather than TIFF, has become the popular choice for electronic document distribution.

- **MyDelivery**

The goal of this project, seen as a successor to DocView, is to develop a new collaborative tool to improve the delivery and exchange of medical and health information, especially information contained in very large files. MyDelivery is intended to enable biomedical researchers, administrators, librarians, physicians, patients, hospitals, and other health professionals to exchange medical information, regardless of the size of the electronic file in which it resides. This communication method is expected to be fast, easy, reliable, safe, and secure.

The MyDelivery project seeks to overcome three significant obstacles: (1) transmission of large electronic files (e.g., document images, digitized photographs and x-rays, sonograms, CT and MRI scans, and digital video) over the Internet; (2) sending files reliably and securely; and (3) complying with requirements of the Health Insurance Portability and Accountability Act (HIPAA). To solve all three problems, the MyDelivery project focuses on the development of server-based software running on a cluster of Internet-based servers, and the development of client software for use by collaborators. MyDelivery allows two client computers to exchange large files through an intermediary server via a user interface similar to email. Potential applications of this new technology include a secure way for researchers to easily exchange research data of virtually any size. Part of the development of MyDelivery has been to create a method of automatically recovering from communication failures due to reduced signal strength.

DocMorph—The DocMorph system is a Web site that allows the conversion of more than 50 different file formats to PDF, for instance, to enable multi-platform delivery of documents. Since 1999, DocMorph has served both browser-based users (16,600 to date: 2200 more than last year) and client-based MyMorph, of which there are more than 8,000 registered users. Many of the users are biomedical document delivery librarians. DocMorph also provides conversion of files to TIFF, text and synthesized speech. By combining text extraction with speech synthesis, DocMorph enables the visually impaired to use library information. It has been used by librarians for the blind and physically handicapped to convert documents to synthetic speech recorded onto audio tapes for patrons. DocMorph is available at <http://docmorph.nlm.nih.gov/docmorph>.

- **MyMorph**

MyMorph replaces the browser for DocMorph users who have large numbers of files to convert. Version 2 of MyMorph was released in April 2007

with the capability of mass converting PDF documents to TIFF. This became necessary because many document delivery library services, including that of NLM, use commercial document transmission systems that deliver TIFF, rather than PDF, documents. NLM and many libraries now use our current version of MyMorph to more easily deliver documents to their patrons.

Language and Knowledge Processing

Terminology Research and Services—LHNCBC research staff build and maintain the SPECIALIST Lexicon, a large syntactic lexicon of medical and general English that is released annually with the Unified Medical Language System (UMLS) Knowledge Sources. New lexical items are continually added using a lexicon building tool; the SPECIALIST lexicon contains over 360,000 records. The UMLS Lexical tools, including lexical variant generator (LVG), wordind, and norm are distributed with the UMLS as are text processing tools which analyze documents into sections, sentences, and phrases. The SPECIALIST lexicon, lexical tools, and text processing tools are released as open source resources and available under an unrestrictive set of terms and conditions for their use.

LexBuild is an evolving lexicon building tool designed to aid the lexicon building team by facilitating entry of lexical information and providing real time quality control. The SPECIALIST lexicon release tables are annually generated using the LexBuild tool. The SPECIALIST lexicon and tools are UTF-8 compliant and capable of dealing with non-ASCII characters. MMTx, the Java implementation of the MetaMap algorithm is a major application of the SPECIALIST lexical and text tools.

LHNCBC researchers have created a Java implementation of the Journal Descriptor Indexing (JDI) tool for release as part of the UMLS lexical tools. The JDI tools provide an element of context that can be useful for word sense disambiguation and other natural language processing tasks. LHNCBC research staff also develop and maintain the UMLS Knowledge Source Server (UMLSKS) that provides Internet access to the UMLS knowledge sources through application programs and a user interface. UMLSKS is updated quarterly to accommodate quarterly UMLS releases. A Grid/Web services implementation of the UMLSKS backend and an implementation of the user interface as a portal consisting of user-chosen “portlets” representing different parts and views of the UMLS data are being released with the 2008 UMLS release.

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, including the Unified Medical Language System (UMLS).

This year, the research team focused on biomedical information integration. First, in the domain of oncology, where the resources used to annotate data include the International Classification of Diseases for Oncology (epidemiology data), SNOMED CT (clinical data) and the NCI Thesaurus (research data), we studied the limitations of automatic mapping between these resources and checked their consistency. In a different domain, we enhanced RxNav, transforming it from an interface to the RxNorm drug terminology system to an environment integrating additional drug information, including pharmacologic actions and product label information. Finally, we contributed to a demonstration created by the Semantic Web for Health Care and Life Sciences interest group, consisting of the integration of about 20 resources in the domain of neurosciences.

As part of the research on ontology alignment, new methods were developed for validating matches and identifying mismatches between anatomical ontologies, including the Foundational Model of Anatomy and GALEN. We also carried out a study of criteria for evaluating biomedical vocabularies in caBIG on behalf of the National Cancer Institute. The method was applied to the evaluation of the NCI Thesaurus. Finally, the mapping of UMLS concepts to MeSH used in the Indexing Initiative was improved and it was adapted to mapping it to the International Classification of Diseases.

The research team continues to work on the creation of an ontology of relationships, as it is one critical element of a repository of biomedical knowledge supporting knowledge discovery and reasoning. We continue to participate in the progress of the Semantic Web for Health Care and Life Sciences and collaborate with leading ontology and terminology centers, including the National Center for Biomedical Ontology and the International Health Terminology Standards Development Organization.

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

MetaMap was recently improved by the addition of a word sense disambiguation (WSD) feature that chooses the best concept for a text phrase, guided by the surrounding context. The WSD method relies on Semantic Type Indexing and selects the concept having a semantic type most consistent with concepts in its immediate neighborhood.

The development of SemRep is based on viable strategies for effective natural language processing. Extending linguistic coverage is at the core of this research, and recent work focused on assertions about risk factors for disease in MEDLINE citations. Semantic predications produced by SemRep support continued work in biomedical information management. Application areas include automatic summarization and visualization of text from MEDLINE and ClinicalTrials.gov.

As a practical implementation, the Semantic MEDLINE application integrates PubMed searching, advanced natural language processing, automatic summarization, and visualization into a single Web portal. This program helps users manage the results of PubMed searches by normalizing core assertions in the citations retrieved. These normalized forms constitute computable knowledge accessible to further manipulation, including automatic summarization. The summarized output of Semantic MEDLINE is visualized as an informative graph with links to the original MEDLINE text. Convenient access is also provided to relevant knowledge resources, such as Entrez Gene, Genetics Home Reference, and the UMLS Metathesaurus. In collaboration with the National Heart, Lung, Blood Institute (NHLBI), Semantic MEDLINE is being enhanced to help conduct literature searches for the creation of clinical practice guidelines.

UMLS and Focused Clinical Vocabularies

Unified Medical Language System (UMLS)—The most recent quarterly release of the UMLS Metathesaurus contains over 1.3 million concepts (+18%) and 6.4 million concept names (+20%). The UMLS provides the only way for the US health care community to obtain SNOMED CT, the largest HIPAA standard clinical vocabulary, under the U.S. government license. Center staff have been principals in the successful transition to NLM's Office of Computer and Communications Systems (OCCS) aspects of the UMLS Metathesaurus project relating to: software and processes for the Metathesaurus Editing and Maintenance Environment (MEME); software for the editing and workflow management; and merging the Metathesaurus release data with other knowledge sources, and packaging the final product for distribution. Researchers continue to work with colleagues in the MEDLARS Management Section (MMS) of Library Operations, to ensure the quality of the Metathesaurus and help with knowledge transfer for customer support. Staff continue to be responsible for the development, testing and quality assurance of the

Metathesaurus installation and customization program, MetamorphoSys, and continue to enhance the product to support and anticipate the growing needs of the user community, such as investigating different output formats (TREF, XML), improving the search and display tool for customized subsets, performance, and user interface issues such as Section 508 compliance.

UMLS-CORE Project—The goal of this new project is to identify a CORE (Clinical Observations Recording and Encoding) subset of the UMLS to support consistent high level encoding of clinical information. The creation of the CORE subset will be based on real clinical data to ensure adequate coverage of commonly occurring clinical conditions. The CORE subset will promote and facilitate the use of standard clinical terminologies by helping users to identify the most frequently used portion of these terminologies. It will also enhance data interoperability by reducing coding variability. For terminology developers, the subset will help them identify gaps in coverage and focus their quality improvement efforts. Datasets that will be very helpful in creating the CORE subset have been obtained from the Mayo Clinic, Intermountain Health Care, the Regenstrief Institute, and the Hong Kong Hospital Authority.

Terminology Representation and Exchange Format (TREF)—The goal of TREF is to serve as a standard publishing format for single-sourced terminologies. Its use will facilitate the exchange of terminologies and the sharing of terminology related tools. TREF will also simplify the task of inversion of source terminologies into the UMLS editing environment. The relational and XML specifications of TREF have been finalized. In an experimental version of MetamorphoSys, the functionality of generating TREF output from the Metathesaurus has been developed and tested.

UMLS User and Usage Statistics—The Web forms for UMLS users to submit their annual reports were updated in FY2007. The annual report collection process started in February and ended in April. A total of 2,300 users submitted their annual reports. The data will be analyzed and compared with that from previous years. The information obtained will guide future developments of the UMLS.

UMLS Archive—Work has begun on constructing the UMLS archive, a data repository of all 35 UMLS releases, from 2007AB back to 1990. Both the original file contents (as released to the public) and a database version of the archive are available for research. The archive is expected to grow to about two terabytes. The physical archive, consisting of the distributed media, CD-ROM, DVD, documentation green books, etc., is also being gathered. Both ORF (Original Release Format) and RRF (Rich Release Format) versions of the data since 2004AA are being archived. Data from 1990 and 1991 was in a pre-ORF format and is being converted to ORF.

To ensure comparability, data modifications involve: (1) consistency in assignment of UMLS identifiers; for example, identical strings should have the

same SUI (String Unique Identifier); (2) consistent use of column and attribute names (e.g., CODE instead of SCD in MRSO); (3) recreating metadata (MRSAB, MRDOC, MRRANK, etc) where possible; (4) consistency in use of versionless SAB (source abbreviations) values; and (5) recomputing index files by using a single, consistent version of the LVG program. These “comparable” releases will exist separately from the original releases so analyses can be done on either version or across versions. Current work is focused on making ORF comparable releases. Creating RRF comparable data will be a greater challenge as the ORF lacks atomic information present and tracked in the RRF. The archive will provide opportunities for research in many areas. A few examples:

- **Concept history** – ways to visualize the additions, deletions and transformations in a concept’s life cycle. A meta-Metathesaurus to capture and identify changes in source level data aggregations (codes, concepts, descriptors) over multiple edit cycles within UMLS concepts.
- **Inter- and intra-release gathering of statistical data and visualization.** Examples are: characterizing high-level differences between releases, semantic profile of sources, content overlaps between sources, unique content from a source (i.e., areas where there is no overlap with other sources in the Metathesaurus). Some of this work is already underway.
- **A prototype viewer** is being built in Perl. It is archive-aware and understands and highlights cross-release differences.
- **Cross-release summary and metadata** are being gathered to help with this work. For example, there are tables with keys (CUI, SUI, STR, etc) and accompanying bit strings representing each release that they are present in.

Identifying and separating “core” functionality in MetamorphoSys from add-on or plug-in functionality will also facilitate research in many areas. Such a Software Development Kit (SDK) for MetamorphoSys can be released to the UMLS user community for collaborative research. This would ideally involve collaboration and harmonization with other extant UMLS object models, e.g., MetamorphoSys, MEME, UMLSKS. User and programmer documentation, UML diagrams and

reference implementations would be required before this can be made public. Once completed, the core MetamorphoSys maintenance and development can be transitioned to OCCS while research that uses the SDK continues in LHCBC.

Training Opportunities

Working towards the future of biomedical informatics research and development, the Lister Hill Center provides training and mentorship for individuals at various stages in their careers. The LHCBC 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 FY2007, the Center provided training to 46 participants from 13 states and nine countries. Participants worked on research projects including medical image processing, consumer health informatics, document analysis, grid computing, information retrieval, machine learning, medical illustration, micro-pathology, medical terminology research, natural language processing, medical ontology research, telemedicine, and ubiquitous computing. The program maintains its focus on diversity through participation in programs supporting minority students, including the Hispanic Association of Colleges and Universities and the National Association for Equal Opportunity in Higher Education summer internship programs.

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

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

David Lipman, MD
Director

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

NCBI is charged with providing access to public data and analysis tools for studying molecular biology information. Over the past 19 years, the ability to integrate vast amounts of complex and diverse biological information created the scientific discipline of bioinformatics. The flood of genomic data, most notably gene sequence and mapping information, has played a large role in the increased use of bioinformatics. NCBI meets the challenge of collection, organization, storage, analysis, and dissemination of scientific data by designing, developing, and distributing the tools, databases and technologies that will enable genetic discoveries of the 21st century.

The Center meets these goals by:

- Creating automated systems for storing and analyzing information about molecular biology and genetics with evidence in biomedical literature;
- 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;
- 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.

In FY2007, NCBI significantly expanded its breadth of public resources by developing the database of Genotypes and Phenotypes, or dbGaP. This database archives and distributes data from genome-wide

association studies, medical resequencing, molecular diagnostic arrays, and associations between genotypes and phenotypes. dbGaP and related NCBI databases are important elements in providing a powerful discovery system in which users will be able to glean information from many areas of genetics from a single online search.

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

Molecular Biology Information Resources

NCBI's molecular biology information resources are based on sequence repositories and built upon by assembly, curation, integration, and classification of sequence data. Information ranges from genetic sequence data to entire genomes, protein sequences and structures to chemical structures and assays, and clinical data paired with genotypes. An integral part of the molecular biology information is also made up of computer/user support and biology research in genomic analysis.

GenBank

GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. NCBI is responsible for all phases of GenBank production, support, and distribution, including timely and accurate processing of sequence records and biological review of both new sequence entries and updates to existing entries. Integrated retrieval tools allow seamless searching of the sequence data housed in GenBank and provide links to related sequences, bibliographic citations, and other related resources. Such features allow GenBank to serve as a critical research tool in the analysis 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 to NCBI from individual scientists and genome sequencing centers, and substantial staff and resources are devoted to the analysis and curation of sequence data. Thousands of sequence records are 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.

GenBank is comprised of two divisions, traditional nucleotide sequences and Whole Genome Shotgun (WGS) sequences, which are contigs (overlapping reads) from WGS projects. Annotations are allowed in these assemblies and records are updated as sequencing progresses and new assemblies are computed. The nucleotide database is divided as well into specialized components consisting of Expressed Sequence Tag (EST) and Genome Survey Sequence (GSS) records. The remaining records reside in the "CoreNucleotide" division.

The traditional and WGS divisions of Genbank combined contain over 97 million sequence records. The traditional GenBank division base count rose from 65 billion in August 2006 to 79 billion in August 2007. The WGS division contains 101 billion bases and 25 million entries as of August 2007. This year, 298 Whole Genome Shotgun (WGS) projects were released, with a total of 627 publicly available.

The Third Party Annotation (TPA) database created in conjunction with international partners EMBL and DDBJ supports third party annotation of sequence data already available in public databases. In order to be included in the TPA database the analyses must be published in a peer-reviewed scientific journal. TPA records are divided into two sections, TPA:experimental and TPA:inferential. TPA:experimental includes annotation of sequence data supported by peer-reviewed, wet-lab experimental evidence and TPA:inferential contains annotation of sequence data by inference where the source molecule or its products have not been the direct result of experimentation.

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. New full releases of GenBank are distributed every two months; daily updates are made available via the Internet and the World Wide Web.

NCBI is continuously developing new tools, and enhancing existing tools, to improve access to, and the utility of, the enormous amount of data stored in GenBank. Sequence data, both nucleotide and protein, are supplemented by pointers to abstracts and publishers' full-text documents as available. GenBank provides links to other Entrez databases and outside sources, such as biological databases and sequencing centers. The availability of such links allows GenBank to serve as a key component in an integrated database system that offers researchers the capability to perform comprehensive and seamless searching across all related biological data on the NCBI web site.

Sequin, NCBI's stand-alone submission tool, allows for updating as well as submission of a large number of GenBank sequences. The submission tool Sequin MacroSend allows submitters to upload a Sequin

file from their computer directly to the GenBank indexing staff where their submission is immediately given a temporary identification number. Guides for specialized submissions, such as genomes, batch sequences, and alignments, are also available. BankIt, another sequence submission software tool, is now in its thirteenth year of use. BankIt is useful for small submissions that can be uploaded directly to NCBI. Sequin and BankIt are updated continuously to improve ease of submission and to accommodate new data types and formats.

Genome Information Resources

NCBI has developed a suite of genomic resources, specialized tools and databases to support comprehensive analysis of entire genomes and sequence data. NCBI maintains an expanding collection of specialized, yet integrated, database repositories that collectively capture the biological relationships between genome sequences, expressed mRNAs and proteins, and individual sequence variations. Integrated repositories include BLAST, dbSNP, RefSeq, Map Viewer, Gene, Probe, UniGene, HomoloGene, and GEO. Links to outside information are also available such as Linkage and Physical Maps, TaxPlot, and chromosome-specific mapping information.

NCBI is also responsible for collecting, managing, and analyzing genomic data generated from the sequencing and genome mapping initiatives of sequencing projects as well as playing a key role in assembling and annotating genome sequences. Through the cooperation of sequencing centers and scientists worldwide, these efforts involve international coordination.

The Reference Sequence (RefSeq) database is a comprehensive, integrated, non-redundant set of sequences, including genomic DNA, gene transcript (RNA), and protein products for major research organisms. These standards serve as a basis for medical, functional, and diversity studies by providing a stable reference for gene identification and characterization, mutation analysis, expression studies, polymorphism discovery, and comparative analysis. The September 2007 full release of NCBI RefSeq records contains 4,167,224 proteins representing 4,646 organisms.

The Entrez Gene database offers a wide range of gene-specific data, integration with other NCBI databases, and enhanced options for query and retrieval from the Entrez system. Entrez Gene integrates information about genes and gene features annotated on RefSeqs and collaborating model organism databases, making it easier for researchers to find and interpret gene-specific information. The database provides information for more than three million genes from nearly 4,500 organisms.

Entrez Gene usability and content was enhanced in FY2007. Several new fields, displays, and properties were added, making it easier to retrieve and scan for specific information. For example, Entrez Gene now reports a gene's location on a chromosome on the

summary page and supports retrieval by chromosome position. Significant improvements were made to the automated process of identifying genes within genomic DNA sequences.

NCBI's Map Viewer is the primary resource for visualization of large genomes. Genes or markers of interest can be found by submitting a query against a whole genome, or by querying one chromosome at a time. Increased standardization of map features better supports cross-species comparison, and multiple-species queries with maps from other sequencing centers are also available. The results table includes links to a chromosome graphical view where a gene or marker can be seen in the context of additional data. The Evidence Viewer is a feature that provides graphical biological evidence supporting a particular gene model and the Model Maker allows users to build a gene model using selected exons. During FY2007, ModelMaker was expanded to 20 organisms, and users were given more options for viewing current and past genome builds in MapViewer for mouse and human and many other species.

A supporting database, UniGene, provides non-redundant clusters for the highly redundant sets of transcript sequences of expressed genes. UniGene continued to expand its coverage, adding 20 organisms during FY2007; it now includes genes for 90 animals, plants and fungi. A complement to Entrez Gene and UniGene is the HomoloGene database of homologous genes, covering 18 model animal and plant genomes. HomoloGene provides a reliable and comprehensive database of gene homologs, offering statistics on inter-species sequence and protein domain conservation. HomoloGene is linked to the genome-wide views available in MapViewer, and Entrez Gene, as well as to the information on gene expression found in UniGene.

The Protein Clusters database was launched in FY2007 and contains Reference Sequence (RefSeq) proteins from the complete genomes of prokaryotes, plasmids, and organelles clustered and annotated based on sequence similarity and protein function. These clusters are used as a basis for genome-wide comparison as well as to provide simplified BLAST access by using one reference sequence from each cluster to represent the entire cluster. Updated weekly, results are presented to the public via ftp releases and through Entrez. Currently the collection includes 2.7 million proteins from 540 genomes.

The Consensus CoDing Sequence (CCDS) database identifies a core set of human protein coding regions that are consistently annotated and of high quality. Annotated genes included in CCDS are given a unique identifier similar to the GenBank system (accession.version). CCDS was expanded this year to add data for mouse and to include an update for human. The CCDS IDs available for human grew from 14,795 (13,142 Genes) in March 2005 to 18,290 (16,008 Genes) in

February 2007. This year, 13,374 CCDS IDs (13,014 Genes) were added to the database for mouse.

The dbSNP database of genetic variation is a comprehensive catalog of common human polymorphisms. dbSNP continues to experience rapid growth, containing over 34 million submissions of human data that have been processed and reduced to a non-redundant set of 12 million refSNP clusters. Forty-two other organisms are represented in the SNP database, with 37 million submissions curated to 28 million refSNP clusters.

The Probe Database, also part of the Entrez system, stores molecular probe data together with information on success or failure of the probes in different experimental contexts. Nucleic acid probes are molecules that complement a specific gene transcript or DNA sequence and are useful in gene silencing, genome mapping, and genome variation analysis. The database contains almost 8 million probes as of October 2007. The RNA interference (RNAi) resource stores the sequences of RNAi reagents and experimental results using those reagents, such as extent of gene silencing and a variety of phenotypic observations.

Comparative Genome Data

NCBI genome resource guides provide information on diverse organism-related resources from multiple centers including sequence, mapping, and clone information. Genome resource guides for 31 organisms provide easy navigation to organism-specific BLAST and MapViewer pages and other NCBI resources, as well as to outside resources such as documentation, maps and sequence information, annotation projects, assembly updates, and comparative genomic sites. Genome specific resources pages released in FY2007 include zebrafish, horse, opossum, and platypus.

NCBI currently provides access to approximately 2,500 viral genomes, more than 550 complete bacterial genomes, and over 150 eukaryotic genomic assemblies. The Entrez Genome database increased significantly in FY2007, to more than 4400 taxa and almost 3,000,000 genes.

Current assemblies of the human and mouse genome (Build 37) have been improved upon with increased annotation, including access to alternate assemblies of the genomes previously unavailable and reference sequences for two alternate haplotypes. NCBI's past experience with curation of human and mouse genomes has benefited the annotation pipelines for many other organisms.

The Entrez Genome Project database is based on cellular organism-specific genomic information, including but not limited to genome sequencing, such as whole genome shotgun or BAC ends sequencing projects, large-scale EST and cDNA projects, and assembly and annotation projects. The database is organized into organism-specific overviews that function as portals from

which all projects in the database pertaining to that organism can be browsed and retrieved. This design allows the collection of disparate data that all refer to a single organism, conveniently displayed for easy access to all subprojects.

Fungal Genomes Central is a portal to information and resources about fungi and fungal sequencing projects from NCBI and the fungi research community. Plant Genomes Central is an integrated, web-based portal to plant genomics data and tools. It provides access to large-scale genomic and EST sequencing projects and high resolution mapping projects. The Viral Genomes website provides a convenient way to retrieve, view and analyze complete genomes of viruses and phages.

The Microbial Genome Annotation Pipeline was developed this year for annotation of prokaryotic genomes. Approximately 100 genomes have been annotated in-house and NCBI is working with 10 outside groups who submit their data. The program is available via the web and by request.

Specialized Databases and Tools

The Influenza Virus Resource aims to rapidly sequence flu viruses from samples collected from all over the world, from birds, pigs, and other animals, as well as humans. About 20,000 new influenza virus sequences were entered into the Influenza Sequence Database in FY2007. Approximately 11,700 of them were from the NIAID Influenza Genome Sequencing Project, the St. Jude Influenza Genome Project, Centers for Disease Control and Prevention, and Air Force Institute for Operational Health, and were processed by the NCBI flu annotation pipeline. Presentations and demonstrations have been given at national and international meetings on the NCBI Influenza Virus Resource and papers were published to illustrate its application to influenza research and discovery.

New functionalities were added to the influenza virus sequence annotation tool, which include the ability to annotate sequences of influenza B virus, multiple sequences as input, and drug-resistance prediction. The Flu Dataset Explorer provides an interactive tool for preliminary analysis of protein sequences from the NCBI Influenza Sequence Database or from a user's own file.

The NCBI Trace Archive, which holds the raw single-pass reads of DNA sequences generated from large-scale sequencing projects, grew to over 1.75 billion traces in FY2007, representing over 980 species and 2,365 influenza virus strains. The trace data can be scanned using a rapid nucleotide-level cross-species sequence similarity search tool called cross-species MegaBLAST. Using the visualization tools of the related Assembly Archive, researchers can examine an assembly of trace data from which finished genomic nucleotide

sequence has been derived and determine, for instance, whether a crucial nucleotide base change associated with a disease is supported by the sequence evidence.

A Short Read Archive (SRA) was created to archive information from next-generation technologies from companies such as 454, Solexa, ABI, and Helicos. The structure and volume of the new data required the development of a new archive separate from the Trace Archive. Most data is included in the batch or "run" information rather than individual traces. Currently, this information is available in the SRA FTP site.

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. GEO grew to 16 million gene expression profiles in FY2007. GEOarchive is a spreadsheet submission format for preparing large batch submissions to the GEO database. These files are suitable for loading into spreadsheet applications.

The Red Blood Cell database (dbRBC) combines the well-established Blood Group Antigen Gene Mutation Database (BGMUT) with tools and interlinked NCBI resources. dbRBC provides open, publicly available genomic, protein, and structural information linked to red blood cell antigens and clinical data related to red blood cells.

A highly specialized database of variation, the Database for the Major Histocompatibility Complex (dbMHC), contains variations found only in alleles of the Major Histocompatibility Complex, a highly variable array of genes that plays a critical role in determining the success of organ transplants and is largely responsible for an individual's susceptibility to infectious diseases. The database supports six major projects.

The NCBI Taxonomy project provides a standard classification system used by the international nucleotide and protein sequence databases. The taxonomy group continues to curate the rapidly growing Taxonomy database to include the names of species for which sequence has been submitted to the protein and nucleotide databases. Tools have been developed for representing alternate, externally maintained taxonomies and cross-mapping them with Taxonomy database entries as well as a database of biological material collections used to enhance links between NCBI sequence entries and the corresponding specimen entries.

The Taxonomy browser allows searches for information on an organism or taxon's lineage. Searches of the NCBI Taxonomy database may be made on the basis of whole, partial, or phonetically spelled organism names, with direct links to organisms commonly used in biological research. The Taxonomy system also provides a "Common Tree" function that builds a tree for a selection of organisms or taxa.

Chemical Information

PubChem is a key component of the NIH Roadmap project in Molecular Libraries and Imaging. The PubChem database is 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. Three databases -- PubChem BioAssay, PubChem Compound, and PubChem Substance -- contain information on millions of small molecules, including their activities, structures, and properties. The three PubChem databases each grew by a factor of two or more and the website is visited by 30,000 users per day. PubChem has been extremely successful in obtaining data from outside organizations; 67 groups have voluntarily contributed data, including 29 NIH-supported or government groups, 18 academic or non-profits, and 20 commercial groups.

PubChem BioAssay allows users to examine descriptions of each assay's parameters and readouts, with links to substances and compounds. The number of BioAssays grew to 587 in August 2007.

PubChem Compound provides unique chemical structures and validated chemical depiction information describing substances in the PubChem Substance division. PubChem Substance contains chemical substance records and associated information. Currently, PubChem contains nearly 11 million compound records and 20 million substance records.

PubChem contains an extensive set of links within its own sets of data as well as to Entrez databases. Many compounds have literature citations to PubMed as well as links to the proteins and/or genes representing a protein to which they bind. Links between substances and compounds characterize chemical constituents. Links between substances and bioactivity indicate a substance was tested in a particular assay. Compound-compound links correspond to similarity relationships. Compounds are searchable by chemical structure, chemical properties, and bioactivity. PubChem adds approximately 27,000 new substance records per day to the database, resulting in many new links within the Entrez system. Currently, there are more than 11 billion links resulting from 19.6 million substance records.

To keep up with the steady growth of PubChem, the need to rapidly process and load data into the archive is paramount. As such, the system, once able to load two million records per day, can now load four million. Other infrastructure improvements include creation of a completely new system for daily updates of the production data system using a backup and restore scheme. This change noticeably increased performance of our production data access systems and overall fault tolerance.

To better support automated query and download of non-Entrez accessible data, the PubChem Power User Gateway (PUG) was launched in FY2007. PUG is a programmatic interface to PubChem that can perform

structure searching and structure downloads. Also, support for PubChem depositions using IUPAC International Identifiers (InChIs) was added. New methodology was created for performing similarity searches of 3D chemical structures by using representative shapes. A web-based tool was made available for clustering chemical structures based on structure 2D similarity. Its features include the ability to collapse clusters by interactively setting up thresholds and selecting sub-clusters, or selecting compounds of interest and querying Entrez on the fly.

Also, in FY2007 a BioAssay deposition system was developed for tracking updates on assay descriptions, adding additional test results, and redefining test result reports. A new "BioAssay Preview" service allows both the depositor and curator to check submitted assay data before publishing in PubChem. A process was developed to identify related BioAssays based on activity overlap and protein target similarity. Functions were added to the BioAssay Summary service that allow views of multiple biological activity testing results and provide functionality for grouping test results based on same chemical connectivity.

A BioActivity Summary tool was developed that provides an overview of biological testing results and activity profile for chosen compounds. Also, a Structure-Activity Analysis tool was developed, which classifies compounds and assays simultaneously based on similarities of chemical structure, activity profiles and target similarities. This tool allows one to derive structure-activity relationships, or to examine activity profile of inhibitors across protein targets.

Protein Structure

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

MMDB continues to grow in size and now has 44,700 unique experimentally determined 3D structure records. MMDB is updated weekly, with the source PDB data checked for consistency in the purported chemistry, sequence and 3D co-ordinates. Annotation of taxonomic source for three-dimensional structures in MMDB, and of derived biomolecular sequence data, has been largely automated, eliminating the need for laborious manual validation. A data model for a new interaction-tracking database has been established. Major changes have been implemented in the calculation, processing, storage, and sharing of 3D-structure links and conserved domain-based annotation for protein sequences in NCBI's Entrez database. For that purpose, the NCBI structure group

maintains a mirror of the Entrez protein sequence dataset, which is updated hourly. For novel sequences, domain annotation is calculated immediately and shared with IEB's dataflow group. The new system has greatly reduced the latency time for computational annotation on protein sequence data.

NCBI's three-dimensional structure viewer, Cn3D, provides easy interactive visualization of molecular protein structures from Entrez. Cn3D also serves as a visualization tool for sequences and sequence alignments. What distinguishes Cn3D is its ability to correlate structure and sequence information. Cn3D features custom labeling options, coloring by alignment conservation, and a variety of file export formats that together make Cn3D a powerful tool for structural analysis.

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

Curation of the Conserved Domain Database has resulted in carefully curated representations of protein domain families, many of which are organized into hierarchies of related domains. More than 3,000 models curated at NCBI are available on public servers, with approximately 1,500 additional models at various stages in the curation pipeline. Curated hierarchies now contain up to several hundred individual models, reflecting the inherent functional diversity of large protein domain super-families. As of this year, redundant models are no longer suppressed in CDD. Research has been conducted on algorithms to name and describe those super-family clusters and on how to label domain-based sequence annotation as either specific (meaning that a particular sub-family and associated functional annotation has been identified with high fidelity) or as non-specific, indicating a coarse classification into a conserved domain super-family.

Calculation of RPS-BLAST results for proteins submitted to NCBI's CD-Search service has been standardized so that uniform database size parameters are used and do not change with the actual search set selected by the user or with database updates. This is a first step towards incremental updates of the CDD database, as it

will allow incremental updates of protein-CDD neighboring data.

CDTree, together with Cn3D, is the main application used by CDD curators. CDTree and Cn3D function as helper applications for Web browsers and enable users of the CDD resource to study protein domain families and achieve accurate functional classification of proteins. The program suite is a very powerful tool for building up and annotating representations of conserved protein and protein domain families. The CDTree v3.0 and Cn3D v4.2, the main applications used for the curation of the Conserved Domain Database, have been distributed to users of MS Windows platforms. A Macintosh version of CDTree/Cn3D has been developed and is currently being tested. A draft tutorial has been written for CDTree. A basic collection of computer code has been established to enable users of CDTree to record and store evolutionary relationships between domain models that entail fusion events, deletion events, and circular permutation.

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.

BLAST Suite of Sequence Comparison Programs

Comparison, whether of morphological features or protein and DNA sequences, lies at the heart of biology. BLAST has made it easier to rapidly scan huge sequence databases for similar sequences and to statistically evaluate the resulting matches. In a matter of seconds, BLAST compares a user's sequence with up to a million known sequences and determines the closest matches.

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

In FY2007, a BLAST home page redesign was released after months of beta testing. Improvements in the new design include a Recent Results feature providing links to recent BLAST search results. Saved Strategies allows users to save customized BLAST forms for later use, with a MyNCBI account. Two new BLAST databases were launched during FY2007 containing RefSeq transcript and genomic sequences arising from NCBI annotations of the human and mouse genomes, respectively. These databases combine two of the most informative types of sequence, omitting redundant non-

RefSeq sequences to reduce noise in the results. Searches of the two databases generate a new, interactive tabular display that partitions the BLAST hits by sequence type (genomic or transcript) and allows sorting by BLAST score, alignment length, or percent identity within the alignment.

Integration of Clinical, Genetic, and Environmental Databases

During FY2007, NCBI launched the database of Genotypes and Phenotypes (dbGaP) to archive and distribute data that correlates genetic characteristics with observable traits. The database will include data from a variety of study types, including large-scale genome-wide association surveys, medical sequencing, molecular diagnostic assays, and surveys of association between genotype and non-clinical traits. The ability to share genotype-phenotype association data among medical researchers will greatly accelerate efforts to understand the genetic determinants of disease. However, because the data also include individual-level phenotype and genotype information, its distribution raises serious issues of privacy. The dbGaP data are therefore offered at two levels of access, “open” and “controlled”, in order to strike a balance between the rapid release of non-sensitive data, and the oversight and investigator accountability required for datasets containing individual-level information. The interface to the open data provides access to an integrated set of study protocols, subject questionnaires, and summaries of measured parameters with corresponding clinical variables. This open-access data can be browsed online or downloaded from dbGaP without prior permission or authorization.

Controlled-access data can only be obtained if a researcher has been authorized by the appropriate Data Access Committee at the National Institutes of Health. Data available to authorized investigators include de-identified phenotypes and genotypes for individual study subjects, pedigrees, and some pre-computed associations between genotype and phenotype that are not available via the open-access interface to dbGaP.

dbGaP was launched in late 2006 with two studies: the National Eye Institute’s Age-Related Eye Disease Study (AREDS) and the National Institute of Neurological Disorders and Stroke’s Parkinson’s Disease Study. A number of additional studies were added during the year, including the first studies under the Genetic Association Information Network (GAIN); the Framingham SHARe study, which adds genotype data to the landmark Framingham Heart Study; and the Genetics of Kidneys in Diabetes (GoKinD) study. dbGaP is expected to grow rapidly, as NIH’s trans-NIH GWAS policy, issued in August 2007, names the database as the repository for all NIH-funded GWAS.

dbGaP was created as NIH was formulating its policy on GWAS studies, and discussions that went into forming the policy helped shape the ultimate structure of

dbGaP. In addition to obtaining advice from NIH and other parties involved in developing the policy, NCBI formed a dbGaP working group from its Board of Scientific Counselors to provide advice. The working group is charged with guiding NCBI in development and implementation of the database, including issues of data security, project scope, post-submission analyses, and presentation of information within dbGaP. The working group held their first two meetings in 2007.

Entrez Retrieval System

Entrez, the major database retrieval system at NCBI was originally developed for searching nucleotide and protein sequence databases and related MEDLINE citations. With Entrez, users can quickly and easily search gigabytes of sequence and literature data. A key feature of the system is the concept of “neighboring,” which permits a user to locate references or sequences that 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. Entrez currently consists of 35 integrated databases, including ones on sequences, taxonomy, genes, chemicals and biomedical literature. Entrez Global Query enables users to search all of the Entrez databases simultaneously in seconds, displaying the number of hits for each database on one page.

Discovery Initiative

NCBI has embarked on a new program to help users better explore the myriad of data contained in NCBI’s databases. The Discovery Initiative aims to improve the usefulness of information resources by drawing users’ attention to related data that may be of interest.

The technical groundwork has been laid by restructuring and replacing the backend for the Entrez retrieval system with a “portal” system. This system duplicates former functionality but also allows for more “on-the-fly” searches with up-to-date related resources. Various results coming from different portlets can occupy different regions of the display.

The system allows a more flexible and robust approach to development by enabling new features to be added and displays to be revised without risking problems that sometimes arise when new coding is introduced. Presentation of results is being improved to more readily draw users to related information in other databases as well as the one searched. The project also involves analyzing web logs and other usage metrics to incorporate knowledge of user behavior into database modifications, and improving search technology to better identify the most useful related information.

The Entrez system was officially switched to the portal architecture in April 2007 after months of testing on a beta site. All 35 Entrez databases have been

converted to implement search and retrieval via the portal system.

Literature Information Resources

PubMed is a web-based literature retrieval system developed by NCBI to provide access to citations and abstracts for biomedical science journal literature. PubMed is comprised primarily of journals indexed in NLM's MEDLINE database plus a limited number of journals that are formally outside the scope of MEDLINE. PubMed is the bibliographic component of the Entrez retrieval system and provides links to full-text journal articles at web sites of participating publishers, as well as to articles available in full text through NCBI's PubMed Central database.

In FY2007, PubMed celebrated a milestone 10-year anniversary and added its 17 millionth citation. Links to full-text journals increased from 5,156 in July 2006 to 5,880 in July 2007. Over 1.2 million OLDMEDLINE (1950 – 1965) records were converted to MEDLINE and updated in PubMed. User options were enhanced by making PubMed available as a search engine add-on on the search bar for FireFox 2.0 and Internet Explorer 7.0. PubMed automatic term mapping was updated to include investigator (collaborator) names. Unique Identifiers for PubChem records are now included in PubMed when publishers send the IDs in their electronic citation submission. Other PubMed enhancements include: Limits page revision, Comment/Corrections additions, additional publication date search options, and enhanced displays of Web page title.

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. The LinkOut for Libraries program continues to provide biomedical libraries the ability to link patrons from a PubMed citation directly to the full text of an article available through their library subscription program.

In FY2007, the number of organizations participating in LinkOut increased to almost 2,300, representing a growth rate of 15% over the past year. These include 1,800 libraries, over 260 full-text providers, and 240 providers of non-bibliographic resources such as biological databases. Together they provide links to 45 million Entrez records, including links to the full text of 48% of PubMed records from over 5,800 journals. In addition, participation to Outside Tool, a related service to LinkOut that also links users to external resources, jumped to more than 400 institutions, representing a 33% increase from last year. LinkOut enhancements include an automatic quality checking procedure to improve the quality of links, an RSS feed to improve communication with link providers, and migration of LinkOut to the new

backend (Portal) system to improve flexibility for future expansion.

PubMed Central (PMC) archives, indexes, and distributes peer-reviewed journal literature in the life sciences and provides free and unrestricted access to full-text journal articles. This repository is based on a natural integration with the existing PubMed biomedical literature database of indexed citations and abstracts. Back-issue journal scanning for online access has greatly enhanced the value of PubMed Central as a digital archive, providing material that was originally published as early as 1865.

As of July 1, 2007, more than a million articles and related items were available from the PMC journal archive. That is an annual increase of 65% for the second consecutive year. The additions have come from newly published material as well as from digitizing back issues that previously were only available in printed form. The number of life science journals for which PMC provides free and unrestricted access now numbers over 350. Unique users (as measured by unique IP addresses) approached 3.3 million per month in March 2007, which is more than double the level of March 2006. Articles deposited by NIH-funded researchers under the NIH Public Access Policy continue to be processed efficiently and made available to the public via PMC. A new feature was introduced to allow journal publishers to submit manuscripts on behalf of their NIH-funded authors.

A UK version of PMC sponsored by the Wellcome Trust, British Library and associated organizations began operating in January 2007. UKPMC is the first of potentially many PMC International (PMCI) centers; the system is based on software from USPMC. Agreements have been negotiated with dozens of publishers to allow their material from USPMC to be made available to UKPMC and future PMCI centers. The long-term goal of PMCI is to create a network of digital journal archives that share content, leading to a stable archive and better access to the biomedical literature.

The NCBI Bookshelf gives users access to the full text of over 75 textbooks in the clinical and research areas of biomedicine. In addition to textbooks from commercial publishers, the Bookshelf includes monographs, tutorials, and help documents authored by NCBI, NLM, and NIH staff. Thirteen new books were added to the database in FY2007 covering a wide range of topics, for example *MedMaster's Patient Drug Information* and *Webvision: the Organization of the Retina and Visual System*. Other existing books and collections were updated and expanded, including the Eureka Bioscience Collection, the HSTAT Collection, Molecular Imaging and Contrast Agent Database (MICAD), and Coffee Break.

Other key texts in the field of genetics include Gene Reviews, Genes and Disease, OMIM and OMIA. GeneReviews is based on Gene Tests, a genetics database

produced at the University of Washington and widely used by genetics counselors and physicians for its comprehensive genetic testing information and genetic disease descriptions. 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. OMIM is an electronic version of Dr. Victor McKusick's *Mendelian Inheritance in Man*, a catalog of human genes and genetic disorders. The database, produced at Johns Hopkins School of Medicine, contains over 17,850 records for more than 11,300 genes and 5,500 phenotypes. Functions are under development to complete the connections between descriptions of allelic variants to NCBI's dbSNP database. LinkOuts have been added to other locus-specific databases. Online Inheritance in Animals (OMIA), authored by Dr. Frank Nicholas, is a database of genes, inherited disorders and traits in animal species other than human and mouse. It contains textual information and references, as well as links to other relevant records from OMIM, PubMed, and Gene.

Research

NCBI's research program focuses on computational approaches to a broad range of fundamental problems in evolution, molecular biology, genomics, biomedical science and bioinformatics using theoretical, analytical, and applied mathematical methods. The Computational Biology and Information Engineering Branches are the main research branches of NCBI, with the latter branch concentrating on applications.

The basic research in the Computational Biology Branch (CBB) has strengthened NCBI applications and databases by providing innovative algorithms and approaches (e.g. BLAST, VAST, and the Conserved Domain database) that form the foundation of numerous end-user applications. Moreover, researchers in the group continue to make fundamental biological and biomedical advances by applying NCBI databases and software to sequences, genomes, and other large-scale data, and by developing experimental strategies in collaboration with NIH and extramural laboratories. CBB consists of 91 senior scientists, staff scientists, research fellows, postdoctoral fellows, and students.

The Computational Biology Branch's basic research includes over 25 projects listed in the NIH Intramural Annual Reports. Projects include new computer methods to accommodate the rapid growth and analytical requirements of genome sequence, molecular structure, chemical, phenotypic, and gene expression databases and associated high-throughput technologies. In other projects, computational analyses are being applied to particular human disease genes and the genomes, evolution, and functional biology of several pathogenic bacteria, viruses and other parasitic organisms.

Another research focus is the development of computer methods for analyzing and predicting macromolecular structure and function. Recent efforts include improvements to the sensitivity of alignment programs, analysis of mutational and compositional bias influencing evolutionary genetics and sequence algorithms, investigation of gene expression regulation and other networks of biological interactions, analyses of genome diversity in influenza virus and malaria parasites related to vaccine development and evolution of virulence, evolutionary analysis of protein domains, and new mathematical text analysis and retrieval methods applicable to full-text biomedical literature. New research projects also were initiated in support of the PubChem molecular libraries project, a major component of the NIH Roadmap. CBB scientists continue to collaborate with other NIH institutes and several extramural and worldwide institutes on many of these research projects.

The high caliber of the work of the CBB group is evidenced by the number of peer-reviewed publications: over 80 publications this year with more in press. CBB scientists gave numerous presentations and posters at scientific meetings, molecular biology research companies, universities, and other government institutes. Presentations were also given to visiting delegations, oversight groups, and steering committees. CBB's weekly lecture series also serve to share information about research projects as well as hosting many guest speakers. The NCBI Postdoctoral Fellows program provides computational biology training for doctoral graduates in a variety of fields.

A Board of Scientific Counselors, comprised of extramural scientists, meets twice a year to review the research and development activities of NCBI. The BoSC's twenty-ninth meeting was held in April 2007.

Bioinformatics Training and Support

Outreach and Education

The educational component of NCBI's mission is recognized both by its advisory board (the Board of Scientific Counselors) and internally as an essential activity to ensure that the research community is aware of all NCBI services and is trained to make effective use of those services. In addition, feedback from the user community is vital to providing services that are meeting the actual research needs of users and for identifying future requirements. The audience for NCBI databases is very broad and the resources are used not only by molecular biologists and health professionals, but students, educators, librarians, and science writers, as well as the general public.

NCBI continues to expand its outreach and education programs to increase awareness of its myriad of public databases and specialized tools and services. Over the past year, NCBI staff exhibited at 21 scientific conferences; presented at numerous 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 increasing demand for education and training in the continually growing array of NCBI products and services, the course "A Field Guide to GenBank and NCBI Resources" is taught at NIH and throughout the United States upon request. The course consists of a three-hour lecture, a two-hour hands-on practicum, and optional one-on-one sessions. An expanded two-day course entitled "Enhanced Field Guide" provides extended and in-depth coverage of BLAST, structure, and genomic resources, and includes an advanced hands-on session. The eleven-member teaching staff presented 40 courses to over 3,000 people in the past year.

New classes added in FY2007 include Field Guide II, which provides update sessions for sites hosting the standard Field Guide course and includes information about more specialized databases. The course was presented four times in FY2007 to 120 participants.

"Gene Expression Resources" and "Principles of PubChem" also debuted in FY2007. The former classer was presented four times to over 100 participants and features the UniGene, Probe, GenSat and GEO databases. "Principles of PubChem" includes lectures and computer workshops on effectively using the databases, search services and analysis tools that constitute the NCBI PubChem system. In FY2007, the course was taught 10 times to over 200 participants.

Education: Mini-Courses and Lecture Presentations

NCBI offers 11 bioinformatics mini-courses at NIH and outside institutions to provide a practical introduction to various resources. The two-hour courses are both problem-based and resource-based and include review and hands-on sessions. This year, over 140 mini-courses were offered to approximately 3,800 participants.

Several NCBI MiniCourses were given at NLM to students in Puerto Rico using distance-learning technology. These new sessions, given to groups of 10 students each, were conducted in collaboration with staff at NLM's distance learning collaboratory.

Education: Technical Workshop Series

The PowerTools NCBI Technical Workshop Series consists of two courses lasting two-to-four days each. "NCBI Power Scripting" includes lectures and workshops on effectively using the NCBI Entrez Programming Utilities (e-Utills) with scripts to automate search and retrieval operations across the entire suite of Entrez databases. The "NCBI 4-Pack" course provides information on practical applications of bioinformatics

resources. Each course is offered quarterly at the NCBI Training Center.

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 NCBI bioinformatics tools provided to the research community. 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 fifteen research institutes at NIH, conducting eight nine-week training programs since the program began in 2001, and one six-week program in the past year. Thirty-nine update sessions and two special topic sessions for the institute representatives have also been held. One-on-one consultations with NCBI faculty are available on an ongoing basis for NIH scientists in the NCBI Learning Center, the NIH Library, and the NCI-Frederick Cancer Research and Development Center.

The NCBI Learning Center (38A, B2 level) is available for the nine-week CoreBio training and update sessions, the 11 NCBI Minicourses, and the one-on-one consultations. An NCBI Seminar and Training Room with individual computer stations (also located in 38A, B2 level) is used for larger NCBI workshops, such as the four-day PowerScripting and the two-day four-pack workshops, and for larger groups.

Education: Extramural Educational Collaborations

The sixth "NCBI Advanced Workshop for Bioinformatics Information Specialists" was held during FY2007. The educational collaborators ("Educollab," who work with NCBI to develop and teach the course) and the course alumni offer a variety of year-round services at their universities, including workshops on NCBI resources, individual research consults and support, and web portals. Many of the workshops are based on materials presented in the Advanced Workshop, thereby extending the impact of those materials. Together, the collaborators and course alumni form the growing Bioinformatics Support Network (BSN), a group supported by NCBI that has been established for the purpose of communication and continuing education among members.

In FY2007, eight new members joined the BSN as a result of the workshop. BSN members, and in particular Educollab members, have significant influence on the growing movement of medical libraries to establish

high quality bioinformatics support programs. The Educollab group continued the process of converting instructional slides for the advanced workshop from HTML format to PowerPoint format, and the Educollab coordinator continued re-engineering parts of the website to adapt to the new format. This approach achieves a balance between facilitating revisions of the course materials by a diverse group of contributors while preserving the essential components of the course web site to facilitate access by library staff nationwide.

The regional training program for the three-day introductory course continued in FY2007. Eight of NCBI's educational collaborators served as regional instructors to present the course at five locations across the country. A total of 92 individuals attended these regional courses and the on-site course offered at NLM.

The purpose of both the introductory and advanced workshop, as well as the Educollab program, is to train the trainers, who then provide assistance with NCBI resources to thousands of end-users across the country. The web-based materials that exist for both courses also serve as a reference for those who have taken the courses. Members of the Educollab group developed a symposium proposal on "State of the art: successes, challenges, and realities of library involvement in bioinformatics." The proposal featured presentations by leaders from the research, clinical, and consumer health communities on the genetics/molecular biology information needs of each community, followed by panel discussions among library leaders on management issues relevant to the successful delivery of information support services to each community. The proposal was submitted to the Medical Library Association Continuing Education Committee for consideration for the 2008 annual meeting. A new Educollab member and the Educollab coordinator, working in an advisory capacity, began to lay the foundation for a study to assess the awareness, current uses, and needs of the clinical community with respect to

genetic/molecular biology information resources. In a parallel effort, the Educollab coordinator also began to collect and organize a list of approximately 25 NCBI resources that have current or potential relevance to clinical practice, clinical research, or molecular medicine.

Outreach: User Guides for NCBI Resources

NCBI has continued to develop a comprehensive set of fact sheets that describe the services and databases offered by NCBI. These fact sheets and guides are available for printing as part of the "About NCBI" site. In addition, a number of other informational and educational resources are available on the NCBI Web site. Interactive tutorials may be found for a number of databases and search and retrieval tools, including Entrez, PubMed, Structure, and BLAST. A new project to produce 3-minute Flash-based tutorials covering key NCBI databases and services was launched. To date 10 such tutorials have been completed with over 30 additional tutorials planned.

Several "Announce" e-mail lists give users the opportunity to receive information on new and updated services and resources from NCBI. For example, "NCBI Announce" provides updates on all NCBI services and education while "Books Announce" provides information regarding the Bookshelf. RSS web feeds are also available for updates and announcements on many of NCBI's databases.

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. Access to the newsletter and its archives via the NCBI Web site has increased dramatically as more people have become aware of its availability online.

EXTRAMURAL PROGRAMS

Milton Corn, MD
Associate Director

The NLM Extramural Programs Division (EP) continues to receive its budget under two different authorizing acts: the Medical Library Assistance Act (MLAA), which is unique to NLM, and Public Health Law 301, which covers all of NIH. The funds are expended mainly as grants-in-aid to the extramural community in support of the Library's research and training goals in informatics and knowledge management. Review and award procedures conform to NIH policies.

EP awards six categories of grants, all of which pertain to biomedical computing, informatics, and the management and dissemination of biomedical knowledge. Some grant programs, such as Grants for Scholarly Works in Biomedicine and Health, are unique to NLM, while others are multi-Institute initiatives or interagency partnerships. Each year, NLM makes new and/or continuing awards in the following categories:

- Research grants for basic and applied research in biomedical informatics, information sciences, bioinformatics and public health informatics;
- Research Resource grants, to support unique research resources for informatics and bioinformatics;
- Resource grants for knowledge management, often involving online information tailored for specific audiences;
- Training, fellowship and career development grants, for informaticians and informationists;
- Scholarly Works and Conference grants, to enhance scientific and scholarly communication; and
- SBIR (Small Business Innovation Research)/STTR (Small Business Technology Transfer) grants, to support informatics innovations in small businesses.

The EP Web site <http://www.nlm.nih.gov/ep/funded.html> lists grants awarded since 1997, with links to abstracts provided in the NIH CRISP (Computer Retrieval of Information on Scientific Projects) database and, when available, links to project Web sites.

A significant feature of EP activities in FY2007 has been a focus on improving success rates for NLM applicants and on helping NIH meet its goals for new investigator awards. In FY2006, EP focused on counteracting falling success rates, both by closing programs and by suspending participation in selected cooperative ventures. Those closures, combined with a special infusion of funds that resulted from the NIH Reauthorization bill, increased the pool of funds available for the core programs. The overall success rate for NLM

applicants (all types of grant programs) climbed from 10.5% in 2006 to 18.7% in 2007. Success rates for R01 research grants climbed from 10.1% in 2006 to 17.1%.

A new feature of FY2007 EP grant activities was the establishment of target grant commitment levels to support NIH goals for improving career development of young scientists. NIH defined four classes of applicants as high priority. For two of them, target award levels were set: applicants for K99/R00 Pathway to independence awards (NLM target was five), and new investigators applying for their first R01 research grant (NLM target was nine). Two other categories with no quotas were: (1) previous new investigators coming back for their first competitive renewal; and (2) vulnerable funded investigators for whom NIH is the sole/most significant source of research funding. EP was able to achieve its target of nine for new investigator awards, but only three K99 applications that received fundable scores were assigned to NLM. For the other two priority categories, those without quotas, NLM relaxed the payline level as needed for investigators in category 1; no applications fell into the second category.

Success Rates of Grant Applicants

Table 11 shows success rates from 2004 to 2007 for NLM's core grant programs. The success rate for research grants increased to 16%. As they did in 2006, award decisions in 2007 continued to favor early career support, although the success rate dipped slightly due to transitioning from the Early Career Development Award (K22) to the Pathway to Independence Awards (K99/R00). These applicants are most often trainees from NLM's informatics training programs who are now moving into independent research careers. Success rates are computed by dividing the number of awards by the number of applicants in a fiscal year.

Table 11

Success Rate, Core NLM Grant Programs, FY2004-FY2007

Grant Type	FY2004	FY2005	FY2006	FY2007
Research	15%	9%	10%	16%
Knowledge Management	18%	12%	8%	16%
Scholarly Works	26%	19%	15%	17%
Career Transition	32%	36%	37%	29%

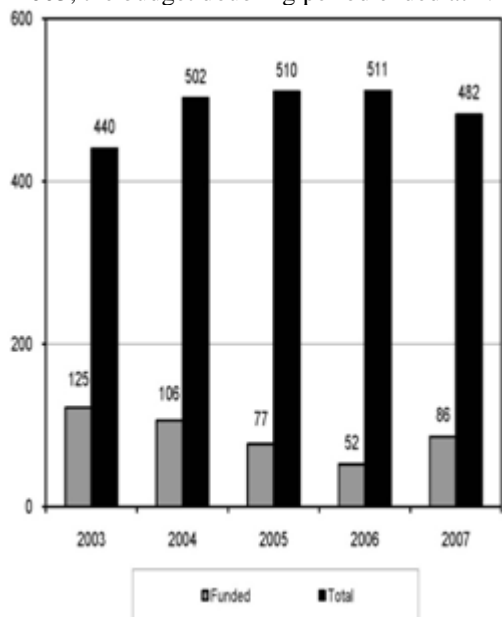
Two major factors continue to shape success rates at NIH: the large number of applications received; and the fact of essentially flat budgets. Table 12 shows the steep increase in the number of applications received for a five-year period, a trend that began to reverse itself in 2007. If this downward trend continues, its positive effect on

success rates will be compounded as completed awards release funds for new projects. However, application trends toward multiple principal investigator grants and larger consortium awards, as well as normal inflationary trends in research costs, make it difficult to lower average cost of awards. The combination of flat or even reduced budgets as applications increase inevitably reduces success rates to a level that could discourage the applicant community, and, over time, diminish academic support and health of the informatics professoriate.

Table 12

Applications and Awards, FY2003-2007

(In 2003, the budget doubling period ended at NIH.)



Research Support for Biomedical Informatics and Bioinformatics (PHS 301)

Extramural research support is provided through a variety of grant mechanisms that support investigator-initiated research. EP’s research grants support both basic and applied projects involving the applications of computers and telecommunication technology to health-related issues in clinical medicine and in research.

Research Grant Program

The R01 research grant program at EP has traditionally had two “branches:” biomedical informatics (computing and knowledge management for clinical care, health services research, and public health), and bioinformatics (computing and knowledge management for basic biomedical research areas such as systems biology, genomics and proteomics). In 2007, public health informatics and translational informatics that links

phenomic and genomic information are emerging as new branches. While many of EP’s research grant applications come from the biomedical informatics research community, an increasing number come from computer science, engineering and basic biomedical science fields.

As of FY2007, all R01 research grant applications are received electronically. In addition, starting November 2006, EP began using the Parent R01 omnibus program announcement (PA-07-070), replacing the NLM-specific program announcement (PA-06-094). This change of PA did not affect EP’s scope of interest and priorities for informatics research.

- 140 reviewed R01 applications (119 in FY2006)
- 24 awarded R01 applications (12 in FY2006)

Small Grant Program

In 2003, EP began offering the R03 small research grant, which provides modest support for “start-up” research projects and pilot studies.

- 36 reviewed R03 applications (27 in FY2006)
- Four awarded R03 applications (three in FY2006)

Exploratory/Developmental Grants

EP’s R21 exploratory/developmental grant supports high risk/high yield projects, proof of concept, and work in new interdisciplinary areas. This grant mechanism seemed better suited to informatics/engineering proposals than are the standard R01 research grants, which are judged in terms of hypothesis-based science.

- 42 reviewed R21 applications (44 in FY2006)
- Six awarded R21 applications (3 in FY2006)

Resource Grants for Biomedical Informatics/Bioinformatics

The P41 program announcement to support scientific research resources was deactivated in FY2005, due to the high cost and long duration of these resources. Only the five existing P41 awardees remain eligible to apply to the program for continuation funding.

- Two reviewed P41 applications (two in FY2006)
- Two awarded P41 applications (one in FY2006)

Conference Grants

Support for conferences and workshops (R13) is offered by almost all the Institutes and, for NLM, is intended to provide relatively small amounts to scientific communities convening workshops and meetings in focused areas of biomedical informatics and bioinformatics. Applicants must obtain approval from EP program staff before they can apply. Of the three applications reviewed in FY2006, one was funded. Only electronic applications are now accepted.

- Five reviewed conference grants applications (three in FY2006)
- Two awarded conference grants applications (one in FY2006)

IAIMS Testing and Evaluation Grants

Although a component of the IAIMS suite of grant programs, the Testing and Evaluation grant is considered a research grant and funded through PHS 301. As of FY2007, only electronic applications are accepted.

- Seven reviewed IAIMS Testing & Evaluation applications (eight in FY2006)
- Zero awarded IAIMS Testing & Evaluation applications (one in FY2006)

Small Business (SBIR/STTR)

All NIH research grant programs allocate fixed percentages of available funds every year to Small Business Innovation Research (SBIR) and Small Business Technology Transfer Research (STTR) grants. These projects may involve a Phase I grant for product design, as well as a Phase II grant for testing and prototyping. SBIR and STTR applications are reviewed by CSR. In FY2007, 33 applications were “unscored” (out of 45 applications), indicating reviewer assessment that they were not competitive for funding.

- 45 reviewed small business applications (62 in FY2006)
- Four awarded small business applications (two in FY2006)

Resource Grants (MLAA)

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

Knowledge Management and Applied Informatics Grants

This program is a refocused continuation of NLM’s former Information Systems Grant program. The new program emphasizes knowledge management, and application projects that “translate” informatics research into practice. As of FY2007, only electronic applications are accepted.

- 64 reviewed KM & AI applications (65 in FY2006)
- 10 awarded KM & AI applications (five in FY2006)

Integrated Advanced Information Management Systems (IAIMS) Planning Grant

Applications to the IAIMS Planning Grant dropped sharply in FY2007, as expected, due to the deactivation of the IAIMS operations grant program. As of FY2007, only electronic applications are accepted.

- 10 reviewed IAIMS Planning applications (21 in FY2006)
- Two awarded IAIMS Planning applications (three in FY2006)

Integrated Advanced Information Management Systems (IAIMS) Operations Grants

The IAIMS operations grant program was deactivated in FY2006, due to poor success rates, applicant confusion about the purpose of the program, and budget constraints. One application remained “in the pipeline” but was not funded. NLM’s KM & AI grants can provide a small amount of funding to support IAIMS operations projects.

- One reviewed IAIMS Operations applications (11 in FY2006)
- Zero awarded IAIMS Operations applications (zero in FY2006)

Grants for Scholarly Works

The Scholarly Works grant program continues to be very popular. Because NLM, alone among the Institutes, is authorized to support book publications, this program continues to play a key role in important areas of biomedical scholarship, particularly in the history of science and medicine. As of FY2007, only electronic applications are accepted.

- 58 reviewed Scholarly Works applications (68 in FY2006)
- 10 awarded Scholarly Works applications (10 in FY2006)

Training and Fellowships (MLAA)

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

NLM’s University-Based Biomedical Informatics Research Training Programs

Five-year institutional training grants support pre-doctoral, post-doctoral, and short-term trainees in 18

programs across the country. (See Table 13.) Seven of these programs were funded for the first time in 2002, while 11 are continuations of previously funded programs. Three of the 18 are consortial programs that support training at multiple institutions.

This program is re-competed every five years. The latest applications were received in March 2006, and awards were made in FY2007. Eighteen awards were made, two of them to new programs, at the University of Colorado and at the University of Virginia. Two existing programs at the University of Minnesota and the Medical University of South Carolina will no longer receive funds from NLM. By NLM custom, all trainees already matriculated into the programs facing sunset will be supported until completion of their training.

Collectively, the programs emphasize training in health care informatics (14 programs), bioinformatics and computational biology (14 programs), Clinical research translational informatics (13 programs), and public health informatics (10 programs). EP receives some co-funding from the National Institute of Dental and Craniofacial Research (NIDCR)-supported training in dental informatics at Pittsburgh. The National Institute of Biomedical Imaging and Bioengineering (NIBIB) provided funding for special one-year projects for four

pre-doctoral and two postdoctoral students located at three of NLM's training sites (Stanford, Wisconsin and Virginia). The National Heart, Lung and Blood Institute provided special one-year awards for two pre-doctoral students, one each at Pittsburgh and Vanderbilt.

- 36 reviewed T15 training grant applications (28 in FY2002)
- 18 awarded T15 training grant applications (18 in FY2002)

Table 13

T15 Trainees for FY2007

<i>Trainee type</i>	<i>FY2007</i>
Pre-doctoral	121
Post-doctoral	93
Short-term	16
Total	230

Table 14

Map of EP-Supported Training Programs



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

Every summer, all NLM-supported trainees attend a national informatics training conference. On June 26–27, 2007, the meeting took place at Stanford University, in Palo Alto, CA. Research projects were presented in plenary and semi-plenary sessions by 42 informatics trainees. An additional 37 trainees presented posters at the meeting. There were 340 attendees, including: directors, faculty, and staff from all NLM-funded informatics training programs; holders of NLM individual fellowships; faculty and trainees from the Department of Veterans Affairs (VA) informatics training sites; and NLM staff and guests.

In 2005, NLM/EP and the Robert Wood Johnson Foundation (RWJF) formed a partnership to lend increased emphasis to training in public health informatics. Through a \$3.6 million grant from the Foundation to EP (through the Foundation for NIH), four existing training sites received supplemental awards to develop formal training tracks in public health informatics and to support trainees in these tracks. The four selected sites were Columbia, Johns Hopkins, Utah, and Washington. Trainees in this initiative meet twice each year for special “cohort” experiences supported by RWJF. The first meeting was held at the AMIA meeting in November 2005, with subsequent meetings held twice a year in conjunction with the NLM training conference and the fall AMIA meeting. NLM staff work with RWJF contractor Public Health Informatics Institute (PHII) to plan these events. At the June 2007 NLM Training Conference, the NLM/RWJF cohort event was held as a pre-meeting, on June 25. The meeting agenda included introducing the PHII staff and approach, plus two speakers on public health preparedness topics.

Individual Fellowships

Training for Informationists

The F37 Informationist fellowship is EP’s only remaining individual fellowship program.

- Four reviewed F37 informationist applications (two in FY2006)
- One awarded F37 informationist applications (one in FY2006)

Career Support (MLAA)

K99/R00 Pathway to Independence

In January 2006, NIH announced a new career transition program, the NIH Pathways to Independence (PI) award (K99/R00), which combines a two-year mentored period with a three-year unmentored research period (the latter being similar to NLM’s former K22 program). Although applications to this new program are not restricted to NLM’s informatics trainees, applicants from our own programs are particularly welcome.

- Four reviewed K99/R00 applications (two in FY2006)
- Three awarded K99/R00 applications (zero in FY2006)

Early Career Development Awards

The K22 Early Career Development program was suspended in FY2006 and replaced by the K99/R00 Pathway to Independence awards. Three grants were “in the pipeline” in FY2007, but none were funded.

- Three reviewed K22 applications (19 in FY2006)
- Zero awarded K22 applications (seven in FY2006)

Loan Repayment Program

EP participates in the NIH loan repayment program by identifying applications from informaticians involved in research related to clinical medicine. These applications are reviewed for merit by a Special Emphasis Panel. For FY2007, EP funded five of 12 applications.

- 12 reviewed Loan Repayment Program applications (16 in FY2006)
- Five awarded Loan Repayment Program applications (six in FY2006)

Trans-NIH Projects

EP and NIH Roadmap Activities

In 2007, NIH Roadmap activities are being managed by the new Office of Portfolio Analysis and Strategic Initiatives (OPASI), which also administers the NIH Common Fund and decides on new initiatives that draw upon the Common Fund. A plan for Roadmap 1.5 was issued in May 2007, citing four new areas of action: Microbiome and Epigenetics (five-year programs); Protein Capture Tools and Phenotyping Services/Tools (staged implementation programs); and Genetic Connectivity Map (pilot study). EP does not have a role in these specific topic areas. However, a new Roadmap Coordination Committee on Bioinformatics has been formed, to be chaired by the NLM director. The EP director serves as co-chair of an Informatics committee for the Roadmap Clinical Translational Science Award (CTSA) Centers.

NCBC and BISTI

Although conceptually related to the NIH Biomedical Information Science and Technology Initiative (BISTI) program, the National Centers for Biomedical Computing (NCBC) program is distinct and funded through NIH Roadmap grants under renewable cooperative agreements for its first 5 years. The funds for the first five-year came primarily from the NIH Roadmap initiative, but several

ICs, including NLM agreed to contribute additional funds to the pool.

NLM provided an additional \$800,000 per year for five years. This initial commitment ends in FY2008. An external evaluation of the NCBC program's accomplishments was held in 2007. The evaluation committee recommended that the program continue for another five years. However, as the Roadmap Common Fund is now seen as a "start up" fund, the extent to which Roadmap funds will be available for the second five-year period has not yet been announced. Planning is underway for an open competition to provide funding for at least one additional five-year cycle of NCBC centers. EP administers one NCBC center, "Informatics Integrating the Bench and Bedside (i2b2)," based at Harvard's Brigham and Women's Hospital. EP program officers have scientific advisory roles in two other NCBC centers: MAGnet at Columbia University; and NCIBI at University of Michigan.

Multi-Institute Grant Programs

Budget constraints and the importance of protecting NLM's own grant programs have increased EP's selectivity when considering the funding of applications to multi-Institute initiatives; EP participation is confined to programs which do not duplicate its existing grant programs. As of 2006, the multi-Institute programs NLM participates in are: Understanding Health Literacy, several specialized SBIR/STTR programs; Research and Exploratory/developmental BISTI grants; Behavioral and Social Science Research on Understanding and Reducing Health Disparities; Advancing Novel Science in Women's Health Research. The applications for these programs are reviewed by the Center for Scientific Review (CSR), and then participating Institutes select grants for full or shared funding. These sources represent five to 10 percent of applications assigned to NLM. They are included in the listing for payline decisions. Links to the multi-Institute initiatives in which EP participates are incorporated into the grant programs list on the EP Web site, at <http://www.nlm.nih.gov/ep/Grants.html>.

Shared Funding for Research and Training

EP contributed approximately \$1 million in collaborative co-funding agreements to grants in FY2007. The listing below outlines the projects and amounts for research co-funded with other NIH Institutes and Centers. Of these, only the final three listings are new.

- Co-funding with the National Institute of Neurological Disorders and Stroke (NINDS) for a project titled "A Mature Brain Architecture Knowledge Management System" (\$23,704). 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.

- 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 (ITK) that will be used in working in a grid computing environment (\$164,306). This grid computing application will allow for computation of very large datasets that otherwise could not be analyzed effectively.
- Co-funding for two grants in the Fogarty International Center's Informatics Training for Global Health, administered by the FIC (\$250,000).
- Co-funding with National Human Genome Research Institute (NHGRI) for a conference grant entitled "Structural, Functional, & Evolutionary Genomics" (\$5,000). This Gordon Research Conference will focus on analysis of biological systems on a broad scale to include genomes and proteomes with a full review of gene and protein function in the context of biological networks.
- Co-funding for the "Comparative Toxicogenomics Database" (CTD) (\$200,000). This resource is being developed for public availability in order to promote understanding about the effects of environmental chemicals on human health, as a component of the NIGMS Pharmacogenetics Initiative.
- Co-funding with the National Institute of General Medical Sciences (NIGMS) for continued support of a cooperative agreement for the Stanford Pharmacogenetic Research Network and Knowledge Base (\$301,777).

NLM also receives co-funding from other Institutes for its research, resource and training grants. In FY2007, the co-funded NLM grants included:

- NLM requested, and received from the NIH, bridge funding support for a grant entitled, "Clinical Decision Support in Community Hospitals: Barriers & Facilitators" (\$461,255). This project examines the underlying problems with computerized clinical decision support systems in community hospitals. The focus of this project is to review the gap between community and teaching hospitals in order to develop an understanding what does and does not work in the interest of improving patient safety.
- In its recompetition of its Bioinformatics Training Programs, NLM has received support from the NHLBI, NIDCR, NIBIB, and NCMHD. Support of informatics trainees, including minority support from the NCMHD, covers various fields of training in bioinformatics and biomedical informatics. This support covers the following grantee institutions: University of Pittsburgh;

Harvard University; Stanford University; University of Wisconsin; Vanderbilt University; and University of Virginia.

- The NLM included support from the NIMH for a grant to Stanford University entitled, "A Resource for Biomedical Ontologies and Knowledge Bases" (\$160,000). This grant support was for the development of autism ontologies.
- NLM received minority training support from the NCMHD (\$68,643), for the development of future scientists in bioinformatics at two of its training programs (Harvard and Vanderbilt).
- The NIDCR expressed strong interest in a NLM grant proposal, "Web-Based Resource on Evidence-Based Dentistry" (\$150,000), for which they provided all grant funding. The intent of this proposal is to develop a dental informatics resource to support evidence-based dental care.
- Co-funding support received from the National Institute of Nursing Research (NINR) (\$62,610) for an SBIR grant entitled "Software Leveraging, a Standards-Based Web Service Framework for Decision Support."
- Co-funding support received from the National Institute of Allergy and Infectious Disease (NIAID) (\$86,416), a research grant entitled "MSM: A Multi-scale Approach for Understanding Antigen Presentation in Immunity." From the Multi-scale Modeling Interagency initiative.
- The NLM received support for two of its conference grants: "Life Science Systems and Applications Workshop," NIBIB (\$2,000), covering many disciplines to include public health systems on chip, bioinformatics, drug discovery, translational medicine, medical imaging and neural science; and "Women in Bioinformatics Seminar Series," NCI (\$5,000). Given the lack of visibility of women in science, the goal of this grant is to continue the seminar series in which exceptional research of established and junior women in bioinformatics and computational biology are showcased to inspire younger women to pursue careers in science.
- NLM worked with the NCCR towards funding of a multi-IC "Roadmap Interdisciplinary" project, supporting specialized centers in biocomputing and bioinformatics, subproject: "Hypothesis Web Development for Neuropsychiatric Phenomics," OD/NIH (\$303,654). All support was provided through NLM as a "linked" award to the parent U54-Specialized Center. The sub-grantee proposes development of a software platform called the "Hypothesis Web," which permits collaborative formulation of complex scientific hypotheses. This project represents a novel approach to the way scientific hypotheses are conceived and tested.

Interagency Agreements and Special Initiatives

NLM continues to provide co-funding to the National Science Foundation (NSF) for the Protein Databank at Rutgers University (\$200,000). This databank supplies three-dimensional representations of proteins and is the single worldwide archive of structural data for biological macromolecules.

EP continues to partner with the NSF on a new grant program titled "Dynamic Data-Driven Systems." EP provides funding for one grant in this program: "Dynamic Data-Driven Brain Machine" (\$328,417 in FY2007). The project's goal is to develop computer architecture for computational modeling and training algorithms for robotic arm movement. A clinical application of this basic work relates to improving the brain-machine interface for paraplegics. For this partnership, the grant is administered by NSF, with EP providing adjunctive program officer oversight.

NLM provided funding in support (\$15,000) of a contract for the "International Assessment of Research and Development on Simulation-Based Engineering and Science." This study is aimed at determining the status and trends of research and development in this field by leading laboratories and organizations via an on-site peer review process developed at the World Technology Evaluation Center (WTEC), Inc.

NLM issued two purchase orders in support of two National Academy of Sciences (NAS) studies: "Overcoming the Technical and Policy Constraints that Limit Large-Scale Data Integration" (\$15,000), for an NRC-appointed committee to plan and organize a pair of cross-disciplinary public workshops to explore alternative visions for achieving large-scale data integration in the fields of importance to the federal government; and "Engaging the Computer Science Research Community in Health Care Informatics" (\$30,000), for an agreement that supports a two-phase study to examine information technology (IT) problems and challenges in realizing the emerging vision of patient-centered, evidenced-based, efficient health care using electronic health records and other information technology.

EP Operating Units

Program Office

Grant Program Development—Program activities in FY 2007 were focused on: refining language for grant program announcements in which EP participates; initial planning for transitions to the new SF-424 electronic application form; participation in Research Condition and Disease Categorization (RCDC) fingerprinting activities; and exploratory discussion in support of EP's planned Challenge Grant RFAs. The NIH timetable for transition to electronic applications has stalled for several of EP's grant programs, including P41, K99, F37 and T15. Until the transition is completed, these applications will continue to be submitted on paper.

Program Staff Activities—EP program staff represent EP on various NIH and NLM standing committees, including: Extramural Programs Management Committee; Program Leadership Committee; Training Advisory Committee; Human Subjects Protection Liaison Committee; Tracking & Inclusion Committee; Electronic Research Administration Program Officials Users Group; Electronic Research Administration Population Tracking Users Group; RCDC Policy Committee; ENS Coordinators' Committee; NLM Web Editorial Committee; NLM Intranet Redesign Committee; BISTI; IMAG/Multi-Scale Modeling; Trans-NIH Genomic Working Group; and a number of RCDC fingerprint “expert” committees.

Program Oversight, Management & Evaluation—On-site visits or reverse site visits were performed for the following funded activities: PDB (Protein Data Bank); and UCLA IAIMS. Program staff attended the following scientific meetings: AMIA Spring Congress; AMIA Fall Annual Meeting; and IEEE/NLM Joint Workshop on Life Science Systems and Applications. EP initiated contracts with Humanitas, Inc. for two program assessment activities. One compared and evaluated NLM’s informatics training program graduates, and the other evaluated achievements of NLM-funded fellows and R01-funded postdoctoral students. These contracts concluded September 31, 2007. A second-year contract will continue the evaluation work on NLM’s T-15 training program.

In FY2007, EP hired a new program officer, Dr. Jane Ye, specializing in computational biology and bioinformatics. Dr. Ye came to NLM from NHLBI and serves as EP’s representative to BISTI and several other trans-NIH committees.

In the summer of 2007, EP’s intern, Mable Cao, prepared several analyses, including a detailed review of sources of funding for NLM’s grantees from 1997 to the present and a listing of the NIH “fingerprints,” which include EP grants.

EP Program Class codes were updated once. The Program unit began meeting twice a month to discuss analyses, policy and practices. In accordance with NIH guidance, Subproject records were added to IMPAC for the U54 mechanism. Programmatic analyses of the grant portfolio and grant applications were conducted, in response to specific information requests and on an ad-hoc basis.

An important role for EP Program staff in 2007 has been implementation of EP’s eGrants system for managing electronic grant files. The program analyst is responsible for implementation and quality control of this new resource. In addition program staff are working closely with DEAS and Grants Management to create records in IMPAC Training module for NLM trainees. This is the first time NIH has permitted NLM to enter its trainee appointments.

Dissemination and Staff Activities—Formal presentations on NLM grant programs were made to the following groups: Leadership program, AAHSL; Regional Library

Directors and the orientation for their new staff; NN/LM; and SIS Native American Internship program. In January, EP again offered its three-day training curriculum to the NLM Associate Fellows. In November 2006, EP staff coordinated the networking meeting for the NLM/RWJF trainees, held in conjunction with Fall AMIA. In June 2007, the annual NLM Training Conference was held at Stanford, CA. Presentations were made to the NLM Board of Regents on the following topics: EP 2007 budget; snapshot of NLM T-15 training programs; possible topic areas for EP challenge grants; evolution of NLM grant programs; and shortening the time to award for NLM grants.

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

- Dr. Paul Biondich, K22 recipient. “Comprehensive Decision Support System for Preventive Care for Children” presented to BOR;
- Dr. Margaret Humphreys, G13 recipient. “The Civil War and American Medicine” presented to BOR;
- Dr. Atul Butte, K22 recipient. “Genomic Nosology: Modernizing our 300-year old Description of Disease” presented to BOR; and
- Dr. Gary Marchionini, BLIRC member. “Personal Health Record Usability” presented at NLM lecture series.

The EP Web site was updated with new grant awards for FY2007. Listings of recent awards were provided at each meeting of the BLIRC and BOR, and were sent to NLM’s National Network Office for distribution to the NN/LM. All basic grants pages were restructured to provide a common look and feel across all programs, and to simplify maintenance of the site. Links to grantee project Web sites were added to many active grants. The training program page is being redesigned to provide a clearer picture of the types of training available at each training program site.

Grants Management Office

EP issued 217 grant awards in FY2007 for nearly \$54 million. Staffing levels have been restored in FY2007. NLM provided key support for all grant co-funding agreements of the NLM, interagency agreements in support of grants, large-scale training grants, and its general training and research grants, split over two funding authorizations: the Medical Library Assistance Act and the Public Health Service Act.

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

Scientific Review Office

Grant Review Activities—Overall, 435 applications were reviewed for which NLM was the primary assignment. Eighty-three percent of those (362) were reviewed by NLM. The remaining 73 applications were reviewed by CSR. All but 29 of those reviewed by CSR were SBIR/STTR grant applications. Of the applications reviewed by NLM, 56% were in one of the three research grant mechanisms (R01, R21, R03). Knowledge management grants represent 20% of the applications reviewed, with Scholarly Works at 16% of the applications reviewed.

BLIRC—EP's standing review group, the Biomedical Library and Informatics Review Committee (BLIRC), evaluates grant applications assigned to EP for possible funding for scientific merit. BLIRC met three times in FY2007 and reviewed 177 applications (as compared to 167 in 2006). The Committee (Appendix 1) reviews applications for most biomedical informatics and bioinformatics research applications, knowledge management/applied informatics, career support, and fellowships.

Special Emphasis Panels (SEPs)—Eight Special Emphasis Panels were held during FY2007 compared to 19 in 2006. These panels are convened on a one-time basis to review applications for which the regularly constituted review group lacks appropriate expertise, such as Scholarly Works grant applications, when a conflict of interest exists between the applicant and a member of the BLIRC, or when the number of applications received is simply too large for BLIRC to handle. NLM's SEP panels reviewed a total of 185 applications during FY2007, compared to 285 in 2006. This number is smaller than 2006 because the number of applications received is smaller, and because applications from several smaller mechanisms (R13, R03, R21) were returned to BLIRC, rebalancing the workload.

BOR/EP Subcommittee—A second-level peer review of applications is performed by the Board of Regents. One of the Board's subcommittees, the Extramural Programs Subcommittee, meets the day before the full Board for the review of "special" grant applications. Examples include applications for which the recommended amount of financial support is larger than some predetermined amount, or those with a high program priority but a borderline score. The Extramural Programs Subcommittee makes recommendations to the full Board, which votes on the applications. The Board also votes en bloc for all other applications that meet criteria for further consideration for funding. In 2007, materials presented to the EP subcommittee were revamped, to clarify the committee's role and help prepare members for a pilot using the NIH Electronic Council Book for en bloc voting. The EP subcommittee also reviews rebuttal requests from applicants who challenge the review of their grants. One rebuttal was reviewed and the committee concurred with staff analysis. There were no applications referred by BOR for re-review.

Review Staff Activities—Members of the Scientific Review unit participated in the following NIH committees: Review Policy Committee (RPC); Review Users' Group (RUG); and CSR Receipt & Referral Coordinators.

Administration and Operations Office

In FY2007, the EP Administrative Office continued to do an excellent job of managing the reimbursement payments to all NLM scientific peer reviewers that were formerly paid under the NLM Scientific Review and Evaluation Award Grant (SREA). FY2006, the first year that the program was managed by EP Administration, was fraught with numerous procedural problems and vendor payment issues. However, with the implementation of new program procedures and contractors for air travel and hotels in 2007, SREA expenditures have decreased by more than 40%. Estimated costs for the upcoming fiscal year should remain on par with FY2007 levels. Furthermore, EP ended FY2007 with a payment success rate of 84% with only a handful of reviewers not receiving a reimbursement because they chose not to apply to the government's central contractor registration (CCR) database, which is mandatory for reimbursement.

Grants Administration Support—Support for grants administration continues to be provided by four staff members from the NIH DEAS organization. In addition to staffing changes within DEAS, new practices were implemented for human subjects tracking; document scanning, and adding program class code and program officer assignments. DEAS staff reorganized the grant files room and supply areas, and updated the administrative review listing to reflect current grant programs. DEAS staff has begun the process of transferring all of the Training Appointments from the NLM system into the NIH system, and scanning grant applications and supplement material into eGrants. Several joint meetings were held with DEAS senior staff members Mitzi Diley (Director) and Bonnie Ellis (Assistant Director). Current DEAS Staff at EP include: Pam Beheler, Supervisor; Joyce Campbell; Jackie McEachin; and Tsegaye Mariam. In early 2007, the organizational redesign plan was presented to NIH staff. The intent of that plan is to create job mobility for DEAS staff, and change the supervision structure, involving IC staff more directly in task assignment and work evaluation.

Contractual Activities—At the end of FY2007, NLM/EP extended one of the two contracts awarded to Humanitas, Inc in 2006, for the purpose of analyzing and evaluating services related to training of biomedical informaticians as exemplified by experience and goals of the NLM training programs. This additional funding will go towards the complete abstraction and analysis of 2001 applications; performing additional data collection and analysis on key metrics of interest; and development of profiles of a "typical" successful career path in

bioinformatics, and/or a comparable profile of a “typical” successful training program.

EP Staff Development

All EP staff completed required IT Security and Ethics training. In addition, individual staff attended a variety of NIH training events related to their work assignments, but also attended training on broader topics offered by the NIH. A new emergency telework policy was drafted and telework agreements put in place in the event of local and national emergencies. Also, additional training requirements were added to each employee PMAP plan where it is now required that each staff member take at least one course in management or work effectiveness skills, including planning, communication, project management, supervision, etc.

Facilities & Equipment

FY2007 saw the upgrade of all NLM/EP workstation units with the installation of faster CPUs with larger storage capacity. The implementation of new NIH databases within the NIH IT community necessitated this upgrade. In addition, funds were obligated for the purchase of two new copier machines to replace older model machines, as well as for the redesign of the former grants processing room to be converted into office space for three new employees.

Personnel

As previously mentioned, in May 2007, Extramural Programs welcomed Jane Ye, as a Health Scientist Administrator/Program Officer. Dr. Ye holds a PhD in

molecular biology from Dartmouth College, Hanover, NH. Before coming to NLM, Dr. Ye worked for the National Heart, Lung, and Blood Institute (NHLBI), where she was a Program Director for the Advance Technologies & Surgery Branch. At NHLBI she was responsible for developing, implementing, and managing research grants, cooperative agreements, and contracts that involve application of genomics, bioinformatics, and enabling technologies to cardiovascular diseases. Her current responsibilities at NLM include: administering a portfolio of research program grants in the area of bioinformatics and computational biology; serving as Science Officer for the National Center for the Multi-scale Analysis of Genomic and Cellular Networks (MAGNet), Physics-based Simulation of Biological Structure (Simbios), and the NIH Roadmap National Centers for Biomedical Computing (NCBC); and NLM representative to the Interagency Modeling and Analysis Group and the Multi-scale Modeling Consortium.

In December 2006, Lakeshia Dixon departed her position as Administrative Technician for a new position at the National Science Foundation (NSF). Ms. Dixon came from the NIH/OD Division of Extramural Administrative Support, where she worked as an Extramural Support Assistant. As an Administrative Technician, Ms. Dixon served as the central contact point for receipt and processing of all equipment requests and purchases, supplies, personnel actions, training requests, awards, telephone service requests, work orders, and property transfer requests.

In FY2008, EP plans on hiring a third program officer, specializing in computational biology and bioinformatics, one additional Program Analyst, a second Committee Management Officer, and a new Administrative Technician.

Table 15

Extramural Programs Budget FY2006–2008

EXTRAMURAL PROGRAMS BUDGET																		
FY 2006–2008																		
(Dollars in thousands.)																		
	FY 2006 Actual						FY 2007 Actual						FY 2008 Proposed					
	Non-Competing		Competing		Total		Non-Competing		Competing		Total		Non-Competing		Competing		Total	
	No.	Amt. (\$)	No.	Amt. (\$)	No.	Amt. (\$)	No.	Amt. (\$)	No.	Amt. (\$)	No.	Amt. (\$)	No.	Amt. (\$)	No.	Amt. (\$)	No.	Amt. (\$)
MLAA																		
IAMS (G08)	5	1,712	3	450	8	2,162	4	953	1	145	5	1,098	3	408	2	280	5	688
Training programs (T15)	18	14,093	0	0	18	14,093	0	0	18	14,743	18	14,743	18	15,737	0	0	18	15,737
Fellowship/other (F37/F38/D43)	11	950	2	121	13	1,071	9	666	3	147	12	813	3	398	2	124	5	522
Career (K22 & K99)	13	1,975	7	1,047	20	3,022	15	2,245	3	294	18	2,539	9	1,248	3	420	12	1,668
TOTAL TRAINING	42	17,018	9	1,168	51	18,186	24	2,911	24	15,184	48	18,095	30	17,363	5	544	35	17,927
SCHOLARLY WORKS (G13)	10	680	15	1,039	25	1,719	17	1,194	11	790	28	1,984	14	980	10	700	24	1,680
RESOURCE GRANTS (G08)	17	2,189	8	1,135	25	3,324	12	1,519	13	1,554	25	3,073	17	2,213	10	1,350	27	3,563
LOAN REPAYMENT (L30)	0	0	6	422	6	422	0	0	5	232	5	232	0	0	7	422	7	422
INDLM CONTRACTS (H01)	0	0	9	12,701	9	12,701	9	13,201	0	0	9	13,201	9	12,532	2	940	11	13,472
LIBRARY & INFORMATION RESEARCH	0	0	0	0	0	0	0	0	5	1,331	5	1,331	5	1,262	0	0	5	1,262
TOTAL MLAA	74	21,599	50	16,915	124	38,514	66	19,778	59	19,236	125	39,014	78	34,778	36	4,236	114	39,014
PHS 301																		
Biomedical informatics research (R01/R03/R13/R21/R24)	30	7,663	14	4,931	44	12,594	32	9,079	21	6,129	53	15,208	47	15,936	4	1,020	51	16,956
Interagency agreements (Y01, Y02)	2	528	0	0	2	528	2	528	1	60	3	588	3	589	0	0	3	589
TOTAL BIOMEDICAL INFORMATICS RESEARCH	32	8,191	14	4,931	46	13,122	34	9,607	22	6,189	56	15,796	50	16,525	4	1,020	54	17,545
Bioinformatics research (R01/R03/R21/R24)	13	4,388	11	3,079	24	7,467	15	4,123	17	4,441	32	8,564	19	5,008	4	1,113	23	6,121
Bioinformatics resource, <i>committed</i> (P41)	4	3,044	1	242	5	3,286	4	1,894	2	1,214	6	3,108	5	3,048	2	751	7	3,799
BISTI (R01/R21/R33A/54)	6	2,004	0	0	6	2,004	1	302	0	0	1	302	1	311	0	0	1	311
TOTAL BIOINFORMATICS RESEARCH	23	9,436	12	3,321	35	12,757	20	6,319	19	5,655	39	11,974	25	8,367	6	1,864	31	10,231
SBIR/STTR (R43, R41, R42)	0	25	3	759	3	784	1	403	4	381	5	784	1	305	4	473	5	778
TOTAL PHS 301	55	17,652	29	9,011	84	26,663	55	16,329	45	12,225	100	28,554	76	25,197	14	3,357	90	28,554
TOTAL EP	129	39,251	79	25,926	208	65,177	121	36,107	104	31,461	225	67,568	154	59,975	50	7,593	204	67,568

Note: Co-fund/collaborative funding projects are excluded from the numbers of awards, but the dollars are counted.

Table 16

Extramural Programs Budget FY2003–FY2007
(Dollars in thousands)

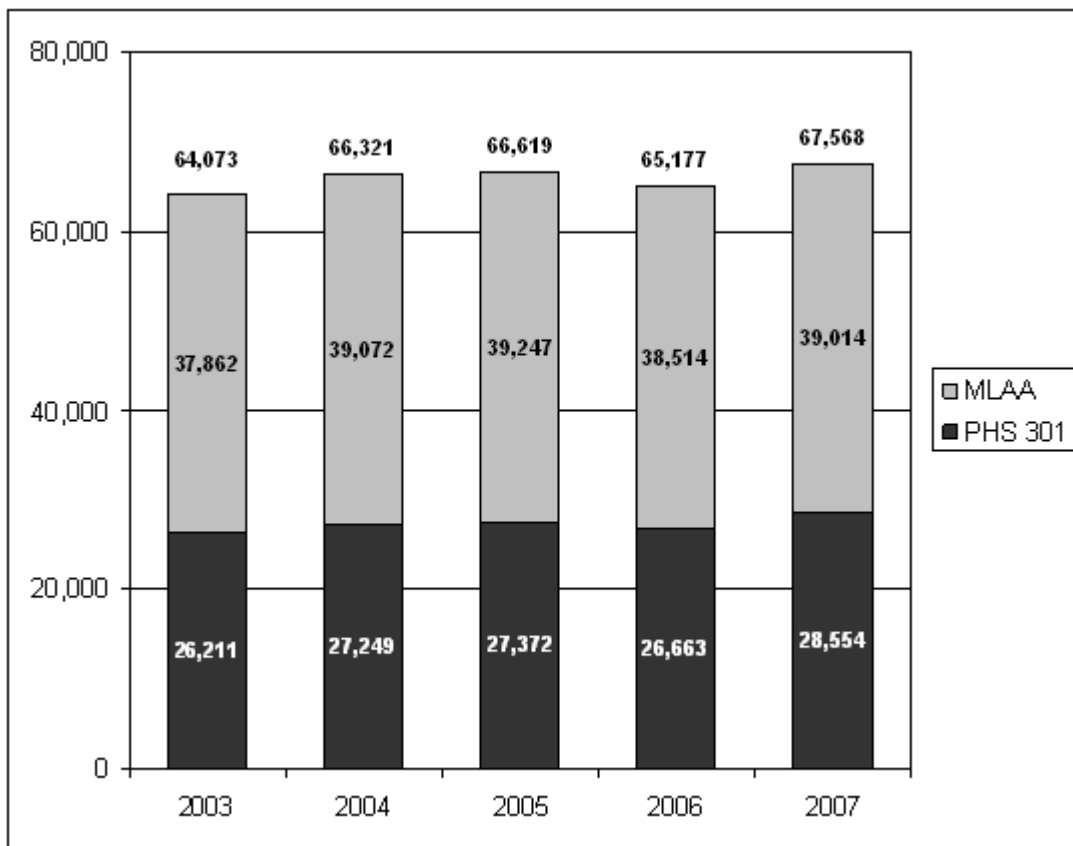


Table 17

EP Budget, FY2007

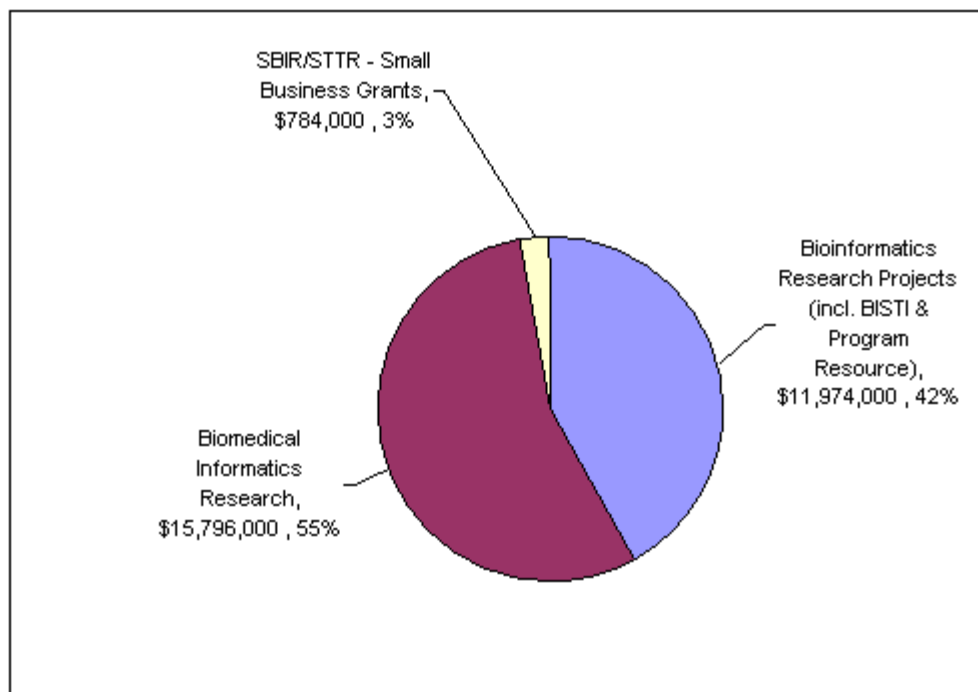
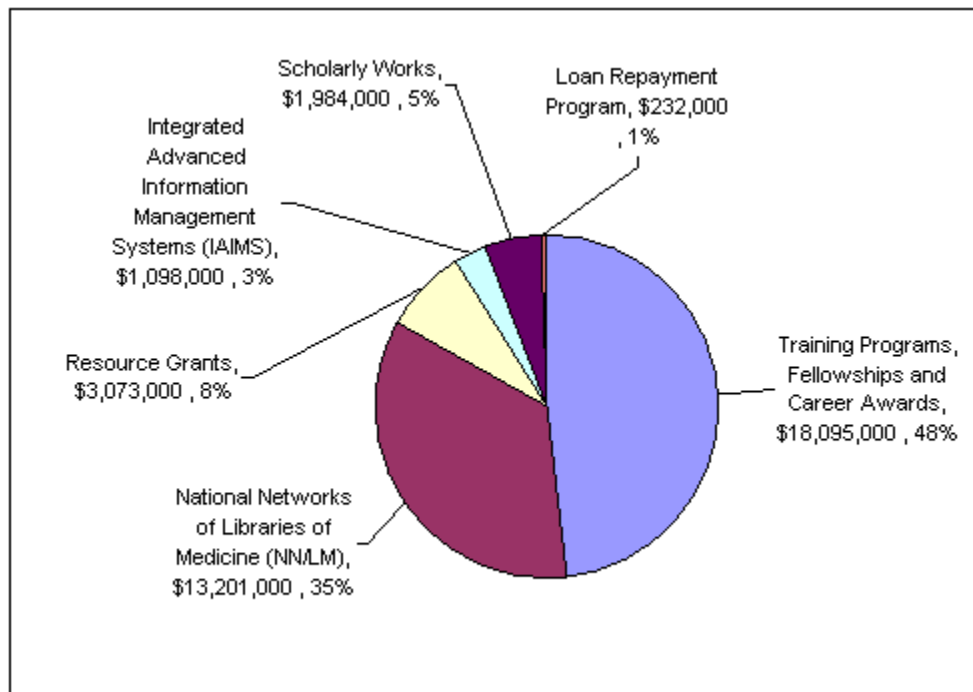


Table 18

RFA/PA Actions in FY2006 and FY2007

Guide #	Program Title	Status
IILM's Core Active Grant Programs		
PA-07-070	Research Project Grants for Biomedical Informatics and Bioinformatics (Parent R01)	Electronic application in 2007; omnibus PA in November 2006
PAR-07-236	NLM Knowledge Management & Applied Informatics Grants (G08)	Electronic application in 2007
PAR-07-237	NLM Grants for Scholarly Works in Biomedicine and Health (G13)	Electronic application in 2007
PAR-07-238	Planning Grant for Integrated Advanced Information Management Systems (IAIMS)	Electronic application in 2007
PAR-05-077	Test & Evaluation Grant for Integrated Advanced Information Management Systems (IAIMS)	Electronic application in 2007
PAR-06-509	NLM Individual Fellowship for Informationist Training (F37)	Open
PA-07-297	NIH Pathway to Independence (PI) Award (K99/R00)	Open
PA-06-180	NIH Small Research Grant Program (Parent R03)	Electronic application
PA-06-181	NIH Exploratory/Developmental Research Grant Program (Parent R21)	Electronic application
PA-06-041	NIH Support for Conferences and Scientific Meetings (Parent R13AU13)	Electronic application
PA 07-281	PHS 2007-02 Omnibus Solicitation of the NIH for Small Business Technology Transfer Grant Applications (Parent STTR [R41/R42])	Electronic application
PA 07-280	PHS 2007-02 Omnibus Solicitation of the NIH, CDC, and FDA for Small Business Innovation Research Grant Applications (Parent SBIR [R43/R44])	Electronic application
GUIDE Announcements Issued by IILM in FY 2007		
NOT-LM-07-001	Genotype and Phenotype Data Now Available from dbGaP Database; Request Process Involves New Procedures for Principal Investigators and Signing Officials	Available May 2007
PAR-07-236	NLM Knowledge Management & Applied Informatics Grants (G08)	Electronic application in 2007
PAR-07-237	NLM Grants for Scholarly Works in Biomedicine and Health (G13)	Electronic application in 2007
PAR-07-238	Planning Grant for Integrated Advanced Information Management Systems (IAIMS) (G08)	Electronic application in 2007
NOT-LM-06-008	NLM Research Grants in Biomedical Informatics and Bioinformatics (R01)	Omnibus PA in November 2006
Multi-IC Announcements in which IILM Participates		
PAR-07-344	Innovations in Biomedical Computational Science and Technology (BISTI) (R01)	Electronic application in 2007
PAR-06-411	Exploratory Innovations in Biomedical Computational Science and Technology (BISTI) (R21)	Electronic application in 2007
PA-06-042	Academic Research Enhancement Award (AREA) (Parent R15)	Electronic application
PA-04-126	Supplements to Promote Reentry into Biomedical and Behavioral Research Careers	Open
PA-05-015	Research Supplements to Promote Diversity in Health-Related Research	Open
PAR-07-020	Understanding and Promoting Health Literacy (R01)	Electronic application in 2007
PAR-07-019	Understanding and Promoting Health Literacy (R03)	Electronic application
PAR-07-379	Behavioral and Social Science Research on Understanding and Reducing Health Disparities Grants (R01)	Electronic application; new in 2007
PAR-07-380	Behavioral and Social Science Research on Understanding and Reducing Health Disparities Grants (R21)	Electronic application; new in 2007
PAS-07-382	Advancing Novel Science in Women's Health Research (ANSWHR) (R03)	Electronic application; new in 2007
PAS-07-381	Advancing Novel Science in Women's Health Research (ANSWHR) (R21)	Electronic application; new in 2007
PA-07-442	Extramural Loan Repayment Programs (L30)	Electronic application
PA-07-440	Extramural Loan Repayment Programs (L40)	Electronic application
IILM Grant Programs Suspended in FY 2006 and 2007		
PA-03-090	NLM Early Career Development Award for Informatics (K22)	Suspended; replaced by PAR-06-133 in January 2006
PA-03-178	Informatics for Disaster Management (R21)	Suspended November 2005
PAR-03-070	Individual Biomedical Informatics Fellowships	Suspended November 2005
PAR-03-109	NLM Senior Individual Biomedical Informatics Fellowships (F38)	Suspended November 2005
PAR-04-014	NLM Senior Fellowship for Informationist Training (F38)	Suspended November 2005
PAR-05-076	Operations Grant for Integrated Advanced Information Management Systems (IAIMS) (G08)	Suspended November 2005
PAR-05-063	Collaborations with National Centers for Biomedical Computing (R01)	Discontinued participation April 2006
PAR-06-223	Exploratory Collaborations with National Centers for Biomedical Computing (R21)	Discontinued participation April 2006
PAR-05-057	Continued Development and Maintenance of Software (P41)	Discontinued participation April 2006
PA-02-011	Bioengineering Research Grants (R01)	Discontinued participation May 2006
PAR-04-023	Bioengineering Research Partnerships (R01)	Discontinued participation June 2006
Small Business (SBIR/STTR) Announcements in which IILM Participates		
PA-06-010	Integration of Heterogeneous Data Sources (STTR [R41/R42])	Electronic application
PA-06-011	Integration of Heterogeneous Data Sources (SBIR [R43/R44])	Electronic application
PA-06-008	Bioengineering Nanotechnology Initiative (STTR [R41/R42])	Electronic application
PA-06-009	Bioengineering Nanotechnology Initiative – SBIR (R43/R44)	Electronic application
PA-06-396	New Technologies for Liver Disease STTR (R41/R42)	Electronic application
PA-06-397	New Technologies for Liver Disease SBIR (R43/R44)	Electronic application
PAR-07-161	Innovations in Biomedical Computational Science and Technology Initiative (BISTI) (STTR [R41/R42])	Electronic application
PAR-07-160	Innovations in Biomedical Computational Science and Technology Initiative (BISTI) (SBIR [R43/R44])	Electronic application
PA-06-012	Manufacturing Processes of Medical, Dental, and Biological Technologies (STTR [R41/R42])	Electronic application
PA-06-013	Manufacturing Processes of Medical, Dental, and Biological Technologies (SBIR [R43/R44])	Electronic application

Table 19**NLM grants awarded in FY2007**

(Sorted by PI name within each grant category)

RESEARCH GRANTS

Ash, Joan S.

2-R56-LM006942-07A1

Oregon Health & Science University

Clinical Decision Support in Community Hospitals:
Barriers & Facilitators

Benos, Panagiotis V. and Kaminski, Naftali

(New Investigators)

1-R01-LM009657-01

University of Pittsburgh, Pittsburgh

Modeling in vivo Protein-DNA Interactions from High-
Throughput Data MP1/1

Bickmore, Timothy W. (New Investigator)

1-R21-LM008995-01A2

Northeastern University

Computational Models of Health Behavior Change Dialog

Cannon Albright, Lisa A.

1-R01-LM009331-01

University of Utah

Analysis of the Familial Component to Disease in a
Biomedical Resource w/Link...

Chapman, Wendy W. (New Investigator)

1-R01-LM009427-01

University of Pittsburgh, Pittsburgh

NLP Foundational Studies & Ontologies for Syndromic
Surveillance from ED Reports

Cimino, James J.

1-R21-LM009638-01

Columbia University Health Sciences

Semantic and Machine Learning Methods for Mining
Connections in the UMLS

Cohen, Aaron M. (New Investigator)

1-R01-LM009501-01

Oregon Health & Science University

Assisting Systematic Review Preparation Using
Automated Document Classification

Craven, Mark W.

2-R01-LM007050-04A2

University of Wisconsin, Madison

Adaptive Information Monitoring and Extraction

Dee, Fred R.

1-R01-LM009121-01A1

University of Iowa

Competencies Training & Assessment in Laboratory
Medicine Using Patient Simulations

Doctor, Jason N.

1-R01-LM009157-01A1

University of Southern California

Detecting Errors in Blood Labs Using Bayesian Networks

Effken, Judith A.; Carley, Kathleen M.; Verran, Joyce A.

(New Investigators)

1-R01-LM009516-01A1

University of Arizona

DyNADS: A Dynamic Network Analysis Decision
Support Tool for Nurse Managers

Fidelis, Krzysztof A.

2-P41-LM007085-08

University of California, Davis

Center for Critical Assessment of Structure Prediction

Floudas, Christodoulos Achilleus

1-R01-LM009338-01

Princeton University

Peptide and Protein Identification via Tandem MS and
Mixed-Integer Optimization

Foran, David J.

1-R01-LM009239-01A1

Univ of Med/Dent NJ-R W Johnson Med School

Image Mining for Comparative Analysis of Expression
Patterns in Tissue Microarray

Garner, Harold R.

1-R01-LM009758-01

University of Texas SW Med Ctr/Dallas

Duplicate Article/Plagiarism Discovery

Halfon, Marc S. (New Investigator)

1-R03-LM008941-01A1

State University of New York at Buffalo

Empirical assessment of analysis methods for DNA
microarrays

Hauskrecht, Milos (New Investigator)

1-R21-LM009102-01A1

University of Pittsburgh, Pittsburgh

Evidence based anomaly detection in clinical databases.

Karp, Peter D. (New Investigator)

1-R01-LM009651-01

Sri International

Pathway Prediction and Assessment Integrating Multiple
Evidence Types

Karras, Bryant T.

1-R21-LM008981-01A2

University of Washington

eHealth Survey System for Public Health: Feasibility and Usability

Karsh, Ben-Tzion
1-R01-LM008923-01A1
University of Wisconsin, Madison
The Value of CPOE in Pediatric Inpatient Units and Its Impact on Safety and Work

Kohane, Isaac S.
1-R01-LM009375-01A1
Children's Hospital Boston
GPing: Informatics for Personal Control of Genomic Data

Lu, Xinghua (New Investigator)
1-R01-LM009153-01A1
Medical University of South Carolina
Automatic Literature-based Protein Annotation

McClure, Marcella A.
1-R13-LM009680-01
Montana State University, Bozeman
Women in Bioinformatics Seminar Series

Miller, Randolph A.
2-R01-LM007995-04
Vanderbilt University
MOMENT (Monitoring for Outpatient Medication Effects and New Toxicities) in TIME

Mooney, Sean D. (New Investigator)
1-R01-LM009722-01
Indiana Univ-Purdue Univ at Indianapolis
Informatic profiling of clinically relevant mutation

Moriyama, Etsuko (New Investigator)
1-R01-LM009219-01A1
University of Nebraska, Lincoln
Efficient and Sensitive Mining System for G-Protein Coupled Receptors

Moult, John
2-R01-LM007174-06A1
University of Md Biotechnology Institute
Analysis of the Functional Impact of Coding region SNPs

Musen, Mark A.
2-P41-LM007885-05
Stanford University
A Resource for Biomedical Ontologies and Knowledge Bases

Ohno-Machado, Lucila
1-R01-LM009520-01
Brigham and Women's Hospital
Assessing quality of individual predictions in medical decision support systems

Parker, D. Stott
1-RL1-LM009833-01
University of California, Los Angeles
Hypothesis Web Development for Neuropsychiatric Phenomics (6 of 8)

Perl, Yehoshua
1-R01-LM008912-01A1
New Jersey Institute of Technology
Taxonomies Supporting Orientation, Navigation and Auditing of Terminologies

Pienta, Amy (New Investigator)
1-R01-LM009765-01
University of Michigan, Ann Arbor
Barriers and Opportunities for Sharing Research Data
Pratt, Wanda (New Investigator)
1-R01-LM009143-01A2
University of Washington
Managing Health Information in Your Life

Rost, Burkhard
2-R01-LM007329-05
Columbia University Health Sciences
Improve predictions of structure and function by PredictProtein

Salzberg, Steven L.
2-R01-LM006845-08
University of Maryland, College Pk Campus
Computational Gene Modeling and Genome Sequence Assembly
Shiffman, Richard N.
2-R01-LM007199-04A1
Yale University
Improving Guideline Development and Implementation

Slonim, Donna (New Investigator)
1-R21-LM009411-01A1
Tufts University, Medford
Evaluating Biological Networks by Mining Public Data

Stojanovic, Nikola
1-R03-LM009033-01A1
University of Texas, Arlington
Computational Analysis of Short Repetitive Motifs in DNA Sequences

Tagare, Hemant D.
1-R03-LM009328-01
Yale University
Constrained Maximum Likelihood Cryo-EM Reconstruction in Proteomics

Tiwari, Hemant K. (New Investigator)
1-R21-LM008791-01A2
University of Alabama, Birmingham
Multiple Imputation in Genetic Studies

Wall, Dennis Paul (New Investigator)
1-R03-LM009261-01A1
Harvard University Medical School
Building a framework for exploration of orthologs and evolutionary distances.

Wong, Stephen T.C.
1-R01-LM009161-01A1
Brigham and Women's Hospital
Neuronal Spines Tracking and Analysis for Time-Lapse, 3D Optical Microscopy

Wong, Stephen T.C.
1-R13-LM009571-01
Brigham and Women's Hospital
Life Science Systems and Applications Workshop

Xu, Xie George
1-R01-LM009362-01
Rensselaer Polytechnic Institute
4D Visible Human Modeling for Radiation Dosimetry

Yu, Hong (New Investigator)
1-R01-LM009836-01A1
University of Wisconsin, Milwaukee
HERMES - Help physicians to Extract and articulate Multimedia information from...

KNOWLEDGE MANAGEMENT

Bakken, Suzanne
1-G08-LM008588-01A2
Columbia University Health Sciences
APN Access to Electronic Resources for Safety & Quality

Cimino, James J.
1-G08-LM009376-01A1
Columbia University Health Sciences
Librarian Infobutton Tailoring Environment

Edgar, Heather J.H. (New Investigator)
1-G08-LM009381-01A1
University of New Mexico, Albuquerque
The impact of an ethnically diverse, Web-based case file in orthodontic education

Frantsve-Hawley, Julie C. (New Investigator)
1-G08-LM008956-01A2
American Dental Association Foundation
Web-based Resource on Evidence-based Dentistry

Gandhi, Rajesh T.
1-G08-LM008830-01A2
Massachusetts General Hospital
eMicrobes: A Digital Library for Learning Infectious Diseases

Gantenbein, Rex E.
1-G08-LM009216-01
University of Wyoming
Western Regional Biomedical Collaboratory

Hunter, Lawrence E.
1-G08-LM009639-01
University of Colorado, Denver/HSC, Aurora
Construction of a Full Text Corpus for Biomedical Text Mining

Ko, Clifford
1-G08-LM009537-01
American College of Surgeons
Plan for Extracting Intraoperative Anesthesia Data to the ACS NSQIP Database

Leahy, J. Michael (New Investigator)
1-G08-LM009094-01A1
Oregon Community Health Inform Network
Oregon Community Health Integrated Data System Project

Pope, Charlene Ann (New Investigator)
1-G08-LM009624-01
Medical University of South Carolina
Carolinas Conversations: A Multiethnic Digital Corpus of Speech for Older Persons

Sasner, Michael (New Investigator)
1-G08-LM009233-01A1
Jackson Laboratory
Enhancing the JaxMice Database Resource

Smith, Thomas J.
1-G08-LM009525-01
Virginia Commonwealth University
Truthful Information about Prognosis and Options for People with Advanced Cancer

Wong, Stephen T.C.
1-G08-LM008937-01A2
Methodist Hospital Research Institute
AFINITI - An Augmented System for Neuroimaging Followup

Zaruba, Gergely V. (New Investigator)
1-G08-LM009262-01A1
University of Texas, Arlington
Teleherence: Monitoring/Increasing Treatment Adherence via Web Telecommunications

SCHOLARLY WORKS

Abir-Am, Pnina G. (New Investigator)
1-G13-LM008872-01A2
Independent Scholar -- Abir-Am, PG

How Scientists Remember a Major Discovery: History and Memory of DNA structure.

Creager, Angela N.H. (New Investigator)
1-G13-LM009100-01A1
Princeton University
Atoms for Peace and Health: Radioisotopes in Postwar Biomedicine

Devries, Raymond G.
1-G13-LM008781-01A2
University of Michigan, Ann Arbor
Medicine and Medical Science in the Age of Bioethics

Janosky, Janine E.
1-G13-LM009372-01A1
University of Pittsburgh, Pittsburgh
Single-Subject Design in Biomedicine

Millett, David E.
1-G13-LM009385-01A1
University of Southern California
Brain waves: EEG and the history of modern neuroscience

Oliver, Elizabeth (New Investigator)
1-G13-LM009399-01
Louisiana State Univ A&M Col, Baton Rouge
The Body Legal in the Dark Ages

Reverby, Susan M. (New Investigator)
1-G13-LM009227-01
Wellesley College
Testifying on Tuskegee: Telling the Tuskegee Syphilis Study Stories

Rogers, Naomi (New Investigator)
1-G13-LM009413-01
Yale University
Sister Kenny, Polio and American Medicine, 1940-1960

Ross, Lainie Friedman
1-G13-LM009096-01A1
University of Chicago
Ethical Issues in Living Donor Transplantation

Saffo, Mary B.
1-G13-LM009396-01
Marine Biological Laboratory
Lives of the infectious and the infected: benign and pathogenic symbiosis

Stellman, Jeanne Mager
1-G13-LM009137-01A1
SUNY Downstate Medical Center
A Data Resource for the Health and Environmental Consequences of the Vietnam War

TRAINING

Altman, Russ Biagio
2-T15-LM007033-24
Stanford University
Graduate Training in Biomedical Informatics

Baldi, Pierre
2-T15-LM007443-06
University of California, Irvine
Biomedical Informatics Training Program

Bui, Alex
2-T15-LM007356-06
University of California, Los Angeles
Medical Imaging Informatics Training Program

Caldwell, Charles W.
2-T15-LM007089-16
University of Missouri, Columbia
MU Biomedical Informatics Research Training Program

Cooper, Gregory F.
2-T15-LM007059-21
University of Pittsburgh, Pittsburgh
Pittsburgh Biomedical Informatics Training Program

Downs, Stephen M.
2-T15-LM007117-11
Indiana Univ-Purdue Univ at Indianapolis
Indiana University/Regenstrief Institute Biomedical Informatics Research Training

Gadd, Cynthia S.
2-T15-LM007450-06
Vanderbilt University
Vanderbilt Biomedical Informatics Training Program

Gorry, Anthony
2-T15-LM007093-16
Rice University
NLM Training Program in Biomedical Informatics

Greenes, Robert A.
2-T15-LM007092-16
Harvard University Medical School
Boston-Area Research Training Program in Biomedical Informatics

Guerlain, Stephanie A. (New Investigator)
1-T15-LM009462-01
University of Virginia, Charlottesville
Systems Engineering Focus on Clinical Informatics

Hersh, William R.
2-T15-LM007088-16
Oregon Health & Science University

Biomedical Informatics Research Training at Oregon Health & Science University

Hripcsak, George M.
2-T15-LM007079-16
Columbia University Health Sciences
Training in Biomedical Informatics at Columbia University

Hunter, Lawrence E.
1-T15-LM009451-01
University of Colorado, Denver/HSC, Aurora
Computational Bioscience Program Training Grant

Lehmann, Harold P.
2-T15-LM007452-06
Johns Hopkins University
Health Sciences Informatics Research Training Fellowship Program

Miller, Perry L.
2-T15-LM007056-21
Yale University
Biomedical Informatics Research Training at Yale

Mitchell, Joyce A.
2-T15-LM007124-11
University of Utah
University of Utah Biomedical Informatics Training

Phillips, George Neal
2-T15-LM007359-06
University of Wisconsin, Madison
Research Training for Computation and Informatics in Biology and Medicine

Tarczy-Hornoch, Peter
2-T15-LM007442-06
University of Washington
Biomedical and Health Informatics Training Program

CAREER DEVELOPMENT

Brasure, Michelle B.
1-F37-LM009759-01
College of St. Catherine
Health Informatics Outreach Training

Folb, Barbara L.
1-F37-LM009471-01A1
University of Pittsburgh, Pittsburgh

Public Health Informationist Training for Academic Public Health Practice

Salzman, Peter
1-K99-LM009477-01A1
University of Rochester
Statistical & Computational Tools for Reconstruction of Gene Regulatory Networks

Shapiro, Jason Scott (New Investigator)
1-K99-LM009556-01
Columbia University Health Sciences
Evaluation of the Impact of Health Information Exchange on Emergency Medicine
Shaw, Pamela L.
1-F37-LM009568-01
Northwestern University
Developing Skills in Research Bioinformatics

Zhang, Zhengdong D. (New Investigator)
1-K99-LM009770-01
Yale University
Gene regulation in metastasis and new methods to analyze its microarray profiles

SMALL BUSINESS GRANTS

Groen, Frank (New Investigator)
1-R43-LM009577-01
Item Software, Inc.
Probabilistic Risk Assessment in Complex Healthcare Settings

Lobach, David F.
2-R42-LM009051-02
Religent, Inc.
Software Leveraging a Standards-Based Web Service Framework for Decision Support

Moore, Steven D.
1-R43-LM009474-01
Science-Approach
NIH Biotechnologist: Authentic Exploration of Biotechnology in Health Science

Quarum, Merrit (New Investigator)
1-R43-LM009190-01A2
Qmedtrix Systems, Inc.
The Use of Mathematic Algorithms in the Prevention of Improper Medical Payments

OFFICE OF COMPUTER AND COMMUNICATIONS SYSTEMS

Simon Y. Liu, PhD
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 capacities, and assists other NLM components in local area networking. The division: provides professional programming services and computational and data processing to meet NLM program needs; operates and maintains the NLM Computer Centers; develops software; and provides extensive customer support, training courses, and documentation for computer and network users.

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

MedlinePlus's greatest achievement this year was the expansion of the MedlinePlus "Go Local" initiative. As of FY2007, the MedlinePlus Go Local site included North Carolina, Missouri, Indiana, Massachusetts, Southern Texas, Utah, Wyoming, Maryland, East Texas, New Mexico, Texas Gulf Coast, Southern Ohio, Arizona, Nevada, Delaware, Vermont, Nebraska, Michigan, Minnesota and Illinois.

In FY2007, MedlinePlus realized an 11% increase in page views, with 885 million pages views and over 109 million unique visitors. In addition, MedlinePlus and MedlinePlus en español received scores of 87 and 85 respectively from the American Customer Satisfaction Index (ACSI)'s E-Government Satisfaction Index, which tracks customer satisfaction trends. Major upgrades and enhancements to MedlinePlus included:

- Building a "What's New" page in both English and Spanish.
- Redesigning the table of contents.
- Developing a new Health Topics page, featuring images, summaries, synonyms and a category labeled, "Start Here."

Multi-layered IT Security Program

OCCS continued its multi-layered IT security program, which successfully detected more than 85,000 probes, scans, denial of service (DOS) attacks, unauthorized access attempts and other security events on a monthly basis. Major accomplishments included:

- Performing a monthly cycle of vulnerability scanning, detection, and remediation to improve NLM security posture.
- Performing automatic virus scanning and signature update mechanisms to combat ever increasing cyber-attacks.
- Implementing software to automate data encryption, securing sensitive information on mobile devices.
- Expanding an automatic patch management system to eliminate security vulnerabilities.
- Strengthening password requirements, in response to directives from the Office of Management and Budget (OMB).
- Successfully passing a network penetration test performed by an independent contractor.
- Leading a successful campaign to achieve 100% participation in information security and privacy awareness training, for NLM employees and contractors.

Unified Medical Language System (UMLS) Project

The Unified Medical Language System (UMLS) provides a common, concept-oriented medical vocabulary and thesaurus based on more than 119 current medical source vocabularies in 17 languages. The UMLS Metathesaurus contains 1.35 million concepts, 5.3 million concept names and 21.9 million relationships. This year, OCCS successfully transferred the inversion and insertion of five major sources, including RxNorm, MeSH, OMIM, LOINC, and SNOMED/CT, and upgraded UMLS computing and storage capacity, which reduced the inversion and insertion computation time from 56 to 28 hours. That change led to the reduction of Metathesaurus release production time, too, from six to four weeks.

DailyMed Project

The DailyMed project is a partnership between the Food and Drug Administration (FDA), the Department of Veterans Affairs (VA), the NLM, medication manufacturers and distributors, and healthcare information suppliers. The project seeks to provide a standard, comprehensive, up-to-date, XML-based capability for labeling the contents of medications. This year, OCCS designed interface prototypes for incorporation of newly modified Drug Label Sections in the Highlights and Navigation Sections and began publishing RxNorm Data for each Structured Product Labeling (SPL) into DailyMed Production for public

usage. In addition, DailyMed contains more than 3,323 approved prescription drug labels for public access, added more than 800 new drug label/package inserts. In 2007, the site recognized more than 1.1 million unique visitors and over 9.1 million page views.

NIH Consolidated Colocation Site (NCCS)

OCCS continued to lead the effort on the NIH Consolidated Colocation Site Project. The NCCS is operational with capabilities as a disaster recovery and load-balancing site. It serves as a disaster recovery/alternate computing site for NLM as well as CIT, NCI, NHLBI, NIDDK, NIAMS, OD/Office of Research Services and HHS/Office of the Secretary.

High Speed Communication Network

OCCS continued improvement of 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 two gigabits per second (Gbps), while the interconnection between NLM and the NIH/CIT campus backbone operates at one Gbps. NLM continues a redundant, diverse fiber connection from NLM to the MAX provided by FiberGate. It provides for increased reliability for this critical network connection. In addition, OCCS:

- Expanded the High Availability Computing Solution, to ensure that critical applications and resources remain available to NLM users.
- Continued Citrix remote access services for NLM flexi-place workers.
- Provided high speed access for indexers and flexi-place workers through Comcast cable servers, as well as Verizon FiOS and DSL.

RxNorm Project

Eight major versions of the RxNorm Editing System were released this year. Version 4.4 was released in December, 4.5 and 4.6 in February, 5.0 in April, 5.3 in May, 5.5 in July, and 5.6 and 5.7 in August. Implementation of these releases included the ability to search on NDC codes, the ability to duplicate atoms to circumvent ambiguity and to create a modified SCD or SBD. The RxNorm application currently contains 14,373 Branded Drug RxNorm Forms and 18,276 Generic Drug RxNorm Forms, for a total of 32,649 RxNorm Form drugs. Additionally, there were 12 monthly releases to RxNorm this year.

Enhanced NIHSeniorHealth

With 33 topics now available in SeniorHealth, many new topics were added this year. These included "Falls and Older Adults," "Depression," "Skin Cancer" and "Talking with Your Doctor." In addition, 13 new flash videos were added to existing topics, such as "Dry Mouth," "Heart

Attack," "Low Vision" and "Osteoporosis." NIHSeniorHealth saw 28 million site hits in 2007, a 41% increase over 2006. The site had over one million unique visitors, 20% of them outside the US.

Computer Facility Reengineering

The NLM computer facility has tripled its use of electrical power, cooling and data transmission capacity over the last five years due to the rapid growth in IT systems. Recognizing that this growth will continue in the years ahead, OCCS began a detailed process for evaluating the safety, reliability and performance requirements of the computer facility. This year's reengineering activities included:

- Completing the design and procurement of cooling infrastructure to increase capacity by 75%, from 170 to 300 tons.
- Completing the design of electrical systems upgrade from 900 to 1500 kilowatts (KW), to support growing NLM information services.
- Continued installation of an overhead Ladder Rack in the computer facility, as a separate pathway for running data networking cables to improve the reliability, availability, and maintainability of data communication services.
- Installation of an environmental monitoring software system that enhances the ability to manage and report environmental incidences. The system provides real-time status of temperature, humidity, and UPS power usage.
- Reconfiguring the computer facility UPS system. The facility is now supported by three 300 KW UPS, two parallel and one stand alone.
- Installation of 336 L5-20 circuits, to support growing electric power demands.

Enhanced Data Creation and Maintenance System (DCMS)

The major event for DCMS this year was the baseline extraction, which is a re-release of all DCMS citations that follows the MeSH Year-end Processing (YEP). A new method for baseline queuing was implemented. It uses a simpler validation query thereby only needing to perform a single pass through the baseline queue table. This new method reduced the required time from FY2006's nine hours to three hours.

High Availability Computing Resources

OCCS continued implementation of high availability computing, storage, and backup services. This initiative significantly improves the reliability, availability, maintainability, security and performance of NLM mission critical systems through database clustering technology. It also increases the NLM mission critical

database storage capacity eight times, and will increase non-database storage capacity four times.

Enhanced Medical Subject Headings (MeSH)

Several new features were implemented for the MeSH Translation Maintenance System (MTMS). These included the generation and distribution of new XML files for 2007 for the French, Italian, and German translations. The development of the Publication Type Maintenance System for Auto YEP was completed for the GCMS System. Additionally, modifications implemented for the M2000 Maintenance System included development and testing of the new Global Record Maintenance System (GRMS) and reduction of the execution time for the processing of HM and EC transactions for YEP.

Enhanced DOCLINE

DOCLINE, the NLM interlibrary loan (ILL) system, supports approximately 3,600 domestic and international libraries in processing more than 2.5 million interlibrary loan transactions a year. Three versions of DOCLINE were released this year, incorporating 82 enhancements created in response to user and Library Operations requests. User page views increased by 20% over the previous year, from three to 3.6 million, and the number of visitors increased by 300%, from 200,000 to 800,000.

Public Health Partners

OCCS re-engineered the Public Health Partners Web site (www.phpartners.org). This resource was created to help the public health workforce find and use information effectively, to improve and protect the public's health. PHP is a collaborative effort of US government agencies (e.g., the Centers for Disease Control and Prevention and the Agency for Healthcare Research and Quality), public health organizations (e.g., the American Public Health Association and the National Association of County and City Health Officials) and health sciences libraries (e.g., NLM and its National Network of Libraries of Medicine). User page views increased by 200% in FY2007, from 500,000 to 1,500,000. The site also saw a 20% increase in the number of visitors, from 131,000 to 156,000.

Pandemic Flu Preparatory Testing

OCCS assisted with the successful load testing of NLM's administrative functions through the Citrix remote access environment and NLM Virtual Private Network (VPN) services. This validated NLM's capabilities and capacities to support planning for a possible pandemic flu outbreak.

Extended Daylight Saving Time

The Energy Policy Act of 2005 extended Daylight Saving Time by four weeks starting in 2007. In support of this

extension, OCCS deployed several software updates to NLM desktops, Blackberry handhelds and home systems, as well as to NLM applications and servers. The transition resulted in no disruption to NLM services.

The following describes in more detail OCCS's FY2007 accomplishments:

Business Continuity and Disaster Recovery

In order to protect NLM's mission-critical systems, CIT and NLM have implemented an NIH Consolidated Colocation Site (NCCS) in Sterling, Virginia. The NCCS is operational with initial capabilities as a disaster recovery and load-balancing site.

At present, all NLM mission-critical systems are under "active/active," "active/passive" or "active/cold-backup" mode, depending on their business requirements. The Business Continuity and Disaster Recovery Plan covers NCCS as the primary resource for system restoration and uninterrupted processing if the primary NLM computing facilities on the NIH campus are rendered unavailable by a disaster or other contingency. In 2007, OCCS procured additional load-balancers and a newer storage system at the NCCS, to provide increased network traffic capacity, more sophisticated traffic routing and higher storage performance. OCCS also performed various other upgrades to the storage systems and servers located at this site.

The NLM computer facility has tripled its use of electrical power, cooling and data transmission capacity over the last five years. Recognizing that this growth will continue in the years ahead, OCCS began a detailed process for evaluating the safety, reliability and performance requirements of the computer facility. This year's reengineering activities included:

- Completion of the design and procurement of cooling infrastructure to increase capacity by 75%, from 170 to 300 tons.
- Completion of the design of electrical systems upgrade from 900 to 1500 kilowatts (KW), to support growing NLM information services.
- Continuation of the installation of an overhead Ladder Rack in the computer facility, as a separate pathway for running data networking cables to improve the reliability, availability, and maintainability of data communication services. The goal is to move all communications cabling from beneath the raised floor, and re-install the cabling in the overhead rack. In conjunction, the computer cabinets will be re-oriented to form hot and cold aisles, for better management of cooling of the rack-mounted systems. Once the new data cabling is installed, the old cabling from underneath the floors will be removed, helping with the flow of cold air.
- Installation of an environmental monitoring software system within the Computer Facility; this

device has enhanced OCCS's ability to manage and report environmental incidences. The environmental system provides real-time status of temperature, humidity, UPS power usage per phase, and CRAC status. Two 42-inch plasma displays are installed in the NOSC to display this information.

- The B1 Computer Facility's UPS system was reconfigured. The facility is now supported by three 300 KW UPS — two parallel and one stand alone.
- To keep up with growing electrical power demands, 336 new L5-20 circuits were installed within the computer facility.

Consumer Health

MedlinePlus

As mentioned above, MedlinePlus's greatest achievement this year was the expansion of the MedlinePlus "Go Local" initiative. Go Local brings local health services to the public by allowing users to search for healthcare providers in their localities while searching MedlinePlus for information on medical conditions or other healthcare issues. MedlinePlus Go Local implemented two releases this year, version 3.0 in October and version 4.0 in June. These releases included the ability for Go Local users to form a group of records and make global updates to these records in the group, thereby updating multiple records at a time. In FY2007, MedlinePlus added Go Local sites serving Nebraska, Michigan, Minnesota and Illinois.

There were also three versions of MedlinePlus released in 2007. Version 20 was released in February, version 20.1 in March and version 20.2 in September. These releases included the addition of several input modules to facilitate the data displayed on the new health topic page. These included: health topic synonym list; health topic summary; health topic image; table of contents; subheading for the subcategory; and different process to handle the current PubMed records.

NIHSeniorHealth

NIHSeniorHealth is a joint project of NLM and the National Institute on Aging (NIA). It provides health information on the Web using modes of delivery video and narration appropriate for older Americans with low vision and/or low hearing, etc. The system includes the Accent "Talking Web" module, developed by OCCS to provide accessibility enhancements, including a selectable range of type sizes and spoken text.

The Accent module received numerous enhancements, including new videos using Flash technology and the establishment of multi-lingual support of Loquendo voice font, including English, French, German, Italian, Portuguese, and Spanish. Also, the native versus Java-based mpeg audio compression units on the speech servers were tested, and adjustments were made to

pronunciations of drug names, disease names and medical terminology.

DailyMed Project

The DailyMed Project seeks to provide a standard, comprehensive, up-to-date, XML-based capability for labeling the contents of medications. This year OCCS:

- Added over 800 new labels/package inserts.
- Designed interface prototypes for incorporation of newly modified Drug Label Sections in the Highlights and Navigation Sections.
- Began publishing RxNorm Data for each Structured Product Labeling (SPL) into DailyMed Production for public usage.
- Added a "Highlights" and "Full Table of Contents" tabs to access DailyMed SPLs with Highlights text and data elements.

IT Security

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

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

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

In order to effectively deal with increasing IT security threats from the Internet, NLM chose BT Counterpane, a managed security company, for 24/7 monitoring of NLM's IDS sensors on incoming and outgoing network traffic. The firm provides real-time monitoring, performs event correlation of security events and analyzes alerts. In addition, they provide global intelligence sharing with NLM by monitoring 550 networks in 38 countries.

In response to NIH and HHS guidance, and guidelines upon directives from the Office of Management and Budget (OMB), NLM has strengthened its password policy and requirements to cope with the evolving sophistication and increased determination of our adversaries. A new minimum eight-character password requirement, as well as a 90-day expiration, is now enforced for all user accounts.

In light of recent security breaches on government-owned laptops, NLM implemented software from Pointsec Mobile Technologies. The software provides for automatic data encryption, securing sensitive

information stored on PCs, laptops, PDAs, smartphones and removable media.

The HHS Inspector General performed a yearly audit of NLM's major IT investments for compliance with the Federal Information Security Management Act (FISMA). No deficiencies were found.

OCCS successfully completed a network penetration test conducted by OnPoint, an independent contract company. This test is performed every three years.

The NLM ISSO participated in the NIH Continuity of Operations (COOP) IT Subcommittee Working Group. The goal of the group was to identify critical IT systems and business processes and owners, assess the existing IT services in COOP plans, and determine requirements in order to meet HHS COOP Initiatives.

OMB requires that HHS computer users complete annual IT security awareness training. NLM has completed 100% of the mandatory FY07 Security Awareness Training for employees, contractors and fellows. A new NLM policy requires that new hires complete the training within their first five work days.

Professional Health Information

Unified Medical Language System (UMLS) Project

The Unified Medical Language System (UMLS) provides a common, concept-oriented medical vocabulary and thesaurus based on more than 119 current medical source vocabularies in 17 languages. One version of the UMLS was released in 2007. Version 2006AD was released in November and included updates to 12 source vocabularies. One new source (the Authorized Osteopathic Thesaurus) was added, and the Advanced Suppressibility Option enhancements for the MetamorphoSys® now allows users the option to remove source term type suppressible data, editor assigned suppressible date, or obsolete data. Version 3 of the UMLS licensing system was released in April.

RxNorm Project

As noted above, eight major versions were released this year for the RxNorm Editing System. Implementation of these releases included the ability to search on NDC codes, to duplicate atoms to circumvent ambiguity, and to create a Modified SCD or SBD. Other enhancements included security fixes for blocking SQL injection and updated synonym editing and concept report view. There were 12 monthly releases to RxNorm this year, and inversion and insertion activities consisted of coding, testing and completion of task, to provide RxNorm data to DailyMed for incorporation into the DailyMed Web site. Other features included the addition of term matching on insertion for ingredients and brand names in sources, and coding, testing and completion of task to provide NDC

code attributes on RxNorm atoms. Three resynchronizations were completed, which included updating the resynchronization scripts to include new core tables in migration to new schema and new sources, Gold Standard and FDA Structured Product Labels.

Data Creation and Maintenance System (DCMS)

The major event for DCMS in FY2007 was the baseline extraction, which is a re-release of all DCMS citations that follows the MeSH Year-end Processing (YEP). The first step in this yearly process is sorting and queuing the citations for extraction. A new method for baseline queuing was implemented. It uses a simpler validation query and thus only has to perform a single pass through the baseline queue table. This new method reduced the required time to three hours from FY2006's nine hours. The 2007 MEDLINE/PubMed baseline database contains over 16.1 million records. In addition, OCCS:

- Added support for "Partial" retractions.
- Added support for the new PubChem Databanks.
- Added the ability to identify individual issues in the "Publication Wait" queue for Online Indexing.
- Reduced the results limit from 20 to 10, to speed up "Related Records" searches.

Medical Subject Headings (MeSH) and Related Systems

MeSH includes an inter-lingual database of translations and a system for extending and maintaining them, namely the MeSH Translation Maintenance System (MTMS). In FY2007, the new XML files for the French, Italian and German translations were generated and distributed. The development of the Publication Type Maintenance System for Auto YEP was completed for the GCMS System. Additionally, modifications implemented for the M2000 Maintenance System included cutover for the 2007 MeSH Supplemental Chemical Records (SCR), development and testing of the new Global Record Maintenance System (GRMS), and reduction of the execution time for the processing of HM and EC transactions for YEP by 60% and 40% respectively.

DOCLINE

DOCLINE, the NLM interlibrary loan (ILL) system, supports approximately 3,600 domestic and international libraries in processing more than 2.5 million interlibrary loan transactions a year. Three versions of DOCLINE were released this year. Version 3.0 was released in October, 3.1 in April and 3.2 in August. These releases included 82 enhancements, some of which included:

- Implementation of links to the library's NCBI LinkOut record.
- Integration of an easy upload function, to send invoice information to EFTS.
- Establishing mail delivery as an optional service to encourage electronic delivery.

- Reorganizing the NLM and RML administrators' section for better usability.
- Implementation of increased security of application code, including moving to secure server protocol.

Voyager Integrated Library System (ILS)

This year, two versions of Voyager were released, Version 2006.2 in March and version 2006.5 in September. Version 2006.2 included the implementation of Unicode and version 2006.5 was a release to fix minor bugs. Base pulls containing all data in Voyager were produced in XML format.

Relais

NLM uses the commercial, off-the-shelf Relais system for electronic document delivery and interlibrary loan (ILL) management. Documents requested via DOCLINE are scanned and automatically delivered using the borrower's requested delivery method. Relais was upgraded to version V2006, which includes a new feature to accommodate longer call numbers and citations. Staff members are now able to reopen a processed request and reprocess it in case a customer requests it.

ScanTrac (PubMed Central Inventory)

PubMed Central (PMC) is NIH's free digital archive of biomedical and life sciences journal literature. ScanTrac now has tracking data for 474 journals (53,354 issues) in various stages of scanning for entry into PMC. Several enhancements were made this year, to include allowing special characters for the "Issues" field, adding a confirmation page when adding new issues records, and picking up five additional fields using NCBI's batch status Entrez query.

Literature Selection Technical Review Committee (LSTRC)

Several modifications were made to the Medline Review application, which is used to review journals for inclusion in MEDLINE, NLM's bibliographic database covering the fields of medicine, nursing, dentistry, veterinary medicine, the health care system, and the preclinical sciences. Program modifications were made to display the short form for printing whenever a user attempts to access a journal record by ISSN, item number, NLM ID, etc. Also, OCCS added and modified fields in the system-generated e-mail, and modified tables to increase the length of the rating field.

Serials Extract File (SEF)

This year, numerous upgrades and fixes to the *List of Journals Indexed (LJI)* and *List of Serials Indexed (LSI)*,

the annual publications of SEF data, were completed. In addition, modifications were made in several SEF processes, to complement the new Voyager release. OCCS also created and modified several functions that process data from the ILS (Interlibrary Loan System) in USMARC format, and examined processing of when a serial in the ILS is removed or recovered, so that the change is also accurately reflected in the SEF for Serial Extract Process.

Network and Systems Support

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

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

Public Internet connectivity services to NLM are provided through a contract with Level3. Internet connectivity is provided via an OC3 (155 Mbps) circuit to the Level3 network node in McLean, VA. The contract also provides an OC3 link for CIT/NIH to the Level3 network. NLM and NIH collaborate in using these diverse connections to back up each other's Internet connectivity. The service features an automatic failover, in the event of a scheduled or unscheduled outage of one Internet connection. Contract modifications were made with Level3, to upgrade from the OC3 to a Gigabit connection. The upgrade is expected to be in place by the end of 2007.

In response to the NIH Pandemic Flu Continuity of Operations Plan (COOP), OCCS assisted with the successful load testing of NLM's administrative functions through the Citrix remote access environment and NLM Virtual Private Network (VPN) services, in order to validate NLM's capabilities and capacities to support planning for a possible pandemic flu outbreak.

Internet 2 has become an important resource for connection with NLM and the research community. Internet 2 connectivity is provided by a Gigabit Ethernet link to the Abilene high-speed backbone network via the Mid Atlantic Exchange (MAX) at the University of

Maryland. NLM's Lister Hill Center and OCCS work in cooperation, to manage traffic to and from Internet 2. A redundant, diverse fiber connection from NLM to the MAX, provided by FiberGate, ensures the reliability of this critical network connection.

OCCS continued implementation of the High Availability Computing Solution, to guarantee that critical applications and resources remain available to NLM users. OCCS continued work on clustered Oracle server systems and clustered storage systems as NLM's high availability computing resources. (A server or storage cluster is a group of independent computer systems working together as a single system, thereby allowing multiple servers to deliver the same application services, so that if one of the servers becomes unavailable as a result of failure or maintenance, another server immediately begins providing service.)

OCCS procured new clustered systems for local area load-balancing for the development architecture; these new systems have higher capacity for network traffic and also have more complex load-balancing mechanisms and rules for routing traffic.

Major deployments of the Solaris 10 operating system were initiated this year. This operating system features technology for virtual servers and improved security.

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

Desktop Support

OCCS selected a new desktop management software product called Altiris Client Management Suite (CMS). This software provides unique strengths and capabilities for the management of NLM desktops, including software delivery and PC inventory.

OCCS is planning and testing for enhanced desktop security controls known as the Federal Desktop Core Configuration (FDCC). These controls, which are being developed by the National Institute of Standards and Technology (NIST) in order to meet the OMB 7-11 mandate, are being reviewed for use for all Windows XP and Vista installations. Implementation is slated for 2008.

Partnering with the NLM Personal Computer Advisory (PCA) Committee, OCCS prepared six NLM-wide consolidated orders for desktop IT hardware, based on corporate IT standards and specifications. These orders, collected from across the organization, rationalize the hardware selection process and effectively reduce administrative and IT support costs. All classes of hardware were ordered based on standards and specifications revised and realigned to NLM business needs.

OCCS facilitated and coordinated the NLM rollout of NIH's New Business System (NBS) Phase 3 & 4, Property and Procurement modules at NLM. The NBS

modules were introduced as planned by the NIH NBS team.

Fifty-four new Microsoft operating system security patches that were released this year were applied to the roughly 800 OCCS-managed desktop computers on the NLM network. In addition, security updates and patches from other software vendors are also applied shortly after being released by their publishers. These patches are deployed overnight to NLM desktop systems, to avoid user interruption and minimize downtime. Patches are then validated for effective application.

Outreach

Local Legends

"Local Legends," the companion Web site to the "Changing the Face of Medicine" exhibition, features biographies and video clips of outstanding women physicians who have been nominated by Members of the US Congress for their exceptional contributions in research and education, public health, military service, or patient care. This year, the OCCS Web Team received 27 guestbook entry submissions, which were scanned and approved for inclusion on the Web site. The current number of spontaneous Local Legend Physician Recognition Registry entries is now at 33.

Health Services Research Projects in Progress (HSRProj)

There are currently 7,333 projects in the Ongoing and Completed HSRProj file, as well as 8,791 archived projects. The HSRProj Web site was updated with 166 new records and the HSRProj database was updated with over 1,110 new project records. Additionally, 910 records with the Final Date between 1/2002 and 12/2002 were relocated to the archive file.

Health Services and Sciences Research Resources (HSRR) Database

A search option was added to the HSRR database, which allows direct searching of PHSR (Public Health Services Research) topic subset records from the home page. The team is also in the process of adding functions to allow printing of selected records in various formats.

PHPartners.org & HSR Info Central

As noted above, OCCS re-engineered the Public Health Partners Web site. Version 4 was released this year for both PHPartners and HSR Info Central. The enhancements included creation of an RSS Feed. (RSS or Really Simple Syndication is a format for sharing and distributing Web content. It provides an easy way to stay current with news and new content that is delivered directly to a subscriber as a "feed".)

Research and Development Initiatives

Search Engine

A major accomplishment for this year was the selection of NLM's new search engine, Vivisimo. The first release of the NLM Main Web, MedlinePlus and MedlinePlus/Spanish searches was finalized and installed in preparation for pre-release testing by NLM staff and RMLs. Also, performance testing was conducted to assess the ability of the Vivisimo to support peak search loads.

NLM Digital Repository Project

OCCS continued to support the NLM Digital Repository Project (originally the "Digital Archive Project"), by participating in the Digital Repository Evaluation and Selection Working Group (DRESWG), studying and reviewing digital repository tools and technology that may be applicable to NLM's needs, and reviewing experiences of other libraries and institutions in using digital repository tools. Other key accomplishments included:

- Defining a set of Master Evaluation Criteria for use in down-selecting to a small set of digital repository tools that will be installed and tested at NLM.
- Identifying and weighting detailed functional requirements that will be used in testing and evaluation of the down-selected digital repository tools.
- Studying and evaluating 10 leading open source and commercial digital repository tools. Users of several of the tools were consulted during the evaluation. Each tool was rated using the Master Evaluation Criteria. The top three contenders (DSpace, Fedora, and DigiTool) were selected for installation and testing at NLM.
- Beginning preparations to install the down-selected three digital repository tools at NLM and conduct a comprehensive set of functional tests.

Voice Recognition

Two-way voice communication is an emerging technology that enables users to navigate Web sites by means of spoken user input and to interact with Web applications that synthesize speech from the Web server. This year, OCCS implemented a new NLP (Natural Language Processing) design, which replaces the old "sight-only" (What-You-See-Is-What-You-Say) design. The new design accurately recognizes voice commands. Other efforts include resolving the best-microphone-approach for Macintosh OS-X and improvements in the computation of linguistics and NLP.

Database

OCCS completed the build of the 10g RAC development database. Accomplishments implemented this year included:

- Drafting a plan for moving DCMS to the 10g RAC database.
- Implementing new password management process for Oracle user passwords, to allow regular changing of passwords.
- Rebuilding the QA RAC and Prod RAC databases.

ReportNet Migration

OCCS is migrating from Impromptu and Impromptu Web Reports to the COGNOS product ReportNET. ReportNET will provide decision-support reporting capability across the spectrum of NLM activities. The Binding application is near implementation and many other applications, with the exception of Voyager, are in the midst of making the transition. The ReportNet software was upgraded in the development environment.

NLM Web Support

Web Content Management (TeamSite)

NLM uses TeamSite to provide content and application management for Web sites. Several feature requests and bug fixes to improve TeamSite workflow for site contributors were completed in FY2007. Modifications were made to reduce Archival team's workload by eliminating notification for certain types of content. Also, OCCS implemented standardized contact e-mail lists within TeamSite templates, so that Web contributors can be notified when issues with their pages need resolution.

Web Analytics

NLM uses the WebTrends software package to track the number of pages served over time by the sites being managed, and to provide detailed analysis of trends in site usage, audience composition, and other matters. WebTrends migrated from Enterprise 7.5 to 8.0d Advanced Marketing, offering features for better administration of the application and a new license for 1.5 billion page views was activated.

Link Checker

One of the most important challenges for Web administrators is maintaining the validity of Web links across numerous hosted Web sites. This year, OCCS worked on resolving the XML generation process, to ensure the XSLT (eXtensible Stylesheet Language Transformations) does not fail for Voyager LinkChecker.

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 7,500 NLM ticket requests for IT support were entered and tracked. NLM IT Staff resolved 67% of the calls (5,000 tickets), with 38% of support calls being completed by NIH staff.

OCCS conducted over 10 training courses this year, in topics such as "Outlook Overview," "Basic PC FUNDamentals," and "Remedy". Focused training was provided in support of Daylight Savings Time and XP-SP2 migrations, as well as for the Stay-In-School and NLM Associates' programs. Many one-on-one sessions were conducted in relation to Blackberry functionality and other topics.

Computer Facilities Operations

NLM systems continue to be supported in a safe environment in NLM's computer facility, which is available 24/7/365. The Network Operations and Security Center (NOSC), which was established in 2002, continues to serve as a central point in IT system and service monitoring, IT system administration, IT security event monitoring, and after-hours Help Desk support.

The NOSC display system consists of four 32-inch wide-screen plasma displays that are visible to persons 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 MEDLINE/PubMed, NLM services seen by remote users around the world, and near-real-time utilization data for NLM's Internet-1 and Internet-2 data communications links.

Administrative Support Systems

Customer Service Support System (Siebel)

The OCCS Siebel Team delivered multiple defect resolutions, enhancement releases, and system upgrades for Customer Service, Change Request, and Firewall Service Request Management applications. Key accomplishments included:

- Adding a Block SPAM button to the SR screen to allow the staff to block offending e-mail addresses sending SPAM to customer service accounts.
- Creating a third set of backup e-mail boxes in the CIT e-mail server, for disaster recovery on incoming e-mails.
- Changing the Siebel service windows account passwords, Schema and SADMIN passwords to comply with NIH password policy in Dev, QA and Production.

ADMINISTRATION

Todd D. Danielson
Executive Officer

Table 20

Financial Resources and Allocations, FY2007
(Dollars in Thousands)

Budget Allocation:

Extramural Programs	71,542
Intramural Programs	237,707
Library Operations	-86,200
Lister Hill National Center for Biomedical Communications	-57,680
National Center for Biotechnology Information	-79,325
Toxicology Information	-14,502
Research Management and Support	12,121
Total Appropriation	321,370
Plus: Reimbursements	19,448
Total Resources	\$340,818

Personnel

In October 2006, **Martha Fishel** was appointed as the Chief of the Public Services Division of Library Operations. This Division is responsible for directing programs that provide access to the collections, centralized reference and customer service support, and preservation and collection management. Ms. Fishel has been at NLM since 1976, starting as an acquisitions librarian in the Serial Records Section. For the past 15 years, she was responsible for the Public Services Division budget, contracts, and oversight of document delivery activities. Prior to coming to NLM, she worked at the Department of Interior Office of Hearings and Appeals Law Library in Arlington, Virginia. Ms. Fishel received a BA in English and Sociology from the American University in Washington D.C. and a MLS from the University of Maryland, College Park. She is a member of the Medical Library Association and the Mid-Atlantic Chapter of the MLA.

In October 2006, **Clement J. McDonald, MD**, joined the Lister Hill National Center for Biomedical Communications as its new Director. Dr. McDonald is a distinguished physician and scientist and one of the nation's most accomplished and productive experts in the field of electronic health record systems. He was previously Regenstrief Professor of Medical Informatics at the Indiana University School of Medicine and the Director of the Regenstrief Institute for Health Care, a privately endowed research institute internationally

renowned for the study of health-care quality and economic issues. Dr. McDonald is a member of the Institute of Medicine and recipient of the Morris Collen Award of the American College of Medical Informatics, among many other honors. He currently serves on the Board of the American College of Physicians. He is a past-President the American Medical Informatics Association and a past member of the NLM Board of Regents.

In November 2006, **Irene F. Kim, PhD**, converted to Staff Scientist after having been a ComputerCraft contractor for GEO since 2005. Dr. Kim earned a PhD in Molecular Biology from Johns Hopkins University in 1997 for work that included the identification and partial characterization of a novel kelch-like protein in mammalian systems. Dr. Kim continued her interest in molecular biology as a postdoctoral fellow at Johns Hopkins studying the transcriptional regulation of winter flounder antifreeze genes. She first joined NCBI as a ComputerCraft contractor for GenBank in 1999 where she served as a curator for two years. Following a brief stint at the Stanford Microarray Database, Dr. Kim spent 2 1/2 years as Laboratory Director of the Johns Hopkins/NIDDK Gene Profiling Center before rejoining NCBI in the GEO group in 2005. Dr. Kim will continue to serve at the NCBI/GEO system as a microarray database curator.

In November 2006, **Wendy Wu, PhD**, converted to Staff Scientist after having been a ComputerCraft contractor since 2000. She earned a MS degree in Plant Physiology from Wuhan University and a PhD in Cell Biology from Georgetown University. Prior to coming to NCBI, she did postdoctoral work at NCI and NIDDK. She will continue to work on the Reference Sequence Project.

In December 2006, **Siyuan Chen, PhD**, joined the Communications Engineering Branch as a postdoctoral fellow and is doing research on document image analysis using image processing, pattern recognition and machine learning techniques. Dr. Chen received a doctorate degree from the Department of Electrical Engineering at the State University of New York at Buffalo where his research focused on pattern recognition, computer vision and image processing.

In December 2006, **Kerry L. Zbicz, PhD**, converted to Staff Scientist after having been an MSD contractor since March 2006. Dr. Zbicz received his PhD in physiology and pharmacology from Duke University. His thesis work and postdoctoral research focused on the effects of neuropharmacologic agents on the electrical properties of neurons. Prior to coming to NCBI, Dr. Zbicz worked in the research informatics groups at Pfizer and GlaxoSmithKline and as a consultant for Knightsbridge Solutions. As a staff scientist, Dr. Zbicz will continue to

work on the international deployment of the pPMC and pNIHMS systems.

In January 2007, **Boris Fedorov, PhD**, converted to Staff Scientist after having been a TAJ Technologies contractor since 2004. Dr. Fedorov received his PhD degree in biophysics in 1998 from Moscow State University. In his thesis research, he developed methods of computational analysis of electrostatics of globular proteins. His work experience includes development of DNA sequence clustering and assembling algorithms for Genome Project in Informax Inc., as well as microarray data analysis in Molecular Staging Inc. He will continue working on various projects at NCBI, including developing visual presentation of genome analysis using comparative analysis of microbial genomes, gene distributions and protein sequence comparisons throughout complete prokaryotic genomes.

In January 2007, **John Hurdle, MD, PhD**, joined the Lister Hill Center as a visiting scholar and will stay six months. Dr. Hurdle holds an MD degree from University of Colorado and a PhD degree in Computer Science from University of Utah where he is currently a Research Associate Professor in the Department of Biomedical Informatics. He brings a strong background in statistical natural language processing and field work experience with clinical narratives. At the Lister Hill Center, he is exploiting Natural Language Processing techniques to screen progress notes for adverse drug effect signals.

In January 2007, **Lidia Hutcherson** joined the NCBI as a Technical Information Specialist in the Information Engineering Branch. She will be working with Kathy Kwan in the LinkOut Project. Prior to NCBI, Lidia worked in the Technical Services Division of NLM's Library Operations (2005-2006) and she was part of the NLM Associate Fellow class of 2004-2005. Lidia earned her MLIS from the University of Illinois Urbana-Champaign in 2004.

In January 2007, **Paul Kiehl** joined NLM as the Deputy Executive Officer. Mr. Kiehl was previously with the Department of State's Bureau of Overseas Buildings and Operations where he served over the past four years as a Management Analyst. Mr. Kiehl had diverse responsibilities in the areas of management/program analysis, risk management, congressional inquiries, competitive sourcing, and development of proposals and analyses aimed at securing Congressional support to purchase New Embassy Compound Sites. Prior to his employment with the State Department, Mr. Kiehl completed a successful military career as an Army medical services corps officer that included a number of assignments as an Army Executive Officer. After Mr. Kiehl's 20-year military career he held a senior management position with Irving Burton Associates, Inc., a Department of Defense contractor with emphasis on

health care management. Mr. Kiehl has a Bachelor of Science in Biology from Dickinson College and a Masters in Health Care Administration from Baylor University.

In March 2007, **Joyce Backus** was appointed as Deputy Chief, Public Services Division, Library Operations. This Division is responsible for directing programs that provide access to the collections, centralized reference and customer service support, and preservation and collection management. Ms. Backus has over twenty years experience with the National Library of Medicine. She has management experience at both the team and section level and has excellent knowledge of NLM computer systems, products, and services. She has keen technical knowledge and skills in areas related to web design, product usability, search engine functionality, and content management, and she has demonstrated those skills in her ability to move important projects forward using new technologies. Since July 2004, Joyce has served as Head, Reference and Web Services Division, and from January 2002 until July 2004, she led NLM's Web Management Team in the Public Services Division.

In March 2007, **Dr. Erick Ducut** joined the Office of High Performance Computing and Communications as a postdoctoral fellow. At NLM, he is working with Dr. Paul Fontelo on handheld computing for medical and paramedical personnel. Dr. Ducut has an MD degree from University of the Philippines in Pedro Gil, Manila. He is a Clinical Associate Professor of the Department of Otolaryngology, College of Medicine at the University of Philippines in Manila. His main responsibility there is to provide learning opportunities for medical students and residents. He is part of the R&D group involved in updating their current electronic system and converting patient data into an electronic format.

In March 2007, **Jennifer Marill** was appointed as Deputy Chief, Technical Services Division, Library Operations. This Division is responsible for formulating, implementing, and evaluating the NLM collection development policies, selection of books and materials in biomedicine, and other bibliographic services. Ms. Marill has 18 years of experience in library systems and digital library technology, most recently as a Senior Project Manager at the Library of Congress, overseeing the collection and preservation of "born-digital" web content. She previously worked in the Public Services Division at the NLM for five years on a variety of key projects including the design and development of MedlinePlus, redesign of the NLM public website, and development of a technical strategy to ensure permanent access to NLM's published web content. She holds a bachelor's degree in International Studies from Johns Hopkins University, a Master's degree from the University of Michigan in Russian and East European Studies, and MLS from the University of Illinois, and is a certified Project Management Professional.

In March 2007, **Rebecca Williams, PharmD**, joined the Lister Hill Center as the new Assistant Director, ClinicalTrials.gov. Dr. Williams most recently served as a Senior Regulatory Affairs Scientist in the BioPharma Regulatory Science and Technology group at Science Applications International Corporation (SAIC) in Falls Church, VA. She provided expert regulatory advice to pharmaceutical and biologic companies on promotional advertising and labeling issues. Before becoming a consultant, she was the group leader for professional review group I in the Division of Drug Marketing, Advertising, and Communications (DDMAC) at the Food and Drug Administration (FDA). In this capacity, she managed the review of promotional materials for dental, dermatology, medical imaging, neuropharmacology, non-HIV/antiviral, ophthalmology, radiopharmaceutical, reproductive, urologic, special pathogen and immunologic drug products. Prior to her group leader role, she served as a senior regulatory review officer responsible for dental, dermatology, medical imaging, non-HIV/antiviral, ophthalmology and radiopharmaceutical drug products. She received her PharmD from the University of Wisconsin-Madison and is currently pursuing an MPH at the Bloomberg School of Public Health at Johns Hopkins University.

In April 2007, **David P. Gillikin** was appointed Chief of the Bibliographic Services Division in Library Operations. David has nearly 23 years of experience in libraries and scientific publishing. He has been Head, MEDLARS Management Section, NLM, since 2003, where he has been involved in numerous important areas relating to public access, digital content, the evolution of MEDLINE, and usability and accessibility of databases. Prior to joining NLM, David held positions at HighWire Press where he was responsible for managing their Journal Production department and with the AAAS where he was responsible for the development and management of the Science Online Internet products.

In April 2007, **Avi Kimchi** converted to Staff Scientist after having been a Management Systems Designers (MSD) contractor since 2003. Mr. Kimchi obtained a masters degree in computer science from Hebrew University in Jerusalem, Israel, and has over 20 years of experience leading software development of database and algorithmic applications for various industries. Prior to coming to NCBI, Mr. Kimchi was a Director of global software development at Startec Global Communications, where he supervised the development, deployment and maintenance of customer-service, call-rating, billing, e-commerce and office automation systems. Prior to that, he was a senior systems analyst at Westat Inc, where he led software development projects for clinical studies and post-marketing medical studies. He will continue working on various software systems, including Genome Annotation Pipeline, Clone Registry and Clone-ends alignment.

In April 2007, **Yuri Sadykov** converted to Staff Scientist after having been an MSD contractor since 1998. Yuri obtained a master's degree in Automatic and Electronics, from the Moscow Physical Technical Institute. Prior to coming to NCBI, he was a member of the technical staff at Netscape Communication Inc., where he worked in the server division on development of Netscape Media Server. He will continue working on various server software, including Task Manager, component based generic server framework and web/gui based management tools.

In May 2007, **Margaret Cam, PhD**, joined NCBI as Staff Scientist to help translate user requirements into database applications. Dr. Cam earned her Ph.D. in Pharmacology in 1996 at the University of British Columbia (UBC) in Vancouver, studying the insulin-mimetic effects of vanadium in type I diabetes under Dr. John H. McNeill. She completed her postdoctoral training specializing in molecular biology at the Cardiovascular Research Laboratory, Department of Pathology and Laboratory Medicine at UBC, studying the mechanisms of virus-induced diabetes and estrogen receptor expression in cardiovascular disease. In 2000, she came to the NIH to work in Dr. Sam Cushman's group at the NIDDK investigating human insulin resistance using bioinformatics tools for microarrays. In 2002, she started the Genomics Core Laboratory which processed microarrays and real-time PCR analysis for NIDDK and research collaborators and offered bioinformatics and statistical support for investigators of the institute. At NCBI, she will be working primarily on web performance analysis, application software usability, and support NCBI user liaison efforts.

In May 2007, **Carol DeWeese-Scott, PhD**, was appointed to a Staff Scientist position after having been a Computercraft contractor at NCBI since 2002. Dr. DeWeese-Scott earned a Ph.D. in Bioengineering from the University of Washington for work on major histocompatibility complex (MHC) proteins associated with insulin-dependent diabetes mellitus (IDDM) susceptibility. Prior to coming to NCBI, she was a postdoctoral fellow with Dr. John Moulton at the Center for Advanced Research in Biotechnology (CARB), addressing protein structure prediction methods, with an emphasis on algorithms for use in low sequence identity modeling challenges. She joined Computercraft in May 2002 as a contractor for NCBI, working as a Conserved Domain Database (CDD) curator. As a Staff Scientist, she will continue this work on the Conserved Domain Database and other projects in the Structure Group of CBB.

In May 2007, **Noreen R. Gonzales, PhD**, was appointed to a Staff Scientist position after having been a ComputerCraft contractor at NCBI since 2004. Dr. Gonzales earned her PhD degree in Biochemistry from

Griffith University in Queensland, Australia. Her PhD work involved the cloning and expression of protein phosphatase 2A (PP2A), as well as structural studies using multidimensional NMR and molecular modeling. She did a postdoctoral fellowship at NIH with Dr. Eduardo Padlan in the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), working on antibody engineering. Her work involved the generation of anti-tumor scFv's for use in targeted drug delivery. Dr Gonzales continued her work on antibody engineering with Dr. Syed Kashmiri at the National Cancer Institute (NCI), where she was involved in the generation of humanized and high-affinity monoclonal antibodies for use in cancer therapy. She joined Computercraft in September 2004 as a contractor for the National Center for Biotechnology Information (NCBI), where she joined a team of curators for the Conserved Domain Database (CDD) of the Computational Biology Branch (CBB). As a Staff Scientist, she will continue working on the curation of domain families for the Conserved Domain Database and other projects in the Structure Group of CBB.

In May 2007, **Narmada Thanki, PhD**, was appointed to a Staff Scientist position after having been a ComputerCraft contractor at NCBI for a year. Dr. Thanki earned a Ph.D. in Crystallography from Birkbeck College, University of London, for the study of protein hydration using high resolution crystallographic structures. She then did her postdoc with Dr. Alexander Wlodawer at the Frederick Cancer Research and Development Center working on crystallographic structures of several HIV protease-inhibitor complexes as well as characterization and modeling of numerous signal transduction proteins, followed by a postdoctoral fellowship at EMBL, Heidelberg, Germany, working on protein engineering, modeling, and crystallographic studies of monomerized triosephosphate isomerase. She went on to work as a Research Scientist at the National Institute of Standards and Technology (NIST), curating and data mining for the Protein Data Bank project. Later, she joined SAIC-Frederick as a Computational Scientist to work in the Laboratory of Computational Technologies (Screening Technologies Branch) at the National Cancer Institute, where she developed methods to relate known 3-dimensional macromolecular ligand structures to cytotoxicity data of screened compounds in NCI 60 tumor cell lines, in order to elucidate possible targets. She joined Computercraft in June 2006 as a contractor for the National Center for Biotechnology Information (NCBI), working as a curator for the Conserved Domain Database with Computational Biology Branch (CBB) of NCBI. As a Staff Scientist, she will continue to work on the CDD and other projects in the Structure Group of CBB.

In May 2007, **Kimberly A. Tryka, PhD**, was appointed to a Staff Scientist position after having been a TAJ Technologies, Inc. contractor at NCBI since 2006. Dr.

Tryka earned a PhD in Planetary Science from Caltech. After post-doctoral work at Northern Arizona University and the Jet Propulsion Laboratory she returned to school and received her MLIS from the University of Pittsburgh. Before coming to NCBI she worked at the University of Virginia with numerous digital projects, including electronic editions of historical documents. While a contractor at NCBI she has been working with the "document" side of the dbGaP project, work she will continue in her new position.

In May 2007, **Leonid Zaslavsky, PhD**, was appointed to a Staff Scientist after being a TAJ Technologies, Inc. contractor at NCBI since 2004. Dr. Zaslavsky earned a PhD in computational modeling from Russian Academy of Sciences for work on adaptive multigrid computational methods. He was a post-doctoral fellow at the Weizmann Institute of Science, Israel and Howard Hughes Medical Institute-NYU. As a Staff Scientist, Dr. Zaslavsky will continue to work on NCBI web-based dataset analysis tools. In particular, he will continue improving analysis and visualization capabilities of the NCBI Influenza Virus Resource.

In June 2007, **Larissa C. Brown, PhD**, was appointed to a Staff Scientist position after having been a ComputerCraft contractor at NCBI since 2001. Dr. Brown earned a PhD in Microbiology and Immunology from West Virginia University. She did her postdoctoral research as an NRC fellow at NIH where she studied the regulation of the stationary phase sigma factor RpoS in response to changes in ppGpp levels in *Escherichia coli*. She joined ComputerCraft in February 2001 as a contractor for NCBI and will continue to work as a GenBank Indexer.

In June 2007, **Linda Frisse, PhD**, was appointed to a Staff Scientist position after having been a ComputerCraft, Corp. contractor at NCBI since 2003. Dr. Frisse obtained a PhD degree in Molecular Biology and Biochemistry from the University of Missouri-Kansas City where her research focused on molecular evolution studies in nematodes. Prior to joining NCBI, she was a postdoctoral fellow at the University of Chicago in the Department of Human Genetics where her research focused on human population genetics. As a Staff Scientist, Dr. Frisse will continue working as a member of the GenBank Submissions Staff.

In June 2007, **Lisa Lang** was appointed as the new Assistant Director for Health Services Research Information of the National Library of Medicine (NLM) and Head of its National Information Center on Health Services Research and Health Care Technology (NICHSR). From these two perspectives, Ms. Lang will coordinate the NLM's health services research information program. Prior to joining the NLM, Ms. Lang worked in the Centers for Medicare and

Medicaid Services, where she has been at the forefront of agency activities implementing the public-private collaboration to promote public reporting on hospital quality of care, as well as efforts to identify quality measures in support of value based purchasing. Her prior experience includes more than a decade focusing on quality and consumer protection issues as a Senior Health Policy Analyst in the Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services, as well as a visiting fellowship at the Henry J. Kaiser Family Foundation, Menlo Park, California. She received a B.A. from Stanford University in Psychology and Russian Language and a Masters in Public Policy from the John F. Kennedy School of Government, Harvard University.

In June 2007, **Ivan Ovcharenko, PhD**, was appointed to a tenure-track Investigator position. Dr. Ovcharenko earned a PhD in Physics and Mathematics from the Novosibirsk State University in Russia in 1999 for work on computational modeling of molecular magnets. After that, he did a postdoc with Dr. William A. Lester, Jr. at the University of California, Berkeley working on development of quantum Monte Carlo (QMC) methods for modeling of molecular structures and chemical pathways. This was followed by a postdoctoral fellowship and a Staff Scientist appointment at Lawrence Berkeley National Laboratory (LBNL) working on comparative genomics analysis of the gene regulatory landscape of the human and mouse genomes from 2001 to 2003. Later, he joined Lawrence Livermore National Laboratory (LLNL) as Scientist/Principal Investigator in 2003, where his work on establishing the Dcode Comparative Genomics Center (<http://www.dcode.org>) and on computational analysis of the semantics of the gene regulatory code was supported by several grants from the Department of Energy and the LLNL R&D Program. As an Investigator at the Computational Biology Branch (CBB) of NCBI, he will pursue his work on functional characterization of gene regulatory elements and networks in the human and other complex genomes, developing computational methods to predict tissue-specific enhancers and repressors, and identifying disease-causing mutations within the noncoding part of the human genome.

In June 2007, **Jiyao Wang, PhD**, was appointed to a Staff Scientist position after having been a Lockheed Martin Management Systems Designers (LM MSD) contractor at NCBI since 2006. Dr. Wang earned a Ph.D. in BioPhysics from Stony Brook University for work on binding measurements on protein-membrane interactions. He then did a postdoc with Dr. Diana Murray at Weill Medical College of Cornell University working on modeling and calculations on protein-membrane interactions. This was followed by a postdoctoral fellowship with Dr. Benoit Roux at the University of Chicago working on absolute binding free energy

calculations on protein-ligand interactions. He joined LM MSD in May 2006 as a contractor for the National Center for Biotechnology Information (NCBI) working as a Programmer/Analyst on the PubChem project with Computational Biology Branch (CBB) of NCBI. As a Staff Scientist, he will continue to work on the PubChem and other projects in the Structure Group of CBB.

In July 2007, **Jane He**, was appointed to a Staff Scientist position after having been a Lockheed Martin MSD contractor since 2003 at NCBI. Ms. Jane He earned a Masters degree in Computer Science from Hood College. She then worked as Software Engineer at Gannet Company, Inc. to design and implement the proprietary Telematch system. Later on she worked as Senior Software Engineer at Northrop Grumman Company to maintain and improve the National Biomedical Computing System for American Redcross. Ms. Jane He joined Lockheed Martin MSD in March 2003 as a contractor for the National Center for Biotechnology Information (NCBI), working as a database developer for the new Conserved Domain Database project with Computational Biology Branch (CBB) of NCBI. As a Staff Scientist, she will continue to work on the Conserved Domain Database and PubChem project in the Structure Group of CBB.

In July 2007, **Rogneda Khovanskaya, PhD**, was appointed to a Staff Scientist position at the NCBI after having been a contractor for NCBI at TAJ Technologies, Inc. since March of 2006. Rogneda earned a Masters degree in Mathematics from the Pennsylvania State University. Rogneda is a Ph.D. candidate in Applied Mathematics from the Pennsylvania State University. As a contractor at NCBI, Rogneda worked on NCBI Taxonomy database related projects, such as taxonomy tree comparison, restructuring of external taxonomical databases into NCBI Taxonomy database format, Name Bank project (database design, data model design, server package implementation and client library). She will continue to work on various NCBI Taxonomy database related projects, NCBI Task Manager projects, and generic server framework.

In July 2007, **David Webb, PhD**, was appointed to a Staff Scientist position after having been a Computercraft contractor at NCBI since 2004. Dr. Webb earned a PhD in Phylogenetics and Molecular Evolution from Wayne State University for work on the morphological and molecular evolution of piciform birds. He spent two years as a postdoctoral fellow at the University of Michigan, where he investigated the genetic evolution of primate visual, olfactory, and vomeronasal systems. Dr. Webb will continue to work as a curator of the RefSeq and Genome Projects databases.

In August 2007, **Bonnie L. Maidak, PhD, MS**, was appointed to a Staff Scientist position after having been a

ComputerCraft contractor since October 2000 at NCBI. Dr. Maidak earned a PhD in Medical Genetics from Indiana University School of Medicine and an MS in Library and Information Science from the University of Illinois at Urbana-Champaign. She received a postdoctoral fellowship at the Mayo Foundation. She started a career in biological database curation at the Welch Medical Library at Johns Hopkins, working on both the Online Mendelian Inheritance in Man (OMIM) and Genome DataBase (GDB) databases. Additional curatorial experience occurred at the Ribosomal Database Project (RDP), both at the University of Illinois and Michigan State University. She joined ComputerCraft in October 2000 as a contractor for the National Center for Biotechnology Information (NCBI), working as an indexer for the RefSeq Project. As a Staff Scientist, she will continue to work on the RefSeq Project curating data and to provide user-based feedback in the development of relevant software tools.

In August 2007, **Karanjit Siyan**, was appointed to a Staff Scientist position after having been a TAJ Technologies, Inc. contractor at NCBI since 2003. Karanjit Siyan obtained a MS degree from UC Berkeley and an M.Tech. degree from the Indian Institutes of Technology (IIT). Prior to coming to NCBI he worked on numerous projects in compiler development, network management, software design. He will continue working on various software systems, including XML authoring, custom SharePoint interfaces and workflow development.

In August 2007, **Zhen-Yuan (Jack) Wang, PhD**, was appointed to a Staff Scientist position after having been a Management Systems Designers, Inc. (MSD) contractor at NCBI since 2004. Dr. Wang obtained a PhD in Crop Physiology from the University of Illinois. He did a postdoc with Dr. Ed Korn in the National Heart, Lung, and Blood Institute working on unconventional myosins. He then worked for Celera Genomics, where he did Drosophila, human and mouse genome assembly, annotation and comparative alignment. Prior to coming to NCBI, he worked for the Genetics Branch in NCI as a biostatistician, where he did statistic analysis for microarray data and clinical testing results. He will continue working on various projects, including the WGAANALYSIS database, pre-computation of genophno (type) association, annotation of genotyping and genome-wide scan results and GenBank DNA sequence mapping.

In September 2007, **Marilu A. Hoepfner, PhD**, was appointed to a Staff Scientist position after having been a Kevric/IMC contractor (2002) and later, a TAJ Technologies, Inc. contractor (2005) at NCBI. In 1996, Dr. Hoepfner earned a PhD in Molecular Biology from Northwestern University, Evanston, IL for studies on the cloning and characterization of mouse CCAAT-binding factor. After graduation, she worked as a research

technician in the laboratory of Dr. Ueli Grossniklaus at Cold Spring Harbor Laboratory, assisting in molecular genetic studies of Arabidopsis embryogenesis. She then worked as a free-lance developmental editor, editing books in the Laboratory Manual series for Cold Spring Harbor Laboratory Press. Before coming to NCBI, she worked at Cygnet Risk Group, Inc., a consulting firm serving the energy industry, where she gained skills in web and database development. At NCBI, she will continue to work on the Bookshelf project, focusing on the maintenance of XML data, transformation to alternative data formats and conversion issues, and migration of data to the NLM Book DTD.

In September 2007, **Masato Kimura, PhD**, was appointed to a Staff Scientist position after having been a Management Systems Designers, Inc. (MSD) contractor at NCBI since 2006. Dr. Kimura earned a Ph.D. in Mathematics from the University of California at Davis, and a Master's in Computer and Information Sciences at Hood College. From 1995-2001, he held faculty appointments in the Mathematics departments at the University of Wisconsin at Eau Claire, The College of William and Mary, and Hood College. In 2001, he changed academic fields and became an Assistant Professor of Computer Science at Hood College. He joined MSD in July 2006 as a contractor for NCBI and was assigned as a software developer for the dbGaP project, a project which he will continue to work on as a Staff Scientist.

NLM Associate Fellows Program for 2007-08

The NLM Associate Fellowship Program is an annual internship for recent graduates of Masters Degree programs in library and information science. Fellows receive a comprehensive orientation to NLM programs and services during a structured five-month curriculum phase, and then conduct individual projects over the remaining seven-month period. Projects relate to key NLM programs areas and are typically of a research, development, or evaluation nature. Seven new Associate Fellows begin their year at NLM on September 4, 2007.

Grace Ajuwon is from Nigeria and is participating in the program as an International Fellow. She received her degree in librarianship in 1994 from the University of Ibadan. In Nigeria, she is a reference and information services librarian in the E. Latunde Odeku Medical Library, which is in the College of Medicine at the University of Ibadan. Her responsibilities include reference, database training, Web page development, overseeing information technology in the library, and representing the medical library to faculty boards and committees. Prior to her current position, she was a librarian at the Hezekiah Oluwasanmi Library, Obafemi Awolowo University in Nigeria; and at the African Regional Centre for Engineering Design and

Manufacturing in Ibadan, Nigeria. Her undergraduate training was in Archeology and Anthropology.

Sarena Burgess received her MSIS in May 2007 from the University of Tennessee, Knoxville, with a focus on information sciences. She has experience in reference, consumer health, and web site development at the Preston Medical Library at the University of Tennessee. She also has experience at Tennessee's Special Collections Library, including processing of manuscript collections and applying preservation treatments to fragile materials. Her undergraduate training was in Anthropology.

Kate Flewelling received her MLIS in May of 2007 from the University of Wisconsin, Milwaukee. She has experience as an intern at the Gerrish-True Health Sciences Library, Central Maine Medical Center. Ms. Flewelling also has experience as a middle school teacher and a public library children's librarian. Her undergraduate degree is in History.

Joanna Karpinski received her MLIS in June 2007 from Drexel University, with a focus on health informatics. She has experience as a reference intern at the University of Pennsylvania Biomedical Library, as well as experience in technical and public services as a library assistant in public and academic libraries. Her undergraduate degree is in Political Science.

Brenda Linares received her MLIS in June 2007 from UCLA. She has reference and outreach experience as an intern at the UCLA Louise M. Darling Biomedical Library. She also has intern experience in collection development and cataloging at California State University. Her undergraduate degree is in Finance.

Melissa Resnick received her MLS in May 2007 from Queens College, with a concentration in medical librarianship. She has experience in copy cataloging at the Stony Brook University Health Sciences Center Library. Prior to librarianship, she had 10 years experience as a medical transcriptionist, as well as volunteer experience with a hospital and ambulance service. Ms. Resnick also holds an MS in Biopsychology from the Rensselaer Polytechnic Institute. Her undergraduate degree is in Biology.

Emily Vardell received her MLS from Texas Woman's University in August 2007, with a specialty area in health information services. She has health sciences library experience as a practicum student. In 2005, she received a Fulbright Award to study the Austrian health care system and work as an English language teaching assistant in Austria. Her undergraduate degree was in Biological Chemistry and German Studies.

Retirements and Separations

In March 2007, **Jack Snyder, MD, PhD**, resigned his position as Associate Director for Specialized Information Services (SIS). During his tenure, SIS introduced many important new products and programs, including the Radiation Event Medical Management (REMM) Web Site just announced by the Secretary, HHS. He has been responsible for providing direction and leadership in planning, developing, and administering a national toxicological and environmental health information program. Prior to joining the NLM, he held various executive positions in private industry including serving as Senior Vice-President and Chief Medical Officer for DIANON Systems, Inc., and Medical Director, East Region, SmithKline Beecham/Quest Clinical Laboratories. In addition he has held various academic appointments at the Jefferson Medical College/Thomas Jefferson University.

In June 2007, **Robert Mehnert**, the respected Director of NLM's Office of Communications and Public Liaison, retired after more than 43 years of federal service. In a career that has roughly paralleled the computer age, Mehnert adapted his considerable editorial and literary talents to the rise of mainframes, the Internet and Podcasts, to promote and share the vast knowledge contained in NLM's print and electronic holdings with what has become a global user group. Mehnert came to the Library in 1965 and, after a stint in publishing, returned in 1971 to become Public Information Officer. He managed the press office of the world's largest medical library until his retirement. Among many achievements, Mehnert helped launch MedlinePlus.gov in 1998. He also led publicity efforts on behalf of ClinicalTrials.gov, NIHSeniorHealth.gov, Genetics Home Reference and the Household Products Database, rich online resources that reflect the Library's trend of disseminating the biomedical knowledge of the world, not only to medical professionals but also to the public at large. Though honored with the NIH Director's Award and the NLM Director's Award, perhaps the greatest compliments for Bob Mehnert have come from colleagues who hail him as a talented wordsmith and the professional's professional.

In August 2007, **William T. Hole, MD**, retired after 18 years of employment at NLM. He came to NLM in 1989 with academic credentials including an MD from the Ohio State University College of Medicine, two post-doctoral fellowships at Stanford University, and five years of teaching and research on the faculty of the Department of Pediatrics at Brown University. His unusual combination of full medical training with experience in clinical medicine, academic medicine, and medical informatics made Dr. Hole a natural for his role as a principal in the Unified Medical Language Systems (UMLS) initiative at NLM. Deeply interested in the structure and management

of medical information, Dr. Hole headed the group that created, edited, and produced the Metathesaurus, the knowledge source at the heart of the UMLS. He performed this critical function from the first release of the Metathesaurus in 1990 until the transition of its production to the NLM Office of Computer and Communications Systems late in 2006. His able leadership was evident in the creation of inversion, editing, workflow management, and quality assurance systems for working with disparate vocabularies under a unifying conceptual framework. He worked with NLM colleagues and with talented contractors to develop a software infrastructure that leveraged the efforts of human editors working with this massive resource. His group developed *MetamorphoSys*, the installation, subsetting, and browsing system for the Metathesaurus used by UMLS licensees. Dr. Hole's creative approach, his tenaciousness, and his intellectual rigor in approaching extremely difficult problems will be sorely missed.

In September 2007, **Karen Hajarian** retired after 15 years of employment at NLM. Karen was hired at the time the Library had its legislation enhanced to include promotional activities. Her responsibility was to manage NLM's outreach efforts to expand use of its databases and develop new services to increase its base of users. In the early 1990s, Karen's efforts centered around access to MEDLINE data, but with the Library's decision to provide services directly to patients and their families, her efforts centered around services such as *MedlinePlus* and *InformationRx*. Karen introduced new marketing terms to our working vocabulary and new outreach ideas to members of our National Network of Libraries of Medicine. Her creativity and new ideas will be missed.

In Memoriam

On June 20, 2007, **Vera Hudson**, long time staff member of NLM's Specialized Information Services Division passed away. Ms. Hudson had worked at SIS for twenty four years. As an information specialist and biologist, she helped to define and evaluate the effects of toxic chemicals, and was a project leader for the Hazardous Substances Database and other SIS projects. From 1975 to 1983, she worked for the National Institute for Occupational Safety and Health and focused on the effects of occupational chemical exposure. She also worked on the development and maintenance of the institute's registry of toxic effects of chemical substances. She was a biology graduate of the American University of Beirut and received a master's degree in physiology from the University of Kansas at Lawrence. She will be greatly missed by all her friends at NLM.

Awards

The 2007 NLM Board of Regents Award for Scholarship or Technical Achievement was awarded to William T.

Hole, MD for outstanding leadership of the iterative design and development of successive generations of the UMLS metathesaurus, a major force in advancing biomedical informatics research and the state of the art in medical terminology; and Naomi Miller in recognition of her substantial contribution to the high standards and quality of NLM's consumer health information products.

The Frank B. Rogers Award recognizes employees who have made significant contributions to the Library's fundamental operation programs and services. The recipients of the 2007 awards were Jiwon Kim in recognition of exceptional contributions to the education unit of the History of Medicine Division's exhibition program, by building a network of actively involved K-12 educators and developing innovative educational materials for educators and students; and Dianne Sun for significant contributions to NLM's goal of providing access to quality health information by leading the technical development effort of the *MedlinePlus Go Local* project.

The NLM Director's Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to **Kathleen Cravedi** for her outstanding accomplishments in creating outreach opportunities on a national scale for the library that result in a public that is better informed about health and medicine.

The NIH Merit Award was presented to 6 individuals and 1 group. The individual award recipients were **Walter T. Cybulski** for exceptional contributions to protecting the NLM collection from disaster and improving the collection storage environment; **Laurie A. Duquette** for her dedicated management of systems used by Library Operations to provide access to and management of the NLM collections; **Ronald L. Gordner** for commitment to excellence in customer service at NLM; **Karen A. Hajarian** for outstanding contributions to increasing the visibility of the National Library of Medicine and promoting its products and services to health professionals, patients, families and the public in the US and around the world; **Ruth A. Hill** for exceptional contributions to ensuring the quality of services and materials in NLM's Staff Library which collects journals, books and audiovisuals in subject areas related to Library and Computer Sciences; and **L. Rodney Long** for leading the development of advanced imaging tools for cancer research, specifically the exploration of visual aspects of human papillomavirus and cervical neoplasia. The group award recipients were **Gale A. Dutcher**, **David P. Gillikin**, **John P. Rees**, **Paul H. Theerman, PhD**, and **Frederick B. Wood, DBA** as members of the Native American Jobs Creation Scanning Project Group: "For innovative outreach contributing to economic development and improved health on reservations."

The NIH Director's Award was presented to one group. **Patricia R. Carson, Kathleen G. Cravedi, Donald W. King, MD, Robert B. Mehnert, Naomi Miller, Angela B. Ruffin, PhD, and Elliot R. Siegel, PhD,** received the award in recognition of their initiative in creating the quarterly *NIH MedlinePlus Magazine* that is an invaluable tool for presenting research-based health information to the public.

The EEO Special Achievement Award was presented to **Crystal L. Smith** for actively representing NLM at professional events for librarians of color and serving as a mentor to several NLM staff members who are seeking better careers in the Library.

Andrew Fire, PhD, a member of the Board of Scientific Counselors for the National Center for Biotechnology Information (NCBI) and professor of pathology and genetics at Stanford University School of Medicine, won this year's Nobel Prize in Physiology or Medicine. Dr. Fire shares the prize with Dr. Craig Mello of the University of Massachusetts Medical School. He is part of a team of researchers credited with discovering that certain RNA molecules can be used to "turn off" specific genes in animal cells. The revolutionary discovery of this process, called RNA interference or RNAi, denotes the first time that biologists have been able to selectively "silence" the voice of one gene over the tens of thousands of other genes providing a cell with instructions. Dr. Fire and Mello's description of RNA interference began in *Nature* in 1998, unlocking the door to previously inaccessible areas of research and ushering in a new field of biological study. Researchers all over the globe are currently using RNAi techniques to swiftly and randomly silence one gene at a time within swaths of cells.

The 2007 Marcia C. Noyes Award was presented to **Betsy Humphreys.** The Marcia C. Noyes Award is the highest professional distinction of the Medical Library Association. It was established in 1947 and recognizes a career that has resulted in lasting, outstanding contributions to health science librarianship.

The 2007 Lucretia McClure Award was presented to **Renata Geer.** The Lucretia McClure Award honors a professional in health sciences librarianship education whose work has had an impact on the profession.

The 2007 Miles Conrad Award was presented to **Donald A.B. Lindberg, M.D.**

Table 21

FY2007 Full-Time Equivalents (Actual)

Office of the Director	9
Office of Health Information Programs Development	6

Office of Communication and Public Liaison	8
Office of Administration	40
Office of Computer and Communications Systems	51
Extramural Programs	15
Lister Hill National Center for Biomedical Communications	70
National Center for Biotechnology Specialized Information Services	171
Library Operations	35
TOTAL FTEs	271
	676

NLM Diversity Council

The NLM Diversity Council welcomed new members Robin Hope-Williams, Crystal Smith and Tim Valin. The Council also bid farewell to one of its stalwarts, Donald Jenkins, who left the Library. NLM Public Liaison Officer Kathleen Cravedi shifted from a regular member of the Council to ex-officio status. Continuing on the Council are: Patricia Carson and Melanie Modlin, co-chairs; Helen Ochej, secretary; and Carmen Aguirre, Susan Levine, Elizabeth Mullen and Bryant Pegram. The Council continues to receive support from its other ex-officio members: Todd Danielson, Executive Officer, Mehryar Ebrahimi, Office of Administrative and Management Analysis Services, Pamela Oliver and Blandina Peterson from the NIH Office of Equal Opportunity and Diversity Management, and Nadgy Roey the Program Advisor and Ethics Coordinator for NLM, as well as its alumni.

FY2007 Accomplishments

- NLM Director's Employee Education Fund:*

The NLM Diversity Council continued its coordination of the NLM Director's Employee Education Fund. In FY2007, the Fund enabled 40 staff members to take 62 classes. Of the NLM staff who have taken advantage of the Fund, 25 came from the Division of Library Operations, 6 from the Office of the Director, 5 from the National Center for Biotechnology Information, 2 from Specialized Information Services and 2 from the Lister Hill Center. Undergraduate classes made up the majority of classes supported. The school with the largest number of NLM enrollees was the University of Maryland (16 attendees), with Strayer University coming in second (4), and University of the District of Columbia (3) third. Other institutions attended include: American University (2), Art Institute (1), Averett University (1), Bowie State University (1), Capella University (1), Catholic University (1), Gallaudet (1), George Mason University (1), Howard University (2), Johns Hopkins University (1), Montgomery College (2), Mount Saint Mary (1), Prince George's County Community College (1),

University of Phoenix (1), and USDA Graduate School (1), among others. Course disciplines enrolled in included: business, chemistry, communications, computer networking, economics, and mathematics. In addition to traditional classroom instruction, some students took courses 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 current NLM brochure entitled "Working at the NLM," and posters are prominently displayed throughout Library.

- *"In Motion: The African-American Migration Experience:"* As part of its celebration of African American History Month in February, the Council was pleased to sponsor the exhibition of this powerful collection of images and text, lent by The Schomburg Center for Research in Black Culture, New York Public Library. An opening reception allowed NLM staff to learn the many ways that the lives of African Americans, perhaps more than any other population in the Americas, have been shaped by migration.
- *"Opening Doors" Exhibition Event:* Also in commemoration of African American History Month, NLM's History of Medicine Division mounted an exhibition on African American surgeons of the 20th century, "Opening Doors," in 2007. On April 27, 2007, the Diversity Council sponsored a special program at NLM, featuring pioneering cardiac surgeon Dr. Levi Watkins, Jr., who was featured prominently in the exhibition. Dr. Watkins, professor of surgery and associate dean of the Johns Hopkins School of Medicine, spoke on health disparities. He was joined by the Unified Voice Gospel Choir, also from Johns Hopkins. A reception at the Library followed.
- *Communication of NLM Diversity:* The Diversity Council frequently collaborates 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 proved an excellent method 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 and Spanish Language Training:* The Council continues to support an English-language tutoring program, to enable NLM employees to improve their linguistic proficiency speaking and writing English. Following the model used by local literacy programs, the NLM program offers one-on-one tutoring with NLM staff members

who volunteer their time. Several instructors have received formal training and the course is in place. In addition, the Council has purchased and installed Spanish language software, from the Rosetta Stone company, in the NLM staff library, for employees who wish to learn the basics of Spanish.

- *NLM Learning Circle and Mentoring TEAM.* The Diversity Council continues to support these two worthy programs. The Learning Circle is a self-directed lifelong learning program, offering professional development and group mentoring opportunities for NLM staff and contractors at all levels. Programs in FY2007 ranged in topic from genealogy to race relations to time management, among others. The Mentoring TEAM's monthly book discussion studies and group mentoring activities continued to flourish, as mentors and mentees were paired for the transfer of skills and knowledge. Guest speakers offered workshops on mentoring and professional development topics, and these were opened up to NIH staff and other federal employees. Now, thanks to help from OCCS, NLM staff from off-campus divisions can "attend" Mentoring TEAM and Learning Circle sessions, via Web meeting technology.
- *School Supply Drive:* In 2005, the Diversity Council forged a partnership with Rolling Terrace Elementary School in Takoma Park, MD, and that connection continues today. The school serves a multi-ethnic population of about 700, including a high percentage of low-income students. In August of 2007, NLM staff filled and refilled the collection bins, so that each needy child could have a new backpack loaded with notebooks, pencils, scissors and other back-to-school essentials. This year's collected supplies filled almost 30 large boxes.
- *Winter Clothing Drive:* In November and December 2006, the Council was pleased to collect 55 bags and boxes of coats, hats, gloves and other outer wear, as well as sweaters and other cold-weather gear for The Carpenter's Shelter in Silver Spring. This non-profit, community-based organization provides food, clothing, medical care and other essential goods and services to Montgomery County's neediest residents.
- *Food Drive:* In January-February 2007, the Diversity Council continued its collection momentum by soliciting gifts of non-perishable food items from the NLM staff. Diversity Council staff delivered some eight car loads of food to the Manna Food Center in Rockville, MD. Manna is a non-profit, community-based organization that provides nourishing food to those in need in Montgomery County.

- *Getting to Know NLM, The Sequel:* After the tremendous success of the series by the same name, which the Diversity Council of 2000-01 created in order to spotlight the work of NLM's many divisions, this year's Council thought it was time for a revival.

Under the retooled format, programs on different aspects of the Library, such as international programs, are featured. Another popular session featured medical cartoons from the NLM collection. The series will continue in FY2008.

Appendix 1: Regional Medical Libraries

1. **MIDDLE ATLANTIC REGION**
NYU Medical Center
423 East 23rd St
Floor 15 South
New York, NY 10010
Phone: (212) 263-2030 Fax: (212) 263-4258
States served: DE, NJ, NY, PA
URL: <http://nnlm.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://nnlm.gov/sea/>
3. **GREATER MIDWEST REGION**
University of Illinois at Chicago
Library of the Health Sciences (M/C 763)
1750 West Polk Street
Chicago, IL 60612-4330
(312) 996-2464 FAX (312) 996-2226
States served: IA, IL, IN, KY, MI, MN,
ND, OH, SD, WI
URL: <http://nnlm.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) 587-3412
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://nnlm.gov/scr>
6. **PACIFIC NORTHWEST REGION**
University of Washington
Health Sciences Libraries and
Information Center
Box 357155
Seattle, WA 98195-7155
(206) 543-8262 FAX (206) 543-2469
States served: AK, ID, MT, OR, WA
URL: <http://nnlm.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://nnlm.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:

MORTON, Cynthia, PhD (Chair)
W.L. Richardson Professor
Departments of OB/GYN and Pathology
Brigham and Women's Hospital
Boston, MA

CHABRAN, Richard, MLS
Chair
California Community Technology Policy Group
Chino Hills, CA

COHEN, Jordan J., MD
Professor of Medicine
George Washington University
Washington, DC

CONNOLLY, John E., MD
Professor of Surgery
University of California, Irvine
Orange, CA

FRIEDMAN, Carol, PhD
Professor
Department of Biomedical Informatics
Columbia University
New York, NY

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

ISOM, O. Wayne, MD
Terry Allen Kramer Professor of Cardiothoracic Surgery
New York Presbyterian-Weill
New York, NY

JAMES, Bruce R., Honorable
President and CEO
Nevada New-Tech, Inc.
Incline Village, NV

ROSSITER, Louis F., PhD
Director, Schroeder Center for Healthcare
The College of William and Mary
Williamsburg, VA

STANLEY, Eileen H., MLS
Senior Information Specialist
Ecolab, Inc.
Eagan, MN

Ex Officio Members:

Director
National Agricultural Library

Librarian of Congress

Surgeon General
United States Air Force

Surgeon General
Department of the Navy

Acting Surgeon General
U.S. Public Health Service

Assistant Director for Biological Sciences
National Science Foundation

Acting Under Secretary for Health
Department of Veteran Affairs

President
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:

ASH, Joan S., PhD (Chair)
Associate Director
Department of Medical Informatics
Oregon Health Sciences University
Portland, OR

BAKKEN, Suzanne, DNSC, RN, FAAN
Professor
Departments of Nursing and Biomedical Informatics
Columbia University
New York, NY

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

CHUEH, Henry C., MD
Director and Chief
Laboratory of Computer Science
Division of Biomedical Informatics
Massachusetts General Hospital
Boston, MA

HUFF, Stanley M., MD
Chief Medical Informatics Officer
Intermountain Health Care
Information Systems
Salt Lake City, UT

LUMPKIN, John R., MD
Senior Vice President and Director
Health Care Group
The Robert Wood Johnson Foundation
Princeton, NJ

SHNEIDERMAN, Ben, PhD,
Professor
Department of Computer Science
University of Maryland
College Park, MD

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

Appendix 4: Board of Scientific Counselors/ National Center for Biotechnology Information

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

Members:

GINSBURG, David, MD (Chair)
James V. Neel Distinguished University Professor
Internal Medicine and Human Genetics
University of Michigan
Ann Arbor, MI

LEVINE, Arthur S., MD
Senior Vice Chancellor for Health Sciences
Dean, School of Medicine
University of Pittsburgh
Pittsburgh, PA

MACKAY, Trudy F., PhD
Professor
Department of Genetics
North Carolina State University
Raleigh, NC

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

SALEMME, F. Raymond, PhD
President
Imiplex, LLC
Yardley, PA

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

THOMAS, Annette C., PhD
Managing Director and President
Nature Publishing Group
Macmillan Publishers Ltd.
London, United Kingdom

WENG, Zhiping, PhD
Associate Professor
Department of Biomedical Engineering
Boston University
Boston, MA

Appendix 5: Biomedical Library and Informatics Review Committee

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

Members:

SPACKMAN, Kent A., MD, PhD (Chair)
Professor
Department of Pathology
Oregon Health & Science University
Portland, OR

ARONSKY, Dominik, MD, PhD
Assistant Professor
Department of Biomedical Informatics
Eskind Biomedical Library
Vanderbilt University
Nashville, TN

CONSALES, Judith C., MLS
Director
Louise M. Darling Biomedical Library
University of California, Los Angeles
Los Angeles, CA

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

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

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

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

MANDL, Kenneth D., MD
Associate Professor
Division of Emergency Medicine
Children's Hospital Boston
Boston, MA

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

MENDONCA, Eneida A., MD, PhD
Assistant Professor
Department of Biomedical Informatics
College of Physicians and Surgeons
Columbia University
New York, NY

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

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

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

SALTZ, Joel H., MD, PhD
Professor and Chair
Department of Biomedical Informatics
Ohio State University
Columbus, OH

SHEDLOCK, James, AMLS
Director
Galter Health Sciences Library
Feinberg School of Medicine
Northwestern University
Chicago, IL

STATES, David J., PhD, MD
Professor
Department of Human Genetics
University of Michigan School of Medicine
Ann Arbor, MI

TEMPLETON, Etheldra, MLS
Executive Director
Library & Educational Information Systems
Philadelphia College of Osteopathic Medicine
Philadelphia, PA

TONELLATO, Peter J., PhD
Chief Scientific Officer

Pointone Systems, LLC
Wauwatosa, WI

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

WARD, Deborah, MA, MLS
Director
Health Sciences Libraries
University of Missouri-Columbia
Columbia, MO

ZHOU, Z. Hong, PhD
Associate Professor
Departments of Pathology and Laboratory Medicine
University of Texas Health Science Center at Houston,
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 MEDLINE.

Members:

DELCLOS, George L., MD (Chair)
Professor and Director
Division of Environmental & Occupational Health Sciences
University of Texas Health Science Center
Houston, TX

BAUCHNER, Howard, MD
Professor of Pediatrics and Public Health
Boston University School of Medicine
Boston, MA

CHRISTOPHER, Mary M., PhD
Professor of Pathology
Department of Pathology, Microbiology & Immunology
School of Veterinary Medicine
University of California
Davis, CA

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

DU, Chunying, PhD
Assistant Investigator
Stowers Institute for Medical Research
Kansas City, MO

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

HASHIMOTO, Frederick, MD
Distinguished Professor of Medicine
University of New Mexico
Albuquerque, NM
KAPLAN, Jerry, PhD
Professor of Pathology
University of Utah School of Medicine
Salt Lake City, UT

MANNING, Phil, MD
Professor of Medicine Emeritus
(University of Southern California)
Corona Del Mar, CA

NORTON, Catherine N., MLS
Director, Information Technology
MBL/WHOI Library
Marine Biological Laboratory
Woods Hole, MA

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

SOEHNER, Catherine B., MLS
Director, Art, Sciences and Engineering Libraries
University of Michigan
Ann Arbor, MI

SPANN, Melvin, PhD
Retired
Silver Spring, MD

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

Appendix 7: PubMed Central National Advisory Committee

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

Members:

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

ADLER, Prue S., MS, MA
Associate Executive Director
Association of Research Libraries
Washington, DC

ALIRE, Camila, EdD
Dean Emeritus
University Libraries
University of New Mexico
and Colorado State University
Sedalia, CO

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

GREENSTEIN, Daniel, DPhil
University Librarian
California Digital Library
Oakland, CA

HAWLEY, John, BA
Executive Director
American Society for Clinical Investigation
Ann Arbor, MI

KOHANE, Isaac S., MD, PhD
Director, Information Programs
Children's Hospital
Boston, MA

MICHALAK, Sarah, MLS
Professor
School of Information and Library Science

University of North Carolina
Chapel Hill, NC

PARTHASARATHY, Hemai, PhD
Independent Contractor
San Francisco, CA

RYAN, Mary L., MPH, MLS
Director
UAMS Library
University of Arkansas for Medical Sciences
Little Rock, AR

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

SOBEL, Mark E., MD, PhD
Executive Officer
American Society for Investigative Pathology
Bethesda, MD

VELTEROP, Johannes, PhD
Director of Open Access
Springer Publishing
Guildford, Surrey
United Kingdom

WARD, Gary E., PhD
Associate Professor
Department of Microbiology and Molecular Genetics
University of Vermont
Burlington, VT

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

Appendix 8: Organizational Acronyms and Initialisms Used In This Report

AAHSL	Association of Academic Health Sciences Libraries	CHRIS	Consumer Health Resource Information Service
ACORN	Automatically Creating OldMedline Records for NLM	CHIC	Chickasaw Health Information Center
ACP	American College of Physicians	CIT	Center for Information Technology
ACSI	American Customer Satisfaction Index	CMS	Centers for Medicare and Medicaid Services
AG	Access Grid	COOP	(NIH Pandemic Flu) Continuity of Operations Plan
AHRQ	Agency for Healthcare Research and Quality	CORE	Clinical Observations Recording and Encoding
ALTBIB	Alternatives to Animal Testing	CoreBio	Core Bioinformatics Facility
AME	Automated Metadata Extraction	CPSC	Center for Public Service Communication
AMIA	American Medical Informatics Association	CPT	Current Procedural Terminology
AMPA	American Medical Publishers Association	CRISP	Computer Retrieval of Information on Scientific Projects
AMWA	American Medical Women's Association	CSB	Computer Science Branch
APDB	Audiovisual Program Development Branch	CSI	Commission on Systemic Interoperability
APIRE	American Psychiatric Institute for Research and Education	CSR	Center for Scientific Review
ASCCP	American Society for Cervical Pathology and Colposcopy	CT	Computer Tomography
BAC	Bacterial Artificial Chromosome	CTSA	(NIH Roadmap) Clinical Translational Science Award Centers
BGMUT	Blood Group Antigen Gene Mutation Database	CUIs	Concept Unique Identifiers
BISTI	Biomedical Information Science and Technology Initiative	DART/ETIC	Developmental and Reproductive Toxicology/Environmental Teratology Information Center
BLAST	Basic Local Alignment Search Tool	dbGaP	Database of Genotypes and Phenotypes
BLIRC	Biomedical Library and Informatics Review Committee	dbMHC	Database for the Major Histocompatibility Complex
BMT	Boundary Marking Tool	dbRBC	Database of Red Blood Cells
BOR	Board of Regents	DDBJ	DNA Data Bank of Japan
BSD	Bibliographic Services Division	DCMS	Data Creation and Maintenance System
BSN	Bioinformatics Support Network	DEAS	Division of Extramural Administrative Support
CANDHI	Central American Network for Disaster and Health Information	DHHS	Department of Health and Human Services
CAS	Collection Access Section	DIRLINE	Directory of Information Resources Online
CBB	Computational Biology Branch	DNA	Deoxyribonucleic Acid
CBIR	Content-Based Image Retrieval	DRAGON	Dynamic Resource Allocation in GMPLS Optical Networks
CCB	Configuration Control Board	DTD	Document Type Definition
CCDS	Consensus CoDing Sequence	EBI	European Bioinformatics Institute
CCR	Central Contractor Registration	EBP	Evidence-Based Practice
CCRIS	Chemical Carcinogenesis Research Information System	ECHO	European Community Humanitarian Office
CDD	Conserved Domain Database	Educollab	Educational Collaborators
cDNA	Complementary DNA	EEO	Equal Employment Opportunity
CEB	Communications Engineering Branch	EFTS	Electronic Funds Transfer Service
CgSB	Cognitive Science Branch	EHR	Electronic Health Record
ChemIDplus	Chemical Identification File	EMBL	European Molecular Biology Laboratory
		EMS	Emergency Medical Services

EnHIOP	Environmental Health Information Outreach Program		Standards Development Organisation
EP	Extramural Programs	IHM	Images from the History of Medicine
EPA	Environmental Protection Agency	ILL	Interlibrary Loan
eRA	Electronic Research Administration	ILS	Integrated Library System
EST	Expressed Sequence Tag	InCHIs	IUPAC International Identifiers
ETIC	Environmental Teratology Information Center	IP	Interactive Publications
FDA	Food and Drug Administration	IRIS	Integrated Risk Information System
FHA	Federal Health Architecture	IRMA	Image Retrieval for Medical Applications
FIC	Fogarty International Center	IT	Information Technology
FNLM	Friends of the National Library of Medicine	ITER	International Toxicity Estimates for Risk
FTE	Full Time Employee	ITK	Insight Toolkit
FTP	File Transfer Protocol	ITP	Informatics Training Program
Gbps	Gigabits per Second	IUPAC	International Union of Pure and Applied Chemistry
GDS	GEO DataSet	JDI	Journal Descriptor Indexing
GEO	Gene Expression Omnibus (database)	JRE	Java Runtime Environment
GENSAT	Gene Expression Nervous System Atlas	LactMed	Drugs and Lactation (database)
GENE-TOX	Genetic Toxicology	LAN	Local Area Network
GHR	Genetics Home Reference	LHC	Lister Hill Center
GIS	Geographic Information System	LHNCBC	Lister Hill National Center for Biomedical Communications
GPS	Global Positioning System	LO	Library Operations
GRMS	Global Record Maintenance System	LOINC	Logical Observations: Identifiers, Names, Codes
GSA	General Services Administration	LSTRC	Literature Selection Technical Review Committee
GSS	Genome Survey Sequences	MARG	Medical Article Records Groundtruth
GUI	Graphic User Interface	MARS	Medical Article Records System
HapMap	Haplotype Map	MAX	Mid Atlantic Exchange, U. of Maryland
HAVnet	Haptic Audio Video Network for Education Technology	MDoT	MEDLINE Database on Tap
HBCU	Historically Black Colleges and Universities	MDT	Multimedia Database Tool
HHS	Health and Human Services	MEDLARS	Medical Literature Analysis and Retrieval System
HIPAA	Health Insurance Portability and Accounting Act	MEDLINE	MEDLARS Online
HLA	Human Leukocyte Antigen	MegaBLAST	Basic Local Alignment Search Tool
HL7	Health Level Seven, Inc.	MEME	Metathesaurus Editing and Maintenance Environment
HMD	History of Medicine Division	MEO	Medical Education and Outreach
HSDB	Hazardous Substances Data Bank	MeSH	Medical Subject Headings
HPCC	High Performance Computing and Communications	MHC	Major Histocompatibility Complex
HPV	Human Papillomavirus	MIM	Multilateral Initiative on Malaria
HRSA	Health Resources and Services Administration	MIRS	Medical Information Retrieval System
HSRProj	Health Services Research Projects	MLA	Medical Library Association
HSRInfo	Health Services Research Information	MLAA	Medical Library Assistance Act
HSRR	Health Services and Sciences Research Resources	MMDB	Molecular Modeling DataBase
HSTAT	Health Services and Technology Assessment Text	MMS	MEDLARS Management Section
IAIMS	Integrated Advanced Information Management Systems	MMTx	MetaMap Technology Transfer
ICMJE	International Committee of Medical Journal Editors	MOU	Memorandum of Understanding
ICs	Institutes and Centers (of NIH)	MTHSPL	Metathesaurus Structured Product Labels
IGSTK	Image Guided Surgery Toolkit	MTI	Medical Text Indexer
IHTSDO	International Health Terminology	MTMS	MeSH Translation Management System
		NA-MIC	National Alliance of Medical Image Computing
		NAS	National Academy of Sciences
		NCBC	National Centers for Biomedical

	Computing		
NCBI	National Center for Biotechnology Information	OCPL	Office of Communications and Public Liaison
NCCS	NIH Consolidated Collocation Site	OCR	Optical Character Recognition
NCHS	National Center for Health Statistics	OD	Office of the Director
NCI	National Cancer Institute	OERC	Outreach Evaluation Resource Center
NCRR	National Center for Research Resources	ORF	Original Release Format
NCVHS	National Committee on Vital and Health Statistics	OHIPD	Office of Health Information Programs Development
NEI	National Eye Institute	OMB	Office of Management and Budget
NGI	Next Generation Internet	OMIA	Online Inheritance in Animals (database)
NHANES	National Health and Nutrition Examination Surveys	OMIM	Online Mendelian Inheritance in Man (database)
NHGRI	National Human Genome Research Institute	OMSSA	Open Mass Spectrometry Search Algorithm
NHLBI	National Heart, Lung, and Blood Institute	OSIRIS	Open Source Independent Review and Interpretation System
NIA	National Institute on Aging	PAHO	Pan American Health Organization
NIAID	National Institute of Allergy and Infectious Diseases	PCA	Personal Computer Advisory Committee
NIBIB	National Institute of Biomedical Imaging and Bioengineering	PCR	Polymerase Chain Reaction
NICHSR	National Information Center on Health Services Research and Health Care Technology	PDA	Personal Digital Assistant
NIDCD	National Institute on Deafness and other Communication Disorders	PDR	Publisher Data Review
NIDDK	National Institute of Diabetes, Digestive, and Kidney Diseases	PDB	Protein Data Bank
NIEHS	National Institute of Environmental Health Sciences	PDF	Portable Document Format
NIGMS	National Institute of General Medical Sciences	PDL	Personal Digital Library
NIH	National Institutes of Health	PHLIP	Public Health Law Information Project
NIH PI	NIH Pathways to Independence Award	PHII	Public Health Informatics Institute
NINDS	National Institute of Neurological Disorders and Stroke	PHP	Public Health Partners
NIOSH	National Institute for Occupational Safety and Health	PHR	Personal Health Record
NIST	National Institute of Standards and Technology	PHS	Public Health Service
NLM	National Library of Medicine	PICO	Patient, Intervention, Comparison, and Outcome
NLP	Natural Language Processing System	PLAWARe	Programmable Layered Architecture With Artistic Rendering
NN/LM	National Network of Libraries of Medicine	PMC	PubMed Central
NNO	National Network Office	PMCI	PubMed Central International
NOSC	Network Operations and Security Center	PRS	Protocol Registration System
NOVA	National Online Volumetric Archive	PSD	Public Services Division
NRCBL	National Reference Center for Bioethics Literature	PUG	PubChem Power User Gateway
NSF	National Science Foundation	QCIM	Quarterly Cumulative Index Medicus
NTCC	National Online Training Center and Clearinghouse	RCDC	Research Condition and Disease Categorization
OAM	Office of Administrative Management	RCSB	Research Collaboratory for Structural Bioinformatics
OCCS	Office of Computer and Communications Systems	RefSeq	Reference Sequence (database)
OCHD	Coordinating Committee on Outreach, Consumer Health and Health Disparities	REMM	Radiation Event Medical Management
		RFA	Request for Applications
		RFP	Request for Proposals
		RHIN	Refugee Health Information Network
		RML	Regional Medical Library
		RNA	Ribonucleic Acid
		RNAi	RNA Interference
		RPS-BLAST	Reversed Position Specific BLAST
		RRF	Rich Release Format
		RSS	Really Simple Syndication
		RTECS	Registry of Toxic Effects of Chemical Substances
		RWJF	Robert Wood Johnson Foundation

SAB	Source Abbreviations	TREF	Terminology Representation and Exchange Format
SBIR	Small Business Innovation Research	TRI	The Toxics Release Inventory
SCR	(MeSH) Supplemental Chemical Records	TSD	Technical Services Division
SDK	Software Development Kit	TT	Teaching Tool
SEF	Serials Extract File	TTP	Turning the Pages
SEP	Special Emphasis Panel	UKPMC	United Kingdom PubMed Central
SIDA	Swedish International Development Agency	UMLS	Unified Medical Language System
SII	Scalable Information Infrastructure	UMLSKS	UMLS Knowledge Source Server
SIS	Specialized Information Services	UPS	Uninterrupted Power Supply
SMART	Scalable Medical Alert and Response Technology	VAST	Vector Alignment Search Tool
SNOMEDCT	Systematized Nomenclature of Medicine Clinical Terms	VHP	Visible Human Project
SOAP	Simple Object Oriented Protocol (formerly Simple Object Access Protocol)	VM	Virtual Microscope
SPER	System for the Preservation of Electronic Resources	VPN	Virtual Private Network
SPIN	Shared Pathology Informatics Network	WAI	WebMARS Assisted Indexing
SPIRS	Spine Pathology Image Retrieval System	WashCAS	Washington Area Computer Assisted Surgery
SRA	Short Read Archive	WebMARS	Web-based Medical Article Records System
STB	Systems Technology Branch	WebMIRS	Web-based Medical Information Retrieval System
STTR	Small Business Technology Transfer Research	Web-STOC	Web-Services Technology Operations Center
STS	Sequence Tagged Site	WGS	Whole Genome Shotgun
SVM	Support Vector Machine	WIISARD	Wireless Internet Information System for Medical Response in Disasters
TEHIP	Toxicology and Environmental Health Information Program	WISER	Wireless Information System for Emergency Responders
TIE	Telemedicine Information Exchange	WSD	Word Sense Disambiguation
TILE	Text to Image Linking Engine	WTEC	World Technology Evaluation Center
TIOP	Toxicology Information Outreach Project	XML	Extensible Markup Language
TOXLINE	Toxicology Information Online	XSLT	eXtensible Stylesheet Language Transformations
TOXNET	Toxicology Data Network	YEP	(MeSH) Year-end Processing
TPA	Third Party Annotation (database)		

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

*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: Cover: “dbGaP.” An online database of genotypes and phenotypes introduced by the National Library of Medicine’s National Center for Biotechnology Information in December, 2006. Illustration produced by NIH Medical Arts. Illustration concept by National Library of Medicine employees Stephen Sherry, Ph.D., and Donald Bliss, M.A.