NATIONAL INSTITUTES OF HEALTH

National Library of Medicine

Programs and Services

Fiscal Year 2008

US Department of Health and Human Services

Public Health Service

Bethesda, Maryland

National Library of Medicine Catalog in Publication

National Library of Medicine (US)
National Library of Medicine programs and services.—
Bethesda, Md.: The Library, [1978-]
1977 –
v.:
Report covers fiscal year.
Continues: National Library of Medicine (US). 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.
Reports for 1997 – present issued also online.
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. Title: National Library of Medicine
programs & services III. Series: DHEW publication; no. 78-256, etc. IV.
Series: NIH publication; no. 80-256, etc.

Z 675.M4U56an

CONTENTS

PREFACE

Fiscal Year 2008 saw important advances on many fronts. I continue to be amazed at the vast possibilities that technology offers for meeting the biomedical information needs of our ever-expanding user group, which now includes people in myriad professions, around the globe.

NLM remains focused on the goals of its 2006-2016 long-range plan, including activities in support of interoperable health records, development of a robust knowledge base for personalized health care and more. Among other highlights of this year's *NLM Programs and Services:*

- *NIH MedlinePlus* Magazine, the free quarterly magazine for consumers, is distributed nationwide to doctors' offices, health centers, clinics and hospitals, medical libraries, Congressional offices and a growing list of individual subscribers. Distribution of the magazine has increased from 50,000 copies of each issue in 2006 to over 500,000 copies in 2008. Today it enjoys a readership of over 5 million people nationwide, and many more read this lively publication online.
- FY2008 saw the 25* anniversary of GenBank, a major milestone for the National Center for Biotechnology Information (NCBI). GenBank began as a collection of 600 DNA sequences for the entire year of 1982. Today, it includes roughly 130 million sequences from 300,000 species.
- MEDLINE/PubMed, the Library's flagship database logged almost a billion searches this year, with over a million users daily. In addition, downloading of genomic data from NCBI has more than doubled in the past year to over five terabytes per day — that's like downloading the entire print holdings of the Library of Congress every four days.
- MedlinePlus continues to expand its vast store of consumer health information, in English and en español. This year, it introduced a multilingual feature, with links to health information in more than 40 more languages, on over 260 topics. In FY2008, five new Go Local sites were added, bringing the total to 27 projects in 24 states covering 40% of the US population.
- The History of Medicine Division introduced a new interactive exhibition, "Against the Odds: Making a Difference in Global Health." Showcasing the remarkable achievers and achievements in world health, it will remain on display at NLM until 2010 and then embark on a national tour.
- "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 Paul Berg, Alan Gregg, Arthur Kornberg and Maxine Singer.
- Our Specialized Information Services Division (SIS) created a Disaster Information Management Research Center (DIMRC) Web site, an Enviro-Health Links page to aid emergency responders and a site with resources on hurricanes, floods and disaster preparedness and recovery. They are also developing a new library subspecialty, "Disaster Information Specialist," to train experts in this increasingly important area.

I trust that this report demonstrates that the National Library of Medicine is successfully carrying out its mandate, to be a worthy steward of the world’s medical information in all its forms and to make that knowledge available to all, for the betterment of the public health. To the extent that we are succeeding, the credit goes to our fine staff and to our many capable consultants and advisors.

Donald AB Lindberg, MD
Director

OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT

Elliot R. Siegel, PhD
Associate Director

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

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

Planning and Analysis

The NLM Long Range Plan remains at the heart of the Library'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* is available in print and on the NLM Web site. 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 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 and for senior citizens, Native American communities, and the Spanish-speaking public. An NLM-wide Coordinating Committee on Outreach, Consumer Health and Health Disparities (OCHD) plans, develops, and coordinates NLM outreach and consumer health activities. The OCHD is chaired and staffed by OHIPD.

One activity, the Physician Information Prescription Project ("Information Rx"), was initiated with the American College of Physicians and reported in previous years. In FY2008, NLM initiated a new collaboration with the National Medical Association (NMA), the principal medical organization representing 30,000 African American physicians and 111 affiliated medical societies, to introduce the Information Rx concept to their membership. Other organizations included in the Information Rx project include 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. 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, were successfully concluded and are being prepared for publication. OHIPD staff are also pursuing several outreach initiatives intended to encourage underrepresented minority high school students to pursue careers in medicine and the health sciences, carried out in collaboration with other divisions of NLM. The DeBakey Symposia, commemorating the work of Dr. Michael DeBakey, focus on health topics for high school

students and take place in selected US metro areas, including New York City and Houston, Texas. A new initiative with ExploreHealthCareers.com and Mentoring In Medicine is supporting work to promote sustainable student interest and participation in health professional careers, both by personal intervention and online education resources.

A 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 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 FY2008, 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 analysis of NLM Web site log data; and access to Internet audience measurement estimates based on Web usage by user panels organized by a private sector company.

During FY2008, OHIPD continued to coordinate NLM's use of the online Web user survey known as the American Customer Satisfaction Index (ACSI). The ACSI provides ongoing user feedback to NLM's Web site manager. Also, in FY08, OHIPD senior staff co-authored a peer-reviewed paper on the successful trans-NLM ACIS Project, now complete, that involved about 60 NIH Web sites and about 28 different NIH operating units (including NLM). The paper was published in the *Journal of Medical Internet Research.*

Native American Outreach

In 2008 OHIPD again participated in the NIH American Indian Powwow Initiative to demonstrate the range of NLM information resources for consumer audiences and to enhance awareness of the resources. This included exhibiting at eight powwows, mostly in the Mid-Atlantic area. An estimated 4,000 persons visited the NLM booth over the course of these powwows. 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 prior years, and have continued into FY2008:

- North Dakota—Cankdeska Cikana Community College (via the Greater Midwest RML), Spirit Lake Nation, Ft. Totten, ND, a 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. Current projects include converting thousands of abstracts from AIDS-related conferences from paper and CD ROM format into NLM-approved XML electronic formats, and transcribing taped interviews related to the planned NLM exhibition on "Native Concepts of Health and Illness."
- Hawaii—Papa Ola Lokahi (via the Pacific Southwest RML), two Native Hawaiian Community Health Education Projects were completed:
  - *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 FY2008, OHIPD helped coordinate and facilitate a major consultation in Seattle, WA with tribal leaders on the planned NLM exhibition on "Native Concepts of Health and Illness." A broad spectrum of urban 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 meeting, NLM staff conducted site visits of nearby Indian reservations, Native villages, and historical museums. OHIPD staff participated in the consultations and related outreach activities. This was the third such consultation, with two others having been conducted in FY2007 in Anchorage, AK and Santa Fe, NM.

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 FY2008, NLM continued its support for the South Texas High School for the Health Professions, known as MedHigh, a magnet health high school in the Lower Rio Grand Valley of Texas. The MedHigh VIVA! Peer Tutors Program is an award winning effort to involve high schools students in teaching their peers about online health information. The peer tutors also conduct outreach to the local community and sponsor annual online virtual conferences open to interested faculty, librarians, and students from high schools around the country. In 2008, the Magnet Health High School Initiative was initiated to extend the MedHigh project and replicate the health information peer tutoring program in other magnet high schools in the Lower Rio Grande Valley. MedlinePlus en español is being emphasized where applicable.

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, spent 10 months in Uganda as a Fulbright Scholar. She was based at Makerere University, Kampala, Uganda, where she worked closely with medical students, researchers, professors, and librarians to identify information needs, strengthen medical libraries in Africa, introduce students to the use of NLM databases, and test MedlinePlus tutorials developed specifically for African use in a village context.

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 African tutorials focus on tropical disease issues in developing country contexts. The first two tutorials were 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. Following field testing as part of the medical students' curriculum, the tutorial on malaria was published into poster and booklet formats and deployed by medical students to 20 districts throughout the country through the Faculty of Medicine's Community based Education and Service Program (COBES) in FY2008. The tutorial was also translated into Luo and Japadhola, in addition to Rukiga and Luganda. A workshop featuring the electronic version of the tutorial was carried out with patients at a health center in Mifumi village in Eastern Uganda where the project had conducted a baseline survey of 100 residents regarding their knowledge of malaria. The project leaders reported that the students enjoyed using the tools and were especially pleased on seeing the community's positive response.

MIMCom and Beyond

MIMCom, a project of the Multilateral Initiative on Malaria (*http://www.mimalaria.org/eng/index.asp*) 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 19 research sites in 13 countries, NLM's current focus is on products and databases to aid the efforts of malaria research. NLM continues to provide technical consultation as needed, carried out by African technical experts who were part of the original MIMCom project.

MIMCom News and Web Sites

MIMCom News, a weekly newsletter reaching more than 3000 researchers around the world, has recently been spun off and is now a privately run publication known as *Malaria World*. *Malaria World* provides the latest information on malaria every week to 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 *JAMA*, *BMJ*, *Lancet*, *EH*, and *AJPH*, and the Council of Scientific Editors. NLM contributed to technical capacity building by providing site visits by experienced IT experts from Africa and helping to purchase equipment, including computers, printers, scanners and software. Staff from each African journal visited the offices of its partner journal for one to two weeks. African editors reported these site visits to be extremely useful for observing the editorial and publishing practices of another journal.

With the support of the Partnership Project, African journal editors organized a series of training workshops for editors, authors, reviewers, researchers, and journalists. The workshops provided hands-on experience and lectures emphasizing international standards for writing and a systematic approach for reviewers. International trainers helped facilitate some of these workshops, and an element of training the trainers was incorporated into many of them. Workshops were well attended and feedback has been positive from both participants and facilitators. Some of the editors have already noticed improvements in the quality of their contributors' work. Three of the original four journals are now indexed in MEDLINE. Four new journals have joined the project: *Zambia Medical Journal*, *Ethiopian Journal of Health Sciences*, *Annals of Internal Medicine*, and *New England Journal of Medicine*.

Network of African Librarians

NLM continues its commitment to utilizing and expanding the leadership of a growing network of African librarians who have received training as NLM Associate Fellows or Cunningham Fellows. The objectives of supporting this network is 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. The network currently consists of seven librarians from Kenya, Zambia, Mozambique, Mali, Nigeria, Uganda, and Zimbabwe. Since the initial planning meeting hosted by NLM in May 2007, in FY2008 these librarians have carried out a successful training session at the Association for Health Information and Libraries in Africa (AHILA) 11th Biennial Congress in Maputo, Mozambique in FY2008. Currently, they are planning a meeting in 2009 that will bring together deans and decision makers from their universities for the purpose of constructing and approving a course in medical database searching as a permanent part of required medical school curricula. All of these African 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.

A project with Ethiopian librarians provides training in the use of NLM resources for librarians, undergraduate students, and graduate students at Ethopian universities. This initiative grew out of a three-day workshop organized by NLM in 2006, the first of its kind at the Ethiopian Civil Service College/Global Distance Learning Center in Addis Ababa.

Visitors

In FY08, the Office of Communications and Public Liaison and the History of Medicine Division's Exhibition Program arranged 376 tours — 113 regular daily 1:30 tours and 263 specially-arranged group tours and programs. In FY08, there were 8930 visitors from the following 80 countries:

Afghanistan, Albania, Argentina, Armenia, Azerbaijan, Bahrain, Bangladesh, Bhutan, Bolivia, Botswana, Brazil, Bosnia and Herzegovina, Canada, China, Columbia, Costa Rico, Croatia, Czech Republic, Egypt, Eritrea, Ethiopia, Finland, France, Georgia, Germany, Greece, Guatemala, Haiti, India, Indonesia, Israel, Italy, Japan, Jordan, Kazakhstan, Kenya, Republic of Korea, Kosovo, Kyrgyzstan, Mali, Marshall Islands, Malawi, Macedonia, Mexico, Moldova, Mongolia, Morocco, Myanmar, Namibia, Nepal, Nigeria, Peru, Philippines, Romania, Russia, Rwanda, Saudi Arabia, Scotland, Sierra Leone, Slovak Republic, Spain, South Africa, Sudan, Switzerland, Taiwan, Tajikistan, Tanzania, Thailand, Turkey, Turkmenistan, Uganda, Ukraine, United

Kingdom, United Arab Emirates, United States, Uzbekistan, Vietnam, Yemen, Zambia, 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

NLM's Library Operations (LO) Division is responsible for ensuring access to the published record of the biomedical sciences and the health professions. LO acquires, organizes, and preserves NLM's comprehensive archival collection of biomedical literature; creates and disseminates controlled vocabularies and a library classification scheme; produces authoritative indexing and cataloging records; builds and distributes bibliographic, directory, and full text databases; provides national backup document delivery, reference service, and research assistance; helps people to make effective use of NLM products and services; and coordinates the National Network of Libraries of Medicine to equalize access to health information across the United States. These basic services support NLM's outreach to health professionals, patients, families and the general public, as well as focused programs in AIDS information, molecular biology, health services research, public health, toxicology, environmental health, and disaster planning.

Library Operations also develops and mounts historical exhibitions; provides educational programs with a focus on K-12 education; carries out an active research program in the history of medicine and public health; collaborates with other NLM program areas to develop, enhance, and publicize NLM products and services; conducts research related to current operations; directs and supports training and recruiting programs for health sciences librarians; and manages the development and dissemination of national health data terminology standards. LO staff members participate actively in efforts to improve the quality of work life at NLM, including the work of the NLM Diversity Council.

The multidisciplinary LO staff includes librarians, technical information specialists, subject experts, health professionals, historians, museum professionals, and technical and administrative support personnel. LO is organized into four major Divisions: Bibliographic Services, Public Services, Technical Services, and History of Medicine; three units: the Medical Subject Headings (MeSH) Section, the National Network of Libraries of Medicine Office, and the National Information Center on Health Services Research and Health Care Technology (NICHSR); and a small administrative staff. The activities of all these components receive essential support from a wide range of contractors.

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

Program Planning and Management

LO sets priorities based on the goals and objectives in the NLM Long Range Plan 2006-2016, and the closely related NLM Strategic Plan to Reduce Racial and Ethnic Disparities. In FY2008, LO began the development of a new strategic plan. In November 2007, 40 LO staff representing all Divisions, including supervisors and non-supervisors, met with a consultant for an all-day retreat. As a result of the ideas shared at the retreat, the LO participants identified three major areas of strategic focus and formed working groups to undertake the first phase of the planning process. The groups are: Bibliographic Control; Collections, Preservation and Access to Information; and Workforce for the Future.

In FY2008, LO continued to review and revise policies, procedures, services, and organizational lines to address 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 through the development of an NLM Digital Repository. An LO-wide group which includes members from OCCS and NCBI was created to identify, test, and recommend software for the repository. Work also continued on the Indexing 2015 initiative, an NLM-wide research and development effort to improve indexing performance and productivity which is being led by LO. In the Technical Services Division (TSD) the Cataloging Section was reorganized from four to three cataloging units and the Medical Text Indexer (MTI) was implemented for catalogers in the production environment. In order to ensure that critical services continue to be provided in the event of a disaster, LO became the first NLM Division to begin work on a Continuity of Operations Plan.

Although many LO efforts are devoted to dealing with electronic information and supporting NLM's high priority outreach initiatives, LO must also devote substantial resources and attention to the care and handling of physical library materials and the space and environment for staff, patrons, and physical and electronic collections. Some areas of the collection are already out of space and the remaining space will be completely filled by 2010. While NLM's plans for a new building include increased space for the growing collection, funds have not yet been appropriated for the building. LO, with input from the Office of Administration, developed a plan for expanding existing space. The plan involves strengthening the B2 floor in the collection area as well as the ceiling of B3 below, installing new sprinklers, lighting, and HVAC, and installing compact shelving in

phases. Implementation of the plan will accommodate collection growth until approximately 2030. In FY2008 offsite space in an NIH leased facility in nearby Rockville, Maryland was obtained and 10,460 linear feet of compact shelving was installed. This space will house some of the Library's less used collections in order to create "swing space" to move onsite collections temporarily out of areas to be strengthened and allow infrastructure improvements to be made.

In FY2008, LO's Administrative Office continued to provide guidance and assist managers, supervisors, and staff with a wide range of cross-functional administrative requirements, including acquisition, budget, human resources and other key areas. During the past year, the Administrative Office began phasing in several fundamental changes to day-to-day business practices including regular monthly meetings with each Division, management reports, and assigned roles for the administrative staff who service LO. Library Operations continued to encourage staff to take advantage of flexiplace work arrangements as appropriate. More than 100 LO employees work at home at least one day per week.

Collection Development and Management

NLM's comprehensive collection of biomedical literature is the foundation for many of the Library's services. LO ensures that this collection meets the needs of current and future users by updating NLM's literature selection policy; acquiring and processing relevant literature in all languages and formats; organizing and maintaining the collection to facilitate current use; and preserving the content for subsequent generations. At the end of FY2008, the NLM collection contained 2,594,276 volumes and 9,296,688 other physical items, including manuscripts, microforms, pictures, audiovisuals, and electronic media.

Selection

Selectors worked on a variety of projects to enhance the breadth and depth of the NLM collection. Several new international sources for acquisitions were identified including a new vendor from Taiwan. Notable historical publications identified during the year included materials concerning Japan's biological warfare experimentation during World War II, works on leprosy, and a book on the eradication of malaria in Taiwan. The Joint Collection Development Policy on Veterinary Science and Related Subjects with the Library of Congress (LC) and the National Agricultural Library was revised and posted on NLM's Web site. Collection Development staff worked with an NLM Associate Fellow on a small assessment project to identify Web-published grey literature on disaster management for possible addition to the NLM collection. Because of the time required to search and evaluate these electronic-only resources, TSD will be testing Web harvesting software to assist with the task in the future.

Acquisitions

The Technical Services Division received and processed 160,632 contemporary physical items (books, serial issues, audiovisuals, and electronic media) which is slightly higher than last year's total. Electronic publishing has not yet had a significant impact on the number of physical items that NLM acquires. Net totals of 35,303 volumes and 376,927 other items, including nonprint media, manuscripts, and pictures acquired by the History of Medicine Division (HMD), were added to the NLM collection.

LO uses subscription agents and book vendors to acquire current literature published around the world. In FY2008, TSD worked with the LC East Africa office to expand acquisitions from countries not covered by the arrangement. The LC New Delhi office will also begin to supply NLM with medical and health care titles published by the Institute of Economic Growth (University of Delhi Enclave).

HMD acquired a wide variety of important printed books, manuscripts and modern archives, images, and historical films during FY2008. Among them were two notable books on hospitals: Soeurs de l'Hôpital General & de l'Hôtel-Dieu, *Reglement* (Le Mans, 1696), containing the rules governing the nurses at Le Mans' two main hospitals, and *Avertissement Concernant la Fondation et les Reglemens de l'Etablissment Bienfaisant pour les Malades à St. Petersbourg* (St. Petersburg, 1791), describing the establishment of the first hospital, statistical information, and lists of doctors and pharmacists who worked there. HMD also acquired Hieronymus Bock, *Kreutter Buch* (Strasbourg, 1580), one edition of the great German herbal first published in 1539 which included more than 530 woodcuts. The New York Academy of Medicine also donated several boxes of historical books.

Recently acquired archives and modern manuscript collections included: a Benjamin Rush letter donated by Bruce Fye; the papers of George Palade, who won the Nobel Prize in Physiology or Medicine in 1974; and the papers of John Daley (chemical ecology and pharmacology), Robert Ledley (medical imaging), and Donald Robinson (medical journalism). New prints and photographs acquisitions included six boxes of medical ephemera and pharmaceutical ephemera donated by William Helfand, and a medical stamp collection donated by the widow of Jacob Shanberge.

Preservation and Collection Management

LO carries out a wide range of activities to preserve NLM's archival collection and make it easily accessible

for current use. These activities include: binding, copying deteriorating materials onto more permanent media, conservation of rare and unique items, book repair, maintenance of appropriate environmental and storage conditions, and disaster prevention and response.

In 2008, NLM evaluated the work of two companies who provide mass digitization services for library materials. Several staff visited a shared use installation of the Internet Archive at the Library of Congress. In July, Kirtas Technologies set up digitization equipment at NLM and more than 80 staff and visitors attended demonstrations of their processing.

NLM began a multi-year inventory of the serials collection in FY2006. By the end of FY2008, the contractor, CBase Solutions Inc., had inventoried 17,072 serial titles. Among the completed titles are all journals indexed in *Index Medicus* and *Index Catalogue*, all currently received English-language titles, and the top 500 non-IM/MEDLINE titles requested on interlibrary loan via the DOCLINE system. The inventory, to be completed in FY2010, will make NLM's collection more accessible by providing accurate holdings information at the issue level to staff and onsite users as well as to libraries and individuals worldwide.

In FY2008, LO bound 18,469 volumes, repaired 2,412 items in the onsite repair and conservation laboratory, made 537 preservation copies of films and audiovisuals, conserved 31 prints and photographs and scanned 34 items. A total of 678,438 items were shelved. This figure is a 6% decrease from the FY2007 total of 720,658 items shelved and reflects the decline in interlibrary loan and Main Reading Room requests.

Permanent Access to Electronic Information

NLM's approach to addressing the unique challenges of preserving electronic information is to use its own electronic products and services as test-beds and to work with other national libraries, the Government Printing Office, the National Archives and Records Administration, and other interested organizations to develop, test, and implement strategies and standards for ensuring permanent access to electronic information. LO collaborates with other NLM program areas on activities related to the preservation of digital information.

PubMed Central (PMC), a digital archive of medical and life sciences journal literature developed by NCBI, is NLM's vehicle for ensuring permanent access to electronic journals and digitized backfiles. LO assists NCBI in soliciting participation of additional journals, particularly in the fields of clinical medicine, health policy, health services research, and public health. LO's Public Services Division (PSD) continued to work closely with NCBI to scan and add digitized backfiles of journals depositing newly published articles in the archive. PSD prepares back issues for scanning, ships them to the scanning contractor, and manages the human review portion of the quality control of scanned images, accompanying OCR data, and XML-tagged citations for articles that pre-date current MEDLINE/PubMed coverage. Because the bindings are cut to make scanning more efficient, NLM does not use volumes from its archival collection in this effort. Instead, NLM solicits copies from publishers and other libraries. In FY2008, twelve new titles were added including: *American Journal of Pathology; Annals of the Royal College of Surgeons of England; British Journal of Sports Medicine; British Medical Journal; Journal* [of the] *Boston Society of Medical Sciences;* and *Journal of Medical Research*. By the end of FY2008, 1,149,219 articles were in the PMC database.

Continuing work begun in FY2006 on the development of an NLM Digital Repository, the Digital Repository Evaluation and Selection Working Group (DRG) completed installation and testing of three repository systems—Dspace, DigiTool, and Fedora, and established an internal wiki to share their work. Charters for the Digital Projects Technical Group and Digital Projects Selection Group were also developed.

Vocabulary Development and Standards

LO produces and maintains the Medical Subject Headings (MeSH), a subject thesaurus used by NLM and many other institutions to describe the subject content of biomedical literature and other types of information; develops, supports, or licenses for US use vocabularies designed for use in electronic health records and clinical decision support systems; and works with the Office of Computer and Communications Systems (OCCS) and Lister Hill Center (LHC) to produce the Unified Medical Language System (UMLS) Metathesaurus, a large vocabulary database that includes many vocabularies, including MeSH and several others developed or supported by NLM. The Metathesaurus is a multi-purpose knowledge source licensed by NLM and many other organizations for use in production systems and informatics research. It serves as a common distribution vehicle for classifications, code sets, and vocabularies designated as standards for US health data.

LO represents NLM in federal initiatives to select and promote use of standard clinical vocabularies in patient records and administrative transactions governed by the Health Insurance Portability and Accountability Act of 1996 (HIPAA). In this capacity, LO staff members: serve on the Department of Health and Human Services Data Council; provide staff support to the National Committee on Vital and Health Statistics (NCVHS) Standards and Security Subcommittee; participate in the Consolidated Health Informatics e-government initiative and the Public Health Data Standards Consortium; and contribute to the Federal Health Architecture (FHA) initiative and the 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 clinical terminology standards. In FY2008, NLM worked with the HHS Office of the Secretary and Office of the National Coordinator for Health Information Technology 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 in support of our role as the national coordinating center for clinical terminology standards.

The 2009 edition of MeSH contains 25,186 descriptors and more than 180,447 supplementary records for chemicals and other substances. For the 2009 edition, the MeSH Section added 446 new descriptors, replaced 123 descriptors with more up-to-date terminology and deleted 26 descriptors. Of special note are developments for 2009 MeSH related to fungi and to processes and phenomena. A major revision of the "Fungi" trees was undertaken to reflect consensus opinion regarding higher level fungi taxa (*Mycology Research* 111 (2007) 509-547). Additionally, two tree categories were changed to separate disciplines from processes and phenomena studied by those disciplines. Category G became Phenomena and Processes and Category H became Disciplines and Occupations. These rearrangements were undertaken to improve the usability of the vocabulary in the G and H trees. The previous (pre-2009) organization was established in 1975. At that time non-biological phenomena and techniques were categorized by the discipline named by a descriptor in the H category. The Biological Sciences discipline descriptors were treed in locations in both G and H tree categories. Phenomena, Related techniques and similar descriptors were treed side by side with sub-disciplines. Many descriptors involving concepts related to phenomena, being of interest to more than one discipline, would be in more than one sub-discipline tree. This arrangement was difficult to maintain and to use. In FY2008, an agreement was completed with the NIH Office of Rare Diseases on incorporating rare disease names into MeSH.

Clinical Vocabularies

The MeSH Section and its contractors also produce RxNorm, a clinical drug vocabulary that provides standardized names for use in prescribing. It is released through the UMLS. RxNorm was designated as a US government-wide target clinical vocabulary standard by the Secretary of the Department of Health and Human Services as one of a suite of standards for use in US federal government systems for the electronic exchange of clinical health information. It represents the information that is typically known when a drug is prescribed, rather than the specific product and packaging details that are available at the time a medication is purchased or administered, and provides a mechanism for connecting information from different commercial drug information services. In FY2008, RxNorm editors prepared and released monthly updates to the clinical drug vocabulary. The coverage of more than 33,000 clinical drugs is made up of over 18,000 branded clinical drugs and 15,000 generic clinical drugs. New dose forms were added to RxNorm: Metered Dose Inhaler, Nasal Inhaler, Dry Powder Inhaler, and Prefilled Syringe. RxNorm editors also began creating normal forms for Drug Delivery Devices.

Through LO's National Information Center on Health Services Research and Health Care Technology (NICHSR), NLM supports the continued development and free distribution of LOINC (Logical Observation Identifiers Names and Codes) by the Regenstrief Institute. LOINC is a clinical terminology important for laboratory test orders and results. In 2003, LOINC was designated by the Secretary of the Department of Health and Human Services as one of a suite of standards for use in US federal government systems for the electronic exchange of clinical health information. In 2005, the Secretary also proposed adoption of LOINC as a HIPAA standard for some segments of the Claims Attachment transactions. In 2008, NLM established a new contract to ensure ongoing distribution and maintenance of this important standard.

In FY2008, NLM continued to support and pay the annual fees for the US-wide license for the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT). SNOMED CT is a comprehensive clinical terminology. In 2004, SNOMED CT was designated by the Secretary of the Department of Health and Human Services one of a suite of standards for use in US Federal Government systems for the electronic exchange of clinical health information. In April 2007, the newly established International Health Terminology Standards Development Organization (IHTSDO) assumed ownership and responsibility for maintenance and distribution of SNOMED CT. The mission of IHTSDO is to significantly promote global standardization of health information. NLM, on behalf of HHS, participated in the negotiations and is now the US 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 US 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.

HHS has set a goal for the nationwide implementation of an interoperable health information technology infrastructure to improve the quality and efficiency of health care. Achieving this goal will require that key clinical data elements are captured or recorded in detailed, standardized form (using standard vocabularies,

codes, and formats) as close to their original sources (patients, health care providers, laboratories, diagnostic devices, etc.) as possible. If these standardized clinical data can also be used to generate HIPAA-compliant billing transactions automatically, this will provide another incentive for adoption of clinical data standards. For automated generation of bills from clinical data to become a reality, robust mappings from standard clinical terminologies to the HIPAA code sets must be created. HHS has given NLM the responsibility for funding, coordinating, and/or performing official mappings between standard clinical terminologies and HIPAA code sets. Several mappings are in various stages of development and technical validation.

UMLS Metathesaurus

The MeSH Section and its contractors are responsible for content editing of the UMLS Metathesaurus using systems developed originally by the Lister Hill Center (LHC) and now managed by OCCS. Responsibility for the production of the Metathesaurus rests with LO/OCCS. The MEDLARS Management Section (MMS) plays a major role in Quality Assurance and Documentation, and MeSH continues its supervision and training for Metathesaurus editing. The MeSH staff also assumes responsibility for monitoring vocabulary updates, the Metathesaurus production schedule, vocabulary licenses, and other agreements. Working with OCCS, a Metathesaurus production coordination group began meeting regularly to coordinate the production efforts, including regular review of inversions and insertions of updated and new vocabularies to the Metathesaurus. The Metathesaurus has shifted to twice yearly releases in response to user requests. The releases in 2008 took place in April and November.

Bibliographic Control

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

Cataloging

LO catalogs the biomedical literature acquired by NLM to document what is available in the Library's collection or on the Web and to provide cataloging and name authority records that minimize the cataloging effort required in other health sciences libraries. Cataloging is performed by TSD's Cataloging Section, staff in HMD, and contractors. The Cataloging Section is responsible for the NLM Classification, coordinates the development and maintenance of the standard NLM Metadata schema for Web documents, and also performs name authority control for selected NLM Web services.

In FY2008, the Cataloging Section cataloged 21,507 contemporary books, serial titles, nonprint items, and cataloging-in-publication galleys, a 5% decrease from the previous year. The Section implemented the use of the Medical Text Indexer (MTI) for catalogers in the production environment. With the Library of Congress and the National Agricultural Library, NLM issued a joint statement regarding plans to test *Resource Description and Access*, the anticipated successor to *AACR2* upon its release in 2009. NLM, LC and NAL will be testing and evaluating RDA prior to making a decision about implementation of the new code.

The Cataloging Section released the 2008 online edition of the massive *NLM Classification*. Twenty-six new class numbers were added to the schedules and 72 MeSH terms were added to the index. The WA (Public Health) and WB (Practice of Medicine) schedules were reviewed and no major revisions were necessary.

Significant progress was made in providing cataloging records for NLM's historical and special collections. HMD cataloged 4,298 early monographs, an increase of 28% over the previous year. HMD also cataloged 243 linear feet of manuscripts, 34,745 pictures, and 3,732 audiovisuals.

Indexing

LO indexes 5,319 biomedical journals for the MEDLINE/PubMed database to assist users in identifying articles on specific biomedical topics. A combination of Index Section staff, contractors, and cooperating US and international institutions indexed 671,904 articles in FY2008, bringing the total number of MEDLINE citations to over 15 million. Publisher-supplied electronic records accounted for 89% of the citation data added this year. Previously indexed citations were updated to reflect 178 retracted articles, 7,756 published corrections, and 33,238 comments found in subsequently published notices or articles.

In FY2008, indexers created 74,252 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 was a 43% increase over last year.

LO continues to work with other NLM program areas to identify, test, and implement ways to reduce or eliminate tasks now performed by human indexers. A successful Web-based Indexer Training Tool has enabled training of new indexers as the need arises rather than at infrequently scheduled classroom training sessions. The

time experienced indexers spend on training has also been reduced. One-on-one mentoring of trainees is done rather than classroom teaching. The Index Section continues to increase the number of articles indexed from the online version of journals. At the end of FY2008, 1,861 journals were indexed from an online version, including online-only journals and those with a print version. Freeing the print version for immediate use onsite and for fulfilling interlibrary loan requests, indexing from an online version has also enabled the elimination of duplicate print issues, saving the cost of processing and shelving those issues.

Indexers perform their work after the initial data entry of citations and abstracts has been accomplished. Over the past 10 years, great strides have been made in improving the efficiency of data entry. By the end of FY2008, more than 87% of all citation data entry consisted of XML-submitted data from publishers. The remaining citations were created by scanning and optical character recognition (OCR). A total of 46,093 more citations were received from publishers compared with the previous year for a grand total of 625,386 XML citations.

NLM selects journals for indexing with the advice of the Literature Selection Technical Review Committee (LSTRC) (Appendix 6), an NIH-chartered committee of outside experts. In FY2008, LSTRC reviewed 423 journals and rated 116 of them highly enough for NLM to begin indexing them.

Information Products

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

Databases

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

Use of MEDLINE/PubMed, which now includes over 18 million citations, registered 776 million searches in FY2008. The implementation of Entrez 2.0 in June of 2007 resulted in changes to the statistical reports that do not allow comparison with the FY2007 data. LO extended the coverage of PubMed to include another 26,000 citations derived from the PubMed Central Back Issue Scanning Project and 5,000 abstracts were added to existing PubMed citations. Additional progress was made in adding or enhancing citations to older articles. Seventy-seven percent of OLDMEDLINE subset citations have original key words mapped to current MeSH. More than 55,000 citations from 1949 *Current List of Medical Literature* (CLML) were added to PubMed. BSD staff also assisted NCBI with the design, development, and testing of many enhancements to PubMed, notably a Revised Automatic Term Mapping that broadens retrieval, a Citation Sensor that detects bibliographic information and performs a citation search rather than a subject search, the addition of an Advanced Search page, and the inclusion of PMCID and MID (manuscripts identifiers) in various displays. In addition, multiple experimental "Discovery" features highlight links to related information such as the Patient Drug Information links to MedlinePlus on the PubMed AbstractPlus display and the Drug Sensor links to a Bookshelf publication on the Summary display.

Use of MedlinePlus and MedlinePlus en español as measured by the number of unique visitors continued to increase to more than 174 million from a total of 156 million last year. The number of page views cannot be compared with FY2007, as several changes occurred which affect the counts. To conform to accepted industry practice, NLM stopped counting redirects to information outside MedlinePlus. In addition, a new search engine was implemented and the search results page was redesigned. Both changes were made to get users more quickly to the information they seek. MedlinePlus and MedlinePlus en español continue to receive high ratings from customers in the American Customer Satisfaction Index (ASCI), ranking among the top government news/information sites. Subscribers to the MedlinePlus Listserv now number more than 104,000.

PSD and OCCS continued to expand and improve the content and features of the English and Spanish versions of MedlinePlus. At the end of FY2008, the site featured 771 health topics in English and 733 in Spanish. A collection of health information in 43 languages was released covering 262 health topics. In addition, 10 new disaster topics were added and access to dermatology information and images was improved by adding links to more than 200 VisualDxHealth pages.

Under the direction of NICHSR, NLM continues to expand and enhance its databases for health services researchers and public health professionals. The number

of serials on topics related to health services research that are indexed in MEDLINE continued to increase during FY2008. NICHSR worked with NCBI to add Centers for Disease Control and Prevention publication *Health US 2007* to the Entrez Bookshelf. Other additions to HSTAT (Health Services and Technology Assessment Text) on the Bookshelf included 54 new documents produced by the Agency for Healthcare Research and Quality, primarily evidence reports, technology assessments, and evidence syntheses.

NICHSR continued to contribute to the field of health services research through its support of NLM databases containing information about newly-funded research and accessible data collection tools and resources. With the assistance of AcademyHealth and the Sheps Center at the University of North Carolina, Chapel Hill, the content of HSRProj (Health Services Research Projects in Progress) continued to expand incorporating work funded by additional foundations, states, and other organizations. Organizations contributing data for the first time in FY2008 included the American Diabetes Association, National Cancer Institute of Canada (NCIC), and the Nova Scotia Health Research Foundation. A new search interface that enables users to search states using geographic information systems (GIS) enhancements was made available in FY2008.

The Health Services Research Resources (HSRR) database also continued to expand to cover additional datasets, surveys, other research instruments, and software packages used with datasets. In addition to working with the librarian, research, and policy communities to improve the usability and value of its HSR Web offerings, NICHSR oversaw the publication of two new user guides: *HSRProj and HSRR: Tools for Developing Research Funding Priorities* and *HSRProj and HSRR: Tools for Conference Planning.*

DailyMed, a Web site which presents high quality information about drugs, including the FDA approved packaging information (labels) for drugs, grew throughout the year to approximately 4000 labels. The DailyMed site provides links to several other sources of drug information, including information available through MedlinePlus, trial information from ClinicalTrials.gov, and literature searches through PubMed. The Web site receives over one million uses per month with an average time per page view of more than 7 minutes.

MyMedicationList, a prototype medication list manager for patients, has been developed. The manager uses input from patients and stores the list in standard vocabulary (RxNorm) in a standard format (the HL7 Continuity of Care Document.) Users are aided in finding the correct names and dosages with the assistance of images and drop down lists. The medication list itself is stored in the patient's computer or hard drive, assuring privacy for the patient. Testing of MyMedicationList in a multicultural environment was begun at the Queens Public Library.

Machine Readable Data

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

Demand for MEDLINE/PubMed data in XML format continues to increase. At the end of FY2008, there were 513 licensees of MEDLINE data, a 17% increase over the previous year. The majority use the data for research and data-mining.

At the end of FY2008, there were 4,143 UMLS licensees, an increase of 36% over the previous year. NLM services and support to licensees of UMLS were enhanced by the introduction of the UMLS Community Web page for resource sharing; improved Quality Assurance reports; the implementation of a UMLS News RSS Feed; the presentation of six Web casts; the first Quick Tour for distance education on how to use the MetamorphoSys RRF Browser Restrict Searches and Views Option; the first Users' Meeting at the AMIA in November 2007; and the launch of the new UMLSKS beta version of the Knowledge Source Server.

Web and Print Publications

NLM's databases and Web sites are its primary publication media. In FY2008, NLM's main Web site use showed 11.7 million unique visitors, a 26% increase in visitors over the previous year's total. Clustering of search results was added to the NLM main Web, MedlinePlus and MedlinePlus en español Web sites. Publications available on the main Web site include recurring newsletters and bulletins, fact sheets, technical reports, and documentation for NLM databases. BSD's MEDLARS Management Section edits the NLM *Technical Bulletin*, which provides timely, detailed information about changes and additions to NLM's databases and related policies, primarily for librarians and other information professionals. Published since 1969, the *Technical Bulletin* also serves as the historical record of the evolution of NLM's online systems and databases. In FY2008, the separate Technical Notes section was discontinued, a search box was added to the homepage, and Vivisimo was implemented as the search engine.

NLM added four new segments to the Profiles in Science Web site during FY2008, bringing the total to 31. The new sites focus on Paul Berg, a biochemist who received the 1980 Nobel Prize in Chemistry for his protein synthesis and rDNA work; Alan Gregg, a

Rockefeller Foundation officer who profoundly influenced medical education and research; Maxine Singer, a pioneering molecular biologist, an influential science administrator, and a leader in science policy and advocacy; and Arthur Kornberg, an American biochemist who made outstanding contributions to molecular biology through his research on enzymes. Launched in September 1998, Profiles in Science promotes the use of the Internet for research and teaching in the history of biomedical science by making widely available archival collections of leaders in biomedical research and public health. Published and unpublished materials appear on the site, including books, journal volumes, pamphlets, diaries, letters, manuscripts, photographs, audiotapes, and video clips. Altogether the program has digitized more than 179,000 pages of material.

In June 2008, NLM released a redesigned NIHSeniorHealth Web site, featuring a new homepage with improved navigation and content organized by categories as well as alphabetically. NIHSeniorHealth added five new topics in FY2008: Parkinson's Disease (National Institute for Neurological Disorders and Stroke); Eating Well as You Get Older (National Institute on Aging); Kidney Disease (National Institute of Diabetes and Digestive and Kidney Diseases); High Blood Cholesterol (National Heart Lung and Blood Institute); and Participating in Clinical Trials (NLM).

In FY2008, LO staff continued to be involved in the two publications designed for patients, families, and the public. The Director's Comments podcasts bring current health news to listeners. Four issues of *The NIH Medline*Plus Magazine were published in print and online in FY2008.

Direct User Services

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

Document Delivery

LO retrieves documents requested by onsite patrons from NLM's closed stacks and also provides interlibrary loan as a backup to document delivery services available from other libraries and information suppliers. In FY2008, PSD's Collection Access Section processed 488,769 requests for contemporary documents, an 8% decrease from FY2007. HMD handled 11,456 requests for rare books, manuscripts, pictures, and historical audiovisuals, a 7% decrease from the previous year.

The number of onsite users registering to use the collection continued to decline by another 11% from last year and use of NLM's collection by users in the Main Reading Room was also down 12% from the previous year's total to 205,178. Users of the HMD Reading room requested 10,631 items from the historical and special collections, a decrease of 6% from last year. Paid printing at Main Reading Room workstations also decreased 6%, from 365,909 pages in FY2007 to 341,703 pages in FY2008, a trend that reflects the increased availability of electronic journal content.

The Collection Access Section (CAS) received 283,464 interlibrary loan requests, a 5% decline from FY2007 with a fill rate of 83%. The number of requests processed in 12 hours declined slightly to 96%, with 98% processed within one day of receipt. NLM continues to deliver 96% of interlibrary loan requests electronically. CAS conducted a study of 4,000 requests that were "not on shelf" to determine why the items were not retrieved. Results of the findings showed that 39% of the requests were available and should have been retrieved; the remaining 61% were not available. Of these, most had no item record or were charged out.

A total of 3,102 libraries use DOCLINE, NLM's interlibrary loan request and routing system. DOCLINE users entered 1,969,656 requests in FY2008, a 6% decline from last year; 93% of requests were filled. DOCLINE requests are routed to libraries automatically based on holdings data. At the end of FY2008, the holdings database contained 1,528,307 holdings statements for 56,785 serial titles held by 3,027 libraries. In FY2008, individuals submitted 429,339 document 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 fulltext journals.

NCBI and the staff at the Regional Medical Libraries continued to promote the use of PubMed's LinkOut for Libraries and Outside Tool, the open-URL services that allow libraries to link directly from PubMed to a wide range of resources beyond the Entrez system. Using these tools, libraries can create custom displays of their electronic and print holdings for their primary clientele. The number of libraries participating in LinkOut increased by 8.5% in FY2008 to 2,013; there are 529 libraries participating in the Outside Tool option, an increase of 16% over last year.

NLM and the Regional Medical Libraries continued to encourage network libraries to use the Electronic Funds Transfer System (EFTS), operated for the NN/LM by the University of Connecticut, as a mechanism to reduce administrative costs associated with interlibrary loan service billing. During FY2008, EFTS participation increased to 1,299 participants. Participants receive either a single net consolidated bill or a net consolidated payment each month.

Reference and Customer Services

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

A total of 93,599 inquiries were received in FY2008, up 4.7% from FY2007. The number of onsite inquires declined 18% to 8,124, reflecting the decline in the number of onsite users. The number of remote inquiries increased 7.5% to 85,457 with the overwhelming majority arriving via e-mail.

PSD conducted a survey with users of the Main Reading Room on service satisfaction; the last one was conducted in FY2000. Users completed 186 surveys providing a 90% confidence rate with an error margin of + or -6%. The 2008 survey recorded the highest satisfaction scores yet with 85% of users giving the highest marks for service.

Outreach

LO manages or contributes to many programs designed to increase awareness and use of NLM's collections, programs, and services by librarians and other health information professionals, historians, researchers, educators, health professionals, and the general public. LO coordinates the National Network of Libraries of Medicine which attempts to equalize access to health information services and information technology throughout the United States; serves as secretariat for the Partners in Information Access for the Public Health Workforce; participates in NLM-wide efforts to develop and evaluate outreach programs for underserved minorities and the general public; produces major exhibitions and other special programs in the history of medicine; and conducts training programs for health sciences librarians and other information professionals. LO staff members give numerous presentations, demonstrations, and classes at professional meetings and publish articles that highlight NLM programs and services.

For the fifth 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. AFP participants from Kenya, Mali, Malawi and Mozambique were joined by a librarian from University of Ibadan in Nigeria who was a Fellow in the FY2007-2008 Associate Program. 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 continued in FY2008: *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*. This year, two of the participating journals, *Malawi Medical Journal* and *Mali Medical* were accepted for indexing in MEDLINE. *African Health Sciences* was accepted for indexing a few years ago.

National Network of Libraries of Medicine

The NN/LM works to provide timely, convenient access to biomedical and health information for US health professionals, researchers, and the general public irrespective of their geographic location. With more than 5600 full and affiliate members, the Network is the core component of NLM's outreach program and its efforts to reduce health disparities and to improve health information literacy. Full members are libraries with health sciences collections, primarily in hospitals and academic medical centers. Affiliate members include some smaller hospitals, public libraries, and community organizations that provide health information service, but have little or no collection of health sciences literature. LO's NN/LM Office (NNO) oversees network programs that are administered by eight Regional Medical Libraries (RMLs) under contract to NLM. In FY2008, a major initiative was the development of a national NN/LM Emergency Preparedness Plan. Activities under this initiative included the development of a toolkit for use by network libraries in preparing for and responding to emergencies, conducting a tabletop exercise to test the plan, and visits to each RML to conduct emergency preparedness and response training.

The Regional Medical Libraries assisted in the identification and initial funding of several Go Local projects in FY2008. Go Local was expanded with the release of five new sites, bringing the total to 27 sites in 24 states, with partial coverage in the states of Colorado, Ohio, and Texas. Go Local now covers 40% of the US population. The Go Local Evaluation Study, completed this year, yielded suggestions for improving the Go Local input system, the Go Local public pages and the Go Local participant experience.

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. Virtually all of these projects involve partnerships between health sciences libraries and other organizations, including public libraries, public health departments, professional associations, schools,

churches, and other community-based groups. In FY2008, the NN/LM initiated 173 outreach projects which target rural and inner city communities and special populations in 35 states and the District of Columbia.

With the assistance of other NN/LM members, the RMLs do most of the exhibits and demonstrations of NLM products and services at health professional, consumer health, and general library association meetings around the country. LO organizes the exhibits at the Medical Library Association annual meeting, the American Library Association annual meeting, some of the health professional and library meetings in the Washington, DC area, and some distant meetings focused on health services research, public health, and history of medicine. In FY2008, NLM and NN/LM services were exhibited at 37 national and 256 regional, state, and local conferences across the US. 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 13-member public-private agency collaboration initiated by NLM, the Centers for Disease Control and Prevention, and the NN/LM in 1997 to help the public health workforce make effective use of electronic information sources and to equip health sciences librarians to provide better service to the public health community. The NICHSR coordinates the Partners for NLM; staff members from the National Network Office, SIS, and the Office of the Associate Director for Library Operations serve on the Steering Committee, as do representatives from several RMLs.

The Partners Web site (PHPartners.org), managed by NLM with assistance from the New England RML at the University of Massachusetts, provides unified access to public health information resources produced by all members of the Partnership, as well as other reputable organizations. In FY2008, the Web site was expanded with more than 650 new links. Two new topic pages on Nutrition and Workforce Development were also added to the Web site. One of the most popular resources on the site is the Healthy People 2010 Information Access Project (HP2010 IAP). For every focus area of Healthy People 2010, the IAP resource includes four or more objective-specific evidence-based PubMed search strategies and links to MedlinePlus topics. The PubMed search strategies for all 28 focus areas were checked and revised with MeSH 2008 vocabulary changes. Search strategies were developed and posted to the Web site for Objective 23-13: *Increase the proportion of Tribal, State, and local health agencies that provide or assure comprehensive laboratory services to support essential public health services.* In addition, in FY2008, NICHSR awarded one new purchase order under the Partnership to the Association of Schools of Public Health to manage the national tour of *Against the Odds*, NLM's exhibition on global health, at schools of public health.

Special NLM Outreach Initiatives

LO participates actively in the Library's Committee on Outreach, Consumer Health, and Health Disparities and in many NLM-wide outreach efforts designed to expand outreach and services to the public as well as to address racial and ethnic disparities. PSD staff, together with the National Center for Minority Health and Health Disparities (NCMHD) created a special MEDLINE/PubMed search and resources page on minority health. The search retrieves over 57,000 citations to literature in many languages and research from many countries. The resources page also provides links to other disparate resources, such as links to US Government agencies, associations, foundations, research centers, and grant information.

In FY2008, BSD continued to work with other NLM components and the NN/LM to provide ongoing support to the Information Rx Project by maintaining the InformationRx.org Web site for ordering materials. Information Rx provides physicians with materials to write prescriptions for information from MedlinePlus for their patients. Users have placed 6,375 orders for these materials since the program began in 2004.

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 on display, *Against the Odds: Making a Difference in Global Health*, looks at the revolution in global health that is taking place in towns and cities around the world. It presents a look at the public health problems posed by Hurricane Katrina. It showcases the barefoot doctors program, which trained over 1 million young people to treat the common ailments of residents of rural China in the 1960s and 1970s. The exhibition also profiles a campaign for oral rehydration in Bangladesh that was so successful that it has been adopted in Afghanistan as well. In another example of nation-to-nation collaboration, *Against the Odds* shows how the Pholela Health Center in South Africa inspired the community health center movement in the US.

The exhibition opened on April 16 with a ribbon cutting and special program featuring several of the persons portrayed in the exhibition. "The National Library of Medicine has long been more of an ***international*** library than a national one," observed NLM Director Dr. Donald A.B. Lindberg. "The fact that the Internet has

made our vast holdings accessible to people around the globe certainly helps in the fight against the complex and widespread health challenges facing the world today," he continued. "*Against the Odds* captures many of the successes in world health policy, such as the eradication of smallpox, and cautions us about potential pitfalls, like the kind of discrimination that can take place when people don't understand the facts about the transmission of HIV/AIDS."

"People all over the world share a commitment to a better life and a healthier future for all," commented Dr. Elizabeth Fee, chief of NLM's History of Medicine Division. "This exhibition highlights some of their achievements as well as the challenges that remain, and encourages each of us to join the fight for health and human rights." As with many exhibitions, *Against the Odds* features online educational resources for K-12 teachers and students. Designed by secondary and post-secondary educators, the education component consists of instructional templates for lesson plans and seminar modules, online activities, and research resources. One activity helps students explore diverse careers and/or health information related to the exhibition.

Previous NLM exhibitions live on through heavily used Web sites, printed catalogs, DVDs, and touring traveling versions. The traveling version of *Changing the Face of Medicine: Celebrating America's Women Physicians*, funded by the NIH Office of Research on Women's Health and NLM, continued touring libraries in the US through a collaborative arrangement with the American Library Association.

NLM is using a flexible and inexpensive banner format to expand its traveling exhibition services. *Against the Odds* is one exhibition that was adapted to this banner format. Launched at the annual meeting of the American Public Health Association (APHA), the exhibition won best in show out of 550 entries. Working in conjunction with the National Information Center on Health Services Research and Health Care Technology (NICHSR) and the Association of Schools of Public Health (ASPH), the traveling banner exhibition will be featured at schools of public health and public and university libraries across the country.

Besides the major exhibitions mounted in the rotunda, HMD installed several mini-exhibits. *Everyday Miracles: Medical Imagery in Ex-Votos* examined the relationship between faith and healing in Italian and Mexican traditions. *Most Horrible and Shocking Murders: An Exhibition of "True Crime" Murder Pamphlets in the Collection of the National Library of Medicine* featured murder pamphlets, which have been hawked on street corners, town squares, taverns, coffeehouses, newsstands, and book shops for more than five centuries. The murder pamphlets in the National Library of Medicine were mostly collected in the mid- and late 19th century and generally deal with cases catalogued under the subject category of "medical jurisprudence" or "forensic medicine." *Health for the People: Continuity and Change in Asian Medicine* featured materials from NLM's diverse collection of East Asian historical materials, including items from the Library's 20th century Chinese public health collection and materials on traditional medicine from Japan, Korea, Tibet, and Mongolia dating from the 17th to the 20th centuries. *To Paint the Lily: A Selection from NLM's Illustrated Herbals, 1486-1793* included seven of the Library's most important historical botanical works featuring images of lilies, including Leonhard Fuchs' *De Historia Stirpium* (Basel, 1542) and Elizabeth Blackwell's *A Curious Herbal* (London, 1737).

To accompany a banner exhibition in the NLM lobby, *Harry Potter's World: Renaissance Science, Magic, and Medicine*, HMD featured a case display of the historical works discussed in the exhibit. Although a fantasy story, the magic in the Harry Potter books is partially based on Renaissance traditions that played an important role in the development of Western science, including alchemy, astrology, and natural philosophy. Incorporating the work of several 15th- and 16th-century thinkers, the seven-part series examines important ethical topics such as the desire for knowledge, the effects of prejudice, and the responsibility that comes with power. Plans are underway to travel *Harry Potter's World* in conjunction with the American Library Association later in 2009.

During May and June 2008, NLM presented a six-lecture series, *Genomics in Perspective*, which explored this complex and often confusing issue for an audience of scientists, physicians, policy makers, and the general public. For some, genomics promises a radical and abrupt transformation in medical practice; others suggest that the new genetics has not and will not revolutionize the way common diseases are identified or prevented. Some welcome genomics as ushering in a golden age of new and more effective treatments, better diagnostic interventions, and more powerful means of biological investigation through bioinformatics, genetic analysis, measurement of gene expression, and determination of gene function. Others caution against over-optimism, and point to the importance of culture, society and history to an understanding of the complexity of interaction between biology, genes, and environment. Featuring a lecture by a historian or social scientist, a response by a physician or scientist, and an audience discussion period, the six sessions stimulated discussion of the social, historical, and cultural meanings and uses of genomics.

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

Training and Recruitment of Health Sciences Librarians

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

In FY2008, the MEDLARS Management Section (MMS) and the NTCC trained 983 students in 65 classes covering PubMed, the NLM Gateway/ClinicalTrials.gov, TOXNET, and the UMLS. A new edition of the PubMed training workbook was completed which includes the redesigned My NCBI feature. A new edition of the NLM Gateway and ClinicalTrials.gov workbook was also completed. MMS continues to offer some training via distance broadcasts. In July, a 30-minute online Search Clinic: PubMed Update on Automatic Term Mapping, Citation Sensor, and Advanced Search was attended by 235 participants and later made available on the NLM Web site. An average of about 16,000 unique users visited the Web-based PubMed Tutorial about 23,000 times each month. Three new animated Viewlet QuickTour tutorials for targeted PubMed search features were created: "Retrieving Your My NCBI Username or Password," "Creating Your Bibliography," and "Changing Saved Searches."

The UMLS for Librarians course and the UMLS Tutorial continue to represent some of the NLM training courses useful in preparing librarians for new and expanded roles. Although NCBI made the decision to discontinue offering the "Introduction to Molecular Biology Information Resources" course, it was recorded for ongoing public use. NICHSR continues to add to its suite of courses on health services research, public health, and health policy. A continuing education course on the Community Health Status Indicators was taught to 12 librarians at the Medical Library Association meeting on May 17. The four-hour course sponsored by NICHSR, was taught by staff of the Public Health Foundation and University of Michigan.

The NLM Associate Fellowship program had 11 participants in FY2008: four second-year fellows at sites across the country (the University of Michigan, Massachusetts Institute of Technology, Uniformed Services University of the Health Sciences, and Yale University) and seven first-year fellows, who completed their year at NLM in August 2008. Included in the first year group was an international fellow from the E. Latunde Odeku Medical Library, College of Medicine, University of Ibadan, Nigeria. Second-year placements were arranged for five of these fellows at NLM, Inova Fairfax Hospital, Vanderbilt University, University of Miami, and University of North Carolina, Chapel Hill. Four new fellows began the 2008-09 Associate Fellowship year at NLM in September. Efforts to recruit fellows from underrepresented groups have been successful in attracting diverse groups of fellows to the program, including Hispanic representation in the 2007-08 and 2008-09 cohorts.

NLM works with several organizations on librarian recruitment and leadership development initiatives. Individuals from minority groups continue to be underrepresented in the library profession and a high percentage of current library leaders will retire within the next five to 10 years. LO has provided support for minority student scholarships available through the American Library Association, the Medical Library Association, and the Association for Research Libraries (ARL). LO also supports the NLM/AAHSL Leadership Development Program which provides leadership training, mentorship, and site visits to the mentor's institution for an annual cohort of five mid-career health sciences librarians. AAHSL contracts with ARL for the leadership training portion of the program. Recruitment efforts have emphasized and been successful in attracting minority candidates. In FY2008, NLM began participating in an ARL-directed Institute of Museum and Library Services (IMLS) grant to recruit minority students enrolled in graduate library and information science programs to work in research libraries.

Table 1

Growth of Collections

Collection	*Previous Total (9/30/07)*	*Added FY2008*	*New Total (9/30/08)*
Book Materials			
Monographs:			
Before 1500	594	1	595
1501-1600	6,005	13	6,018
1601-1700	10,284	26	10,310
1701-1800	24,775	45	24,820
1801-1870	41,760	50	41,810
Americana	2,341	0	2,341
1871-Present	789,296	16,177	805,473
Theses (historical)	288,091	0	288,091
Pamphlets	172,021	0	172,021
Bound serial volumes	1,345,922	21,641	1,367,563
Volumes withdrawn	-112,116	-2,650	-124,766
Total volumes	2,558,973	35,303	2,595,276
Nonbook Materials			
Microforms:			
Reels of microfilm	148,768	123	148,891
Number of microfiche	456,781	0	456,781
Total microforms	605,549	123	605,672
Audiovisuals	81,884	1,953	83,837
Computer software	2,549	1	2,560
Pictures	69,285	0	69,285
Manuscripts	8,161,457	374,850	8,536,307
Nonbook items added	8,920,734	376,927	9,297,661
Nonbook items withdrawn	-973	-0	-973
Total nonbook items	8,919,761	376,927	9,296,688
Total book & nonbook items	11,478,734	412,230	11,890,964

Note: Audiovisuals total for FY2007 corrected from 81,974 to 81,884.

Table 2

Acquisition Statistics

	FY2006	*FY2007*	*FY2008*
Serial titles received	20,815	20,165	20,901
Publications processed:			
Serial pieces	134,020	136,543	138,893
Other	22,368	21,905	21,739
Total	156,388	158,448	160,632
Obligations for:			
Publications	$8,715,869	$10,329,902	$8,778,048
(For rare books)	-$337,386	-$402,153	$299,764

Table 3

Cataloging Statistics

	FY2006	*FY2007*	*FY2008*
Completed Cataloging	21,662	22,520	21,507

Table 4

Bibliographic Services

Services	*FY2006*	*FY2007*	*FY2008*
Citations published in MEDLINE	623,000	670,900	671,904
Journals indexed for MEDLINE	5,020	5,194	5,319
Journals indexed for Index Medicus	4,416	4,530	4,660
Total items archived in PubMed Central	755,860	1,115,778	1,683,664

Table 5

Consumer Web Services

Services	*FY2006*	*FY2007*	*FY2008*
NLM Web Home Page			
Page Views	61,000,000	63,600,000	54,600,000
Unique Visitors	9,000,000	9,300,000	11,700,000
MedlinePlus			
Page Views	820,000,000	906,000,000	754,000,000
Unique Visitors	95,000,000	120,000,000	132,300,000
ClinicalTrials.gov			
Page Views	129,673,811	479,926,975	563,956,116
Unique Visitors	6,105,822	6,188,562	10,466,607
DailyMed			
Page Views	842,977	9,187,248	16,436,561
Unique Visitors	96,522	1,298,775	1,795,937
Genetics Home Reference			
Page Views	20,181,044	39,829,732	27,252,298
Approximate Unique IPs	1,397,406	1,906,055	2,772,477
Household Products Database			
Page Views		12,457,074	15,142,059
Unique Visitors		1,143,883	1054,447
Tox Town			
Page Views	7,271,692	5,134,230	5,039,964
Unique Visitors	261,938	170,360	224,762

Table 6

Circulation Statistics

Activity	*FY2006*	*FY2007*	*FY2008*
Requests Received	573,828	530,656	488,769
Interlibrary Loan	328,661	298,616	283,591
Onsite	245,996	232,040	205,178
Requests Filled	467,143	442,731	405,475
Interlibrary Loan	265,562	246,902	234,020
Onsite	201,581	195,829	171,455

Table 7

Online Searches—PubMed and NLM Gateway

Data is processed completely differently from past years, and should not appear with previous data. They are not comparable.

PubMed Statistics	*FY2008*
Page Views:	4,100,051,973
Interactive Sessions*	229,571,548
Searches	775,504,557

* Interactive Sessions are a set of hits that represent one person's activity in a day on our system.

Gateway Statistics	*FY2007*	*FY2008*
Page Views	4,230,910	6,391,604
Unique Visits	312,690	709,259
Searches	746,219	1,128,932

Table 8

Reference and Customer Services

Activity	*FY2006*	*FY2007*	*FY2008*
Offsite requests	76,582	79,502	85,457
Onsite requests	15,202	9,919	8,142
Total	91,784	89,421	93,599

Table 9

Preservation Activities

Activity	*FY2006*	*FY2007*	*FY2008*
Volumes bound	19,317	18,684	18,583
Volumes microfilmed	1,509	Ceased in 2006	
Volumes repaired onsite	1,814	2,313	2,142
Audiovisuals preserved	863	318	537
Historical volumes conserved	154	163	82

Table 10

History of Medicine Activities

Activity	*FY2006*	*FY2007*	*FY2008*
Acquisitions:			
Books	1,401	938	583
Modern manuscripts	783 (lin. ft.)	372 (lin. ft.)	820 (lin. ft.)
Prints and photographs	9,945	839	13,226
Historical audiovisuals	1,864	476	4
Processing:			
Books cataloged	3,952	3,339	4,298
Modern manuscripts cataloged	374 (lin. ft.)	1,276 (lin. ft.)	243 (lin. ft.)
Pictures cataloged	6,342	1,274	34,745
Citations indexed	856	5,134	2778
Public Services:			
Reference questions answered	23,818	22,237	22,506
Onsite requests filled	14,140	12,588	13,183

SPECIALIZED INFORMATION SERVICES

Dr. Steven Phillips
Associate Director

The Division of Specialized Information Services (SIS), National Library of Medicine (NLM), creates information resources and services in toxicology, environmental health, chemistry, and HIV/AIDS. SIS also has an Outreach and Special Populations Branch, which seeks to improve access to quality and accurate health information by underserved and special populations. This year a new Office for Disaster Information Management Research was established in SIS to manage the new NLM Disaster Information Management Research Center (DIMRC).

The Toxicology and Environmental Health Information Program (TEHIP), known originally as the Toxicology Information Program, was established more than 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 underserved communities through implementation of innovative information access-enabling approaches and dissemination of NLM's resources. This past year has seen the beginning of many disaster information resource activities as the new DIMRC was established.

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, special population health information, and disaster information resources; 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 FY2008, SIS continued to balance efforts to enhance and re-engineer existing information resources with efforts to provide new services in emerging areas. such as disaster information resources. 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 2008 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 FY2008 customer surveys 86% of the respondents reported that they would "return to this site" and "recommend it to others." In FY2008, enhancements to TOXNET were based on user feedback/requests and routine upgrades/additions of data and capabilities. Databases in TOXNET include: Drugs and Lactation (LactMed), which 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, and is now expanding its coverage of chemical compounds of interest in monitoring potential terrorist activities.
- **IRIS** (Integrated Risk Information System), a database from the US Environmental Protection Agency (EPA) containing carcinogenic and non-carcinogenic health risk information on over 542 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 655 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 US 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 3.57 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 1987-2006 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 290,000 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. In collaboration with the Department of Labor, tasks and chemicals associated with work at the Department of Energy hazardous sites. are now included in Haz-Map.
- **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 4.1 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 during FY2007. Usage among first responders and others continued to grow with over 47,000 downloads of WISER onto PDA's (Palm and Pocket PC) and Windows-based desktop/laptops over FY2008. Total number of WISER downloads is now over 150,000.
- **REMM** (**R**adiation **E**vent **M**edical **M**anagement**)** 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 of 2007. REMM, a system intended for use by emergency physicians and related emergency health care providers, 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 FY07, the REMM system was released, following peer review by experts in the radiation domain, and changes based on user feedback. A version of REMM for mobile platforms was released in 2008. Plans have been made to begin a similar collaborative project focusing on medical management in mass-casualty incidents involving chemicals.
- **Tox Town** was enhanced with new content (in English and Spanish) in the neighborhoods of Tox Town, Tox City, Tox Farm, Tox Port and a US Mexico Border scene. A new information resource on potential toxicity of gardens and plants was added this year. 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 4.02.01 in FY2008, featuring interface improvements and names and locations of hospital that appear when users zoom to the detailed map.
- **Enviro-Health Link** pages continue to be useful to our users, especially the *Dietary Supplements* page, with links to many sources of relevant information and the Pesticide *Exposure* page with links to Web sites 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 FY06. 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. This year a prototype museum kiosk featuring ToxMystery was developed with the Carnegie Science Center in Pittsburgh.

Disaster Information Management Research Center

The National Library of Medicine (NLM) has a long history of providing health information during times of disaster. Recognizing the potential for the use of libraries as major untapped resources during disasters, and responding to the current increased need for disaster health information, the *NLM Long Range Plan 2006-2016* called for the creation of a Disaster Information Management Research Center (DIMRC) to aid the nation's disaster management efforts. DIMRC is tasked with the effective collection, organization, and dissemination of health information for natural, accidental, or deliberate disasters.

In FY2008, the Office of the Disaster Information Management Research Center (ODIMRC) was formed within SIS to serve as the coordinating office for disaster health information management activities across NLM. A trans-NLM working group was established to coordinate disaster activities across NLM. ODIMRC began a number of new initiatives this fiscal year including the development of a Web site (http://disasterinfo.nlm.nih.gov) and joining the Bethesda Hospitals' Emergency Preparedness Partnership, along with the NIH Clinical Center, Suburban Hospital Healthcare System and the National Naval Medical Center.(NNMC). NLM is coordinating the research and development initiatives of BHEPP and currently is working on eleven projects ranging from investigating a variety of back-up communication systems to the development of a family reunification system, and the use of virtual reality for training emergency personnel. In addition, NLM participates in the annual disaster drill coordinated by NNMC.

In an effort to explore ways in which librarians can help emergency personnel prepare for, respond to, and recover from disasters, ODIMRC initiated a pilot Disaster Information Specialist project with four diverse medical institutions, including military, government, university, and a community hospital library. In response to requests by other libraries to participate in the project, a Disaster Information Outreach Listserv was established and now has several hundred members exchanging comments and ideas. As part of the NNO effort to encourage libraries to develop relationships and agreements with other libraries to assist in times of disaster, the Bethesda Medical Libraries Emergency Preparedness Partnership, consisting of five medical libraries, was established.

As described under the Toxicology and Environmental Health section, NLM continued to promote and enhance several resources for emergency medical personnel, including WISER, REMM, and several Enviro-Health Link Web guides on disaster topics (Wildfires, Hurricanes, Biological and Chemical warfare agents, etc.). SIS also began investigating the development of a tool to screen for Post Traumatic Stress Disorder and Traumatic Brain Injury. In late FY2008, SIS began discussions with HHS' Assistant Secretary for Preparedness and Response on the development of a Chemical Hazard Emergency Medical Management (CHEMM) system. CHEMM would be similar to REMM and focus on the medical diagnosis and management of exposure to chemical threat agents and provide much needed guidance to medical personnel, most of whom do not have extensive training or knowledge in this subject.

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 2008 score for AIDSinfo is 84, which places is among it at the top government Web sites. The ACSI was also implemented for InfoSIDA in both Spanish and English.

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 FY2008, NLM made 12 awards. The decrease was due to a lower level of available funding.

Evaluation Activities

In FY2008, SIS conducted professional evaluation of a number of its Web products and outreach programs. Several of these were funded via the NIH Evaluation Set-Aside Program mechanism. These include a usability study of the Asian American Health Web portal; two focus groups with health consumers and professional users of Asian American Health Web sites; two focus groups with health consumers and professional users of American Indian Health Web site and one focus group with professional users of Refugee Health Information Network (RHIN). The set-aside mechanism was also used to fund development of a detailed evaluation plan for the NLM/UNCFSP-HBCU Access Project to promote disease prevention and wellness at Historically Black Colleges and Universities (HBCU). In FY2008, SIS also submitted to the Evaluation Branch a proposal for conducting six focus groups assessing elementary and middle school teachers' needs in locating and using Web resources for teaching environmental health. The project was approved by the Evaluation Branch in FY2008, with the allocation of 2009 resources. User satisfaction with TOXNET, AIDSinfo and InfoSIDA continued to be measured by the American Customer Satisfaction Index (ACSI) surveys. Over the past three years, other SIS Web products have also been professionally evaluated: Radiation Event Medical Management (REMM) Web site, Tox Mystery, TOXMAP, 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 FY2008 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 FY2008 and awarded 4 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 Mr. Dwayne Ashley, CEO, Thurgood Marshall College Fund, and additional presentations by Dr. Yvonne Maddox, Deputy Director, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Dr. Janine Smith, Deputy Director NIH Office of Research on Women's Health, and Dr. Sidney McNairy, Associate Director, National Center for Research Resources.

Adopt-a-School program with Woodrow Wilson Senior High School, Washington, DC, encourages students to take an active interest in consumer health and promotes interest in science. In addition to providing summer internship for Wilson students, this year NLM also had summer interns from Flowers High School in Prince George's County, MD.

The mission of the **Environmental Health Information Partnership (EnHIP)** 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. NLM hosted an EnHIP meeting in January 2008. The Environmental Health Information Outreach Program (EnHIOP) changed its name to the Environmental Health Information Partnership to signify a new relationship with NLM firmly based in bi-directional partnership. EnHIP also developed a strategic plan based upon NLM's new strategic plan. EnHIOP meetings included representation from 14 HBCUs, three tribal colleges and three Hispanic-serving institutions.

Chickasaw Health Information Center (CHIC), a project that was initiated by the Sacred Root Tribal Information Fellows from the Chickasaw Nation, is fully operational. It is located in the Carl Albert Indian Health Facility in Ada, Oklahoma. The room has two workstations and a printer along with brochures and other NLM and CHIC materials. It is staffed full-time by a trained tribal member. In addition, the CHIC has a mobile kiosk that is moved to various clinics for use there. CHIC has developed oversized prescription pads with links to MedlinePlus topics specific for 10 different clinics in the facility. The topics were chosen by the directors of those clinics. NLM also sent staff to present a one-day program to middle and high school students in the Nation's science summer camp.

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. The partnership was expanded with the addition of the Association of Refugee Health Coordinators. A members only section of the Web site was developed to support discussion of refugee issues and to review new materials. A project carried out by an NLM Associate Fellow resulted in useful information about information needs and seeking related to new refugee populations. A RHIN symposium was held, which provided input into the further

development of the site as well as in outreach to promote the site to potential users.

SIS maintains several special population Web resources in collaboration with organization that provide expertise about the particular population groups. The most recent additions to these are the Women's Health Information Resources developed with the NIH Office of Research in Women's Health. The American Indian Health Web site, Asian American Health Web site, and the Arctic Health Web site continue to be redesigned and expanded.

Minority Health Professional and Health Information 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. NLM also worked with NMA to initiate the Information Rx project with their membership. 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.

NLM continues to reach out to minority communities and the librarians who serve them through collaborations with the American Library Association's Office of Literacy and Outreach Services. One of these projects launched the "Good Health Information @ Your Library" site to assist in introducing library staff and communities in rural areas to NLM resources, specifically MedlinePlus and Tox Town. Other presentations, also designed to introduce NLM or health resources to minority communities, were delivered at the American Library Association Annual Conference and at the Reforma National Conference, and for the first time a health track of nine programs was incorporated into the programming of the National Diversity in Libraries Conference.

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 US 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, and Panama and Costa Rica. The CANHDI 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 14,000 full-text documents available online. During FY2008, each center received a small amount of funding from NLM to enhance their disaster information services. Projects included development and enhancement of Web sites, development of training materials and conducting training sessions with emergency personnel, collection and digitization of local disaster documents, and upgrading equipment. During FY2008, the CANDHI Web site was updated and maintained (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 work closely with national emergency preparedness and response agencies in their countries and provide critical information services to disaster managers, health care professionals and the public.

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 was initially developed by SIS staff, and has been transferred to a partnership of outside organizations working in toxicology. This will become a project in which voluntary representatives from participating nations provide crucial input and feedback to assure credible and high-quality sources of information. NLM will continue to provide funding for the next several years. The World Library was initially populated with information resources from more than forty countries and collaborations with many other countries are in progress.

Another resource under development was the **Dietary Supplements Database**, a resource of comprehensive information on supplements used by US 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.

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 FY2008, 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 led an NLM-wide collaborative initiative to produce a **Drug Information Portal** that makes it easier for consumers and health professionals to find drug information in the NLM and other governmental resources. The system uses ChemIDplus drug records for searching and resource locator selection, but operates with a user-friendly, simple interface. The portal was released in early 2008 and brought many users to resources in the area of drug information. A pill image databases, under development in SIS, will be linked to the Drug Information Portal.

In these and other new initiatives, SIS continues to search for new ways to be responsive to user needs in acquiring and using specialized information resources on such topics as toxicology and environmental health, HIV/AIDS, disaster preparedness and response, among others.

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 the NLM. The Center continues its active research and development, seeking to improve access to high quality biomedical information for individuals around the world. 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 new *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 the NLM, other Institutes at the 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 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); 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 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)

A Personal Health Record (PHR) is an electronic medical record whose contents are controlled and managed by the person whose data it carries. Having attracted much press attention in recent years, the US IT industry, health industry, and federal government envision the PHR as a possible solution to the information sharing and efficiency problems in health care.

The NLM has embarked on the development and deployment of a PHR in order to study and improve their utility, reduce barriers to their adoption, identify best practices, and provide a platform and test bed for advanced PHR applications. The development of the NLM PHR is based on a set of existing health care message and vocabulary standards that have either been developed by, or are supported by, the NLM and for the most part are also part of the accepted standards of the Secretary of HHS.

The PHR is a pure Web application. It is based on a forms generator (developed by NLM) that produces Web input forms on the fly. These forms include built-in skip logic (e.g. if the person is male, it does not show the question about pregnancy history), performs edit-checks and auto-completion of user entries. It uses AJAX server access techniques, so the response times are fast. The PHR uses Ruby on Rails (an open source Web management system) as the software for the application server and MySQL or Oracle as the database server. The PHR makes heavy use of JavaScript on the client in order to approximate the speed and capabilities of a desk top application. It also borrows heavily from the JavaScript open source world using Scriptaculous and Dojo.

The PHR provides tools for managing clinical information by different family members, so, e.g., a mother can maintain immunization records for each of her children and/or keep track of her ailing father's medications. It provides for the recording of medications, medical problems, allergies, surgeries and implants, immunizations (vaccines), measurements such as blood pressure and laboratory results (e.g. serum glucose), and questions to remember to ask one's care provider.

A key feature of the PHR is its emphasis on the coding of the names of medications, drugs, problems, surgeries, immunizations, and allergies that a user enters.

When a user types in one of these names, the PHR presents a menu of the names that "match" with what the user has typed so far. The menu gets shorter with each additional key stroke. The user can either select from this short menu or keep typing until there is only one choice. The computer matches the input string to any word in its list of vocabulary entries so it can find heart failure whether the user types in "heart" or "failure." It also has synonyms for its internal vocabulary to account for the many ways that people name things. So it will find "heart attack" whether one types in "heart attack" or "myocardial infarction."

The computer encodes the user's entries into NLM-supported coded vocabularies: RxNorm for drugs, LOINC for measurements and laboratory tests, CDC's vaccine codes for vaccines and SNOMED CT for problems. To ease the entry of medication records, we developed a special subset of RxNorm, called RxTerms. This vocabulary is used by the Centers for Medicare and Medicaid Services (CMS) for their post-acute care project and by the PHR. The PHR borrows its data type and much of its data structure from Health Level Seven (HL7).

The orientation toward automatic coding enables the PHR to provide two special capabilities: decision support and one-click access to information recorded in the record. Decision support in the PHR is based on predefined rules that the patient conditions to care recommendations. The PHR compares the patient's data against rules for preventive care and reminds the individual about interventions that are due, such as mammograms and colonoscopies.

The PHR provides one-click links to information about most of the items (drugs, problems, medications, allergies surgeries) that the user records on the form. This content comes from Medline plus, CDC, the American Academy of Allergy Asthma and Immunology, and AHQR, depending upon the topic. When the PHR suggest preventive or other care, it also provides one-click context-sensitive information from the US Preventive Services Task Force Web site.

The user can print a paper copy or download an electronic copy of their PHR (to their local computer or to a thumb drive). Currently, the electronic version is delivered as a spreadsheet in which separate tabs are dedicated problems, medications, allergies, tests, etc. The spreadsheet will provide text guidance about how to best keep the information private. The spreadsheet format provides an easy way for the users to review the contents on their own machine in a familiar media. We have plans to provide other export formats in the future.

We have invested substantial work in the development of mechanisms for user registration, management of passwords and user IDs and logging on. We do this without asking for user name, address or the other usual identifying information. We have also expanded the database support to both Oracle (new support) and MySQL. We moved to Oracle in order to obtain the levels of encryption and security required by NIH rules.

This young project addresses the longstanding NLM interest of facilitating health care management and is part of the NLM strategic plan. It will help tune the message and vocabulary standards that NLM has supported and also provide another consumer entry point to a rich trove of patient-oriented data. Early research projects will focus on user needs, usability, and usage patterns to guide the next round of development and research.

De-identification Tools

De-identification can unlock the research potential of long term clinical records. No well-supported and freely available de-identification tools exist. Taking advantage of past efforts and experience with de-identified procedures (NCI Shared Pathology Informatics Network (SPIN) grant) and existing Lister Hill Center tools that can recognize sensitive content such as dates, person names, locations, and numeric identifiers, LHC researchers initiated an effort to develop an open source text de-identification tool.

The system uses more than 700,000 clinical records from the Clinical Center for testing and validating current work (under IRB exemptions). Currently, developers are identifying and scrubbing sensitive information in the clinical text and labeling these items by type (e.g. personal name and postal address, etc.). To accomplish the task, the system utilizes several data source, including databases of first and last names along with frequency counts derived from the Social Security database of 480 million persons. US street names and information about cities, states, and ZIP codes were obtained from the US Postal Service and 2000 US Census databases. Statistical information about the usage of clinical words, common English words, and word co-occurrences have been extracted from multi-billion word corpora such as Wikipedia and Core clinical journal article abstracts.

Developers presented preliminary results to the LHNCBC Board of Scientific Counselors in April 2008. They plan to release a beta version of the software package for evaluation in the next few months and a complete version of the software in 2009.

Clinic Database Research and Development

As part of our medical record research and development we have embarked on the creation of a general purpose longitudinal database structure that could be used for many purposes, e.g.:

a) as a major part of the database for an open source EHR as has been proposed in some pending legislation

b) as the structure for the clinical data that might be needed for an EHR oriented to disasters
c) as part of a systems for statistical analysis of de-identified longitudinal research data

We have used a number of sources to inform this data structure. We have looked at the structure of a number of longitudinal data models. These have been especially instructive:

1) the data model employed by the HL7 version 2 message segments
2) the Regenstrief SPIN data model (1 billion results)
3) the Women's Health Initiative (WHI) database for the data collected over decades (500+ million results)

We have also obtained a de-identified clinical data set from a university (under a restricted use MOU). This data set carries clinical data for over 15,000 ICU encounters, including laboratory, data, radiology reports, discharge summaries and electrophysiological data – more than 100 million individual records, in toto. This data set has the complexity and size we need to test and tune the database design and it includes clinical content rich enough to test the utility of applications that we might attach to the database statistical analysis tool.

We have successfully loaded his de-identified clinical data into our early stage database system – achieving a major goal for this year. We have also developed a linkage to the R statistical system. R is an open source analysis system that encapsulates very sophisticated statistical knowledge. The current interface will perform simple statistical analysis and data manipulation. R has powerful facilities for analyzing and graphing data. One of our goals is to experiment with innovative ways to present and analyze an individual patients' data. But it may also provide a greater window into research databases.

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. The development platform is Ruby on Rails, an open-source Web application framework. Developers are using this tool in the data capture function of personal health records. They are also using several terminology resources from the UMLS (e.g. RxNORM, ICD9-CM) in data entry fields that require a set of controlled terms. Further development will involve work with very large databases of de-identified patient data. The goal is to create additional reusable software tools, some of which will involve biostatistical analysis with the R package.

Collaboration with Centers for Medicare and Medicaid Services (CMS)

LHC has assisted CMS in the development of many aspects of Medicare's Post Acute Care data collection project 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. LHC proposed and demonstrated a Web approach to auto-completion of entered text, much of which has been adopted by Medicare for this demonstration project. LHC has proposed and delivered the full content of many of the look-up tables that Medicare will use in this project, including the RxTerms database, now in the third release, 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

Through collaborations with a major university and the NIH Clinical Center the LHC Natural Language Processing, researchers have gained access to over 400,000 progress notes, 15,380 discharge summaries, and 178,126 radiology reports which are de-identified. These resources are instrumental in the ongoing effort to identify the clinical concepts in a narrative clinical report and map them into UMLS Concept Unique Identifiers (CUIs). Clinical text presents the additional challenge of dealing with grammatically incomplete sentences, significant numbers of acronyms and abbreviations (often idiosyncratic to a particular clinical unit), and distinguishing positive from negative statements about a finding, disease or symptom. Tools for finding important clinical concepts in narrative records 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 for converting strings found in text into concepts that can facilitate this effort.

Standards for Identifying Clinical Observations, Forms and Panels

As part of an effort to ease the mapping of local laboratory systems to a universal standard (LOINC) in collaboration with other NLM divisions and our contractors, we have identified a sample of the 800+ most commonly reported test results and embedded them in an application that lets users map entries in their lab to the LOINC database. We have also produced a guide for mapping the tests that HL7_eLINCs chose as the most important. They have used that guide to shape their standard but did not incorporate it into their standard.

To make LOINC terms more accessible and directly understandable to users in the same collaboration,

we have developed tools that systematically convert the formal names to short names and to long common names. The long common names are a brand new development that produces names that replace the formal names with conventional names, e.g. "Prothrombin time" instead of the more formal "Coagulation surface induced." It also eliminates parts of the formal name that are not distinguishing in context and not part of common usage, and includes prepositions and connectors to make them more readable, e.g. "Prothrombin time in blood by coagulation method." These will be distributed in the next release and encompass at least 90% of the laboratory tests.

In collaboration with Office of the National Coordinator for Health Information Technology (ONC) and the Healthcare Information Technology Standards Panel (HITSP), we engaged in an effort to create LOINC terms and standardization for two clinical areas of high priority to the secretary of HHS. The first is prenatal screening with a major focus on the reporting of newborn screen laboratory results, especially those produced by mass spectrophotometry testing. We produced a set of LOINC terms and panels to accommodate the reporting of quantitative values for all newborn screening laboratory tests. We have also begun work on newborn screening for hearing loss.

The second of these efforts was the development of an approach to the reporting of clinical genetics studies. This was a collaborative effort with HTSP, Partners of Boston, Intermountain Healthcare, HL7, and NCBI. It resulted in the creation of 12 LOINC panels and about 100 new LOINC terms and a balloted HL7 implementation guide that describes exactly how to report the results of such studies. The approach is a general one that uses a repeating series of panels, one per region of interest, one per genetic marker noted, and a few others to carry the full information in the report.

The LOINC vocabulary system continues to grow in usage and size. More than 3,000 new terms have been added in the last year. The LOINC mapping tool now supports both search and display of those non-English languages that have been submitted (Spanish, Chinese and French).

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).

Interactive Publications Research (IPR)

The IPR project is in line with Recommendation 13: Illuminate the value of multimedia-rich interactive publications ... to a broad readership that would reuse the media content for analysis. This project intends to demonstrate a type of highly interactive multimedia document that could serve as a model for next-generation publishing in biomedicine. The project focuses on the standards, formats, authoring and reading tools necessary for the creation and use of such interactive publications (IP) containing many media objects relevant to the biomedical literature: text, video, audio, bitmapped images, interactive tables and graphs, and clinical images such as x-rays, CT, MRI, and ultrasound. These objects are in a wide range of file formats: e.g., text in MS Word or PDF, animations in Flash, spreadsheets in Excel and clinical images in the DICOM format. Following a definition of authoring procedures, we have created and demonstrated two prototype documents. One focuses mainly on still and moving clinical images and animations, and the other focuses on large stores of tabular research data subject to statistical analysis.

The LHC has developed the first version of an IP viewer named Panorama, an Eclipse-based Java module, which can present different kinds of multimedia, tables, graphs, and image modalities in multiple panes. A reader may easily move among them, or perform analysis on a graph in one pane and display results in tabular form in another, for example. A component of Panorama is **iTAG**, a tool for selecting data subsets, creating graphs and charts from tables (and the reverse), performing basic statistical analysis, and exporting data in formats required by SAS or R for further analysis. In 2009, developers will enhance Panorama with additional features, such as volume rendering from 2D clinical images.

Research in FY2009 will also address a serious challenge to the network delivery of IP arising from their large file size, conceivably in the hundreds of megabytes. This research will seek strategies to overcome the barriers posed by network bandwidth limitations in the delivery of IPs, e.g., progressive transmission in which small text files are delivered for immediate reading, while larger media files are transmitted in the background. In addition, we will develop **Forge**, an authoring tool in Java. Forge will enable authors to create multimedia-rich interactive articles using a set of wizards while permitting them to write content in conventional word processors, e.g., MS Word or LaTeX.

Developers are also working with the Optical Society of America (OSA) to produce four interactive issues of regularly published OSA journals during the next year. The articles in each of the special interactive

issues will contain image data sets with which the reader can interact. The journals will be accompanied by a free interactive publication reader similar to the free Adobe PDF Reader.

Multimedia Database R&D

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

These tools are under development as part of the collaboration with NCI. This suite of tools earned the Internet2 IDEA award in Spring 2008. The tools include the following:

Multimedia Database Tool (MDT)

The **Multimedia Database Tool (MDT)** accommodates the new text/image database currently being created for the collection of 100,000 uterine cervix images from NCI, but also existing databases of x-rays and associated text from NHANES. It has a flexible database schema and GUI that allows new databases to be incorporated easily by a database administrator without software modification. The system allows not only data dissemination but also distributed (remote) data collection.

Developers plan to add capabilities to the **Multimedia Database Tool (MDT)** for users to write to the database, add levels of user privileges for user access, and add support for flexible schema and for simultaneous display of multiple images related to a single subject.

In FY2009, research will be extended beyond the uterine cervix images to other NCI image sets, e.g., histology images. Techniques are to be developed to support the dissemination of these very large files using image tiling approaches.

Boundary Marking Tool (BMT)

The **Boundary Marking Tool (BMT)** is a Web-accessible system for the accurate and efficient collection of descriptive data and boundary data for specific structures of biomedical interest in digitized uterine cervix images. It is in active use by NCI for about a dozen studies, e.g., to select biopsy sites in colposcopic images. The BMT Study Administration Tool (BSAT) is a recent addition that allows users to create their own studies by uploading their own images and configuring the screens to collect boundary data particular to their studies.

Virtual Microscope (VM)

The **Virtual Microscope (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.

V**irtual Microscope (VM) and Virtual Slides (VS)** are an archive of virtual slides has been developed from the teaching set of glass slides from the Department of Pathology of the Uniformed Services University (USUHS) and other collaborating institutions. An entire slide is digitized, segmented and processed to simulate an examination of a glass slide under the microscope but with a Web browser. The collection preserves the specimen for posterity and allows viewing by users worldwide anytime. Annotations and automatic linking to MEDLINE/PubMed is planned. A related collection of images from the AFIP fascicles allows users to search images and automatically link to MEDLINE citations.

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 question 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 NCI Visual Triage Study concluded in FY2008. This study had the objective of estimating the effectiveness of "screen-and-treat" cervical cancer prevention programs using HPV testing and cryotherapy. The study used image and diagnostic data collected from 552 HPV-positive women in the NCI Guanacaste Project, and had three phases.

First, using both image and diagnostic data collected during the Guanacaste work, two expert gynecologists independently determined, whether the women were treatable (at the time of enrollment as participants in the Guanacaste Project) for HPV infection by cryotherapy, or whether they required referral for advanced treatment; these expert opinions were then

reconciled into the final "truth" standard for the Visual Triage Study. Second, 12 midwives in Peru performed a simulated visual triage on each of the 552 women by viewing images for each of them and judging, based on visual criteria, whether a woman was treatable by cryotherapy or required referral for advanced treatment. The work of the midwives was carried out at Internet cafes in the Amazon region of Peru. Finally, five gynecologists, located in Costa Rica, Ghana, Peru, and Thailand, repeated the triage carried out by the midwives. The cumulative data was used to compare the effectiveness of the midwives, and of the gynecologists, in assessing treatability by visual methods. The conclusion of the study was that the performance of the visual triage was suboptimal, and that such screen-and-treat programs might be ineffective. The study was used for a major part of the successful PhD dissertation in epidemiology of Julia C. Gage of NCI, and has been submitted for publication to the *International Journal of Gynecology and Cancer*, with the title, "An Evaluation by Midwives and Gynecologists of Treatability of Cervical Lesions by Cryotherapy Among HPV-Positive Women."

The Visual Triage Study software developed by LHNCBC staff used a Web-browser interface to display both images and related text, and to collect responses from the study observers (midwives and gynecologists). Images were provided in a range of pixel sizes so that the physical images appeared on all monitors at a uniform size, regardless of the monitor's screen size. This was done to simulate the appearance of the cervix area when viewed through a colposcope. Secondary, larger versions of the images were also provided to simulate the appearance of the image after magnification. For most of the study participants, images were accessed in real-time over the Internet. For two of the very low bandwidth participants, images were provided on DVD; the Visual Triage Study Software was written to accommodate both options.

Other Tools Include the Following:

WebMIRS (Web-based Medical Information Retrieval System)

Developed some years ago and still in active use, WebMIRS continues to provide access to spinal x-ray images and associated text from nationwide surveys conducted by the National Center for Health Statistics.

At present, there are 568 users of WebMIRS in 58 countries. Forty-three percent of the users are from academia, the rest from government agencies, corporate and medical organizations. More than half the users are in the US, but relatively high numbers are from Canada, UK, India and China. 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.

Content Based Image Retrieval (CBIR)

We continue our research in **Content Based Image Retrieval (CBIR)**, using text descriptors, image examples, sketches, or combinations of these to create queries. As part of this work, we will incorporate user-relevant feedback and image indexing trees to enhance the accuracy of image retrieval in CBIR systems. Based on this research, a next-generation prototype CBIR system will be developed with capability of demonstrating *image retrieval by content* to professional biomedical groups for initiating collaborative work with biomedical subject matter experts. The goal is to validate the image data and to acquire technical critiques of the usefulness of our design approaches for biomedical research and/or clinical practice. CBIR techniques are also used in the **Image Text Indexing** project in which the illustrations in medical articles are indexed by processing figure captions and mentions in the text using natural language processing techniques, as well as by image features in the illustrations.

In FY2009, research will continue into computer-assisted image segmentation using efficient and accurate manual boundary point marking, Active Contours, Active Shape Modeling, the Generalized Hough Transform, Active Appearance Models, Live Wire, Level Sets, and general deformable template methods. We will test these methods focusing on the NHANES II x-ray images and on other image collections and modalities, such as digitized color images of the uterine cervix. Developers are working to improve the accuracy of these methods, integrating them into a practical system for shape segmentation, and making them useful for efficient production-level segmentation of large collections of images by shape. These functions are being incorporated into the new Web-based **Spine Pathology and Image Retrieval System (SPIRS)**.

In FY2009, research will continue toward establishing a gold standard for evaluating segmentation

algorithms, as they relate to both the spine x-rays and NCI uterine cervix images, that will be useful to the biomedical community.

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. 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.

FY2008 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.

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. Based on these findings, a project entitled A Knowledgebase of Human Variation was started during FY2008. The goal is to build a database of the parameters and variances which would define the normal range of all human anatomical structures and the dependencies and covariances between them. The initial data will be gleaned from the existing anatomical literature. The database will then be confirmed using data obtained from radiological scans of normal human anatomy. A Web 2.0 paradigm, inviting the participation of the interested community, will be used to glean and collect the needed data.

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 US 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.

Beyond data collection, this group is engaged in aggressive image analysis programs in partnership with the High Resolution Microscopy Laboratory at the National Cancer Institute. 3D Informatics staff members are developing new techniques using the **Insight Toolkit** from the Visible Human Project to study sub-cellular structure from 3D dual-beam scanning electron micrographs and 3D transmission-electron tomograms of cultured cells exhibiting critical pathologies such as melanoma and HIV/AIDS. This effort includes analysis as well as visualization and rendering of complex microscopy data, integrating the HPCC facilities in 3D printing to support intramural, trans-NIH research.

3D Telepresence for Medical Consultation

Completed in 2008, this project involved testing the efficacy of 2D versus 3D representations of video data transmitted in real time in remote clinical consultations. The technology infrastructure continues to be developed at the University of North Carolina and its efficacy continues to be researched there with help from colleagues at other institutions. The research team made substantial progress in implementing the technology infrastructure. A prototype portable camera unit was added to the stationary one and calibrated. The PDA application was completed and all the basic components of the system proposed are in place. Of special concern were 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. Real time 3D video is a very difficult problem to solve and, while the team did not solve the problem completely, substantial advances were made. Initial programs rendered the video as computer graphics while the later renderings were of video quality. Moreover, the team completed research comparing 2D video with a 3D proxy condition where the consulting ER physicians advised first responders at a distance via high quality 2D video or in the same room, but in constrained conditions mimicking 3D video communication. The consulting doctors could move to get different views and could point with a laser pen, but could otherwise only talk to the responder. Responders made 11 critical life threatening errors when unassisted, 7 when given advice via 2D video and only one in the proxy condition. In addition, they had greater self-efficacy and confidence in performing the airway task. Consultants had to ask fewer questions because the view was less obstructed, allowing responders to concentrate more on the task. The higher level of questioning in the 2D context made many responders feel the consultant lacked confidence in their performance and they had less self-efficacy as a result.

Advanced Network Infrastructure for Distributed Learning and Collaborative Research

This project built on previous work with HAVnet (Haptic Audio Visual Network for Educational Technology) and was collaboration between Stanford University and the University of Wisconsin at La Cross. The project was completed in 2008 and focused on developing visual and haptic applications for anatomy and surgical training and included aspects of self scaling technology, self-optimizing end-to-end, network aware, real time middleware, wireless technology, and GIS. The technology was developed and refined in the context of teaching anatomy and surgical skills and addressed issues concerning network bandwidth and latency and the integration of 3D visualization, haptic, and real time online collaboration tools. The project delivered: 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 that suggested 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. The testing on many of the visualization tools was done as part of an iAnatomy collaboration with the Northern Ontario School of Medicine involving the use of the stereo viewer for anatomy teaching and distance learning.

Insight Tool Kit (ITK)

The Insight Toolkit, a research and development initiative under the Visible Human Project, is now in its seventh year with a recent official software release of ITK 3.8 in July 2008. 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 Insight Toolkit 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 1500 active subscribers to the global mailing list for the project. ITK is an essential part of the software infrastructure of such projects as Osirix, an open-source diagnostic radiological image viewing system available from a research partnership between UCLA and the University of Geneva. ITK is also having an impact in other scientific fields, appearing in the Orfeo Toolbox (OTB) from the Centre Nationale d'Etudes Spatiales, the French National Space Administration. ORFEO Toolbox (OTB) is distributed as an open source library of image processing algorithms. OTB is based on the medical image processing library ITK and offers particular functionalities for remote sensing image processing in general and for high spatial resolution images in particular.

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. 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 2008, 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. At the current time, the Optical Society of America is adopting VolView as their free 3D data viewing software as part of a joint NLM/OSA project in Interactive Publication, intended to distribute open access journals accompanied by open data. 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.

InfoBot (formerly Medline on Tap – MDoT)

The **InfoBot** project is in line with Recommendation 27: Promote the development of just-in-time, patient-relevant knowledge bases that link scientific and clinical information within personal health records. Its goal is a system that will enrich an institution's existing EMR system with useful information from NLM resources. The InfoBot software would run as background agents, both at the institution and at NLM. APIs would be supplied to the institution to allow them to integrate the search setup and to display and store results in their existing EMR system, in accordance with their own preferences. Part of this project is to automatically generate a repository of key facts extracted from the biomedical literature to support informed clinical decision making. This subproject is RIDeM (Repository for Informed Decision Making). Practitioners of Evidence Based Medicine (EBM) advocate informed decision making that combines the clinician's expertise and judicious use of current best evidence. EBM provides guidance in both how to best find needed information, and appraise information found in literature. The three basic components of EBM are [1] Clinical Task (e.g., prevention, therapy, etiology, etc.) [2] The PICO framework for question formulation and document analysis (PICO = **P**atient/**P**opulation, **I**ntervention, **C**omparison, and **O**utcome); [3] Strength of Evidence. Tools will extract these items automatically from MEDLINE records to populate the RIDeM database.

Turning The Pages Information Systems

In line with *Recommendation 12: Educate the public about the historical development of biomedical sciences and technology*, the Turning The Pages project aims to provide the lay public a compelling experience of historically significant and normally inaccessible books, e.g., medieval volumes on anatomy and zoology. In a photorealistic manner, this project allows users to turn and view page images on touch-sensitive monitors in kiosks, as well as "click and turn" in an online version. In the kiosk version of TTP, techniques and tools are explored to optimally capture the original pages, enhance these page images to improve quality, animate the page images, and import them into a software environment that allows user interaction. We developed a 3D wireframe model in Maya, a commercial modeling and animation system to replace the labor-intensive manual process used earlier in collaboration with the British Library. Within Maya, each pair of page images is texture-mapped to both sides of the wireframe model of a turning page, with a multisource lighting model that provides realistic highlights and shadows. Our technique allows the automated generation and rendering of 15 intermediate animation frames which are then imported into Macromedia Director for viewing and interaction by library patrons.

While retaining the eye-catching appearance and easy touch-and-turn characteristics of a virtual book, the TTP project has also prototyped an extension of this *exhibit* product to an *information system* (TTP+) by linking the displayed pages to relevant biomedical information sources such as databases of clinical trials information, anatomic images and biomedical literature citations. Blackwell's *Herbal* and Vesalius' *Anatomy*, both rare books normally inaccessible to a library patron, has been extended to the TTP+ metaphor.

Two significant developments in 2008 included the creation of a 3D model for flat scrolls leading to the construction of the Edwin Smith Papyrus in TTP form ("touch and scroll"), and the release of a French language version of TTP Online ("Tournez les Pages"). Plans are under way to increase the TTP collection with other rare books, selected in collaboration with NLM's historians.

Video Retrieval and Reuse Project

The FY2009 budget request for APDB includes funding for four major project areas: core resources (COREAPDB), NLM support (AVSNLM), LHNCBC research support (AVRES), and NLM Media Assets (AVSNLM/Media Assets). The NLM media assets and the NLM support project contribute to the NLM-wide AV support area in the NLM Long Range Plan. The LHC research support and the core resources contribute to ongoing LHC information services projects.

The core resources continue to provide the equipment, software, support and training required for the video graphics and animation facility, the interactive multimedia development projects, and the video production capability of the NLM. These video and graphics communications areas depend on technologies which are changing rapidly and are driven by major market requirements not influenced significantly by biomedical needs. Therefore, technology transformation and adaptation are major branch activities pursued along with the development of high quality facilities and support resources required to produce the educational and informational materials essential to a major emphasis by NLM of providing health information to the consumer and to encourage the use of high quality information by health professionals and the public.

In FY2008, the Media Assets project area continued the upgrade of the high definition (HD) editing systems and the transition to digital video and high definition production workflow. The changeover to a completely high definition video production facility remains a high priority for FY2009. Much of the post production area of the facility has been converted to the DTV standard, including accommodation of the HD portion of the standard. The network integration of the four existing nonlinear editing systems and the development of a media asset management system within a networked content creation environment, allows more efficient manipulation of video for faster editing turn around, and the ability to include revisions and updates much more quickly. In FY2009, staff will continue the transition to the HD standard and provide the IT (information technology) convergence for both the production and postproduction functions. This will allow the efficient use of HD still graphics, animation, and full motion HD video in post production editing. Network access to branch-created high definition graphic and animation materials will assist the expanding role of high quality digital images in support of media development.

The NLM still image, graphics, and video support team provides ongoing capability in these areas to all of the NLM and includes the production, post-production and authoring services for the development of Internet video, kiosk interactive multimedia and DVDs. The number of requests for content creation continues to increase and has exceeded the in-house resources available. In addition to meeting the requests through the application of advanced technology, support from contract sources is essential. This area of the budget also contains some equipment funding to maintain the audio, video and multimedia capability in the NLM board room, auditorium and other conference areas supported through branch project management and technical resources.

The fourth project area is LHNCBC research support (AVSRES). A number of LHNCBC development projects require video graphics, interactive multimedia development, imaging or video production as part of the overall project objectives. A major effort for FY2009 in this project area is the improvement of rendering times for video graphics, and 3D visuals and animations for DVD and other interactive multimedia productions. These animations are rendered on state of the art workstations enhanced with specialized graphics engines and require continual development and upgrading of current modeling and rendering software modules. Some of these 3D animations have been integrated into Web-based materials and into kiosks equipped with DVD systems and touch screen monitors. Resources are required to continue to improve image creation, graphic design, rendering times, and interactive multimedia application development.

Biomedical Image Transmission via Advanced Networks (BITA)

In FY2009, the **BITA** project will continue to have both an R&D as well as a technical support role in NLM's Next Generation Internet (NGI) activities. It is in line with *Recommendation 10: Pilot test and evaluate new ways to use the digital infrastructure to enhance access to online health information by minority and underserved persons and communities*. Technical support for the NIH's Malaria Initiative program in Africa will continue, with a focus on characterizing end-to-end performance measurements over the MIMCom network, 802.11a and 802.11b wireless network implementations, and networks exhibiting narrow bandwidths, high latency and high jitter.

In addition, the BITA project will continue to deploy and test basic connectivity over advanced implementations of Internet2. Developers are collecting and examining data for speed, error, and QOS performance and analysis of advanced network infrastructures including Packet-Over-Sonet, Wavelength Division Multiplexing, and ATM using different techniques, e.g., FTP, multisocket, IP over ATM, native ATM, as well as client server systems such as WebMIRS.

With the worldwide increase in the Internet Protocol addresses needed for network devices of all kinds, research will focus on IPv6, in light of its emergence as a viable architecture for wide area networking. The project aims to investigate the speed, error and QoS performance of this protocol in the transmission of medical images, of fundamental importance in telemedicine and digital libraries.

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. These systems address the NLM Goal 1: *Seamless, Uninterrupted Access to Expanding Collections of Biomedical Data, Medical Knowledge, and Health Information*.

Medical Article Records System (MARS)

The **MARS** project and its several spinoffs address NLM's Long Range Plan *Recommendation 4: Support the development of data mining techniques to integrate access and repackage information*. The projects in this group aim to develop and operate systems that efficiently extract bibliographic information from the paper-based or online medical journal literature to build MEDLINE citations. This is done by a combination of document scanning, optical character recognition (OCR), and rule-based or machine learning algorithms. The MARS production system currently in operation extracts and organizes bibliographic data by using advanced algorithms for automated zoning, field identification, and syntax reformatting. In addition, biomedical lexicons are used to implement pattern matching algorithms to correct errors in OCR-detected affiliation information, and reducing incorrectly highlighted words for increased operator productivity.

WebMARS

Developed as part of the MARS project, WebMARS is a system to automatically extract data from Web-based online journals. Besides serving this purpose, in FY2009, the basic WebMARS technology will be used to help achieve goals of NLM's Indexing 2015 Initiative, by the development of the Publisher Data Review (PDR) system. This system will provide operators data missing from the XML citations sent in directly by publishers such as databank accession numbers and NIH grant numbers, thereby lowering the manual effort in completing citations for MEDLINE. In addition, incorrect data sent in by the publishers can be corrected by PDR. The current manual effort in filling in missing data and correcting wrong data from the publishers is considerable since the operators generally have to look through an entire article to find this information, and then key them in.

In FY2009, while the PDR system will be completed and placed into daily production for NLM indexers, we will continue all engineering support for the offsite production facility: installation of upgraded modules, testing, maintenance and operation of all hardware and software for servers, clients and networks, and the necessary system administration.

Analysis of Images for Data Extraction (AIDE)

These projects support the goals of the NLM *Indexing 2015 Initiative* and the *NLM Long Range Plan Recommendation 1: NLM Resources and Infrastructure for the 21st Century*. The goal of this research is to increase the efficiency of creating citations for MEDLINE in order to accommodate the expected doubling of the citation rate within the next few years.

ACORN

Conducted within the **AIDE** project is the underlying research in image analysis and lexical analysis that contributes to the continual improvements in the MARS system, as well as the creation of new initiatives in which these techniques could find application. An example is the ACORN initiative (Automatically Creating OLDMEDLINE Records for NLM) which aims to capture bibliographic records from pre-1960 printed indexes (e.g., IM, QCIM, QCICL, etc.) for inclusion in NLM's OLDMEDLINE database, thereby creating a complete record of citations to the biomedical literature since Index Medicus appeared in the late 19th century. Manual data entry has proven to be too slow and costly. In FY2009 we will continue our investigation of scanning, image enhancement, OCR, image analysis, pattern matching, and related techniques to extract unique records from the printed indexes, and develop a prototype system for real-world testing and as a precursor to a production system.

Validated Test Set for Document Image Analysis

As part of the AIDE project, a ground truth database, Medical Article Records Groundtruth (**MARG**), has been released for use by the international computer science and informatics communities for research into advanced

algorithms for data mining. It has attracted more than 16,500 visits from 96 countries. The MARG database consists of document images and the corresponding OCR data, zones, labels and verified data obtained from the normal operation of the MARS production system. In 2009 MARG will be expanded and tools provided for easy use.

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. Image Indexing Initiative (I3) has used interactive MetaMap successfully for automated mapping of terms suggested by subject matter experts to Metathesaurus concepts. I3 will continue efforts to improve the precision and recall of these mappings with additional features of the MetaMap API.

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 HIV/AIDS, health sciences research, and space life sciences.

Indexing Initiative (II) development focuses on testing and updating recently added functionality to the Medical Text Indexer (MTI) system such as the inclusion of subheading attachment recommendations and the addition of an explanation facility to inform indexers how MTI arrived at specific MeSH recommendations. Improvements to MTI will benefit the Indexing 2015 project through its MTI subproject. A secondary focus will be to test the application of MTI to NLM Cataloging and to make any necessary modifications to the cataloging version of MTI. System-related objectives for II include completing the final testing of the migration of our systems to a Linux environment.

The major objectives for the Indexing Initiative in FY2009 are the development of the ***Basic Medical Text Indexer (MTI):***

- Test and update MTI's subheading attachment function and continue to coordinate with NLM's Document Creation and Management System (DCMS) team.
- Test and update MTI's explanation facility and coordinate its inclusion in DCMS.
- Assess catalogers' acceptance of MTI's adaptation for NLM Cataloging and to modify the cataloging version of MTI as needed.
- Improve MTI's ability to handle citations without an abstract, i.e., title-only citations.
- Develop MetaMap 3D, a tool for colorizing and otherwise highlighting features of biomedical text, accounting for both literature and clinical views.
- Explore further Word Sense Disambiguation (WSD) algorithms for improving MTI's accuracy.
- Complete the code modifications that have been necessitated by the migration ***to new machine and secure network architecture.***

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 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 (DPR)

This project is in line with Recommendation 1: Build a repository that employs advanced technology for the storage and preservation of large quantities of print and digital material. It addresses an important problem for libraries and archives, viz., to retain electronic files for posterity, both documents in multiple formats (e.g., TIFF, PDF, HTML) as well as video and audio resources. For document preservation, our objective is to develop a prototype **System for Preservation of Electronic Resources (SPER)** that builds on open source systems (e.g., DSpace from MIT) while incorporating in-house developed modules that implement key preservation functions: ingesting, automated metadata extraction and file migration. One role of SPER is as a testbed to evaluate alternative approaches to these functions and to select the optimum ones. Another role is to use it to preserve a collection of historic medico-legal documents that NLM has acquired from the FDA. This effort will continue in 2009.

While the accuracy of the automated metadata extraction (AME) module is reasonably high, further work is planned for 2009: enhancing AME's layout classification and pattern recognition modules; investigating other SVM implementations for use in SPER; modularizing AME to extend its application to other NLM collections.

In 2009 we will continue to collaborate with OCCS and LO in an effort to create a production-level NLM archive, drawing upon our design experience with SPER. The target collections to be preserved will be selected in collaboration with NLM's preservation managers and curators, and could include Profiles in Science, medical pamphlets, or some other collection.

Also, in 2009, as the production system is developed, the SPER prototype will be used as a testbed system to investigate promising approaches to implement critical preservation stages, for eventual inclusion in the operational system. Issues to be researched: selecting optimal file formats for the long term; maintaining unique and reliable references to migrated files; feature selection, quantization and clustering in a Bayesian Learning approach to automatically classify digital documents; extending the learning algorithms to Web and video resources; intelligent decision making for file migration.

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 63,000 protocol records sponsored by the US Federal government, pharmaceutical industry, academic and international organizations from all 50 states and in 158 countries. Some 41% of the trials listed are open to recruitment, and the remaining 59% are closed to recruitment or completed. ClinicalTrials.gov receives over 52 million page views per month and hosts approximately 800,000 unique visitors per month. Data are submitted by over 5,800 study sponsors through a Web-based Protocol Registration System, which allows providers to maintain and validate information about their studies.

ClinicalTrials.gov was established by the National Library of Medicine (NLM) in 2000 in response to the Food and Drug Administration Modernization Act of 1997 and to support NLM's mission of disseminating biomedical knowledge and advancing public health. ClinicalTrials.gov was enhanced in 2007 - 2008 to implement initial requirements of Section 801 of the Food and Drug Administration Amendments Act of 2007 [Public Law 110-85]. The law required the expansion of the registry and the addition of a results database. In response to the law, the ClinicalTrials.gov registry was expanded in November 2007 to provide members of the public, health care professionals, and researchers with additional descriptive, recruitment, location, contact, and administrative information about ongoing and completed applicable drug and device clinical trials. As a consequence of this law, new registrations have from December 2007 to September 2008 increased by 44% (average 360 per week) compared to an average of 250 per week from December 2006 to November 2007) and modifications to existing registrations over the same time periods increased by 300% (average rate of 2,400 per week compared to an average of 800 per week). Also in response to the law, links were added from registration records to related results on the FDA Web site (e.g., Drugs@FDA) and links were enhanced to NLM's Medline and DailyMed Web sites. ClinicalTrials.gov also researched, designed, tested and implemented a results database which is complementary to the registry. The results database is required by law and is the first-of-its-kind. It includes results information on primary and secondary outcomes of registered trials, as well as

information on the patient populations studied. The results database also includes a module for reporting serious and frequent adverse events observed in a clinical trial, although this component is not required by law until September 2009. Preliminary versions of the database were made available for public testing and comment and the final version was implemented on September 22, 2008. The expanded registration requirements as well as the results database will be further implemented through rulemaking and NLM is working with the Food and Drug Administration (FDA) on drafting of the registration rule. ClinicalTrials.gov continues to work on other aspects of the law, including but not limited to expansion of the results database, a pilot quality control study, hosting a public meeting, and consulting with risk communication experts. When fully implemented, the registry and results database will become a unique resource for scientific and clinical information that can assist in providing patients, healthcare providers, and researchers more comprehensive information about ongoing and completed research.

ClinicalTrials.gov was actively involved in educating the public on the new law and continuing to promote standards of transparency in clinical research through trial registration. This information was communicated to a broad range of US and international stakeholders via presentations and peer-reviewed publications. As a result of increasing awareness of the law and the importance of trial registration, nearly 27,000 new registrations were received over the last calendar year. ClinicalTrials.gov continues to collaborate with other registries and professional organizations, working towards developing global standards of trial registration.

Genetics Home Reference

Genetics Home Reference (GHR) is an online resource that offers basic information about genetic conditions and the genes and chromosomes related to those conditions. This resource provides a bridge between the public's questions about human genetics and the rich technical data that has emerged from the Human Genome Project and other genomic research. Created for the general public, particularly individuals with genetic conditions and their families, the site currently includes summaries of more than 325 genetic conditions, more than 500 genes, all the human chromosomes, and information about disorders caused by mutations in mitochondrial DNA. The Web site also includes a handbook, *Help Me Understand Genetics*, which introduces users to fundamental topics in human genetics including mutations, inheritance, genetic testing, gene therapy, and genomic research. Usage of the GHR Web site, as measured by the daily unique visitors, increased almost 50 percent in the past year, and the site continues to be recognized as an important health resource.

GHR specifically targets NLM's goal of advancing scientific knowledge in molecular biology by providing information about hereditary conditions and their underlying genetic causes. The LHNCBC continues to investigate a variety of ways to make the results of the Human Genome Project more readily available to the public through the Genetics Home Reference (GHR) Web site. Because GHR is designed for patients, families, and the general public, the genetic information is written in a consumer-friendly format. Brief summaries of genetic conditions and genes will continue to be added in FY2009 and existing topics will be reviewed and updated. Strategies for creating "just-in-time" links to salient resources for additional consumer information (e.g., MedlinePlus, support/advocacy groups, and recent literature) will continue to be developed. Additionally, new types of content are under development, including pages that introduce families of related genes.

The GHR Web site celebrated its fifth anniversary in 2008 with a combined total of more than 800 user-friendly summaries of human genetic conditions, genes, and chromosomes. In the past year, the project expanded its genetics content for consumers, added new features, and continued with outreach activities to increase public awareness of the Web site. Specifically, GHR staff added new summaries of 94 genetic conditions and 121 genes to the Web site. On average, about 20 new summaries were added per month, which is more than twice the monthly average from the previous year. Staff intend to continue this rate of production in FY2009, with a goal of covering as many Mendelian genetic disorders as possible. In the past year, staff members developed a new feature about gene families that allows users to find out how genes are related to one another. The GHR Web site currently includes almost 20 gene family summaries, with more under development.

GHR continues to support the Information Rx initiative, a free program that enables doctors and nurses to write "prescriptions" directing patients to the GHR Web site for an explanation of genetic disorders and related topics. In other outreach activities, GHR staff presented the Web site to several visiting groups, including educators and students, and represented the project at local and national meetings. Staff members will continue to educate others about this important resource in FY2009.

Profiles in Science® Digital Library

The Profiles in Science Web site provides researchers, educators, and potential future scientists worldwide access to extraordinary, unique biomedical information previously accessible only to patrons able to make an in person visit to the institutions holding the physical manuscript collections. "Profiles" also serves as a tool to attract scientists to donate their collections to archives or repositories in order to preserve their papers for future

generations. Profiles in Science decreases the need for handling the original materials by making available high quality digital surrogates of the items. Standardized, in-depth descriptions of each item make the materials widely accessible, even to individuals with disabilities. The growing Profiles in Science digital library provides ongoing opportunities for future experimentation in digitization, optical character recognition, handwriting recognition, automated image identification, item description, and search and retrieval.

The Profiles in Science Web site showcases digital reproductions of items selected from the personal manuscript collections of prominent biomedical researchers, medical practitioners, and those fostering science and health. Profiles in Science provides researchers, educators, and potential future scientists worldwide access to extraordinary, unique biomedical information previously accessible only to patrons able to make an in person visit to the institutions holding the physical manuscript collections. "Profiles" also serves as a tool to attract scientists to donate their collections to archives or repositories in order to preserve their papers for future generations. Profiles in Science decreases the need for handling the original materials by making available high quality digital surrogates of the items. Standardized, in-depth descriptions of each item make the materials widely accessible, even to individuals with disabilities. The growing Profiles in Science digital library provides ongoing opportunities for future experimentation in digitization, optical character recognition, handwriting recognition, automated image identification, item description, digital preservation, emerging standards, digital library tools, and search and retrieval.

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. Several collections have been donated to NLM and contain published and unpublished materials, including manuscripts, diaries, laboratory notebooks, correspondence, photographs, poems, drawings and audiovisual resources. The collections of Arthur Kornberg, Maxine Singer, Alan Gregg and Paul Berg were added this year. An additional 6,192 digital items composed of 12,052 image pages were also added to the twenty-six existing Profiles in Science collections. Presently the Web site features the archives of 27 prominent individuals:

*Christian B. Anfinsen *Donald S. Fredrickson *Joshua Lederberg *Wilbur A. Sawyer *Virginia Apgar *Edward D. Freis *Salvador E. Luria *Maxine Singer *Oswald T. Avery *Alan Gregg *Barbara McClintock *Fred L. Soper *Julius Axelrod *Michael Heidelberger *Marshall W. Nirenberg *Sol Spiegelman *Paul Berg *C. Everett Koop *Linus Pauling *Albert Szent-Györgyi *Francis Crick *Arthur Kornberg *Martin Rodbell *Harold Varmus *Rosalind Franklin *Mary Lasker *Florence R. Sabin

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.

In addition to releasing new Profiles in Science collections during FY2008, LHNCBC staff made several enhancements to the Profiles in Science systems. Among these was the addition of the ForeSee Results American Customer Satisfaction Index (ACSI) survey. Survey results are expected to identify who is using Profiles in Science as well as what collections users want NLM to add. Staff also increased the accessibility of the Profiles in Science Web site. They extracted OCR text from the Web site's PDF documents and made the OCR available as an alternative format to PDF, developed algorithms to identify documents whose OCR is particularly bad or good, and made more noticeable the transcripts of the Web site's digitized files. LHNCBC staff developed software to automate the review and addition of items to existing Profiles in Science collections. Developers also experimented with various conversion software and created mockups of the Data Entry program in Java Swing and ASP.NET. They completed most of the new Linux-based Profiles in Science, including synchronizing data with the current Solaris-based system and displaying search results. Developers also migrated machines to new networks and migrated applications to new machines.

Nursing Home Screener

In line with Recommendation 9: Work to reduce and eliminate health disparities by providing underserved populations...with access to high quality health information that is understandable, the Nursing Home Screener is a tool for the public to judge the quality of nursing homes in keeping with the NLM goal for customized personalized health information. This Web 2.0 tool will deliver the Quality Indicators for such facilities, derived from publicly available data from the Center for Medicare and Medicaid Services, in an easily navigable geographic graphical interface. The Nursing Home Screener locates homes on a Google map. It allows users to survey nursing home quality, indicated by map icons, in any of four major categories: staffing, fire safety deficiencies, healthcare deficiencies, and quality of care inferred from residents' health. Options can be tailored within each category and other filters may be set to selectively hide home markers of less interest.

Based Medicine - PubMed for Handhelds

PubMed for Handhelds was publicly released in FY2003. 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. An algorithm to derive "the bottom line" (TBL) of published was recently added for a clinician's quick reading at the point of care.

User Focused Portals

NLM Gateway

The **NLM Gateway** is an ongoing production system that provides results from 23 NLM information resources with a single search. Since these resources are "moving targets" frequently updated, improved, and otherwise modified, the Gateway must change with them. Periodic changes to the NLM DTD (Document Type Definition), to MeSH and the MeSH Mapping File, and to the UMLS Metathesaurus are accommodated each year. More than 100,000 meeting abstracts are indexed using the tools of the Indexing Initiative. The most recently added resource to which the Gateway provides access is NLM's Profiles in Science.

A formal usability study of the NLM Gateway was performed in-house in FY2007, leading to a comprehensive evidence-based redesign of the system. The redesigned Gateway was placed into production late in February 2008. We are pleased to report that usage is up substantially (more than 50%) since that time.

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 was completed in 2008. It involved the use of open source, cross-platform technologies based primarily on grid technologies in general and the Access Grid (AG) in particular. The research was 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 the AG for 3D teleradiology; and using AG for volume rendering of patient image data in the operating room with wearable (e.g., eyeglass-like) environment. The latter allows surgeons to view the 3D data and to share it with colleagues and consultants while working on a patient.

In FY2008, the research team completed testing the scenarios, including those implementing color algorithms for 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 were employed in a haptic environment allowing surgeons to rehearse liver operations. In addition, the technology was tested in simulation labs for team training and in classroom settings for anatomy teaching. Several additional successful wide area wireless demonstrations of transmitting video and other patient data from ambulances using 3G and mesh cellular technology were completed. The University of Chicago has patented the imaging algorithms.

Scalable Information Infrastructure Initiative

The Scalable Information Infrastructure (SII) Project, which ended during FY2008, encouraged 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 included wireless and geographic information system (GIS) techniques.

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 (iHDTV). In 2008, staff successfully implemented the technology within the Collab and are

now exploring options for transmitting the video to distant end points. These options include implementing multiplexing and optical network switching and upgrading NLM's bandwidth from its current 1 gbps. Staff are also exploring UltraGrid technology as a method of transmitting uncompressed video as well as ways to adjust the bandwidth usage of iHDTV. This work is being undertaken with an eye toward testing the technology in telemedicine settings and, possibly, performing a clinical trial. 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. In 2007-2008, the program was expanded to include a high school serving Native Americans in Kotzebue, 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 presentations quite highly. Over the years the ratings of the distant presentations are sometimes higher than the face to face ones at King Drew. These differences are only two tenths of a point (5 point scale) apart. The Alaska students, who receive all presentations at a distance, rated the presentations higher overall than the students at King Drew.

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 to offer NCBI's biotechnology training at a distance. A follow up implementation was done with the Charles R. Drew University of Medicine and Science. Unfortunately, budget constraints have forced NCBI to cut back all training, but the methodology was validated.

A study of collocation as a factor in synchronous learning was completed with the University of Alabama at Birmingham in which 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 indicate that there are no significant differences between groups except for perceived interaction, which was much higher for the dispersed group. Observations confirmed that those groups experienced higher levels of actual interaction because the technology required all to work together. The collocated students simply interacted with the person next to them, if at all. The results are being written for possible publication.

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. A more formal assessment has been undertaken with CPSC and the Medical University of South Carolina and data is being collected of patient, provider, and interpreter judgments of the quality of communication in clinical encounter when interpretation is provided by video, by phone or in person. Depending on the outcome, the use of video technology beyond the medical center to remote rural clinics may be explored, given a) interest and b) in place network infrastructure.

Both the Web casts of the bi-monthly Washington Area Computer Assisted Surgery 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.

OHPCC Collaboratory for High Performance Computing and Communication

The "Collab" was established primarily as a resource for researching, testing, and demonstrating imaging, collaboration, communications and networking technologies related to NLM's Next Generation Network initiatives. This infrastructure is also used by staff to keep abreast of and test new technologies of possible interest to NLM (and others in biomedical informatics) and to conduct ongoing imaging, collaboration and distance learning research within OHPCC. The technology infrastructure is used to collaborate with researchers outside the NLM and, when appropriate, it is leveraged to support other activities and programs of the NLM. The facility can be configured to support a range of technologies, including 3D interactive imaging (with stereoscopic projection), the use of haptics for surgical planning and distance education, and interactive imaging and communications protocols applicable to telemedicine and distance education involving a range of interactive video and applications sharing tools. The latter enable staff to collaborate with others at a distance and, at the same time, demonstrate much of the internal and external

work being done as part of NLM's Visible Human and advanced networking initiatives. The collaboration technologies include a complement of tools built around the H.323 and MPEG standards for transmitting video over IP and open source technologies such as the Access Grid.

BabelMeSH

BabelMeSH is a multi-language and cross-language search tool for healthcare personnel who prefer to search MEDLINE/PubMed in their native languages. Journals' language of publications can be selected. Through international collaborations, including WHO Eastern Mediterranean Regional Office in Cairo, users can now search in Arabic, Chinese, Dutch, French, German, Italian, Japanese, Korean, Portuguese, Russian, Spanish, Swedish and English. Some specialty organizations are using BabelMeSH as a tool to search their collection of images.

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.

Computing Resources Projects

The Computing Resources (CR) has a variety of core projects to build, administer, support, and maintain an integrated secure infrastructure that facilitates the LHC research activities and thereby leverages the overall effectiveness of research staff members. The integrated secure infrastructure contains network management, security management, facility management, and system administration support for large numbers of individual workstations and shared servers.

The network management team plans, implements, deploys, and tests high speed network connectivity over Internet and Internet-2. One core project is studying the implementation of OMB requirements in the next generation Internet version 6 (IPv6). The security management team incorporates firewall administration, patch management, anti-virus management, intrusion monitoring, security scanning, and vulnerability remediation to ensure a safe working environment in overall security perspectives. The facility management team facilitates the product deployments, including power planning, network connection, cabling and space allocation in the computer room as well as the NCCS co-location. The system administration team provides center IT services, such as DNS, NIS, backup, printing, and remote access to ensure an efficient operation across the NIH campus.

Core projects also provide critical project planning and support for FISMA security mandates, such as yearly Certification and Accreditation (C&A) audit. Additionally, core projects provide operation assistance and troubleshooting functions for shared communication resources for public access to production systems.

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

The goal of the **DocView** project is to conduct R&D on advanced tools allowing libraries and users to access biomedical information. 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 195countries have downloaded it since it was released.

MyDelivery

In FY2009, research will primarily focus on the development of **MyDelivery**, a new Internet communications system designed to solve an important biomedical and health sciences communications problem. Health science applications often require the use and exchange of information contained in very large electronic files (e.g., digitized x-ray images, sonographic images, digital video files, MRI, CT scans, PET scans, and scanned document images). The delivery of this information should be fast, easy, reliable, safe, and secure (HIPAA-compliant). Because no current Internet communications technology (e-mail, FTP, instant-messaging and Web delivery) fully meets these criteria for large file communication, especially over potentially unreliable wireless networks used by an increasingly mobile user population, we are designing and developing a new communications system that will provide these capabilities to the biomedical and health science

communities. Targeted for use in clinical, research, administration, and library environments, the **MyDelivery** system will be capable of reliably communicating biomedical information contained in files of any size over networks of all types, including potentially unreliable ones.

DocMorph

As part of the DocView project, research and system engineering design will continue to maintain and improve the operation of **DocMorph**, a Web-based server providing users remote image and information processing capabilities via the Internet. DocMorph serves as a test bed for evaluating new image and library information processing algorithms, as well as a public service for document format conversion. This system now accepts more than fifty file formats, including black and white images, grayscale and color images, text and word processing files, to produce four outputs: PDF files, TIFF files, text, and synthesized speech. DocMorph averages 1,000 conversions daily, and 1,000 unique users monthly. It is used by several hundred libraries, including NLM, mostly in their interlibrary loan operations.

MyMorph

While DocMorph is generally accessed via a Web browser, the **MyMorph** client software allows users to perform *large scale* conversion of thousands of files at a time. MyMorph has more than 9,800 registered users, many of whom are document delivery librarians in small libraries around the country, using MyMorph as part of their daily document delivery operation. In 2009, we will expand MyMorph's role to digital preservation, both for bulk migration of archived files as well as conversion of ingested files to canonical forms such as PDF/A. PDF/A is an ISO-proposed standard file format for long-term electronic document preservation that promises to be of importance in preserving part of NLM's collection. Initial investigation has started in modifying both MyMorph and DocMorph to produce PDF/A files from document images.

Image Storage and Transmission Optimization (ISTO)

The **ISTO** project *(Recommendation 28: Develop software for acquisition, organization and access to ... digitized biomedical images)* addresses the problems of efficiently storing and delivering large biomedical image collections through research into advanced compression and transmission techniques. The focus is on Visible Human images, spinal x-rays from the NHANES surveys, and digitized color images of the uterine cervix from NCI. In FY2009, we will research the best approaches for using advanced compression techniques, such as Wavelet Transform compression, for the storage of the NHANES images in our multimedia database, as well as efficient decompression algorithms. The result of this effort will be executable modules for Wavelet decompression and multiscale display to be incorporated into the dissemination system for the NHANES x-rays. This work will position NLM to consider large file size image databases as service components for telemedicine as well as future Health Networks.

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 as one of the Unified Medical Language System (UMLS) Knowledge Sources. The SPECIALIST Lexicon consists of over 360,000 records describing the syntactic, morphological and orthographic properties of words. These records are released in a unit record format as a set of relational tables. New lexical items are continually added by a team of lexicon builders using LexBuild a lexicon building tool developed by LHNCBC staff. 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.

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. LHNCBC researchers have also released a text classification system which is a JAVA port of the Journal Descriptor Indexing tool originally developed by LHNCBC staff in LISP. The JDI algorithm is an unsupervised text statistical classification method that can provide context for word sense disambiguation and other natural language processing tasks.

LHNCBC research staff also develop and maintain the UMLS Knowledge Source Server (UMLSKS), an application that provides Internet access to the UMLS knowledge sources. UMLSKS is updated quarterly to accommodate quarterly UMLS releases. A beta version of the Grid/Web services implementation of the UMLSKS backend and portlet-based user interface has been released and is undergoing usability testing.

Medical Ontology Research (MOR)

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 from the perspective of translational research. Effective data integration of data repositories created by different communities (e.g., basic research and clinical care) is often realized through the integration of the terminologies used for the annotation of data in these repositories. We evaluated semantic integration among such terminologies, more specifically between SNOMED CT and the NCI Thesaurus, and between LOINC and SNOMED CT. We also evaluated methods based on Semantic Web technologies for data integration, with application to nicotine dependence.

RxNav, the standalone browser for **RxNorm**, NLM's drug terminology integration database, was extended in two different directions. First, we created an Application Programming Interface, making it possible for developers to integrate RxNorm information in their programs. Second, we broadened the scope of RxNav to include clinical information about drugs.

The research team continues to work on assessing the quality of biomedical terminologies and ontologies. Investigated this year were RxNorm and the UMLS (categorization of polysemous concepts, alignment of relationships between Metathesaurus and Semantic Network). Recommendations were made for evaluating the quality of vocabularies for use in NCI's caBIG.

We continue to collaborate with leading ontology and terminology centers, including the National Center for Biomedical Ontology and the International Health Terminology Standards Development Organization.

The major objective of the Medical Ontology Research project is to develop methods whereby ontologies could be acquired from existing resources (including the Unified Medical Language System), as well as validated against other knowledge sources.

Specific objectives for FY2009:

- To integrate biomedical information from various knowledge sources using Semantic Web technologies.
- To format biomedical terminologies (e.g., MeSH) for use in the Semantic Web.
- To evaluate the use of UMLS concept identifiers as a source of permanent identifiers (Uniform Resource Identifiers) for the Semantic Web.
- To assess similarities and differences between the UMLS Semantic Network and other top-level biomedical ontologies.
- To integrate publicly available drug information sources through RxNav, including RxNorm, NDF-RT and MedlinePlus.
- To continue providing the service of mapping between vocabularies to client projects such as ClinicalTrial.gov and The Indexing Initiative.

Semantic Knowledge Representation (SKR)

The **Semantic Knowledge Representation** project provides a context for basic research in natural language processing based on the UMLS knowledge sources. Research focuses on development of SemRep and MetaMap to extract semantic predications from text to support innovative information management applications in biomedicine. We are currently developing Semantic MEDLINE, a Web application which exploits semantic predications to help users manage the results of PubMed searches. Research is being conducted to adapt the application to support clinical practice guideline development (in cooperation with NHLBI) and scientific portfolio management (in cooperation with OD/OPASI). Further collaboration with NLM/SIS is extending the technology to medical aspects of disaster information management.

The context of the SKR project is articulated in the NLM Long Range Plan, especially 1.6.1 (Discovery initiative: Facilitate scientific discovery through computational methods which identify relevant linkages across a variety of information resources) and 1.6.3 (Advanced literature search tools for finding articles and facts for targeted purposes, including decision support and guideline development).

Major objectives for the planning year include:

- Continue to expand SemRep effectiveness, concentrating on recall.
- Develop an algorithm to efficiently process large amounts of biomedical text (MEDLINE citations, grant applications, and Web documents) to accommodate Semantic MEDLINE.
- Collaborate with the Disaster Information Management Research Center in the Division of Specialized Information Services to expand SemRep to medical aspects of disaster information management, focusing initially on influenza

epidemics, burns management, and post-traumatic stress disorder.

- In continued collaboration with NHLBI and OPASI enhance Semantic MEDLINE as an adjunct to traditional information retrieval systems for helping biomedical researchers managing large amounts of text.
- Expand the use of Semantic MEDLINE for literature-based discovery.
- Conduct research to combine SemRep processing with MedLEE for effective processing of clinical text.

UMLS and Clinical Vocabulary Standards

Unified Medical Language System (UMLS)

The most recent release of the UMLS Metathesaurus contains over 1.5 million concepts and 7.7 million concept names. After the successful transition of the production of the Metathesaurus to NLM's Office of Computer and Communications Systems (OCCS), staff are focusing on the research and development aspects of the Metathesaurus. A NLM-wide UMLS Priorities and Services Working Group is convened to formulate proposals for priorities of development in the coming 3 -5 years. The group solicits input from recent surveys of UMLS users, questions and comments received via the UMLS listserv and at professional meetings, customer inquiries; other NLM staff; knowledgeable people in other Federal agencies and in standards organizations; and known heavy users. Staff continue to provide assistance to the Library Operations (LO) team in providing user support; and in various terminology related projects initiated or supported by NLM nationally and internationally.

UMLS-CORE Project

The UMLS-CORE (Clinical Observations Recording and Encoding) Project has finished the data collection phase and data analysis is ongoing. The goal of this project is to identify a clinical subset of the UMLS to support consistent high level encoding of clinical information (e.g. in discharge summaries or problem lists). Lists of terms and their actual frequencies of usage in real-life clinical systems are collected from large healthcare providers including: Kaiser Permanente, Mayo Clinic, Intermountain Health Care, Regenstrief Institute, Nebraska University Medical Center and the Hong Kong Hospital Authority. These terms are mapped to the UMLS and their pattern of usage and overlap are studied. 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 qualify improvement efforts on the most frequently used terms.

Terminology Representation and Exchange Format (TREF)

The specification of the Terminology Representation and Exchange Format (TREF) has been finalized. The purpose 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 simplify the task of inversion of source terminologies into the UMLS editing environment. There is an ongoing collaborate with the National Center for Health Statistics to produce a TREF version of ICD9CM. This will be the first major terminology published in the TREF format and will fill the need for a machine-readable version of ICD9CM.

UMLS User and Usage Pattern Study

Ongoing collection and analysis of UMLS user and usage information through the Web-based annual report application. This information will help guide the NLM-wide UMLS Products and Services Working Group which will recommend priorities for UMLS development in the next three to five years.

RxTerms

RxTerms is an innovative solution to a common problem in the development of clinical applications that capture medication or prescription information. The lack of a publicly available interface terminology for drugs has meant that application developers must either use proprietary terminologies or build from scratch. RxTerms fills this gap by providing a free, user-friendly, and efficient drug interface terminology that links directly to RxNorm, the national terminology standard for clinical drugs which is also developed by NLM. Efficiency of data entry is enhanced by systematic segmentation of RxNorm clinical drug names and aggressive pruning of drugs that are not available in the US. Additional synonyms from sources outside RxNorm further enhance the user-friendliness of RxTerms. RxTerms is currently being used in one of CMS's applications in the post-acute care environment (CARE). It will also be used in the NLM Personal Health Record. RxTerms will be freely available for download from NLM's Web site for wider testing and feedback.

Disaster Information Management

RxHub Medication Reconciliation Project

Funded by the Bethesda Hospitals' Emergency Preparedness Partnership, a collaboration formed by three Bethesda area healthcare facilities (National Naval Medical Center, NIH Clinical Center, and Suburban Hospital), this project will study the feasibility and value of using prescription history information from RxHub in a disaster situation. RxHub access points will be set up in these three local Bethesda hospitals and RxHub data will be compared with that obtained by the traditional manual medication reconciliation process in terms of coverage, accuracy and completeness. This external source of medication information that can be obtained automatically could be both time and life saving in disaster circumstances.

Our hypothesis is that when a patient has medication information in the consortium of Pharmacy Benefit Managers (PBM) databases, that information will be more complete and more precise than the corresponding information collected manually from the patient. We also hypothesize that the PBM medication history will be obtained more quickly and with less effort than the manual history. The value of such data for a population will depend upon the proportion of patients in the population who have medication information within the consortium database. The PBM consortium will have no information about patients without insurance or those with insurance who have not been processed by the PBM consortium. We will set a binary variable to identify patients who had no data in the PBM consortium and will collect demographic, administrative variables (arrival mode), and insurance class on all patients. Then we will model the existence of PBM consortium data on these attributes. We will use this model to predict the proportion of patients with data in the PBM and to assess policies that might eliminate this gap.

Lost Person Finder (LPF)

The **Lost Person Finder (LPF)** is included as a part of BHEPP program and funded by interagency agreement. This partnership seeks to create systems that would be used in the event of a disaster, either natural or manmade. The LPF system addresses the problem of missing people, a common problem in the chaos of a disaster event. At registration, pictures may be taken of patients/disaster victims by registrars or volunteers using digital cameras or cell phones. These pictures are uploaded to the database along with other information. The LPF matches these with pictures and descriptive information from the general public searching for lost children, spouses, or friends. The LPF displays pictures and some descriptive information on large screens situated at the hospitals, both indoors and at key outside locations. In addition, through remote computers or handheld devices the public may access LPF to search for this information.

Training Opportunities

Working towards the future of biomedical informatics research and development, the Lister Hill Center provides training and mentorship for individuals at various stages in their careers. The LHNCBC Informatics Training Program (ITP), ranging from a few months to more than a year, is available for visiting scientists and students. Each fellow is matched with a mentor from the research staff. At the end of the fellowship period, fellows prepare a final paper and make a formal presentation which is open to all interested members of the NLM and NIH community.

In FY2008, the Center provided training to 37 participants from 13 states and five countries. Participants worked on research projects including 3-D informatics research, personal health record research, medical image processing, image & text retrieval, InfoBot research, interactive publication research, information retrieval, document analysis, natural language processing, ontology research, question answering research, grid computing, medical terminology research, 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. Participants in the program have been authors on 24 of the 64 manuscripts (38%) published by LHC researchers in the period from June 2007 through June 2008.

In FY2008, we started a new Clinical Informatics Postdoctoral Fellowship Program to attract young physicians to NIH to pursue research in informatics. This program is run jointly by the Lister Hill Center and the Clinical Center to bring postdoctoral fellows to labs throughout NIH. Funding is from the LHC. Our first Clinical Informatics Fellow arrived in May.

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 provides 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 providing the public with the tools, databases, and technologies that will enable genetic discoveries of the 21st century.

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 and apply the results of their research to the development of public information resources.

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

- Creates automated systems for storing and analyzing information about molecular biology and genetics with evidence in biomedical literature.
- Performs research into advanced methods of computer-based information processing for analyzing the structure and function of biologically important molecules and compounds.
- Facilitates the use of databases and software by researchers and healthcare personnel.
- Coordinates efforts to gather and disseminate biotechnology information worldwide.

Molecular Biology Information Resources

NCBI's molecular biology information resources are based on sequence repositories upon which curated and annotated sets of data resources are built. 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 infrastructure is also made up of computer/user support and biology research in genomic analysis.

GenBank

The basis for NCBI sequence data is GenBank®, the NIH genetic sequence database. GenBank is 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. This year, NCBI celebrated 25 years of GenBank service with a two-day meeting that included presentations by distinguished molecular biologists.

Important sources of GenBank data are direct sequence submissions from individual researchers and scientists as well as institutions, such as genome sequencing centers. Thousands of sequence records are submitted prior to publication. Records submitted to NCBI's international collaborators—EMBL (European Molecular Biology Laboratory) in the UK and DDBJ (DNA Data Bank of Japan)—are shared through an automated system of daily updates. Other cooperative arrangements, such as those with the US Patent and Trademark Office for sequences from issued patents, ensure that the collection contains all available relevant data.

GenBank is comprised of two divisions of sequences, traditional nucleotide sequences and Whole Genome Shotgun (WGS) sequences. WGS sequences are contigs (overlapping reads) from WGS projects. Annotations are allowed in WGS assemblies, and records are updated as sequencing progresses and new assemblies are computed.

The traditional nucleotide database is divided as well into three specialized components consisting of Expressed Sequence Tags (EST's), Genome Survey

Sequence (GSS) records, and the "CoreNucleotide" group.

A new division, the Transcriptome Shotgun Assembly (TSA), was included in release 165 in April of this year. TSA sequences are shotgun assemblies of primary (mRNA) sequences deposited in dbEST, the Trace Archive, or the Short-Read Archive (SRA).

The Third Party Annotation (TPA) database, created in conjunction with international partners, 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 contains data supported by peer-reviewed, experimental evidence. TPA:inferential contains data by inference where the source molecule or its products have not been the direct result of experimentation.

The amount of data submitted to GenBank grows continuously. In fact, from 1982 to the present the number of bases in GenBank has doubled approximately every 18 months. GenBank's two divisions combined now contain over 127 million sequence records. At the end of FY2008, the database reached a milestone of over 200 billion basepairs. The traditional nucleotide sequences division base count rose from 79 billion in August 2007 to 95 billion in August 2008. It contains 92 million entries. The WGS sequences division base count rose from 101 billion in August 2007 to 118 billion in August 2008. It contains 40 million entries. 1066 public WGS projects have been released by GenBank with 345 new projects in the last year and 59 updated reassemblies. 140 sequenced genomes were processed by NCBI this year. New full releases of GenBank are distributed every two months. Daily updates are made available via the Internet and the World Wide Web.

Substantial resources are devoted to the analysis and curation of sequence data. GenBank indexers with specialized training in molecular biology create the records, applying rigorous quality controls. NCBI taxonomists consult on organism classification, and, as a final step, senior NCBI scientists review the records for biological accuracy.

NCBI has developed various tools for GenBank data submission. Sequin is a stand-alone tool that updates and submits large groups of sequences. Sequin MacroSend allows submitters to upload a Sequin file from their computer directly to the GenBank indexing staff. Upon submission they receive a temporary identification number. Another tool, BankIt, is now in its fourteenth year of use. BankIt is useful for small submissions that can be uploaded directly to NCBI via the Web. Sequin and BankIt are continuously updated to improve ease of use and to accommodate new data types and formats. Guides for specialized submissions such as genomes, batch sequences, and alignments are available online.

In order to simplify access to, and improve the quality of the enormous amount of data stored in GenBank, NCBI is continuously developing new tools and enhancing existing products and methods.

Sequence data, both nucleotide and protein, are supplemented by pointers to abstracts and publishers' full-text documents as they become available. Links are provided to other NCBI and outside resources, such as biological databases and sequencing centers. The links enables GenBank to serve as a key component in an integrated database system that allows researchers to perform comprehensive and seamless searching across all related biological data on the NCBI Web site.

NCBI has been involved in the Barcode of Life Project, which is creating a public collection of reference sequences from vouchered specimens of all species of life. A barcode sequence is a short nucleotide sequence from a standard genetic locus for use in species identification. NCBI developed a barcode submission tool (BarSTool) that facilitates the submission of barcode sequences to GenBank.

Genome Information Resources

NCBI plays a key role in assembling and annotating genome sequences. A suite of genomic resources, specialized tools, and databases have been developed to support the comprehensive management, mapping, and analysis of entire genomes and sequence data. In addition, NCBI maintains an expanding collection of integrated resources that identify the biological relationships between genome sequences, expressed mRNAs and proteins, and individual sequence variations. The genomic information databases include: BLAST, dbSNP, RefSeq, CCDS, MapViewer, Entrez Gene, Probe, UniGene, HomoloGene, and GEO. These networked systems also link to outside information such as Linkage and Physical Maps, TaxPlot, and chromosome-specific mapping data.

The Reference Sequence (RefSeq) database is a comprehensive, integrated, non-redundant set of sequences for major research organisms. RefSeq sequences include genomic DNA, gene transcript (RNA), and protein products that 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 curated RefSeq collection contains 5,859,684 proteins representing 5,513 organisms. The RefSeq curation group supports the whole genome annotation process flow for updating existing genomes and processing new submissions. The group also provides curated transcripts and proteins.

Extensive testing, quality assurance, and documentation are essential to the release of data in Map Viewer, BLAST, and Entrez Gene, as well as documentation for Web sites that support the scientific community's access and use of NCBI resources. In

FY2008, several new genomes were annotated and updates were provided to existing genome assemblies for 27 eukaryotic species including human, fruit fly, zebrafish, and several protozoan and fungal species.

Map Viewer is the primary tool for visualization of large genomes. Genes or markers of interest are found by submitting a query against a whole genome or by querying one chromosome at a time. Cross-species comparison is supported by increased standardization of map features. Maps from outside sequencing centers are utilized for multiple-species queries. Query results are viewed in a results table that includes links to a chromosome graphical view where a gene or marker is seen in the context of additional data. During FY2008, a new Map Viewer homepage was released that allows for expansion and contraction of organism kingdoms and divisions. The Evidence Viewer is a Map Viewer feature that provides graphical biological evidence supporting a particular gene model. Model Maker allows users to build a gene model using selected exons.

The Entrez Gene database makes it easier for researchers to find and interpret gene-specific information by providing a unified query environment for genes defined by sequence and/or genes included in the Map Viewer. It integrates information about genes and gene features annotated in RefSeq and collaborating model organism databases. Entrez Gene continued to be enhanced in usability and content in FY2008. For example, several new indices and properties were added to facilitate retrieval by chromosome location, identify human proteins that interact with HIV proteins, and retrieve genes by number of exons. Also, more reports were integrated for the human genome in order to facilitate identification of genes of clinical interest. Significant improvements were made to the automated process of indentifying genes within genomic DNA sequences. Several new genomes were annotated including pig and pea aphid, bringing the total number of organisms represented in RefSeq to 18. A new sequence viewer, released in FY2008, has improved zooming capacity and horizontal browsing of graphical records. Overlapping features can be displayed simultaneously, such as assembly details, genes, transcripts, and coding regions. A multi-panel display can show views of varying scales and a text sequence display allows for copying and pasting to other applications.

Collaboration between the Entrez Gene team and the Human Genome Epidemiology Network (HuGENet) of the CDC resulted in the integration of thousands of new connections between Entrez Gene and PubMed, particularly in the area of genetic epidemiology and association studies (HuGE Navigator). Collaboration with NHGRI resulted in tagging genes associated with complex disorders. Collaboration between Entrez Gene, RefSeq, and many stakeholders in the human genetics community resulted in over 550 genomic sequence standards (RefSeq Gene) generated for human genes that are often genetically tested.

The Consensus Coding Sequence (CCDS) database identifies a core set of consistently annotated, high quality human and mouse protein coding regions. This year, annotation was updated for human and mouse coding regions resulting in 6,202 new CCDS entries.

The Genome Reference Consortium is a new international collaboration formed with the goal of updating and improving the mouse and human genome assemblies. The project plan is to correct the small number of regions in the reference sequence that are currently misrepresented, close as many of the remaining gaps as possible, and produce alternate assemblies of structurally variant loci when needed. NCBI is providing informatics support for the project, including tracking of tiling path files, overlaps between adjacent clones, and curation. NCBI will also generate the final assembly after collaboration and will be responsible for quality assurance.

UniGene provides non-redundant clusters for the highly redundant sets of transcript sequences of expressed genes. Expanded coverage over the past year, brings the number of animals, plants, and fungi represented to 105.

HomoloGene, a reliable and comprehensive database of gene homologs, complements to Entrez Gene and UniGene. Covering 20 animal and plant model genomes, it provides statistics on inter-species sequence and protein domain conservation. HomoloGene is linked to the genome-wide views available in Map Viewer 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. They also provide simplified BLAST access by using one reference sequence from each cluster to represent the entire cluster. Currently, the collection includes 2.5 million proteins from 700 genomes. Updated weekly, results are presented to the public via FTP releases and the Entrez system.

The database of single nucleotide polymorphisms (dbSNP) is a comprehensive catalog of common human genetic variation. dbSNP contains over 55 million submissions of human genome data that has been processed and reduced to a non-redundant set of 14 million refSNP clusters. Forty-three other organisms are represented in the SNP database, with 42 million submissions curated to 35 million refSNP clusters.

Like most of NCBI's data repositories, the Probe database, is part of the Entrez system. 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. This

database stores molecular probe data, together with information on success or failure of the probes in different experimental contexts. The database contains over 9 million probes as of October 2008. 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 provides guides for comparing organisms on a genome scale. There are currently 33 guides available. The Genome Resource Guides provide information on genome-related tools and repositories available through NCBI and various outside centers and institutions. The guides explain the easy navigation to NCBI resources, such as organism-specific BLAST and Map Viewer pages, and list outside resources that provide sequence, mapping, and clone information. The Guides also list documentation, annotation and comparative genomic projects. New genome guides created in FY2008 include the organisms: jewel wasp, platypus, zebra finch, and water buffalo.

The Entrez Genome database provides views for a variety of genomes, complete chromosomes, sequence maps with contigs, and integrated genetic and physical maps. The database is organized in six major organism groups: Archaea, Bacteria, Eukaryotae, Viruses, Viroids, and Plasmids and includes complete chromosomes, organelles and plasmids as well as draft genome assemblies. It includes, but is not limited to assembly, annotation, and genome sequencing projects, such as whole genome shotgun or BAC ends, large-scale EST, and cDNA projects. The six organism-specific overviews function as portals from which all projects in the database pertaining to that organism can be browsed and retrieved. There are currently 2403 genome sequencing projects completed, in draft form or in progress. The number of species represented in the database is currently 4,835.

The Viral Genomes Web site provides a convenient way to retrieve, view, and analyze complete genomes of viruses and phages. NCBI's viral genotyping tool helps identify the genotype of a viral sequence using BLAST. NCBI currently provides access to 3,180 reference sequences for 2,131 viral genomes and 40 reference sequences for viroids.

Fungal Genomes Central is a portal to information and resources about fungi and fungal sequencing projects. There are currently 151 fungal genomes in various stages of annotation. 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 Microbial Genome Annotation Pipeline was developed for annotation of prokaryotic genomes. Approximately 767 genomes have been annotated in-house and NCBI is working with ten outside groups who submit data. During FY2008, the Microbial Genome Annotation Pipeline annotated 545 microbial genomes: 187 complete genomes and 358 whole genome shotgun draft assemblies.

Specialized Databases and Tools

The Influenza Virus Resource is a comprehensive collection of flu sequences. Samples collected all over the world include viruses obtained from birds, pigs, humans, and other species. Data is obtained from the NIAID Influenza Genome Sequencing Project and from GenBank. Links are provided to other flu resources containing sequences, publications, and general flu virus information. About 20,580 new influenza virus sequences were entered into NCBI's Influenza Sequence Database in FY2008. About half of them were from the NIAID Influenza Genome Sequencing Project, the University of Hong Kong, the Centers for Disease Control and Prevention, and the Air Force Institute for Operational Health, and were processed by the NCBI flu annotation pipeline.

The Flu Dataset Explorer provides an interactive tool for preliminary analysis of protein sequences from the Influenza Sequence Database or from a user's own file. New functionalities were added to the Influenza Virus Resource, including an improved multiple sequence alignment tool and a new version of the phylogenetic tree building tool. Presentations and demonstrations on the NCBI Influenza Virus Resource were given at national and international meetings, and papers were published using the resource to further enhance flu research.

The NCBI Trace Archive is a permanent repository of DNA sequence chromatograms (traces), base calls, and quality estimates for single-pass reads from various large-scale sequencing projects. The trace data can be scanned using a rapid nucleotide-level cross-species sequence similarity search program called cross-species MegaBLAST. Using the visualization tools of the related Assembly Archive, researchers can examine an assembly of trace data from which a finished genomic nucleotide sequence has been derived. They can determine, for instance, if a crucial nucleotide base change associated with a disease is supported by the sequence evidence. The Trace Archive currently holds over two billion traces representing over 980 species.

The Short Read Archive (SRA) was created to archive information generated by next-generation technologies from companies such as 454, Solexa, ABI, and Helicos. It presents data by sequencer runs rather than individual traces from new massively parallel sequencing technologies. Created in June 2007, SRA contains 222 studies and 4 terabytes of data.

The Gene Expression Omnibus, or GEO, is a high-throughput gene expression/molecular abundance

data repository providing curated online storage and retrieval of gene expression data. Profiles are submitted via GEOarchive, a spreadsheet format for large batch submissions. For the Short Read Archive, GEO has established submission and brokering standards and procedures for high throughput sequencing data. GEO also added clustering of arrays, based on chromosomal coordinates. A new proteomics database and submission pipeline was developed and will be released soon. The GEO Web site receives about 100,000 Web hits and 12,000 data downloads per day. In FY2008, new submissions reached 97,000, bringing the total number of records to over 300,000. Six billion new individual data points were also accepted this year, bringing the total number to over 16 billion.

The Molecular Imaging and Contrast Database (MICAD) contains information on in vivo molecular imaging agents and is a key component of the NIH Roadmap. MICAD is part of the NCBI Bookshelf. In FY2008, in order to expand the database, a Guest Author Program was initiated to encourage members of the imaging community to prepare chapters in order to increase the breadth and depth of the database. Also, a Supplemental Information Section was added to each chapter to augment the database with state-of-the-art unpublished research information from the community.

The database for the Major Histocompatability Complex (dbMHC), contains variations found only in alleles of the major histocompatability complex (MHC), a highly variable array of genes that play a critical role in determining the success of organ transplants. The MHC region is largely responsible for an individual's susceptibility to infectious diseases. The dbMHC supports six important research projects.

The NCBI Taxonomy Project provides a standard classification system used by the international nucleotide and protein sequence databases. NCBI's rapidly growing Taxonomy database is curated to include the names of species for which sequences have been submitted to the protein and nucleotide databases. Tools have been developed for representing alternate, externally maintained taxonomies and cross-mapping them with the Taxonomy database entries. A database of biological material collections has been developed to enhance links between NCBI sequence entries and the corresponding specimen entries. The Taxonomy database browser can be used to view position in the taxonomic tree or retrieve data in any Entrez database for a particular organism or group. Searches may be made on the basis of whole, partial, or phonetically spelled organism names. The Taxonomy system also provides a 'Common Tree' function that builds a tree for a selection of organisms or taxa.

The UniVec database is used to quickly identify segments of nucleic acid sequences that are of vector origin or vector contamination. It was significantly expanded (Version 5.0) in FY2008, to enable the VecScreen tool to detect more foreign sequence contamination in nucleotide sequences.

Chemical Information

PubChem is organized as three linked Entrez databases. These are PubChem Substance, PubChem Compound, and PubChem BioAssay. Together, they form a complete information resource for millions of small molecules, including their bioactivity data, structures, and properties. The PubChem databases are a key component in the Molecular Libraries and Imaging initiative of the NIH Roadmap.

PubChem BioAssay allows users to examine descriptions of each assay's parameters and readouts, with links to substances and compounds enhanced by a queuing system and caching mechanism. The volume of biological data doubled in FY2008, with over 30 million test results. The number of bioassays increased from 587 in August 2007 to over 1,000 in July 2008.

The BioAssay deposition system provides a tool for tracking updates on assay descriptions, adding additional test results, and redefining test result reports. In FY2008, the deposition system was significantly enhanced with a redesigned Web interface, shortened processing time for assay publishing, and the addition of assay submission via FTP. Presentation of Entrez search and retrieval results have been improved as well. Links to other databases are now based on specific information categories. A redesigned portal interface has improved discoverability of annotation for small molecules with bioassay data, and integration between different information resources. More data can be downloaded and new analysis tools are available. Annotation for biological test results such as "active concentration" has been added to over 200 bioassays. Annotation for protein targets are provided by integrating bioassay summaries with the Conserved Domain Database (CDD) and related protein structure data. These annotations have also improved bioassay searches and test result comparisons. A specialized BLAST search feature was created to search against biological targets to help users discover activity data. The BioActivity Summary tool provides an overview of biological testing results and activity profile for chosen compounds. A Structure-Activity Analysis tool classifies compounds and assays simultaneously based on activity profiles and target and chemical structure similarities.

The PubChem Compound database provides unique chemical structures and validated chemical depiction information describing substances in the PubChem Substance database. There are now more than 19 million compound records. The PubChem Substance database contains chemical substance records and associated information. Currently, there are nearly 41 million substance records. The number of records in these two databases has doubled since last year.

PubChem contains an extensive set of links to related information within its own sets of data as well as to other Entrez databases and outside resources. 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. Many compounds have literature citations to PubMed as well as links to the proteins and/or genes representing a protein to which they bind.

Compounds are searchable by chemical structure, chemical properties, and bioactivity. New versions of the Chemical Structure search system were released over the past year. Significant modification and enhancements were made to compound and summary pages including the addition of the NLM Daily Medication information section. Computed 3D coordinates for structures in the PubChem Compound database were released on the PubChem FTP site. Also, a PubChem Standardization Service shows users how PubChem would handle any chemical structure they would like to input

Continuous improvements are required to keep up with the steady growth of PubChem. The need to rapidly process data is paramount. Major infrastructure upgrades included increasing database server disk capacity by a factor of three. Nearly all database server hardware now uses 1 TB disks, which handle must greater loads. These upgrades were implemented without disrupting service. The Web site is visited by 40,000 users per day.

The PubChem Power User Gateway (PUG), deployed in 2007, is an interface that allows users to search and download data that is not accessible to the Entrez system, such as structure queries. This year, interfaces were added to the PubChem Power User Gateway (PUG), which further opened PubChem to data mining. A beta release of the Web-based PUG SOAP (Simple Object Access Protocol) was released, which complements the existing PUG service, integrating it with existing workflow applications as well as many programming and scripting languages.

Protein Structure

NCBI's Molecular Modeling DataBase (MMDB) is the Entrez structure database, a compilation of all the biopolymer structures in the Protein Data Bank (PDB). MMDB is augmented with domain annotations and links to relevant literature; protein and nucleotide sequences; chemicals and conserved domains in the CDD; as well as structural neighbors computed by the VAST algorithm on compact structural domains in the 3D Domains database.

MMDB contains over 50,000 unique, experimentally determined 3D structure records. The database is updated weekly, with the source PDB data checked for consistency in the purported chemistry, sequence, and 3D coordinates.

An interaction tracking database (with the working title of "Intrac") has been established to store intermolecular interaction data as observed in the 3-dimensional structures tracked by MMDB. Intrac holds observed interactions as well as interactions inferred by their structural similarity. The similarity, the corresponding structure superpositions, and the derived molecular sequence alignments are calculated by the VAST algorithm. Intrac focuses on interactions between protein domains conserved in molecular evolution and other protein domains, nucleic acids, and small molecules—also referred to as ligands. Intrac will facilitate computational prediction of binding partners and the associated binding sites for a large number of proteins in the Entrez database, which are closely related to proteins with known three-dimensional structure, providing some hints at biological function where experimental characterization of function is pending. Intrac is currently being refined so that interactions can be displayed on public Web pages maintained by the NCBI structure group.

The Conserved Domain Database (CDD) is the Entrez database of sequence alignments and profiles defining protein domains as recurrent evolutionary modules, ancient domains, and full-length proteins. The CDD annotation staff produces curated hierarchies of models related by descent from a common ancestor, representing the ancient evolutionary history of protein and domain families. The staff use 3D structure information, phylogenetic analysis, Entrez resources and published literature to enhance alignment quality; annotate functional sites; identify relevant links to PubMed and the NCBI Bookshelf; and update domain family summary descriptions.

The most recent version of CDD (Version 2.14), released in spring 2007, now features explicit representation of domain superfamilies in Entrez/CDD. Superfamily clusters are calculated as sets of conserved domain models. These models are imported from outside sources as well as from domain hierarchies curated at NCBI. The clusters help explain relationships between overlapping and partially redundant domain models and provide an approximate count of ancient protein domain superfamilies, i.e., units conserved in molecular evolution which appear to be independent of each other—or not related by common descent. Currently, the CDD resource reflects about 10,000 such superfamilies. Also, CDD now mirrors the Pfam protein family database, Version 22, in its entirety.

CDD Version 2.14 also comes with domain footprint annotation on protein sequences. The new version summarizes information from redundant and homologous domain models at a superfamily level. Domain annotation on proteins is flagged as either "specific" (identifying molecular function with high

confidence) or as 'non-specific' (identifying superfamily membership only).

CD-Search is the Web-interface to CDD which visualizes domain-based annotation on protein sequences. CD-Search also shows the location of conserved sites, as inferred from alignments to NCBI-curated domains models, such as active sites and binding sites, providing direct links to the domain models of the site data and evidence collected by CDD curators.

NCBI's three-dimensional structure viewer, Cn3D, provides an interactive three dimensional graphical image of molecular protein structures from the Entrez system. Cn3D also serves as a visualization tool for sequences and sequence alignments. The ability of Cn3D to correlate structure and sequence information distinguishes it from other viewers. Cn3D features custom labeling options, coloring by alignment conservation, and a variety of file export formats that together make this a powerful and extremely user friendly tool for structural analysis.

CDTree, together with Cn3D, is the main application used by CDD curators to create models. CDTree and Cn3D function as helper applications for Web browsers and can be used to study molecular evolution of proteins and protein domain families, as a powerful interface to the PSI-Blast program, and as a viewer for NCBI-curated conserved domain models and hierarchies.

A new version of CDTree has been released. It features a new viewer that summarizes annotation on a set of models in a hierarchy, supports simultaneous operations on multiple models in the curators workbench, provides greater flexibility in coloring sequence trees, and contains optimized computations that significantly speed up processing.

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

The structure group has begun work on a new database to store information on biological systems, such as metabolic pathways. This database will provide a means to record non-homologous, but functional relationships between biopolymers tracked in the Entrez database, and also between biopolymer sequences and small molecules, as they are tracked in PubChem. The biosystems database will accept depositions from data providers outside NCBI and will facilitate queries and retrieval strategies in Entrez that have not been possible previously. The biosystems database will restrict functionality, but point to Web-services of individual data providers for in-depth analysis and sophisticated visualization of search results.

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 millions of known sequences and determines the closest matches. The NCBI Web interface for BLAST allows users to assign titles to searches, to review recent search results, and to save parameter sets in My NCBI for future use.

The BLAST suite of programs is continuously enhanced for effectiveness and ease of use. In July 2008, a new URL for BLAST searches was made public <blast.ncbi.nlm.nih.gov>. Also in FY2008, a new BLAST report was designed that features more concise organization, grouped BLAST results links, a new "Search Summary" link, downloadable links, and collapsible sections. Gene information was added to BLAST reports including Gene IDs, gene name, and gene entry title. BLAST now offers searches on human and mouse genome databases that are prefiltered to eliminate matches to low-complexity and repeat sequences, thus reducing runtime by two-thirds.

The BLAST tree view option shows a dendrogram that clusters sequences according to their distances from the query sequence. This display is helpful for recognizing the presence of aberrant or unusual sequences or potentially natural groupings of related sequences. Improvements to tree view include new evolutionary distance models, tree downloading, rerooting at any user-selected node, collapsible subtrees, and sequence grouping.

Integration of Clinical, Genetic, and Environmental Databases

The NCBI database of Genotype and Phenotypes (dbGaP) melds genotype and phenotype data. The data is collected from various clinical studies, organized and distributed as an open access subset to the public and a controlled access subsets to researchers.

Open access information allows all users to browse and search projects and studies, protocols, questionnaires, and supporting documents. Users are able to view summaries of the genotype and phenotype data, where permitted. Controlled access data includes de-identified individual phenotypes and genotypes, pedigrees, measured traits, genotype calls, raw genotype data (CEL files), and select analyses during the period of protected use.

Study Submissions

When dbGaP was launched in late 2006, two studies had been published. As of June 2008, there have been 23

studies released. Eighteen of these studies contain both public summary-level data and individual level phenotype/genotype data distributed through the authorized access (AA) system. Five contain public data distributed through the dbGaP FTP site. Within the latter set, three studies are summary-level association results linked to published GWAS and two are GAIN genotype quality control studies that provide genotype data for HapMap samples. HapMap samples have been consented for unrestricted broad public distribution.

Studies with individual-level data submitted in 2008 covered the disease areas of psoriasis, schizophrenia, bipolar disorder, Parkinson's Disease, cerebrovascular disease, ischemic stroke, amyotrophic lateral sclerosis, and motor neuron disease. The Framingham Heart Study was updated. Additional genome-wide association study results were submitted for type 2 diabetes, systemic lupus erythematosus, and neuroblastoma.

Collectively, the studies released in 2008 included measurements for 55,109 total research participants, and GWAS measurements included:

- Over 30,000 unique phenotype traits.
- 624 XML-based phenotype documents linked to 47,894 variable summaries. These document the studies through natural language descriptions, collection forms, or scientific protocols, and also provide models for future research. Links between searchable documents and variable descriptions provide an unprecedented level of functionality and usability.
- 22.4 billion total individual genotypes, i.e. single-nucleotide measurements of participant DNA sequence. These potentially reveal systematic and heritable genetic differences between affected and unaffected individuals.
- 2,819 pre-computed statistical associations between select phenotype traits and participant genotypes that describe locations in the human genome where differences between affected and unaffected participants are statistically significant.

Phenotype variable summary metrics, data dictionaries, summary-level association results, and study documents (XML files) are distributed to the public via anonymous FTP service. Approximately 153,000 file download requests were logged from September 2007 to June 2008.

Authorized Access System Download Activity

Principle investigators (PI's) log in and request access to de-identified individual level data, via the dbGaP authorized access system. Users are approved by institutional signing officials and NIH Institute/Center Data Access Committees (DAC). Requests are routed to one or more NIH Institute/Center Data Access Committees (DAC) for review. DAC review confirms that each proposed research use is consistent with the restrictions placed on the data by study participants during the informed consent process. Approved users return to the dbGaP system to create and download a password-protected copy of the data (e.g. phenotypes and genotypes). Then they are able to download it to their secure local computing environment.

As of August 2008, 437 PI-defined research projects have been created, using the approval system. Of these, 280 started at least one research project with data request, 205 have at least one DAC-approved research project, and 158 have downloaded at least one dataset. Separate requests and DAC approvals are needed for each distinct research use of a dataset.

Data Usability: Tools and Software Development

During 2008, new functionality was implemented for both stand-alone data analysis clients and Web-based data visualization displays. In the first case, the dbSNP software development team created a set of GWAS plug-ins for the NCBI Genome Workbench to support data visualization and exploration in the context of a graphical DNA sequence browser. The plug-ins support four important functions for GWAS data: data file loading, data element filtering, data element integration with Entrez, and dynamic rendering to keep data elements synchronized with zoom scale. In the second case the dbGaP association results browser was enhanced to show multiple analysis tracks in simultaneous view, show local recombination rate data, and support search and navigation functions for dbGaP variables, genes, and SNPs.

Entrez Retrieval System

Entrez, the major database search, retrieval, and indexing system at NCBI, was originally developed for searching nucleotide and protein sequence databases and related MEDLINE citations, but has since expanded to become the indexing and search foundation for all of NCBI's major resources. With Entrez, users quickly and easily search gigabytes of sequence and literature data. A key feature of the system is the concept of "neighboring," which automatically identifies references or sequences that are related to a user's research. The ability to traverse the literature and the molecular sequences via "neighbors" and links, provides an efficient and intuitive way of accessing data. Entrez currently supports and integrates 35 databases, providing sequence, taxonomy, gene, chemical, and biomedical literature and data. Entrez Global Query enables users to search Entrez-supported databases simultaneously, in seconds, displaying the number of hits in each database on a single page view.

Discovery Initiative

NCBI has embarked on a program to help users better explore the myriad of data contained in NCBI's resources. The Discovery Initiative aims to improve the usefulness of NCBI information resources by using automated methods to draw users' attention to related data that do not necessarily appear as part of the original search. For example, users performing searches for medical terms in the PubMed database may not even be aware that separate databases, on genetics or drugs, for example, contain additional relevant information.

To improve the search results pages and to augment document summaries with the most useful related information for users, the Discovery Project analyzes Web logs and other usage metrics and modifies the Entrez response accordingly. New results pages are tested on a subset of users, and if successful, this approach is implemented for all users. The technology behind the Discovery Project is a re-designed retrieval system known as the Portal. The new infrastructure utilizes interrelated portlets that provide a more flexible and robust approach to development.

Recently, a prototype system has been developed that makes suggestions to users based on their query and on stored results of general Weblog mining. This year after beta testing 5% of users, PubMed Summary results pages began showing related information from other resources as well as the regular set of PubMed results.

Literature Information Resources

PubMed

PubMed provides Web-based access to citations and abstracts for the biomedical science journal literature. PubMed is comprised primarily of journals indexed in NLM's MEDLINE database, but also contains a limited number of journals outside the scope of MEDLINE. Links to articles available in full text through NCBI's PubMed Central database are also provided. Serving as the foundation of NCBI's bibliographic information system, PubMed contains over 18 million citations from more than 20,300 journals, some dating back to the 1950s. Links to full text journals indexed in PubMed have increased from 5,880 in July 2007 to 6,324 in July 2008.

PubMed is continually updated and enhanced for better functionality and more precise search results. A beta version of a new advanced search page was released this year, which makes it easier for users to search by author, journal, and publication date. PubMed's automatic term mapping was augmented to recognize terms that do not include search tags such as author names, journal titles, and MeSH terms, thereby providing more comprehensive search results. A Citation Sensor was added to PubMed which recognizes combinations of search terms, characteristic of citation searching, such as author name, journal title, and publication date, then matches them to citations. Collaborator names and PubMed Central identifiers (PMCID) were added to PubMed citations. AbstractPlus was enhanced to include links to MedMaster Patient Drug Information, a new book available on the Bookshelf, and to Related Reviews which are review articles. Lastly, diacritical marks were added to PubMed Summary, AbstractPlus, Abstract, and Citation displays.

My NCBI is an Entrez feature that allows users to store searches and results. It also provides the option of automatically updating searches and sending results via email. My NCBI Version 2.0 was released in September. New features include improved account, navigation, preferences, and filter options and capabilities. New tools include My Saved Searches, My Collections, and My Bibliography, which allows authors to search and collect citations for their publications.

LinkOut

LinkOut is an Entrez feature that provides users with links from NCBI databases to a wide variety of outside resources, including full-text publications, biological databases, consumer health information, and research tools. The LinkOut for Libraries program links patrons from a PubMed citation directly to the full text of an article available through their library subscription program.

During 2008, the number of organizations participating in LinkOut increased to over 2,530, representing a 10% growth over the past year. LinkOut providers include 1,970 libraries, over 300 full-text providers, and 255 providers of nonbibliographic resources, such as biological and chemical databases. Users can now link to 66 million Entrez records, including full text articles of 48% of PubMed records from over 6,300 journals. A new version of the LinkOut Library Submission Utility was released in April 2008. This tool allows libraries to select their online holdings from a list of participating journals and providers. The new version provides a better layout and new functionality to enhance the user experience.

Outside Tool is a related service that also links users to outside resources. Participation in this program increased to over 500 institutions. Usage of LinkOut resources reached over 27 million hits per month, about 1.2 million hits per week day.

PubMed Central

PubMed Central (PMC) archives, indexes, and provides free and unrestricted access to full-text articles from life science journals. This repository integrates with the PubMed biomedical literature database of indexed citations and abstracts. Users from approximately 250,000 unique IP addresses access PMC on an average weekday.

Analyses indicate that the number users per day is estimated to be 1.5 to 2 times the number of unique IP addresses. As of July 1, 2008, more than 1.5 million articles were available from the PMC journal archive. That amounts to an annual increase of 50%, following annual increases of 65% in each of the two preceding years. New material includes recently published articles as well as the scanned and downloaded back issues that date back as far as 1865. An increasing number of articles by NIH-funded researchers are now being deposited and made available to the public via PMC, in response to the NIH Public Access Policy. A number of scientific society publishers have signed NIH Portfolio Agreements, whereby they deposit all articles authored by NIH-funded researchers in their journals into the PMC database.

PubMed Central International

PMCi (PubMed Central international) is an NLM initiative to create an international network of digital archives that operates on principles similar to those of PMC. This program resulted in the production of a portable version of PMC and the NIH Manuscript Submission (NIHMS) system. Sharing resources across borders means that the archives in the network will be richer, the integrity of the material will be maintained, and the quality of each archive will be augmented and improved. The first international "sister network" to NCBI's PMC was successfully deployed in the United Kingdom. The UKPMC service, sponsored by the Wellcome Trust, British Library, and associated organizations, has been operational since January 2007.

Bookshelf

The NCBI Bookshelf gives users access to the full text of over 100 textbooks in the clinical and research areas of biomedicine. In addition to textbooks from commercial publishers, the Bookshelf includes tutorials and help documents authored by NCBI, NLM, and NIH staff. Twenty new books were added to the Bookshelf this year, as well as a new collection entitled, *Drug Class Reviews*. Existing collections have been greatly expanded, most notably, *Gene Reviews*, which now includes over 440 human genetic disorders and is updated on a weekly basis. All new books, including GeneReviews, are displayed in a new books viewer, which features enhanced navigation and a customized look and feel. Gene Reviews is based on Gene Tests, a database produced by the University of Washington and widely used by genetics counselors and physicians for its comprehensive testing information and disease descriptions. It currently contains over 430 disease-specific reviews with approximately 48 added each year. *Genes and Disease* is a collection of articles designed to educate students, as well as the lay public, on how genes are inherited, how they 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 OMIM database, produced at Johns Hopkins School of Medicine, contains over 18,880 records for more than 12,800 genes and 6,000 phenotypes. Connections are being completed between descriptions of allelic variants involving substitutions and dbSNP.

Research

Using theoretical, analytical, and applied mathematical methods, NCBI's research program focuses on computational approaches to a broad range of fundamental problems in evolution, molecular biology, genomics, biomedical science, and bioinformatics. The Computational Biology Branch (CBB) and the Information Engineering Branch (IEB) are the main research branches of NCBI, with the latter focusing on database and software applications.

The research conducted by CBB has strengthened NCBI applications and databases by providing innovative algorithms and approaches (e.g., BLAST, VAST, and the CDD) that form the foundation of numerous end-user applications. By developing experimental strategies in collaboration with NIH and extramural laboratories, researchers in this group continue to make fundamental biological and biomedical advances. CBB consists of 89 senior scientists, staff scientists, research fellows, postdoctoral fellows, and students.

CBB is carrying out basic research on over 20 projects that have been reported for the NIH Intramural Program annual reports of research. Projects include new computer methods to accommodate the rapid growth and analytical requirements of genome sequences, molecular structure, chemical, phenotypic, and gene expression databases and associated high-throughput technologies. In other projects, computational analyses are applied to particular human disease genes and the genomes, evolution, and functional biology of pathogenic bacteria, viruses, and other parasitic organisms. Several of these projects involve collaboration with experimental laboratories at the NIH and elsewhere. Another focus of research is the development of computer methods for analyzing and predicting macromolecular structure and function. Recent advances include: improvements to the sensitivity of alignment programs, analysis of mutational and compositional bias influencing evolutionary genetics and sequence algorithms, investigation of gene expression regulation and other networks of biological interactions, analyses of genome diversity in influenza virus and malaria parasites related to vaccine development and evolution of virulence, the evolutionary analysis of protein domains, the development of theoretical models of genome evolution, genetic linkage methods, and new

mathematical text retrieval methods applicable to full text biomedical literature.

The high caliber of work performed by the CBB is evidenced by the number of peer-reviewed publications generated—over 80 publications this year with more in press. CBB scientists gave numerous presentations and posters at scientific meetings. Presentations were also given to visiting delegations, oversight groups, and steering committees. CBB hosts many guest speakers and shares information about research projects at its weekly lecture series. The NCBI Postdoctoral Fellows program provides computational biology training for doctoral graduates in a variety of fields, including molecular, computational, and structural biology.

The Board of Scientific Counselors (BoSC), comprised of extramural scientists, meets twice a year to review the research and development activities of NCBI and the research programs of senior investigators in the CBB. The BoSC's thirty-first meeting was held in May 2008.

Bioinformatics Training and Support

Outreach and Education

The educational component of NCBI's mission is recognized both by its advisory board (BoSC) 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. The audience for NCBI databases is very broad. The resources are used not only by molecular biologists and health professionals, but by students, educators, librarians, and science writers, as well as the general public. Garnering feedback from the user community is vital in order to provide services that meet their actual research needs and anticipate their future requirements.

Over the past year, NCBI staff exhibited at nine 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 print materials. During FY2008, budget constraints forced NCBI to suspend most of its outside courses and educational programs. However, NCBI continues to maintain a helpdesk for direct user support by phone or email.

Education: NCBI Courses

Thirty-one Field Guide and related courses were presented to over 2,000 people from June 2007 to March 12, 2008. "A Field Guide to GenBank and NCBI Resources" was taught at NIH and throughout the United States. The course consisted of a three-hour lecture, a two-hour hands-on practicum, and an optional one-on-one session. An expanded two-day course entitled "Enhanced Field Guide" provided extended in-depth coverage of BLAST, structure and genomic resources, and also included an advanced hands-on session. This course was presented once in FY2008 (to 30 participants).

Other courses provided training on specific resources and datasets, such as gene expression resources, molecular structures, searching PubChem, and structural alignments. Several hundred participants attended these courses in 2008. NCBI offered 12 bioinformatics minicourses at NIH and outside institutions to provide practical introductions to various resources and basic training on their use. The two and a half hour 'mini-courses' were problem-based and resource-based and included hands-on sessions and reviews. This year, over 83 minicourses were offered to approximately 2,100 participants.

Education: Technical Workshop Series

The PowerTools NCBI Technical Workshop Series consisted of two courses lasting two-to-four days each. "NCBI Power Scripting" included lectures and workshops on effectively using the NCBI Entrez programming utilities with scripts to automate search and retrieval operations across the entire suite of Entrez databases. "NCBI 4-Pack" provided information on practical applications of bioinformatics resources. Each course was offered quarterly at the NCBI Training Center. In FY2008, 117 participants from across the US, as well as Canada and England, attended the courses.

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 offered an intramural Core Bioinformatics Facility (CoreBio) – a network of bioinformatics specialists serving individual institutes within the NIH. Individual CoreBio members were trained over a nine-week period in the use of NCBI bioinformatics tools provided to the research community. CoreBio trained representatives from fifteen research institutes at NIH, conducting eight 9-week training programs since the program began in 2001. Forty-five update sessions and two special topic sessions for the institute representatives were also held.

Outreach: User Guides for NCBI Resources

NCBI provides fifteen "Announce" e-mail lists that 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 has merged with the NCBI Announce e-mail list 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, and a featured resource article on a new or updated resource. Access to the newsletter and its archives has been moved to the NCBI Bookshelf and notices are provided via email and RSS feed.

EXTRAMURAL PROGRAMS

Milton Corn, MD
Associate Director

The NLM Extramural Programs Division (EP) receives Congressional authority for its programs from two different authorizing acts: the Medical Library Assistance Act (MLAA, unique to NLM), and Public Health Law 301 (covers all of NIH). The funds are expended mainly as grants-in-aid 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 & 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 every category:

- **Research grants for basic and applied research**
- Research Resource grants to support unique research resources for research in biomedicine.
- Resource grants for knowledge management, and application of informatics.
- Training, fellowship and career development grants for informaticians and informationists;
- Scholarly Works and Conference grants to enhance scientific and scholarly communication;
- SBIR (Small Business Innovation Research) / STTR (Small Business Technology Transfer Research) grants to support informatics-related business projects.

The EP Web site (below) 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 (http://www.nlm.nih.gov/ep/funded.html).

A significant feature of EP activities in FY2008 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. The overall success rate for NLM applicants (all types of grant programs) is improving. (See Table 11.)

NIH continued in 2008 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 six). 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 exceeded its target of 6 for new investigator awards by three, but only three of eight K99 applications assigned to NLM received fundable scores. For the other two priority categories, those without quotas, NLM extended the payline modestly as needed.

Success Rates of Grant Applicants

Table 1 shows success rates from 2004 to 2008 for NLM's core grant programs. The success rate for research grants increased to 23%. As they did in 2007, award decisions in 2008 continued to favor early career support. These applicants are most often trainees from NLM's informatics training programs who are now moving into independent research careers. Success rates are computed by dividing the number of awards by the number of applicants in a fiscal year.

Table 11

Success Rate, Core NLM Grant Programs, FY2004–2008.

Grant Type	FY 2004	FY 2005	FY 2006	FY 2007	FY 2008[1]
Research	15%	9%	10%	16%	23%
Knowledge Management	18%	12%	8%	16%	7%
Scholarly Works	26%	19%	15%	17%	19%
Career Transition	32%	36%	37%	29%	38%

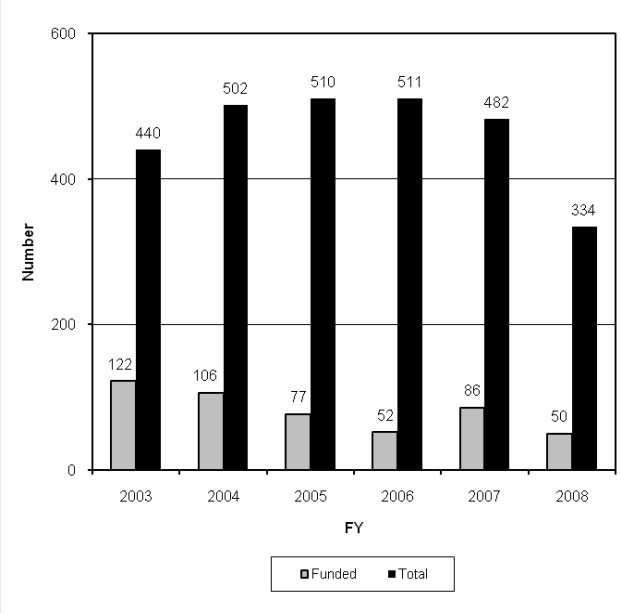
1 FY2008 success rates for FY include nine early concurrence grants from 2008/10 council. In prior FY, success rates were calculated for three regular council rounds only.

Two major factors continue to shape success rates at NIH: the large number of applications received and the fact of essentially flat budgets. Table 2 shows the steep increase in the number of applications received since FY2003, a trend that began to reverse itself in 2007 and 2008. 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, FY2003–2008.
In 2003, the budget doubling period ended at NIH.



MLAA and PHS 301 – Historical Separation of Extramural Programs Grant Budget

Traditionally, EP provides its budget tables broken down into two expenditure categories: MLAA (Training, Career Development and Knowledge resource grants) and PHS 301 (Research and Research Resource grants). NIH grant mechanism tables for budget projection and reporting, which NLM began using in FY2008, do not recognize this distinction, which, in any event, has lost crispness as new programs, including information science research, were developed for the MLAA budget. Comparisons between NLM grant expenditures and those of other Institutes are greatly facilitated if NLM uses standard NIH mechanism terminology. As of FY2008, EP has dropped the distinction between MLAA and PHS 301 for budgeting and reporting purposes.

Research Support for Biomedical Informatics and Bioinformatics

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.
In FY2008, EP issued an Express Research Grant R01 program (PAR-08-080), with a limit of 15 pages for the research plan and a modular budget format. This Express Research Grant program has the same scope of interest and priorities for informatics research as with EP's participation in the Parent R01 omnibus program announcement (PA-07-070) as of November 2006. As of FY2007, all R01 grant applications are now received electronically.

- 67 reviewed R01 applications (140 in FY2007)
- 17 awarded R01 applications (24 in FY2007)

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. As of FY2009, EP has withdrawn participation from the R03 small research grant program.

- 33 reviewed R03 applications (36 in FY2007)
- 7 awarded R03 applications (4 in FY2007)

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 is sometimes a better fit for informatics/engineering proposals than the standard R01 research grant, which is judged in terms of hypothesis-based science. 36 reviewed R21 applications (42 in FY2007) 7 awarded R21 applications (6 in FY2007).

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. In FY2008, the first of two limited competition RFAs were announced for possible continuation funding in FY2009.

- 0 reviewed P41 applications (2 in FY2007)
- 0 awarded P41 applications (2 in FY2007)

Conference Grants

Support for conference 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. Only electronic applications are now accepted.

- 3 reviewed conference grants applications (5 in FY2007)
- 1 awarded conference grant application (2 in FY2007)

IAIMS Testing & Evaluation Grants

The IAIMS Testing and Evaluation grant program was closed in FY2008.

- 3 reviewed IAIMS Testing & Evaluation applications (7 in FY2007)
- 0 awarded IAIMS Testing & Evaluation applications (0 in FY2007)

Small Business (SBIR/STTR)

All NIH Institutes allocate a fixed set-aside of available research funds every year to Small Business Innovation Research (SBIR) (2.5% of research grants budget) grants and Small Business Technology Transfer Research (STTR) (.5% of research grants budget) 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 FY2008, 25 applications were "unscored" (out of 32 applications), indicating reviewer assessment that they were not competitive for funding.

- 31 reviewed small business applications (45 in FY2007)
- 1 awarded small business application (4 in FY2007)

Resource Grants

Resource Grants support access to information, connect computer and communications systems, and promote collaboration in networking, integrating, and managing health-related information. Two of the three 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 third 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. In FY2008, the grant program was suspended, with the intent to issue an RFA for knowledge management grants in FY2009.

- 46 reviewed KM & AI applications (64 in FY2007)
- 3 awarded KM & AI applications (10 in FY2007)

Integrated Advanced Information Management Systems (IAIMS) Planning Grant

Applications to the IAIMS Planning Grant dropped sharply in FY2007 and FY2008, as expected, due to the deactivation of the IAIMS operations grant program. In FY2008, this grant program was closed, although planning-related knowledge management applications may be submitted under the forthcoming RFA for knowledge management in FY2009.

- 5 reviewed IAIMS Planning applications (10 in FY2007)
- 1 awarded IAIMS Planning application (2 in FY2007)

Grants for Scholarly Works

NLM alone among the Institutes is authorized to support book publications, and the Scholarly Works program continues to play a key role in important areas of biomedical scholarship, particularly in the history of science and medicine. In FY2008, the grant program was suspended, with the intent to issue an RFA for scholarly works grants in FY2009.

- 52 reviewed Scholarly Works applications (58 in FY2007)
- 10 awarded Scholarly Works applications (10 in FY2007)

Training and Fellowships

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 informatics research trainees in 18 programs across the country (see Table 3).

This program is re-competed every five years. The latest applications were received in March 2006, and five-year 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. One former NLM training program, at the Medical University of South Carolina, no longer receives ongoing funds from NLM but NLM continues to support its matriculated trainees until completion of their training in 2010

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 co-funding from the National Institute of Dental and Craniofacial Research (NIDCR) supported two training slots in dental informatics at the University of Pittsburgh. In 2008, the National Institute of Biomedical Imaging and Bioengineering provided funding for special one-year projects for four predoctoral 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 predoctoral students, one each at Pittsburgh and Vanderbilt.

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 Fall AMIA meeting in November 2005, with subsequent meetings held twice a year in conjunction with the NLM training conference and the Fall AMIA meeting. At the July 2008 NLM Training Conference, the NLM/RWJF cohort event was held as a pre-meeting, on July 7. In FY2008, RWJF funds supported 8 predoctoral and 4 postdoctoral trainees at the 4 sites.

In 2007, NLM restructured its Short-Term Trainee Program (STTP) to allow on-demand awards for minority or disadvantaged trainees, with the long-term goal of recruiting more minorities into informatics research careers. Twelve of NLM's programs participate in the STTP program. Eighteen STTP trainees were supported at seven of those programs in 2008 (eight full-time equivalent).

Due to budget reductions, in February 2008, 16 of NLM's 18 active training programs were required to leave 21 predoctoral and nine postdoctoral slots unfilled. A summer 2008 Congressional Supplement to NLM of $1.705 million was dedicated to restoring 26 of these slots; the remaining four slots were restored with end-of-year budget funds.

Table 13

T15 Trainees funded by NLM for FY2008: 273 total trainees (full-time equivalent).

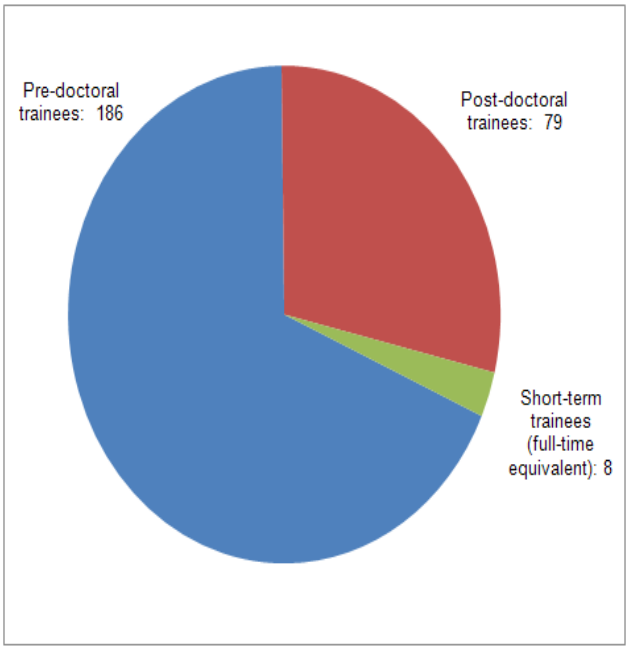


Table 14

In FY2008 NLM Sponsored Trainees at 20 Institutions

1. University of California Irvine (Irvine, CA)
2. University of California Los Angeles (Los Angeles, CA)
3. Stanford University (Stanford, CA)
4. University of Colorado Denver/HSC Aurora (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, MN) [NLM funding ended in 2007]
10. University of Missouri-Columbia (Columbia, MO)
11. Columbia University Health Sciences (New York, NY)
12. Oregon Health & Science University (Portland, OR)
13. University of Pittsburgh at Pittsburgh (Pittsburgh, PA)
14. Medical University of South Carolina (Charleston, SC) [NLM funding ended in 2007]
15. Vanderbilt University (Nashville, TN)
16. Rice University (Houston, TX)
17. University of Utah (Salt Lake City, UT)
18. University of Virginia Charlottesville (Charlottesville, VA)
19. University of Washington (Seattle, WA)
20. University of Wisconsin Madison (Madison, WI)

Every summer, all NLM-supported trainees attend a national informatics training conference. On July 8–9, 2008, the two-day meeting took place at the National Library of Medicine (Natcher Building on NIH campus). Research projects were presented in plenary and semi-plenary sessions by 39 informatics trainees. An additional 45 trainees presented posters at the meeting. There were 354 attendees, including directors, faculty, and staff from all NLM-funded informatics training programs; holders of NLM individual fellowships; faculty and trainees from the Veteran's Administration informatics training sites; and NLM staff and guests.

Individual Fellowships

Training for Informationists

In FY2008, the F37 Informationist fellowship program expired. This program had been EP's only remaining individual fellowship program.

- 1 reviewed F37 informationist application (4 in FY2007)
- 0 awarded F37 informationist applications (1 in FY2007)

Career Support

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 un-mentored 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, they are preferred applicants.

- 8 reviewed K99/R00 applications (4 in FY2007)
- 3 awarded K99/R00 applications (3 in FY2007)

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 FY2008, EP funded 4 of 12 applications.

- 15 reviewed Loan Repayment Program applications (12 in FY2007)
- 4 awarded Loan Repayment Program applications (5 in FY2007)

Trans-NIH Projects

EP and Roadmap Activities

In 2008, NIH Roadmap activities continue to be 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 (5 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 five 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 ended 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

NLM participates in two types of multi-institute grant programs: general and topical. General programs such as the AREA grants, diversity and reentry supplements are fundamental components of NLM's overall grant program. However, budget constraints and the importance of protecting NLM's own grant programs have increased EP's selectivity when considering participation in topical multi-Institute initiatives. EP participation is confined to topical programs which do not duplicate its existing grant programs, and as active funding announcements expire,

EP reconsiders participation in all topical multi-institute grant programs. The active multi-institute programs NLM participates in are listed in Table 8. The applications for these programs are reviewed by 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 & Training

EP contributed approximately $1.25 million in collaborative co-funding agreements to seven grants in FY2008. The listing below outlines the projects and amounts for research co-funded with other NIH institutes and centers.

- Co-funding with the National Institute of Child Health and Human Development (NICHD) for a small business grant entitled "Novel Informatics for Highly Reliable Multi-Locus Allele Calling for Embryo Screening" ($326,777). The purpose of this project is to develop screening technology that can reliably screen for genetic information at a cellular level and one that is more cost efficient and affordable. The proprietary technique developed will be applied to the parents in order to retain DNA matched to reconstruct the embryonic DNA at a statistically high confidence level.
- Co-funding to the National Institute of Nursing Research (NINR) for two small business grants for the development of "Permanent Antimicrobial Medical Plastic" and for "Demand Driven Healthcare Scheduling using Flexible Shifts and Monte Carlo Screening": ($198,041). The purpose of the first grant is to develop an antimicrobial plastic to reduce infections at hospitals and health care facilities due to cross-contamination. The second grant is to develop an efficient health care scheduling system to provide staff when needed and one that can shift to accommodate the demands of the workplace especially during times of peak staffing demands.
- Co-funding for the "Comparative Toxicogenomics Database (CTD)" ($196,000) which is being developed for public availability in order to promote understanding about the effects of environmental chemicals on human health as a components of the NIGMS Pharmacogenetics Initiative.
- 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 the National Institute of General Medical Sciences (NIGMS) for continued support of a cooperative agreement for the Stanford Pharmacogenetic Research Network and Knowledge Base ($301,818).

NLM also receives co-funding from other institutes and centers, $4.7 million, for its research, resource and training grants.

In FY2008, the co-funded NLM grants included:

- NLM requested funding support for two NIH Director's Bridge Awards for grant proposals, "Bioinformatics Linkage of Protein Disorder and Function" and "Causal Discovery Algorithms for Translational Research with High-Throughput Data" ($652,854). The first award is focused on developing a better understanding of intrinsically disordered protein (IDPs). IDPs and their genomic regions, structures and functions have a direct correlation to proteins involved in human disease. The second award's goal is to develop "next-generation" causal algorithms that will help in the discovery of molecular pathways and development of biomarkers related to disease. This will help in the early detection and diagnosis of diseases and impact prognosis and personalized care as well as future drug design.
- 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 funding for two small research grants that are linked awards as part of the Partners In Research Program, $63,625. The linked projects entitled, "Medical Marvels Interactive Translational Research Experience (MITRE) purpose is to increase public knowledge and understanding of translational biomedical and behavioral research through this joint partnership.

The OD/NIH provided ongoing and additional program support for the National Center for Biocomputing (NCBC) entitled, "Informatics for Integrating Biology and the Bedside (RMI), $3,287,929. The supplemental additions to the project were in support of the i2b2 conference and one postdoctoral researcher.

- 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.

- The Office of Research on Women's Health (ORWH), NIH provided support for a NLM "Scholarly Works" grant entitled: "The History of Emergency Contraception", $75,530. This grant will support research leading to a publication on the history of emergency contraception from the 1960s to the present.
- The National Institute of Biomedical Imaging and Bioengineering (NIBIB) supported a NLM conference grant for a workshop entitled, "Life Science Systems and Applications Workshop", $3,000. The primary aim of the workshop is to provide a forum for presenting new systems advancements in emerging life science applications.
- The NLM received support for one Cooperative Roadmap research proposal from the OD, NIH that is linked to a comprehensive National Center for Biocomputing entitled, "Hypothesis Web Development for Neuropsychiatric Phenomics", $ 303,721. This central goal of the Hypothesis Web project with the Consortium for Neuropsychiatric Phenomics is to aid in the development of interdisciplinary hypotheses spanning multiple disciplines of neuroscience.

Interagency Agreements and Special Initiatives

NLM continues to provide co-funding to the 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.

NLM receives ongoing gift funding support from the Robert Wood Johnson Foundation (RWJF) through the Foundation for the NIH (FNIH) to its Bioinformatics Training Programs (Columbia University, University of Utah, University of Washington, Johns Hopkins University) for Public Health Informatics, $784,877. The purpose of this project is to develop future leaders in the field of public health to enter into both the Federal Government and the Health Care infrastructure of the nation.

The NLM received funding from the Agency for Health Care Research and Quality provided cofunding to a NLM grant entitled, "Improving Guideline Development and Implementation", $133,000. This projects specific aim it to improve medical guidelines for clinicians and to operationalize their use in a clinical setting.

The Defense Advanced Research Projects Agency (DARPA), DoD provided support for the further development of the NLM supported small business grant for the, "Medical Emergency Disaster Response Network", $389,685. DARPA funding was to further develop the existing software and to create a package that Federal Agencies within the government could use. This work further enhances the grantees ability to bring a viable product to market.

Extramural Programs Web site

EP updated its Web site in FY2008 to include grantee publication lists. For each funded research project grant and training grant, the awards pages now link to a PubMed search for publications related to that grant. Searching is based on grant number as acknowledged in original publications. Throughout the year, new grant awards for FY2008 were added to the EP Web site, including both competing awards and noncompeting renewals. Links to grantee project Web sites were added to many active grants.

During FY2008, the EP Web site received approximately 222,000 page views, 72,000 visits, and 51,000 visitors, across roughly 100 pages. In FY2009, EP intends to implement a full-scale update of its Web site content, structure, and design. This Web site update will include added resources for existing and prospective trainees.

EP Operating Units

Program Office

Grant Program Development

Program activities in FY2008 were focused on (1) refining language for grant program announcements in which EP participates; (2) increased utilization of the new SF-424 electronic application form; (3) participation in Research Condition and Disease Categorization (RCDC) fingerprinting activities and (4) Complementing EP's planned Challenge Grant RFAs by participation in GM's EUREKA program calling for high-risk, high-reward innovative applications. The NLM component requested application in the area of computational discovery and hypothesis testing.(5) A continuing discussion with national experts on Artificial Intelligence (AI) to help formulate goals for a planned RFA for research in AI. The NIH timetable for transition to electronic applications has stalled for several of EPs grant programs, including P41, K99, F37 and T15). Until the transition is completed, these applications 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 a number of Roadmap-related activities. Program staff attended the following scientific meetings: AMIA Spring Congress; AMIA Fall Annual Meeting; IEEE/NLM Joint Workshop on Life Science Systems and Applications. EP extended for an additional year its contract 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.

EP's 2007 summer intern, Mable Cao, returned for the summer of 2008.

EP Program Class Codes were updated once. The Program unit began meeting twice per 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.

Implementation of EP's eGrants system for managing electronic grant files is fully completed. The program analyst is responsible for implementation and quality control of this new resource. In addition program staff continues to work with DEAS and Grants Management to create records in IMPAC Training module for NLM trainees.

Dissemination and Staff Activities

Presentations by NLM staff were made to AAHSL/MLA Mentor program; RML New staff orientation; the NIH conf on women in biomedical research and AMIA inter alia. In July 2008, the annual NLM Training Conference was held at NIH. Presentations were made to the NLM Board of Regents on the following topics: EP 2008 budget; pilot project on analyzing productivity of NLM research grants; possible topic areas for EP challenge grants; evolution of NLM research grant programs 1996-2006; delegation of authority of BOR to BOR EP Subcommittee to approve applications for early concurrence; conflict of interest issues for principal investigators; changing research themes over time; analysis of Informatics Research Training Program; approval of operating rules. 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.

The following EP grantees or BLIRC members made presentations to the NLM Board of Regents and/or NLM staff: Krzysztof Fidelis, PhD, on Critical Assessment of Structure Prediction; Dr. Daniel Reininger on his SBIR project for Medical Emergency Disaster Response Network; Lisa Cannon-Albright, PhD on Analysis of the Familial Component to Disease in a Biomedical Resource w/Links; Dr. Elizabeth Liddy, BLIRC Member, "Improving Public Health Grey Literature Access for the Public Health Workforce", presentation in the NLM Informatics Lecture Series.

Grants Management Office

EP issued 232 grant awards in FY2008 for nearly $53 million in NLM support. The EP was provided with an additional $9.6 million in cofunding to NLM grants. 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 resource and research grants as well as loan repayments.

In addition, the NLM Grants Management Office provides daily and end-of-fiscal year grant accounting support for the EP budget for all awarding mechanisms, updates on 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. The NLM grant office is now the official point of contact for the management and use of the NIH Conflict of Interest Database when a NLM COI is reported.

NLM Grants Management Service Center

In FY2008 the NLM Grants Management Office became an official service center in support of the Office of the National Coordinator (ONC) for Health Information Technology, Office of the Secretary (OS), DHHS. NLM grants staff was charged with developing an electronic award notice in IMPAC II for issuance of DHHS awards, receipt and review of all grant proposals, terms and conditions for issued awards and the funding and accounting of all DHHS approved proposals. The NLM funded a cooperative agreement to the Logistics Management Inc. (LMI) for the creation of a successor organization to the American Health Information Community (AHIC) - Federal committee, $2,000,000. The newly developed institution, AHIC Successor, Inc., was awarded a cooperative agreement for $3,000,000. The purpose of AHIC Successor, Inc. is to help in the development of interoperable health care for the US This project will also incorporate health care feasibility data - cooperative agreements connected to the AHIC Successor project and six awards were issued in the amount of $598,671 in support of the National Health Information Network (NHIN). The NHIN, as well as the Certification Commission for Healthcare Information Technology

(CCHIT) and the Healthcare Information Technology Standards Panel (HITSP), are integral parts of the overall collaborative effort tied to the cooperative agreement for the AHIC Successor, Inc.

Grants Management Staff Activities

Committee participation for GM staff during FY2008 included the Grants Management Advisory Committee (GMAC), the Board of Survey for the NLM, and an IMPAC II user feedback group. At the Chief Grants Management Officer retreat in Chestertown, MD, the CGMO discussed the current and future direction of grants management at the NIH and DHHS. This year the CGMO also re-certified as Level IV "Grants Management Executive" to fulfill DHHS requirements. GM staff also continued to liaison with the Foundation for the NIH regarding gift funding for public health informatics trainees from the Robert Wood Johnson Foundation.

Scientific Review Office

Grant Review Activities: Overall, 327 applications were reviewed for which NLM was the primary assignment. Seventy eight percent of those (258) were reviewed by NLM. The remaining 69 applications were reviewed by CSR. Most of those reviewed by CSR were SBIR/STTR grant applications. Of the applications reviewed by NLM, 42% were in one of the three research grant mechanisms (R01, R21, R03). Knowledge management grants represent 19% of the applications reviewed, with Scholarly Works at 20% 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 FY2008 and reviewed 158 applications (as compared to 177 in 2007). 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 FY2008 compared to 8 in 2007. 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 100 applications during FY2008, compared to 185 in 2007. This number is smaller than 2007 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, reviews and votes electronically on a list 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 2008, materials presented to the EP subcommittee were revamped, to clarify the reasons for bringing them to the subcommittee for review. In 2008, a new early concurrence process was initiated, in which the subcommittee used the NIH Electronic Council Book for en bloc voting in advance of the regular meeting. 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); CSR Receipt & Referral Coordinators.

Administration and Operations Office: For the first time in nine years the EP Direct Operations program did not see an increase in its annual NLM budget. EP had a very difficult time working with this budget, as the NIH received a higher percent increase in salary than what was originally budgeted for in FY2008. As a result, a significant portion of our EP operations budget had to be allocated to cover payroll expenses. This further resulted in a reduction of funds for other vital areas such as equipment, operating expenses, supplies, Inter-agency agreements, SREA program support, and employee training and travel. The forecast for 2009 appears to be even bleaker as we were asked to submit a 2009 budget with a projected 2% decrease from our FY2008 spending level.

In 2008, NLM/EP participated in the NIH Vital Records Recognition Project. All federal agencies are mandated by the National Archives and Records Administration (NARA) to identify and maintain Vital Records in a systematic manner. This requirement is also found in numerous Federal laws and regulations as well as the National Institutes of Health (NIH) COOP Plan and NIH Manual Chapter 1744 - Vital Records Program. Vital Records are the sum total of all essential agency records documenting legal and financial rights of the Government

and those affected by Government activities and those needed for emergency operations under national security emergencies, other emergencies, or disaster conditions. The major focus of this project was to identify and show proper management and safeguarding of the NLM/EP's Vital Records and to develop and implement plans to continue operations in the event of an emergency.

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 NLM system into NIH system and scanning grant applications and supplement material into eGrants. DEAS staff also has begun tracking the NBS Acquisitions CAN errors. DEAS staff is preparing and setting up Video Conferences. Current DEAS Staff at EP include: Pam Beheler, Supervisor; Joyce Campbell, Renee Jellerette-Stainback and Tsegaye Mariam, Extramural Support Assistants. 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 FY2008, NLM/EP optioned the third year on the purchase order 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. If additional funds become available, the additional funding will go towards the complete abstraction and analysis of 2007 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 new mandatory IT Security and Ethics training, Prevention of Sexual Harassment (PoSH), Privacy Awareness, No-Fear Act, and Alternate Dispute Resolution (ADR) training. These new training requirements were aided with the implementation of the new NIH Learning Management System (LMS). LMS is NIH's new training database and will greatly reduce the need for employees to go to multiple Web sites to complete their mandated 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 office telework policy was drafted and telework agreements put in place for calendar year 2008, with an 80% participation rate.

Facilities & Equipment: FY2008 saw the upgrade and replacement of all NLM/EP Laptop. The increase in the number of Teleworkers on staff necessitated the upgrade of our laptop units as well as units provided for teleworking from home. All computer equipment, i.e. laptops, desktops, and PDA units were upgraded with current NIH PK encryption software. In addition, the EP admin staff installed two new digital copier machines 1 black & white and the other color, in the EP copier room. New modular furniture was also installed in January 2008 to re-invigorate the once drab space. The re-design of the old EP Grants processing room was also completed and already has one occupant. EP now has the much need space to house the new additions to the staff that are expectedin the Fall of 2008.

Table 15

Extramural Programs Grants Budget by NIH Mechanism Groupings, FY2008
Dollars in Thousands

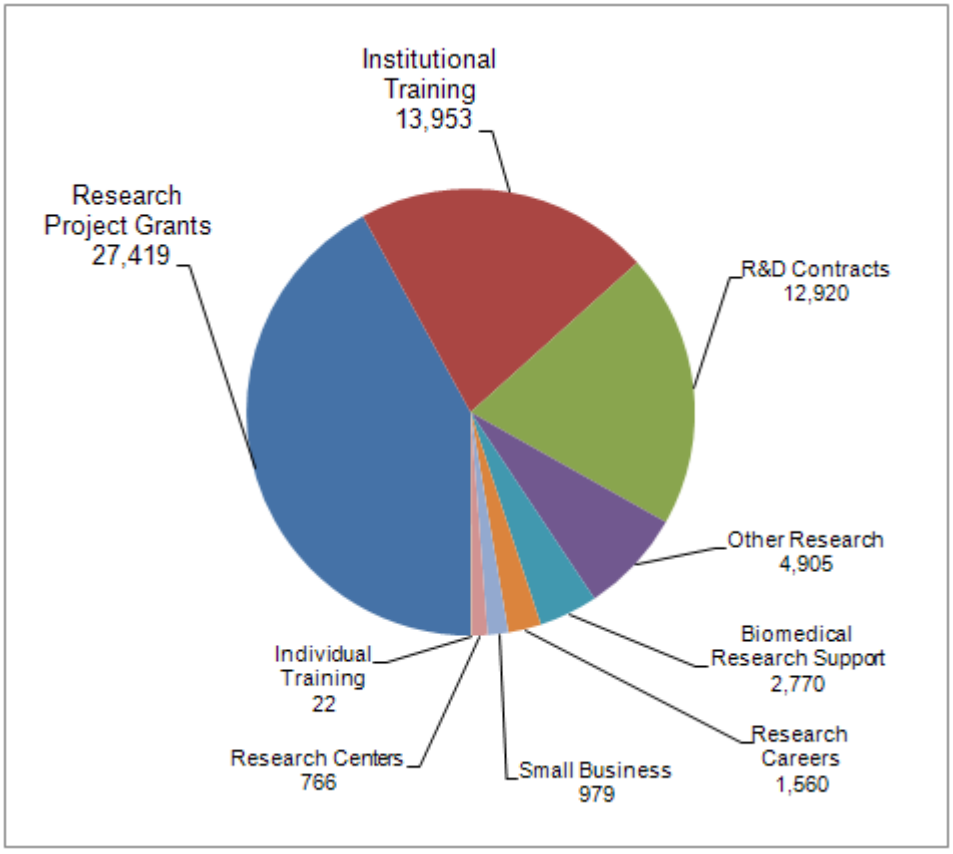


Table 16

Extramural Programs Grants
Budget by NIH Mechanism Groupings, FY2008

(dollars in thousands)

FY 2008 operating budget request by NIH mechanism groupings	FY 2008 Actual[1]	
	No.	Amount
Research Project Grants (R01, R03, R21, R00, R15, RL1, R56, U01)	97	27,419
SBIR/STTR (R41, R42, R43, R44)	6	979
Research Centers - Specialized/Comprehensive (U54)	1	766
Other Research - Research Careers (K22, K99)	12	1,560
Other Research - Biomedical Research Support (P41)	5	2,770
Other Research - Other (G08, G13, R13, R24, D43)	51	4,905
Training - Individual (F37)	1	22
Training - Institutional (T15)	19	13,953
R&D Contracts (L30, L40, N01, Y01)	14	12,920
Total EP grants, excluding Taps and RMS	**206**	**65,294**

[1]FY 2008 budget includes $1,705K supplement for T15 slot restoration and $200K one-time increase for NN/LM.

Table 17

Extramural Programs Grants Budget by Activity Code, FY2008

FY 2008 operating budget request by activity code	FY 2008 Actual[1]	
	No.	Amount
D43: International Training Grants in Epidemiology (cofund)	-	250
F37: Individual Informationist Fellowships	1	22
G08: Knowledge Management & Applied Informatics; Planning Grant for IAIMS	23	2,963
G13: Scholarly Works in Biomedicine and Health	23	1,596
K22: Early Career Development	7	1,025
K99: Pathway to Independence	5	535
L30: Extramural Loan Repayment Program	4	135
L40: Extramural Loan Repayment Program	-	-
N01: NN/LM Contracts	9	12,585
P41: Biomedical Resource Grant	5	2,770
R00: Pathway to Independence		-
R01: Research Project Grants	70	23,821
R03: Small Project Grants	12	794
R13: Conference Grants	5	96
R15: Academic Research Enhancement Award (AREA)		-
R21: Exploratory/Developmental Grants	13	2,362
R24: IAIMS Test & Evaluation		-
R41: Small Business Technology Transfer (STTR)		-
R42: Small Business Technology Transfer (STTR)	2	304
R43: Small Business Innovation Research (SBIR)	3	348
R44: Small Business Innovation Research (SBIR)	1	327
R56: Director's Bridge Award	1	141
RL1: Linked Research Project Grant		-
T15: University Biomedical Informatics Research Training Programs	19	13,953
U01: Cooperative Agreement	1	302
U54: NCBC Roadmap Center	1	766
Y01: Inter-Agency Agreement	1	200
Total EP grants, excluding Taps and RMS	**206**	**65,294**

Table 18

RFA/PA actions in FY2006–2008

Guide #	Program Title	Status
NLM's Core Active Grant Programs		
PAR-08-080	NLM Express Research Grants in Biomedical Informatics and Bioinformatics (R01)	Expires May 9, 2011
PA-07-070	Research Project Grants for Biomedical Informatics and Bioinformatics (Parent R01)	Expires January 8, 2010 (NIH Parent R01)
PA-06-181	NIH Exploratory/Developmental Research Grant Program (Parent R21)	Expires May 2, 2009 (NIH Parent R21)
PA-07-297	NIH Pathway to Independence (PI) Award (K99/R00)	Expires January 3, 2010 (NIH parent for K99)
PA-06-041	NIH Support for Conferences and Scientific Meetings (Parent R13/U13)	Expires May 8, 2011 (NIH parent for R13)
TBA	Knowledge Management /Applied Informatics Grants (G08)	To be issued as PAR or RFA in 2009
TBA	NLM Grants for Scholarly Works in Biomedicine and Health (G13)	To be issued as PAR or RFA in 2009
Multi-IC Active Announcements in which NLM Participates		
RFA-GM-09-008	Exceptional, Unconventional Research Enabling Knowledge Acceleration (EUREKA) (R01)	Single Deadline October 28, 2008
PAR-07-344	Innovations in Biomedical Computational Science and Technology (BISTI) (R01)	Expires May 8, 2009
PAR-06-411	Exploratory Innovations in Biomedical Computational Science and Technology (BISTI) (R21)	Expires May 8, 2009
PA-06-042	Academic Research Enhancement Award (AREA) (Parent R15)	Expires Jan 8, 2009
PA-04-126	Supplements to Promote Reentry into Biomedical and Behavioral Research Careers	Expires September 30, 2011
PA-05-015	Research Supplements to Promote Diversity in Health-Related Research	Expires September 30, 2011
PAR-07-020	Understanding and Promoting Health Literacy (R01)	Expires January 26, 2010
PAR-07-019	Understanding and Promoting Health Literacy (R03)	Expires January 26, 2010
PAR-07-379	Behavioral and Social Science Research on Understanding and Reducing Health Disparities Grants (R01)	Expires September 19, 2009
PAR-07-380	Behavioral and Social Science Research on Understanding and Reducing Health Disparities Grants (R21)	Expires September 19, 2009
PAS-07-382	Advancing Novel Science in Women's Health Research (ANSWHR) (R03)	Expires January 8, 2010
PAS-07-381	Advancing Novel Science in Women's Health Research (ANSWHR) (R21)	Expires January 8, 2010
PA-07-442	Extramural Loan Repayment Programs (L30)	Ongoing; One deadline per year
PA-07-440	Extramural Loan Repayment Programs (L40)	Ongoing: One deadline per year
PAR 08-023	Predictive Multiscale Models of the Physiome in Health and Disease (R01)	Expires Sepbember 15, 2010
Small Business (SBIR/STTR) Active Announcements in which NLM Participates		
PA 07-281	PHS 2007-02 Omnibus Solicitation of the NIH for Small Business Technology Transfer Grant Applications (Parent STTR [R41/R42])	Expires January 8, 2009 (NIH parent for R41/R42)
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])	Expires January 8, 2009 (NIH parent for R43/R44)
PA-06-396	New Technologies for Liver Disease STTR (R41/R42)	Expires January 8, 2009
PA-06-397	New Technologies for Liver Disease SBIR (R43/R44)	Expires January 8, 2009
PAR-07-161	Innovations in Biomedical Computational Science and Technology Initiative (BISTI) (STTR [R41/R42])	Expires May 8, 2009
PAR-07-160	Innovations in Biomedical Computational Science and Technology Initiative (BISTI) (SBIR [R43/R44])	Expires May 8, 2009
PAR-08-201	Technological Innovations for Interdisciplinary Research Incorporating the Behavioral and Social Sciences (STTR [R41/R42])	Expires May 2, 2009
PAR-08-202	Technological Innovations for Interdisciplinary Research Incorporating the Behavioral and Social Sciences (SBIR [R43/R44])	Expires May 2, 2009
GUIDE Announcements Issued by NLM in FY 2008		
PAR 08-080	NLM Express Research Grants in Biomedical Informatics and Bioinformatics (R01)	Issued 1/25/2008
RFA-LM-08-001	Limited Competition for Continuation of NLM Biomedical Informatics/Bioinformatics Resource Grant (P41)	Issued 6/11/2008
NOT-LM-08-001	Correction to Stipend and Allowances, PAR-06-509, National Library of Medicine NLM Individual Fellowship for Informationist Training (F37)	Issued 11/30/2007
NOT-LM-08-002	Notice of Suspension, NLM Knolwedge Management and Applied Informatics Grants (G09)	Issued 9/19/2008
NOT-LM-08-003	Notice of Suspension, NLM Grants for Scholarly Works in Biomedicine and Health (G13)	Issued 9/22/2008
NOT-LM-08-004	Notice of Suspension, Planning Grants for Integrated Advanced Information Management Systems (IAIMS)	Issued 9/23/2008
NOT-LM-08-005	NLM Suspends Participatoin in PA-06-180, NIH Small Research Grant Program (Parent R03)	Issued 9/25/2008
NOT-LM-08-006	Clarification of Eligibility, NLM Express Research Grants in Biomedical Informatics and Bioinformatics (R01)	Issued 9/25/2008
NOT-LM-0-8-007	Notice of Expiration, NLM Individual Fellowship for Informationist Training (F37)	Issued 9/25/2008
NLM Grant Programs Suspended or Expired FY 2006 - 2008		
PAR-07-236	NLM Knowledge Management & Applied Informatics Grants (G08)	Suspended September 2008. RFA pending 2009.
PAR-07-237	NLM Grants for Scholarly Works in Biomedicine and Health (G13)	Suspended September 2008. RFA pending 2009.
PAR-07-238	Planning Grant for Integrated Advanced Information Management Systems (IAIMS)	Suspended September 2008
PAR-05-077	Test & Evaluation Grant for Integrated Advanced Information Management Systems (IAIMS)	Expired May 2008
PAR-06-509	NLM Individual Fellowship for Informationist Training (F37)	Expiration notice September 2008
PA-06-180	NIH Small Research Grant Program (Parent R03)	Discontinued participatoin September 2008
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
PA-06-012	Manufacturing Processes of Medical, Dental, and Biological Technologies (STTR [R41/R42])	Expired September 8, 2008
PA-06-013	Manufacturing Processes of Medical, Dental, and Biological Technologies (SBIR [R43/R44])	Expired September 8, 2008
PA-06-008	Bioengineering Nanotechnology Initiative (STTR [R41/R42])	Expired September 8, 2008
PA-06-009	Bioengineering Nanotechnology Initiative – SBIR (R43/R44)	Expired September 8, 2008
PA-06-010	Integration of Heterogeneous Data Sources (STTR [R41/R42])	Expired January 8, 2008
PA-06-011	Integration of Heterogeneous Data Sources (SBIR [R43/R44])	Expired January 8, 2008

Table 19

NLM grants awarded in FY2008

(Sorted by PI name, within each grant category)

RESEARCH GRANTS

Aliferis, Constantin F.
2-R56-LM007948-04A1
Vanderbilt University
Causal Discovery Algorithms for Translational Research with High-Throughput Data

Archer, Kellie J.
1-R03-LM009347-01A2
Virginia Commonwealth University
Recursive Partitioning and Ensemble Methods for Classifying an Ordinal Response

Aronsky, Dominik
1-R21-LM009747-01A1
Vanderbilt University
An Informatics-Based Guideline Implementation Framework for Asthma Care

Bahar, Ivet
2-R01-LM007994-05
University of Pittsburgh at Pittsburgh
Bridging Sequence Patterns and Structural Dynamics

Bennett, Kristin P. and Yener, Bulent (New Investigator)
1-R01-LM009731-01
Rensselaer Polytechnic Institute
Discovering Hidden Groups Across Tuberculosis Patient and Pathogen Genotype Data

Butte, Atul J.
1-R01-LM009719-01A1
Stanford University
Integrating Microarray and Proteomic Data by Ontology-based Annotation

Cendan, Juan C. and Lok, Benjamin (New Investigator)
1-R03-LM009646-01A1
University of Florida
Immersive Virtual Patients with Abnormal Clinical Conditions for Medical Student

Detmer, Don Eugene
1-R13-LM009769-01
American Medical Informatics Assn
Development of Biomedical and Health Informatics Competencies

Dunker, A. Keith
2-R56-LM007688-05A1
Indiana Univ-Purdue Univ at Indianapolis
Bioinformatics Linkage of Protein Disorder and Function

Embi, Peter J. (New Investigator)
1-R01-LM009533-01A1
University of Cincinnati
Evaluating EHR-based, Point-of-Care Trial Recruitment Across Clinical Settings

Fetters, Michael Derwin (New Investigator)
1-R03-LM010052-01
University of Michigan at Ann Arbor
Medical Marvels Interactive Translational Research Experience (MITRE)

Ganz, Aura (New Investigator)
1-R21-LM008942-01A2
University of Massachusetts Amherst
DIORAMA: Dynamic Information Collection and Resource Tracking Architecture

Guo, Nancy Lan (New Investigator)
1-R01-LM009500-01A2
West Virginia University
A Novel Computational Framework for Individualized Clinical Decision-Making

Hazlehurst, Brian L. (New Investigator)
1-R21-LM009728-01A1
Kaiser Foundation Research Institute
Investigating the Generalizability of Natural Language Processing of EMR Data

Hunter, Lawrence E.
2-R01-LM008111-04A1
University of Colorado Denver
Technology Development for a MolBio Knowledge-base

Hurdle, John F.
1-R21-LM009967-01
University of Utah
POET: Consolidated, Comprehensive Clinical Text Preprocessing

Ilyin, Valentin A. (New Investigator)
1-R01-LM009519-01A1
Northeastern University

Accurate Protein Structural Comparisons by TOPOFIT

Kalet, Adina Luba (New Investigator)
1-R01-LM009538-01A1
New York University School of Medicine
Randomized Trial of Educational Outcomes of Web Initiative in Surgical Education

Kim, Seungchan
1-R21-LM009706-01
Arizona State University-Tempe Campus
Integrating Genomic Data and Biological Knowledge to Learn Context-Specific Gene

Ma, Shuangge (New Investigator)
1-R03-LM009828-01
Yale University
Efficient Microarray Meta Analysis and Cancer Biomarker Selection

Medvedovic, Mario
1-R21-LM009662-01A1
University of Cincinnati
Integrative Probabilistic Models for Identifying Transcriptional Modules

Nakhleh, Luay (New Investigator)
1-R01-LM009494-01A1
Rice University
Evolutionary Analysis of Bacterial Genomes: High-Throughput Computational Tools

Ochs, Michael F.
1-R21-LM009382-01A2
Johns Hopkins University
An Open-Source Algorithm Isolating Overlapping Signatures in Microarray Data

Pakhomov, Serguei V. S. (New Investigator)
1-R01-LM009623-01A2
University of Minnesota Twin Cities
Semantic Relatedness for Active Medication Safety and Outcomes Surveillance

Peterson, Josh F. (New Investigator)
1-R01-LM009965-01
Vanderbilt University
Information Systems for Detecting and Managing Acute Kidney Injury

Prihod, Kevin F. (New Investigator)
1-R03-LM010053-01
Detroit Science Center
Medical Marvels Interactive Translational Research Experience (MITRE)

Rosenbloom, Samuel T. (New Investigator)
1-R01-LM009591-01A1
Vanderbilt University
A Framework Based Clinical Documentation Evaluation Method

Sarkar, Indra N. (New Investigator)
1-R01-LM009725-01A1
Marine Biological Laboratory
Enhancing Organism Based Disease Knowledge Via Name Based Taxonomic Intelligence

Srivastava, Ranjan (New Investigator)
1-R03-LM009753-01
University of Connecticut Storrs
Automatic Generation of Genome-scale Flux Balance Models

Szolovits, Peter
1-R01-LM009723-01A1
Massachusetts Institute of Technology
Capturing Patient-Provider Encounter through Text Speech and Dialogue Processing

Wagner, Michael M.
1-R01-LM009132-01A2
University of Pittsburgh at Pittsburgh
Decision Making in Biosurveillance

Zhao, Zhongming (New Investigator)
1-R03-LM009598-01A1
Virginia Commonwealth University
Investigating CpG islands in mammalian genomes

KNOWLEDGE MANAGEMENT

Eppig, Janan T.
1-G08-LM009693-01
Jackson Laboratory
International Mouse Strain Resource (IMSR)

Little, Charles Medaris (New Investigator)
1-G08-LM009710-01
University of Colorado Denver
A Web-based Hospital Computerized Disaster Information and Management System

Mandl, Kenneth D.
1-G08-LM009778-01A1
Harvard University (Medical School)
Evolving Clinical Information Libraries: Contextualizing Evidence Based Medicine

Triola, Marc (New Investigator)
1-G08-LM009535-01A1
New York University School of Medicine
Web Services for Preventive Health

SCHOLARLY WORKS

Aronowitz, Robert Alan
1-G13-LM009587-01A1
University of Pennsylvania
History of Health Risks in American Society and Medicine

Davis, Frederick Rowe (New Investigator)
1-G13-LM009606-01A1
Florida State University
Pesticides and Toxicology: A Century of Environmental Health

Fairchild, Amy L.
1-G13-LM009241-01A1
Columbia University Health Sciences
The Reach of Ethics: Politics, Bioethics & the Regulation of Social Inquiry

Foley, Paul (New Investigator)
1-G13-LM009863-01
University of New South Wales
Encephalitis Lethargica and Epidemic Influenza: Forgotten, but not Gone

Kisacky, Jeanne (New Investigator)
1-G13-LM009479-01A1
Individual Award
From Pavilions to Skyscrapers: A History of Healthy Hospital Design in New York C

Lerner, Barron H.
1-G13-LM009702-01
Columbia University Health Sciences
Getting Away with Murder? A Social and Cultural History of Drunk Driving

Prescott, Heather Munro
1-G13-LM009242-01A2
Central Connecticut State University
The History of Emergency Contraception

Rosner, David
1-G13-LM009707-01
Columbia University Health Sciences
Sidewalk Asylums: A History of Homelessness and Mental Illness in New York and LA

Tracy, Sarah W.
1-G13-LM009871-01
University of Oklahoma Norman
Health Revolutionary: The Life and Science of Ancel Keys

Wall, Barbra Mann (New Investigator)
1-G13-LM009691-01
University of Pennsylvania
A Comparative History of Twentieth-Century Catholic Hospitals

CAREER DEVELOPMENT

Alterovitz, Gil (New Investigator)
1-K99-LM009826-01A1
Children's Hospital Boston
A Holistic Approach to Information Processing for Biomedical Networks

Chang, Jeffrey T. (New Investigator)
1-K99-LM009837-01A1
Duke University
Functional Components of the Rb/E2F and p53 Pathways

Scotch, Matthew (New Investigator)
1-K99-LM009825-01
Yale University
Informatics for Zoonotic Disease Surveillance: Combining Animal and Human Data

SMALL BUSINESS GRANTS

Zhu, Yuerong (New Investigator)
1-R43-LM009913-01A1
Bioinforx, Inc.
Development of a Highly Automated Microarray Data Analysis System That Allows Re-analyze Deposited Microarrary Data With New Algorithms

Administration

Personnel

NLM/EP workload for program staff continues to increase each year due to new NIH requirements for reporting and analysis, including scrutiny and documentation for public access deposit of grant-related publications, fingerprinting for Research Condition and Disease Categories (RCDC) and writing new solicitations for the transition to RFAs. Administrative office workload has expanded due to new systems and reporting requirements, and the AO has worked without an assistant for more than one year. In FY2008, EP began the recruitment process for two new personnel (one Administrative Officer, and one Program Analyst). We hope to have these two positions posted by December 2008. Also, our Admin office is currently working with NIH/HRM to develop and recruit for two additional personnel in FY09; a Junior Committee Management Officer and one Program Officer, which will put us at an FTE level of 19.

In September 2008, NLM/EP welcomed a new addition to the DEAS support staff, Ms. Renee Jellerette-Stainback, as an Extramural Support Assistant. Ms. Jellerette-Stainback replaced Ms. Jackie McEachin who departed NIH in August 2008.

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' greatest achievement this year was the first public appearance of the collection of health information in multiple languages, which provides access to health information in more than 40 languages. The collection contains over 2,500 links to information and covers nearly 250 Health Topics.

Expanded MedlinePlus Go Local coverage by 22% from 22 areas to 27 areas and increased the percentage of the US population with Go Local coverage by 17% from 34% to 40%. As of FY2008, the MedlinePlus Go Local site includes North Carolina, Missouri, Indiana, Massachusetts, Utah, Wyoming, Maryland, East Texas, New Mexico, Texas Gulf Coast, Southern Ohio, Arizona, Nevada, Delaware, Vermont, Nebraska, Michigan, Minnesota, Illinois, Georgia, South Texas, Central Texas, Arkansas, Iowa North Dakota, Alabama and South Carolina.

Also this year, MedlinePlus saw over 723 million pages views and realized a 28% increase in unique visitors with over 140 million unique visitors. Additionally, MedlinePlus and Spanish MedlinePlus received a score of 85 and 84 respectively on the ACSI E-Government Satisfaction Index, a survey that tracks trends in customer satisfaction. Other major enhancements to MedlinePlus include adding several modules to provide foreign language capability.

DailyMed Project: The DailyMed project is a partnership between the Food and Drug Administration (FDA), the Veterans Administration (VA), the NLM, medication manufacturers and distributors, and 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 released a new version of FDA style sheets in October 2007, version 4.0 of the FDA SPL style sheets in September 2008 and version 4.0.2 of the FDA schema in September 2008. In addition, DailyMed approved more than 4,039 prescription drug labels for public access, added more than 700 new labels/package inserts, and recognized more than 1.7 million unique visitors and over 16.4 million page views.

Enhanced NIHSeniorHealth: With 38 topics now available in SeniorHealth, many new topics were added this year, including Participating in Clinical Trials, High Blood Cholesterol, Eating Well as You Get Older and Parkinson's Disease. Additionally, a new training guide toolkit was launched in November to help older adults find health information online. SeniorHealth recognized over 25 million site hits and over 1 million unique visitors with 18.74% being International visitors.

Multi-layered IT Security Program: OCCS continued its multi-layered IT security program that successfully detected more than 3.3 million probes, scans, denial of service (DOS) attacks, unauthorized access attempts and other security events on a monthly basis. Major accomplishments include:

- Performed a monthly cycle of vulnerability scanning, detection, and remediation to improve NLM security posture.
- Performed automatic virus scanning and signature update mechanisms to combat ever increasing cyber-attacks.
- Implemented software to automate data encryption securing sensitive information on desktop and mobile devices.
- Expanded an automatic patch management system to eliminate security vulnerabilities.
- Strengthened password requirements in response to directives from the Office of Management and Budget (OMB).
- Successfully passed a network penetration test performed by an independent contractor.

- Led a successful campaign to achieve 100% participation in information security and privacy awareness training for NLM employees and contractors.

High Speed Communication Network: Continued improvement of the redundancy of equipment and network paths to eliminate single points of failure in the network. NLM upgraded its Internet-2 connection by more than 10 times from 622Mbps to 10Gigabits and the public Internet connection by more than six times from 155Mbps to 1Gigabit. NLM maintains a redundant 2Gigabit connection to the NIHnet, managed by NIH's Center for Information Technology (CIT) and continues a redundant, diverse fiber connection from NLM to the MAX provided by FiberGate. It provides for increased reliability for this critical network connection. Also:

- 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.

Data Center Reengineering: The NLM data center 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 this growth will continue in the years ahead, OCCS continued a detailed process for evaluating the safety, reliability and performance requirements of the data center. This year's reengineering activities included:

- Increased air conditioning capacity from 170 tons to 380 tons.
- Installed an infrastructure for 40 In-Row Coolers (*RC) which will create a high density zone that will allower power usage in excell of 12KW per cabinet.
- Continued installation of an overhead Ladder Rack in the data center, as a separate pathway for running data networking cables to improve the reliability, availability, and maintainability of data communication services.
- Installed 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.
- Reconfigured the data center UPS system. The facility is now supported by three (3) 300 KW UPS, two parallel and one stand alone.
- Installed 336 L5-20 circuits to support growing electric power demands.

Unified Medical Language System (UMLS) Project: The Unified Medical Language System (UMLS) is a large multi-purpose, multi-lingual vocabulary database that contains information about biomedical and health-related concepts. There are currently, in the 2008AB edition of the UMLS Metathesaurus, 148 source vocabularies in 17 languages. The Metathesaurus contains nearly 1.9 million concepts (with 9.2 million names for those concepts) along with over 43 million relationships between and among these concepts. This represents a nearly 30% growth in concepts since the 2007AB edition a year ago. In the past two editing cycles (2008AA and 2008AB), we have sucessfully inverted and inserted over 53 sources and added seven new sources improving the coverage and richness of the Metathesaurus content.

RxNorm Project: Nine major versions were released this year for the RxNorm Editing System. Version 6.0 was released in December, 6.1 and 6.2 in January, 6.3 in February, 6.4 and 6.5 in March, 6.6 in April, 6.7 in May and 7.0 in August. Implementation of these releases included the ability to create Drug Delivery Device atoms, addition of Dose form authority control module, completion of Unit authority control, and addition of NDC code conflict module. The RxNorm application currently contains 18,502 active generic drugs, 15,206 active branded drugs, 4,135 active ingredients, 10,019 active grand names and 260,234 distinct NDC codes for RxNorm forms. Additionally, there were 12 monthly releases to RxNorm this year.

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. The NCCS serves as a disaster recovery/alternate computing site for NLM as well as CIT, NCI, NHLBI, NIDDK, NIAMS, OD/ORS and HHS/OS.

Enhanced Data Creation and Maintenance System (DCMS): The major event for DCMS this year is the baseline extraction, which is a re-release of all DCMS citations that follows the MeSH Year-end Processing (YEP). The 2008 MEDLINE®/PubMed® baseline database contains 16.8 million records. In addition, loaded and processed over 58,000 "new" OLDMEDLINE records for the publisher year 1949 and completed mapping over 1.4 million of the more than 1.8 million records of the OLDMEDLINE terms to current MeSH terms and switched that status to Medline. Also, the elapsed time between NCBI's delivery and DCMS availability of data to the Publisher Data review staff was reduced to 29 minutes from 155 minutes in the old method.

Enhanced Medical Subject Headings (MeSH): Several new features were implemented for the MeSH Translation Maintenance System (MTMS) which included the generation and distribution of new XML files for 2008 for

the French, German, Italian, Czech and Croatian translations. The development of the Publication Type Maintenance System for Auto YEP was completed for the GCMS System. Additionally, a new program was created called Rare Disease Maintenance System (RDMS), the data was provided by the NIH Office of Rare Disease, and loaded into the existing MeSH database. Also, a new functionality for processing of Preferred String Changes, Registry Number Changes, and Revision Date for YEP was developed.

Enhanced DOCLINE: DOCLINE, the NLM interlibrary loan (ILL) system, supports approximately 3,000 domestic and international libraries in processing more than 2 million interlibrary loan transactions and 400,000 Loansome Doc requests this year. Three versions of DOCLINE were released this year. Version 3.2.5 was released in October, 3.3 in February and 3.4 in July. These releases included 90 enhancements in response to user and Library Operations requests. Also, increased user page views from 3.6 to 45.5 million and increased the number of visitors by over 400%, from 800,000 to 3.5 million.

PHPartner.org: This year, the National Agricultural Library joined as a new partner of Public Health Partner. The Public Health Partner Web site helps the public health workforce find and use information effectively to improve and protect the public's health. This is a joint project among US government agencies (e.g. the Centers for Disease Control and Prevention, Agency for Healthcare Reseach and Quality), pubic health organizations (e.g. American Public Health Association, National Association of County and City Health Officials) and health sciences libraries (e.g. National Library of Medicine, National Network of Libraries of Medicine). Also, user page views increased by 33%, from 1.5 million to over 2 million as well as increased the number of visitors by over 35%, from 156,000 to over 210,000.

The following describes in more detail OCCS accomplishments in FY2008:

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. The NCCS serves as a disaster recovery/alternate computing site for NLM as well as CIT, NCI, NHLBI, NIDDK, NIAMS, OD/ORS and HHS/OS.

At present, all NLM mission-critical systems are either under active/active, active/passive or active/cold-backup mode depending on their business requirements. The Business Continuity and Disaster Recovery Plan covers NCCS as the primary resource for system restoration and uninterrupted processing if the primary NLM computing facilities on the NIH campus are rendered unavailable by a disaster or other contingency. During this year, OCCS 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.

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.

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

Increased air conditioning capacity from 170 tons to 380 tons.

- OCCS installed an infrastructure for 40 In-Row Coolers (IRC). The IRC's will create a high density zone that will allow power usage in excess of 12 KWs per cabinet. Twenty-four of the 40 IRCs have been installed and configured in groups of eight, for a total of four groups. Each group is configured for N+1 redundancy.
- Continued installation of an overhead Ladder Rack in the data center, 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 legacy communications cabling from beneath the raised floor, and re-install the cabling in the overhead ladder rack. In conjunction, the computer cabinets were re-oriented to form hot and cold aisles, for better management of cooling of the rack-mounted systems. Most of the legacy cabling under the floor has been removed, increasing the flow of cold air to computer racks.
- The installation of new cable trays for backbone blown-fiber optic cabling throughout Buildings 38 and 38A continued during the year. Phase one will be a pathway between the Onsite Alternate Computing Facility (OACF) and the Data Center. Later phases will service the key areas of both buildings. This is designed to provide alternate

diverse cable paths in the event that a disaster destroys one path of the cabling.

- An environmental monitoring software system was installed within the data center that has enhanced the ability to manage and report environmental incidences. The environmental system provides real-time status of temperature, humidity, UPS power usage per phase as well as CRAC status. Two 42 inch plasma displays are installed in the NOSC to display this information.
- The B1 data center's UPS system was reconfigured. The facility is now supported by three (3) 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 data center.

Consumer Health

MedlinePlus: MedlinePlus' greatest achievement this year was the first public appearance of the collection of health information in multiple languages, which provides access to health information in more than 40 languages. The collection contains over 2,500 links to information and covers nearly 250 Health Topics. An interactive World Map was built for the MedlinePlus languages release which is a flash-based feature that allows users to view a graphical representation of where nine major languages are spoken worldwide.

MedlinePlus Go Local implemented two releases this year, version 4.1 in November and version 4.2 in October. These releases included a sponsorship page, announcement link, event calendar and MapQuest zoom in/out feature. MedlinePlus Go Local expanded coverage by 22% from 22 areas to 27 areas adding sites serving Georgia, South Texas, Heart of Texas, Arkansas, Iowa and North Dakota. Additionally, the percentage of the US population with Go Local coverage increased by 17% from 34% to 40%.

There were five versions of MedlinePlus released this year. Version 20.3 was released in October of 2007, version 20.5 in January of 2008, version 21.0 in April, version 21.1 in May and version 21.2 in October. These releases included changing the search engine from RecomMind to Vivisimo and adding several modules to provide foreign language capability.

MedlinePlus saw over 723 million page views and realized a 28% increase in unique visitors with over 140 million unique visitors. Additionally, MedlinePlus and Spanish MedlinePlus received a score of 85 and 84 respectively on the ACSI E-Government Satisfaction Index, a survey that tracks trends in customer satisfaction.

SeniorHealth Project: NIHSeniorHealth is a joint NLM and National Institute on Aging (NIA) project that 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. With 38 topics now available in SeniorHealth, many new topics were added this year, including Participating in Clinical Trials, High Blood Cholesterol, Kidney Disease, Eating Well as You Get Older and Parkinson's Disease. A new training guide toolkit was launched in November, from NIA, to help older adults find health information online. SeniorHealth recognized over 25 million site hits and over 1 million unique visitors with 18.74% being International visitors.

The Accent module received numerous enhancements including medical terms, an upgraded US English female voice to help with NLM pronunciation dictionary terms, generated Spanish voice topics summaries and generated the 40 most popular MedlinePlus health topics to speech summaries in English and Spanish.

DailyMed Project: The DailyMed project is a partnership between the Food and Drug Administration (FDA), the Veterans Administration (VA), the NLM, medication manufacturers and distributors, and 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:

- Added over 700 new labels/package inserts.
- Implemented a new version of FDA style sheets.
- Released version 4.0 of the FDA SPL Style sheet in September.
- Released version 4.0.2 of the FDA schema in September.
- Approved more than 4,039 prescription drug labels for public access.
- Recognized more than 1.7 million unique visitors and over 16.4 million page views.

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 120,000 patches on commodity desktops this year fixing known vulnerabilities to software.

NLM implemented an OMB mandate on Federal Desktop Core Configuration (FDCC) for security configurations on the XP and VISTA operating systems. Vulnerabilities are eliminated which allow intruders to gain unauthorized access to government computers and networks, potentially disrupting an agency's work, such as compromising or stealing data.

NLM has strengthened its password policy and requirements in response to NIH and HHS guidance and guidelines upon directives from OMB to cope with evolving sophistication and motivation from our adversaries. A new minimum 8-character password requirement, as well as 60-day expiration, is being enforced for all user accounts.

Due to increased security and privacy threats reported by US-CERT and other Federal security and law enforcement officials, NIH blocked unauthorized Peer-to-Peer (P2P) file sharing traffic. NLM complied with HHS and NIH security policies as well as the National Institute of Standards and Technology (NIST) security controls prohibiting the use of P2P programs.

Web-based application attacks have dramatically increased in recent years. To provide application and development teams the knowledge and expertise needed to improve Web application security, over 80 NLM staff were trained in areas such as Web application security and vulnerabilities, secure coding principles, and threat modeling.

To comply with DHHS' policy on encryption and data security for all computers, NLM installed Pointsec disk encryption software on all laptops and over 40 desktops that contain personally identifiable information (PII).

Due to NIH's support of the issuance of digital certificates under HHS Public Key Infrastructure (PKI), OCCS created a procedure for HHS PKI Certificate Application and Installation. This procedure provides the high level workflow to request, generate and install a HHS PKI digital certificate.

To comply with the Federal Managers' Financial Integrity Act (FMFIA), OCCS prepared an assurance statement on risk identification and corresponding risk management that ensures continuity of NLM computing operations and protects NLM digital assets from malicious Internet attack attempts.

HHS issued an OPM mandate on Role-based training requirements for staff with significant information security responsibilities. NLM met the new requirement.

The Office of Management and Budget (OMB) requires that HHS computer users complete annual IT security awareness training. NLM has completed 100% of the mandatory FY08 Security Awareness Training for employees, contractors and fellows. A NLM policy requires completion of the training within 5 days for new hires.

Controlled Medical Vocabularies

Unified Medical Language System (UMLS) Project: The Unified Medical Language System (UMLS) is a large multi-purpose, multi-lingual vocabulary database that contains information about biomedical and health-related concepts. There are currently, in the 2008AB edition of the UMLS Metathesaurus, 148 source vocabularies in 17 languages. The Metathesaurus contains nearly 1.9 million concepts (with 9.2 million names for those concepts) along with over 43 million relationships between and among these concepts. This represents a nearly 30% growth in concepts since the 2007AB edition a year ago. In the past two editing cycles (2008AA and 2008AB), we have sucessfully inverted and inserted over 53 sources. Seven new sources (ICNP, LNC_Braden, LNC_FLACC, LNC_WHO, Medcin, FMA, NDFRT) were added, improving the coverage and richness of the Metathesaurus content. The SPECIALIST tools and the lexicon were updated in 2008AA. There were no changes to the Semantic Network.

Work is progressing on re-architecting MetamorphoSys to use a plug-in framework. This is planned to be included as part of the 2009AA or 2009AB release of the UMLS. Other accomplishments this year include work on a common domain model for the Metathesaurus and preliminary work with an object/relational mapping framework using Hibernate to assess the feasibility of its use in the MEME, MRD and MetamorphoSys systems.

RxNorm Project: Nine major versions were released this year for the RxNorm Editing System. Version 6.0 was released in December, 6.1 and 6.2 in January, 6.3 in February, 6.4 and 6.5 in March, 6.6 in April, 6.7 in May and 7.0 in August. Implementation of these releases included the ability to create Drug Delivery Device atoms, addition of Dose form authority control module, completion of Unit authority control, addition of NDC code conflict module, manage splits module, merge capability, ability to create "modified" SCD/SBD, performance improvements in component editing, ability to resurrect normal forms from archive, creation of fake source atom, source editing completion tracking page and security related updates.

There were 12 monthly releases to RxNorm this year which included the regular monthly releases and the addition of the new CMS (Centers for Medicaid and Medicare) source. Regular monthly inversion and insertions of the eight regularly released sources, specification, analysis, coding, testing of inversion and insertion for the new CMS source was conducted, 119 sources were inserted and approximately 3,000 atoms were processed for data quality changes. Two

resynchronizations were completed which included support for new merge scenarios and weekly releases during resynchronization, support for Drug Delivery Devices and integration of post tasks into load process.
The RxNorm application currently contains 18,502 active generic drugs, 15,206 active branded drugs, 4,135 active ingredients, 10,019 active grand names and 260,234 distinct NDC codes for RxNorm forms.

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). During FY2008, the MTMS cutover for the French, German, Italian, Czech and Croatian translations were completed. Development of Publication Type Maintenance System for Auto YEP was completed for the GCMS System.

Additionally, implemented modifications for the M2000 Maintenance System which included cutover for the 2008 MeSH Supplemental Chemical Records (SCR), completed development of new QA programs and report interface to monitor discrepancies between M2000 and DCMS_MESH in allowed subheadings, generated and distributed the 2008 MeSH XML file for Descriptors, Qualifiers and SCR and generated the new MeSH MARC file for 2008 MeSH.

Development of the first phase for Journal Descriptor Maintenance System (JDMS) for GCMS was completed. Also, new functionality was developed for processing of Preferred Sting Changes, Registry Number Changes, and Revision Date for YEP.

A new program was created called Rare Disease Maintenance System (RDMS), the data was provided by the NIH Office of Rare Disease, and loaded into the existing MeSH database. Modifications were made to the M2000 Client Interface and M2000 Administrative Module to accommodate the Chemical Class field. The new Chemical Class field will identify the special class of new MeSH records.

Medical Literature Services

Data Creation and Maintenance System (DCMS): The major event for DCMS this year is the baseline extraction, which is a re-release of all DCMS citations that follows the MeSH Year-end Processing (YEP). The 2008 MEDLINE®/PubMed® baseline database contains 16.8 million records. In addition, loaded and processed over 58,000 "new" OLDMEDLINE records for the publisher year 1949 and completed mapping over 1.4 million of the more than 1.8 million records of OLDMEDLINE terms to current MeSH terms and switched the status to Medline.
Also:

- Installed version 1.7 of Loader/Extractor.
- Incorporated over 70,000 OLDMEDLINE citations into DCMS.
- Reduced the elapse time between NCBI's delivery and DCMS availability of data to the Publisher Data review staff to 29 minutes from 155 minutes in the old method.
- Loaded LHC's MTI (Medical Text Indexer) indexing for more than 90,000 records and re-released to the Gateway system, as part of the annual Meeting Abstracts baseline extraction.

DOCLINE: DOCLINE, the NLM interlibrary loan (ILL) system, supported approximately 3,000 domestic and international libraries in processing more than 2 million interlibrary loan transactions and 400,000 Loansome Doc requests this year. Three versions of DOCLINE were released this year. Version 3.2.5 was released in October, 3.3 in February and 3.4 in July. These releases included 90 enhancements some of which included:

- Revising the interface network libraries use to order reports of their serial holdings and optional email notifications when a request is routed to their library.
- Adding a feature so libraries could easily set a range of "out of office" dates.
- Enhancing citation order displays to indicate author manuscripts, Elocation ID's, and LinkOut information for the library's own catalog.

Also, increased user page views from 3.6 to 45.5 million and increased the number of visitors by over 400%, from 800,000 to 3.5 million.

Voyager Integrated Library System (ILS): This years' major accomplishment was the year end processing of the database, which included the reformatting of several fields and upgrading MeSH terms to the 2008 version. Also, the first version of a program to replace the Oracle Forms component in the Voyager Bibliographic Publication process was coded and promoted.

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 v2008 which includes a customized view request report.

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 544 journals (55,125 issues) in various stages of scanning for entry into PMC, and also completed four data loads from APEX.

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 change forms format for printing, look and feel, date format, forms format and text font to the Medline Review form.

Serials Extract File (SEF): Among numerous upgrades and fixes, the generation of 2008 NLM LJI (List of Journals Indexed) and LSI (List of Serials Indexed) publications was completed. Also, changes were made to the Manage Serials Form to allow user maintenance of all 9 URL fields from SEF; revised rules for creation of PMC_Forthcoming_Note, PMCEmbargo, and XML PMC field; transferred URL data in the Manage Serials form when there is a title change; and migrated from the Oracle 9i to Oracle RAC 10g.

NLM Classification System: The NLM Classification System allows public and institutional access to the NLM Classification and related services and includes a Classification Editor. Publication of printed editions ceased with the 5th revised edition in 1999. Currently, a major reengineering and redesign is being conducted to improve the functionality and ease of use of the system.

Network and Systems Support

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

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

Public Internet connectivity services to NLM are provided through a contract with Level3. Internet connectivity was upgraded from an OC3 (155 Mbps) circuit to a 1Gigabit Ethernet (GigE) circuit in May of 2008. This primary circuit, and the redundant diverse backup circuit, connects NLM to the Level(3) Internet point-of-presence in McLean, VA. CIT and NLM have a peering arrangement where, in the event that the primary and backup NLM-Level(3) circuits fail, NLM Internet services will automatically failover to use the CIT Internet connections to Level 3 and Sprint. This failover capability is tested once per month.

Internet-2 has become an important resource for connection with NLM and the research community. Internet-2 connectivity was upgraded this year from 622Mbps to 10 Gigabits and provides a link to the Abilene high-speed backbone network via the Mid Atlantic Exchange (MAX) at the University of Maryland. LHC and OCCS work in cooperation to manage traffic to and from Internet 2. A redundant, diverse fiber connection from NLM to the MAX is provided by FiberGate. It provides for increased reliability for this critical network connection.

The OCCS-managed NLM perimeter network provides a multi-gigabit security boundary to aggregate connections to NIHnet, Internet, Internet-2, and the NCCS. It also provides for gigabit connections to interconnect OCCS, LHC, and NCBI divisional networks.

OCCS continued implementation of the High Availability Computing Solution to ensure 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. This initiative will increase the NLM mission critical database storage capacity eight times and will increase non-database storage capacity four times.

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.

Office Automation

OCCS planned and tested for enhanced desktop security controls known as the Federal Desktop Core Configuration (FDCC). These controls, which were developed by the National Institute of Standards and Technology (NIST) in order to meet the OMB 7-11 mandate, were reviewed for use for all Windows XP and Vista installations. Implementation was completed in February 2008.

Partnering with the NLM Personal Computer Advisory Committee (PCA), 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.

Seventy-one 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.

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 six courses this year, including "PC FUNdamentals", and "Managing your Mailbox" offerings. Focused training was provided in support of the 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 Outlook PST file reduction and other topics.

Public Health

Health Services Research Projects in Progress (HSRProj): There are currently 7,736 projects in the Ongoing and Completed HSRPROJ file, as well as 9,826 archived projects. The HSRProj Web site was updated with 1,451 new records. Additionally, 1,035 records with the Final Date between 1/2003 and 12/2008 were relocated to the archive file. Also, a new release of the Web application was promoted that included a new state map search feature which can be used by clicking on the map icon on the home page to allow users to activate the feature to retrieve projects performed in a certain state.
Health Services and Sciences Research Resources (HSRR) Database – Search & Maintenance: Search version 4 was released this year. Changes include the addition of the PubMed search icon on the display results page for a record with a PubMed search, update view of record display, and several choices of formats for printing retrieved records. Maintenance version 4 was also released which included updated view of record display and new HSRProj field which will hold the URL for HSRProj related searches.

PHPartners.org & HSR Info Central: The Public Health Partner Web site is a site to help the public health workforce find and use information effectively to improve and protect the public's health. This is a joint project among US government agencies (e.g. the Centers for Disease Control and Prevention, Agency for Healthcare Reseach and Quality), pubic health organizations (e.g. American Public Health Association, National Association of County and City Health Officials) and health sciences libraries (e.g. National Library of Medicine, National Network of Libraries of Medicine).
Version 5 was released this year which includes a 'News Features' which archives news items by month, a FAQs for PHPartners.org, and information regarding the National Agricultural Library which has joined as a new partner for PHPartners.

A new security enhancement was added for the input system that will limit the number of user login attempts before locking the users' account.

Also, user page views increased by 33%, from 1.5 million to over 2 million as well as increased the number of visitors by over 35%, from 156,000 to over 210,000.

Outreach and Customer Services

Against the Odds Web Exhibit & Web interactive activities: This project was released in April after 6 months of development. The accomplishments included:

- Design and artwork programming of three education interactive Modules
- Integrating the Web site into NLM architecture.
- Coordinating security scanning and verifying results
- Creating new RSS feature, aggregating health news from multiple government and private sources and presenting it in a unified page

Exhibit Asset Manager: A new database is being created for HMD Exhibitions Program to replace a very old Filemaker pro database. The Exhibit Asset Manager will allow the HMD exhibits team to track all art assets, artifacts and digital reproductions associated with each exhibit project. The new application is being written as a ColdFusion/FLEX application with the initial module (asset entry & contacts). The new application will be completed in FY09.

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:

- Added a Block SPAM button to the SR screen to allow the staff to block offending e-mail addresses sending SPAM to customer service accounts.
- Created a third set of backup e-mail boxes in CIT e-mail server for disaster recovery on incoming e-mails.
- Changed the Siebel service account passwords to comply NIH password policy.
- Added a validation for uniqueness of e-mail address and Last Name to eliminate duplicate contact record creation.

Research and Development

Search Engine: A major accomplishment for this year was moving NLM Main Web, MedlinePlus, MedlinePlus Spanish and GoLocal searches from Recommind to Vivisimo. OCCS upgraded the Vivisimo search cluster to new hardware, providing a 62% improvement in response time for the NLM Main Web and MedlinePlus searches. Spotlight boosting for MedlinePlus Spanish health topic pages was implemented. Production search servers were moved from Solaris to Linux resulting in a significant performance improvement.

NLM Digital Repository Project: OCCS continued to support NLM Library Operations on the NLM Digital Repository Project, by participating in the work of the Digital Repository Evaluation and Selection Working Group (DRESWG). Work during the year focused on the exploration, installation, and hands-on testing of three digital repository tools (DSpace, DigiTool, and Fedora) that had been previously selected by the DRESWG. Key accomplishments in support of the DRESWG included:

Conducted dozens of demonstrations and tutorial briefings on DSpace and Fedora capabilities and features to prepare and guide the DRESWG members in their hands-on testing.

- Briefed the DRESWG on technology roadmaps that describe the expected future evolution of these tools.
- Coordinated with the OCCS security team to conduct Web application security scans of DigiTool and Fedora.
- Prepared the working paper, "Initial NLM Collection Types," to help the DRESWG make decisions needed to begin implementation of the NLM digital repository.

Database: The main accomplishment this year was the upgrade from Oracle 9i to 10g. The following applications were migrated to the production 10g RAC database: JBIS, HSRR_SEARCH, PACS, M2000 and MTMS. Preparation continues for the migration of DCMS, SERIALS and LSTRC as well as implementing database security changes based on recommendations from the security team.

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 main focus has been in building the Voyager model, testing Voyager reports then making changes to the Voyager model.

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. OCCS launched RSS as an alternative to the NLM New Files mailing lists for receiving timely notifications of changes to NLM Web sites. Also added was a full UTF-8 support to the workflow to enable Unicode characters in the RSS messages as well as Vivisimo search to the NLM templates.

Web Analytics: NLM uses the WebTrends software package to track the number of pages served over time by the sites being managed and to provide detailed analysis of trends in site usage, audience composition, and other matters. WebTrends was upgraded from version 8.0d to 8.1a. An additional Web analytics product, Accenture Digital Diagnostics, has been purchased by HHS and will be licensed to all Operational Divisions for use with Section 508 remediation.

ADMINISTRATION

Todd D. Danielson
Executive Officer

Table 20

Financial Resources and Allocations, FY2008
(Dollars in Thousands)

Budget Allocation:	
Extramural Programs	$69,113
Intramural Programs	241,967
Library Operations	(92,535)
Lister Hill National Center for Biomedical Communications	(52,298)
National Center for Biotechnology Information	(81,180)
Specialized Information Services	(15,954)
Research Management and Support	12,307
Total Appropriation	323,387
Plus: Reimbursements	24,703
Total Resources	**$348,090**

Personnel

In September 2007, **Valer Gotea, PhD**, joined the Computational Biology Branch of NCBI on as a Research Fellow. Dr. Gotea earned his PhD from Penn State University on documenting the impact of transposable elements on the evolution of vertebrates. He will be working in the group of Dr. Ovcharenko on the development of novel computational methods and algorithms to identify and characterize gene regulatory elements and gene regulatory networks. He will focus on characterizing the genomic landscape and regulation of mammalian transcription factors, which should lead to a better understanding of global gene regulation in mammals. As a direct extension of his PhD work, Dr. Gotea will also study the impact of transposable elements on gene regulation, with the goal to identify mechanisms of lineage-specific adaptation of gene regulatory networks facilitated by transposable elements.

In September 2007, **Martin Shumway** converted to a Staff Scientist after having been an MSD contractor since July 2007. Martin obtained a MS in computer science from the Colorado State University and has over 20 years of experience leading software development of database and algorithmic applications. Prior to coming to NCBI, Martin was a director of software engineering at the Institute for Genomic Research (TIGR), where he supervised genome sequencing data production, handling, archiving, and analysis, participating in numerous sequencing projects and publications. Before that he was a staff engineer at Hewlett-Packard Company. Martin will be working in Dr. Ilene Mizrachi's team in IEB on trace, short read, and trace assembly archrivals.

In October 2007, **Steven Phillips, MD**, was appointed as the Associate Director for Specialized Information Services (SIS). Dr. Phillips is a graduate of Hobart College and Tufts Medical School. He is board certified both in general and thoracic surgery and practiced cardiac surgery in Des Moines, Iowa, for nearly 30 years. Dr. Phillips is former NLM Deputy Director for Research and Education and also former Acting Associate Director for SIS. Dr. Phillips was previously a member of the NLM Board of Regents, and he served as the Chair in 1997. Dr. Phillips has served as a consultant to NLM, providing expertise in support of NLM's initial efforts to define the requirements for the development of a Disaster Information Management Research Center, a high priority initiative recommended in the NLM Long Range Plan for 2006-2016. Dr. Phillips' expertise and experience in medical research, clinical practice, organization leadership, and disaster management combined with his direct experience with the NLM provides a unique combination of talent for leading and managing an important Division of NLM. Dr. Phillips will be working collaboratively with the NLM senior leadership team, NIH, other HHS and government agencies, academia and industry to develop the NLM disaster health information and research agenda as well as leading the Division of Specialized Information Services. Dr. Phillips will also serve as Chair of the NLM Disaster Information Management Coordinating Committee.

In October 2007, **Vladimir Sarkisov,** PhD, was appointed to a Staff Scientist position after having been a Lockheed Martin MSD contractor since 2002 at National Center for Biotechnology Information (NCBI). Dr. Sarkisov earned a PhD in Applied Physics from Moscow State Academy of Information Science for work on the hydrodynamics of liquid crystal. He joined Lockheed Martin MSD in 2002 as a contractor for NCBI working on the design and development of PubMed Central--the digital archive of full-text journal literature at the National Library of Medicine (NLM). As a Staff Scientist, he will continue to work on the previously developed and other new projects in PubMed Central.

In October 2007, **Ming Xu, PhD**, was appointed to a Staff Scientist position in the Information Engineering Branch's (IEB) dbGaP group. He was appointed to this position after spending several years as an IRTA Postdoctoral Fellow in the Structure Group, working under Lewis Geer and Steve Bryant on developing algorithms and software applications for proteomics studies. Dr. Xu earned a PhD at Florida State University for his work on Ab initio calculation and

molecular modeling. Prior to joining NCBI, he worked on developing algorithms and software within the biotech, IT, engineering and environmental health sectors. As a Staff Scientist at NCBI, he will be responsible for further development of NCBI's repository of genotype and phenotype information, dbGaP. Specifically, he will develop and maintain the dbGaP genome browser, which presents the integration of phenotype and genotype information within the context of the human genome.

In January 2008, **Vasuki Gobu** joined NCBI as a Staff Scientist after having been a contractor from TAJ Technologies. She was the primary software developer for the SKY/M-FISH & CGH and the Entrez Cancer Chromosomes databases. She joined the BankIt development team during the fall of 2004, where she develops software that facilitates online submission of different types of nucleotide sequences. Vasuki Gobu earned her MS in Bioinformatics from George Mason University; her project work included designing preliminary vocabulary for representing controls and replicates in microarrays. At NCBI, she will continue to provide software development, support and maintenance for the projects mentioned above.

In January 2008, **Margaret Haber** joined the National Information Center on Health Services Research and Health Care Technology (NICHSR), on detail from the National Cancer Institute. Ms. Haber is a biomedical informatics specialist with a background in history and languages, and as a certified critical care, bone marrow/organ transplant, and oncology nurse. She joins the Library to help develop a plan for NLM's clinical terminology coordination activities, with particular attention to its responsibilities as the US Member of the International Health Terminology Standards Development Organization, which is responsible for SNOMED. As an NLM contractor, Margaret served as clinical editor for the Library's *Unified Medical Language System*. For the past several years, she has served the Office of the Director, National Cancer Institute, as Co-Director of NCI's Enterprise Vocabulary Services, a set of services and resources that help meet NCI and its partners' needs for controlled terminology. She has also served as Project Officer in charge of Interagency Memoranda with both the FDA and the Veteran's Health Administration; as Representative to the Federal Health Architecture (FHA); Institutional Voting Member to HL7; Vocabulary Facilitator to HL7's Regulated Clinical Research Information Management Technical Committee; as Liaison to the Clinical Data Interchange Standards Consortium; and as formal Liaison to the SNOMED International Editorial Board.

In January 2008, **Hoda Marie Khouri** joined NCBI as a Staff Scientist. She is working with Steve Sherry and Martin Shumway on the 1000 Genomes and the Human Micro biome projects. Prior to her appointment Hoda worked for 10 years at the Institute for Genomic Research (TIGR), now known as the J. C.Venter Institute. She was a Team Leader for finishing Prokaryotic genomes and later a Biotech core Manager. Hoda's background is in Microbiology/Virology. She worked at the American Type Culture Collection (ATCC) as a Research Associate in the Virology department. She earned her Masters' degree from the American University of Beirut in Animal Science. Hoda will graduate in May 2008 with a Biotechnology Masters' degree from the Johns Hopkins University.

In January 2008, **Barbara Rapp, PhD**, assumed the position of Chief, Planning and Analysis, Office of Health Information Programs Development (OHIPD). Dr. Rapp has more than 20 years experience at NLM. Most recently, she served in the Division of Library Operations as coordinator of the postgraduate Associate Fellowship Program. Prior to that, she was in the National Center for Biotechnology Information for 12 years as manager of user services. She also has prior experience in OHIPD, where she participated in an important study of the impact of MEDLINE on clinical decision making. Prior to NLM, Dr. Rapp was on the faculty of the School of Library and Information Science at The Catholic University of America, where she taught courses in information systems and managed the program in health sciences librarianship. Dr. Rapp earned a PhD in Information Science from Drexel University in 1985, and an MS in Library and Information Science from the University of Illinois in 1978. She has numerous publications and presentations on scientific databases and information retrieval systems, and is also an active member of the Medical Library Association.

In January 2008, **Tam Paterson Sneddon, PhD**, joined NCBI as a Staff Scientist following a three-year postdoctoral position with the HUGO Gene Nomenclature Committee (HGNC) at University College London, UK, where he specialized in the naming and curation of segmentally duplicated and copy number variant genes. Tam completed his DPhil at the University of Oxford, UK, characterizing genes homologous to the PKD1 gene and later completed an MRes in Bioinformatics at the University of York, UK. Tam will be working with Deanna Church in the Information Engineering Branch (IEB) of NCBI to develop a large-scale structural variation database.

In January 2008, **Suresh Srinivasan** was appointed Chief, Medical Language Branch, Office of Computer & Communications Systems (OCCS). Mr. Srinivasan has a Masters degree in Operation Research and Statistics, a Masters degree in Transportation Engineering from Rensselaer Polytechnic Institute and a Bachelors degree in Civil Engineering from Indian Institute of Technology, India. Previously, Mr. Srinivasan was a project leader from the Unified Medical Language System (UMLS) at the Lister Hill Center. Prior to joining NLM, Mr. Srinivasan was a Principal Researcher with the Internet Lab at the Thomson Technology Services Group where he led research efforts

in electronic commerce and virtual communities for collaboration.

In February 2008, **Dianne Babski** was appointed as the new section head for the MEDLARS Management Section, part of the Bibliographic Services Division in Library Operations. She has been a member of MEDLARS Management Section since 2006, working on various projects including the PubMed training, testing, and customer service teams, the OLDMEDLINE data conversion project, and the Year End Processing team. Before coming to NLM, Ms. Babski was involved with the Scientific Review Program within the NIAID, working on new peer review tools and database resources for staff. She also worked on the NASA Spaceline data project that supplied citations and indexing to MEDLINE.

In February 2008, **Kathy Cravedi** was appointed Director of the Office of Communications and Public Liaison (OCPL). Ms. Cravedi has more than twelve years experience at NLM where she has served as the Deputy Director of OCPL since 1996 and as NLM's first Public Liaison Officer. During her service with NLM's OCPL, Ms. Cravedi has received numerous awards including the NIH Director's Award for her work on creating the NIH MedlinePlus Magazine, the NIH Director's Award for her contributions to NLM's outreach programs to the public, and the NLM Director's Award for her outreach activities. She brings a wonderful range of communications experience to her new position. From 1975 to 1994, Ms. Cravedi developed health care policies on Capitol Hill. As the Staff Director for the House Select Committee on Aging's Subcommittee on Health and Long-Term Care, she developed for the Chairman legislation that resulted in the establishment of the National Institute on Arthritis and Musculoskeletal and Skin Diseases Research, the National Deafness and Other Communication Disorders Institute, and the NLM's National Center for Biotechnology Information. She is a graduate of the University of Montana and received her Master's of Science degree from American University. She is the author and coauthor of numerous publications and congressional reports.

In February 2008, **Andrew Shkeda** was appointed to a Staff Scientist position after having been a TAJ Technologies contractor since 1999 at NCBI. Mr. Shkeda earned a B.S. degree in Physics from the Turkmenistan State University in 1993. He was employed as a programmer working on geological data management for 5 years, and subsequently in a supervisory account management position for 2 years before joining the NCBI RefSeq project as a contractor. He will continue to work in RefSeq project where his main responsibilities will be designing, implementing and maintaining software tools to facilitate and automate curation, generation, and validation of RefSeq sequences.

In February 2008, **Xiaoli Zhang, PhD**, joined the Lister Hill Center, Communications Engineering Branch to work on machine learning techniques for the automated extraction of bibliographic data from electronic journals. Dr. Zhang received a doctoral degree in Electrical Engineering from Rensselaer Polytechnic Institute, NY in 2007 where her thesis work focused on style-constrained optical character recognition. Her research interests are in document analysis, machine learning, image processing and computer vision.

In March 2008, **Renée Bougard, MLIS**, joined the NN/LM National Network Office as the Outreach Librarian. Ms. Bougard has more than 22 years of experience in health sciences libraries. She has held positions at the NN/LM South Central Regional Medical Library at the Houston Academy of Medicine - Texas Medical Center Library since 1992. Ms. Bougard served as the Network Coordinator for the NN/LM South Central Region (SCR) from 1992-1998 and as the Associate Director of the NN/LM SCR RML since 1998. In her role as Associate Director, she was responsible for the day to day operation of the RML and its very productive network and outreach programs. Prior to these positions, she was the director of the Lake Charles Memorial Hospital Library in Lake Charles, Louisiana, worked in several other hospital libraries, was the manager of user support services at Texas A & M University Medical Sciences Library, and worked at the LSU School of Veterinary Medicine Library. In 2006, she was one of two recipients of the MLA President's Award in recognition of her valuable contributions to the Hurricane Katrina disaster relief efforts. Ms. Bougard earned her MLS in Library and Information Science from Louisiana State University. As the NNO Outreach Librarian, Renee will focus on all outreach activities and projects involving health professionals, veterinarians, and the public health workforce.

In March 2008, **Farideh Chitsaz, PhD**, was appointed to a Staff Scientist position at NCBI. Dr. Farideh Chitsaz joined the staff of the computational Biology Branch (CBB), NCBI as a Research Fellow in June of 2002. She worked on several projects analyzing Plasmodium falciparum genome, the most lethal of the four species of plasmodium, using computational techniques. Dr. Chitsaz received her PhD in molecular and cell biology from University of Maryland. Prior to joining NCBI, Dr. Chitsaz was employed in the Malaria Vaccine Development Unit at NIAID. She joined Conserved Domain Database group at structure unit of CBB in March of 2007. As a Staff Scientist she will continue to work on the Conserved Domain Database as a curator.

In March 2008, **Nisar Hundewale, PhD**, joined the Lister Hill Center, Cognitive Science Branch to conduct research on PHR infrastructure and develop tools for smooth and efficient transactions. Dr. Hundewale received a doctoral degree in Computer Science from Georgia State University, GA in 2007 where his thesis work focused on algorithms and techniques for automated

design and manufacturing workflow for DNA microarray. His research interests are in algorithms, machine learning, and software engineering. His current focus is on Personal Health Record (PHR).

In March 2008, **Anna V. Ivshina, MD**, PhD, was appointed to a Staff Scientist position after having been a MSD contractor since February, 2008. Dr. Ivshina earned her MD from the First Moscow Medical Institute. She went on to receive clinical training in oncology and a PhD in oncological research from Russian Cancer Research Center in Moscow. She worked on developing new prognostic factors in cancer patients and conducted clinical trials. In 1998 she jointed CBER, FDA as IRTA Research Fellow, working on evaluation and control of live virus vaccines including influenza B viruses along with their vaccine reassortants, rotavirus G genotypes and attenuated mumps virus vaccine. Starting in 2004, she worked as Research Scientist in Genome Institute of Singapore (GIS), where she conducted studies aimed at discovering novel clinically relevant gene signatures in breast cancer based on microarray high-throughput gene expression data and class prediction statistical methods. During her tenure at GIS, Dr. Ivshina developed a novel classification of clinically relevant genetic grade breast cancer. She will be working in Dr. Tatiana Tatusova's group as fungal curator for RefSeq project with the NCBI.

In March 2008, **Priyanka Sharma** was appointed in NCBI to a Staff Scientist position after having been a Lockheed Martin contractor since October, 2007. She will be working with Deanna Church in the Map Viewer group on Data Loading and Database management issues. Prior to her appointment (from 2005 – 2007) she worked as an Informatics Engineer at Bioinformatics Labs, LBGM, University of La Reunion (France) where she was responsible for the development and maintenance of Web server that host online bioinformatics resources for public domain. She also was involved in the development of in-house applications. She holds Bachelors of Engineering in Computer science and Engineering.

In March 2008, **Douglas J. Slotta, PhD**, was appointed to a Staff Scientist position after having been a Lockheed Martin contractor since November 13, 2007. Dr. Slotta earned a PhD in Computer Science from Virginia Polytechnic Institute and State University for work on nonparametric statistical analysis of microarray data. This was followed by postdoctoral fellowship with Dr. Sanford Markey in the National Institute of Mental Health where he developed software to aid in the aid in the analysis of mass spectrometry-based proteomics data. He joined the National Center for Biotechnology Information (NCBI) to aid Dr. Ron Edgar's Gene Expression Omnibus (GEO) group in the design and implementation of the Peptides data repository for mass spectrometry-based proteomics data.

In March 2008, **Irena Zarestkaya** was appointed to a Staff Scientist position. She earned a Master's degree in Optical/Electrical Engineering from the Leningrad Institute of Fine Mechanics and Optics in St. Petersburg, Russia. She has extensive experience in the software development life cycle, having worked on various commercial and medical research projects. She joined Ariadne Genomics in December 2004 as a contractor for the National Center for Biotechnology Information (NCBI). At the NCBI she worked in the BLAST (Basic Local Alignments Search Tool) group on the design and implementation of a new Web interface for BLAST and other Web applications. As a Staff Scientist, she will continue to work on BLAST and other Web projects in the Information Engineering Branch.

In April 2008, **Haiying Guan, PhD**, joined the Lister Hill Center for Biomedical Communications as a postdoctoral fellow in the Communications Engineering Branch where she is working on Machine Learning Methods for Image Retrieval from Biomedical Image Databases and the Biomedical Literature. Dr. Guan received her doctorate degree in Computer Science from University of California, Santa Barbara in July 2007. Prior to coming to NLM, Haiying was a senior research engineer at The Digital Medial Solution Lab, Samsung Information Systems America in Irvine, California where she worked on motion detection, motion pattern recognition and skin color detection algorithms.

In April 2008, **Justin Paschall** joined NCBI as a Staff Scientist. His background includes computer science and graduate work in human genetics at Washington University in St. Louis. Mr. Paschall previously worked as a bioinformatics research assistant included building databases for the study of molecular evolution, gene annotation for micro-array experiments, and genetic mapping of gene expression traits. He will work on the NCBI's dbGaP project, a database of human genetic and phenotypic information.

In May 2008, **Krystl Haerian, MD**, joined the Lister Hill Center for Biomedical Communications as a clinical postdoctoral fellow and works with Dr. Jim Cimino on a project aimed at improving the NIH Clinical Center's electronic health record (EHR) system. She plans to work on developing a patient scheduling tab; a customized electronic ICU flowsheet; introduction of clinical alerts into the system; and incorporating infobutton links in the electronic medical record. Dr. Haerian received her medical degree from University of Maryland, Baltimore and an MS in Biotechnology from Johns Hopkins University.

In May 2008, **Wratko Hlavina** was appointed to a Staff Scientist position, following seven years of service at NCBI as a contractor. Mr. Hlavina earned his master's degree in Computer Science from the University of Ottawa, Canada. His thesis, in the area of artificial intelligence, incorporated techniques from the field of information retrieval with research into knowledge

management. His background is further complemented with a BSc in mathematics-science, focused on probability and statistics. Mr. Hlavina's work experience includes research activities with knowledge management systems, and consulting as a Sun Certified Java Developer. He joined NCBI in April, 2001, as a contractor from Lockheed Martin, Inc., and Management Systems Designers, Inc. He became an important member of the NCBI Genome Annotation Pipeline, where he has been and continues to be involved in several stages of processing, from sequence alignments to final annotation of RefSeq genomic products for a range of various organisms. His current role continues this involvement, with a focus on improving the quality, efficiency, and automation of the Pipeline, in both the workflow process and the computational algorithms.

In May 2008, **Yang Huang, PhD**, was appointed to a research fellow position after having been a Visiting Fellow since 2007 at NCBI. He is working with Teresa Przytycka in the research group on algorithmic and graph methods in computational biology. Before coming to NCBI, Dr. Huang completed his PhD in computer science from Rutgers University. In his thesis, he developed efficient graph algorithms for analyzing high-throughput protein-protein interaction and gene expression data.

In May 2008, **Hanzhen Sun, PhD**, was converted to a Staff Scientist position after having been a MSD contractor since March 2001. Hanzhen earned a PhD degree in biological sciences from Columbia University in 2000, for work on splicing signals. Prior to joining NCBI, he had brief postdoctoral training at NCI. He will continue to work in the dataflow team in IEB, generating Genbank and RefSeq update data products, Refseq releases, as well as on other data acquisition, data maintenance, database monitoring and synchronization projects.

In June 2008, **Pertti (Bert) Hakkinen, PhD**, joined the Division of Specialized Information Services as senior toxicologist in the Office of the Director, and will serve as NLM Toxicology and Environmental Health Science Advisor. As a member of the SIS staff, Dr. Hakkinen will provide leadership on the development of new resources in toxicology and enhancements to existing NLM resources in this field. He will also represent NLM on various committees and will provide leadership for NLM's participation in national and international efforts in toxicology information. During his career Dr. Hakkinen has held numerous leadership positions in the field of toxicology and risk assessment. He continues to serve as the Vice-chair of the Scientific Advisory Panel of the Mickey Leland National Urban Air Toxics Research Center. Dr. Hakkinen has also been a Principal at Gradient Corporation, an environmental and human risk assessment consulting company, and served on the staff of the European Commission at the EC's Institute for Health and Consumer Protection, Joint Research Centre, in Ispra, Italy for several years. While at the EC, he worked on the creation of the European Information System on risks from chemicals released from consumer products and articles (EIS-ChemRisks). Prior to joining the EC's staff, Dr. Hakkinen held positions with Toxicology Excellence for Risk Assessment (TERA) in the US and at the Procter and Gamble Company in the US and Japan. Dr. Hakkinen earned a B.A. in Biochemistry and Molecular Biology from the University of California, Santa Barbara, and received his PhD in Comparative Pharmacology and Toxicology from the University of California, San Francisco. Dr. Hakkinen is a member of the Society of Toxicology (SOT) and a charter member of the Society for Risk Analysis (SRA) and the International Society of Exposure Analysis (ISEA). He was a co-editor and co-author of the latest edition of the Encyclopedia of Toxicology, and of the upcoming new edition of the Information Resources in Toxicology book. Dr. Hakkinen has authored and co-authored numerous other publications.

In June 2008, **Larry Smith, PhD**, was appointed to a Staff Scientist position at the National Center for Biotechnology Information, Computational Biology Branch after having been a MSD/Lockheed contractor since 2002 at the NCBI. Dr. Smith earned his PhD in mathematics from Indiana University in 1998 (computer algebra major, probability theory minor). Soon after graduating, he accepted a position at the National Cancer Institute as a staff scientist, where he developed statistical methods and software to analyze, visualize, and link gene microarray data from cancer cell lines. In 2000 and 2001, he worked as a mathematician at the Department of Defense, where he developed computer algorithms for voice and digital signal analysis. In 2002, he received a medical informatics fellowship at the Lister Hill National Center for Biomedical Communications, where he prototyped the "Genetics Home Reference" Web site. Dr. Smith continues to work with Dr. John Wilbur to research and develop mathematically-oriented text mining applications at the National Library of Medicine.

In July 2008, Captain **Mary Chaffee**, Nurse Corps, US Navy, was detailed from the National Naval Medical Center to NLM to serve as the Disaster Research Coordinator in Disaster Information Management Research Center. Her work is in support of the Bethesda Hospital Emergency Preparedness Partnership in which NLM became a partner earlier this year. Captain Chaffee has served in the Navy for 25 years and recently completed a PhD in Nursing at the University of Maryland Baltimore. Previous to this, she served as Director of the Navy Medicine Office of Homeland Security where she led Navy Medicine's worldwide efforts to improve hospital disaster preparedness. A critical care and emergency nurse, Captain Chaffee served in the Federal Emergency Management Agency's Emergency Operations Center in Washington, DC following the terror attacks of September 11, 2001. Captain Chaffee is co-editor of the textbook *Policy and*

Politics in Nursing and Health Care and has published extensively on topics including emergency preparedness, leadership and health policy. Her work has been recognized with the Legion of Merit and other awards.

In August 2008, **Darlene Dodson** joined the National Library of Medicine as the Budget Officer. Ms. Dodson previously served over a decade at the Health Resources and Services Administration (HRSA), HHS. Most recently Ms. Dodson served as the Deputy Associate Administrator for the Office of Federal Assistance Management providing leadership and oversight of HRSA's grants management activities. While with HRSA Ms. Dodson also served as Chief of the Resources and Analysis Branch in the Division of Financial Management; Deputy CIO and Deputy Director of the Office of Information Technology; and Director of the Division of Financial Integrity. Before HRSA, Ms. Dodson held positions within HHS as Special Assistant to the Director for Budget Analysis in the Office of the Secretary; Team Leader in the Office of the Assistant Secretary for Health Financial Management Branch; and Budget Analyst at the NIH. In June 1991, Ms. Dodson earned a BS degree in Business Administration and Accounting from the University of Maryland University College, College Park, MD. In September 1996, Ms. Dodson earned a Graduate Certification in Leadership and Change Management from the Johns Hopkins University, Baltimore, MD.

In August 2008, **Kathel Dunn** was appointed the Associate Fellowship Program Coordinator with oversight of the Library's post-graduate training program for library and information science professionals. Ms. Dunn comes to NLM from the New York University Medical Center where she has served as the Associate Director of the National Network of Libraries of Medicine, Middle Atlantic Region for the past two years. Prior to that, she was the Associate Director of Public Services at the Ehrman Medical Library at NYU. In addition to her extensive experience in reference, document delivery and circulation services, she has been involved in planning, designing and delivering instructional modules and training classes to librarians, staff, and students. She is currently an Adjunct Instructor in Health Sciences Librarianship at Queens College in New York. Ms. Dunn has a bachelor's degree in history from Mary Washington College and earned her MLS from the University of North Carolina at Chapel Hill. She is working on her dissertation "Toward a Theory of the Epistemic Values of Biological Scientists" to complete the requirements for her doctorate in Library Science from Rutgers, the State University of New Jersey.

In August 2008, **Gale Dutcher** was appointed Deputy Associate Director, Division of Specialized Information Services. Previously, she held the position of Chief, Outreach and Special Populations Branch within SIS. Ms. Dutcher initially came to NLM as a Library Associate Fellow. She subsequently worked in several areas of NLM starting in the Cataloging Section. She then went to the Collection Access Section where she served as a systems librarian on the project team that developed DOCLINE. In 1981, Ms. Dutcher moved to the Division of Specialized Information Services where she headed the DIRLINE team and worked on other projects within the Division. Ms. Dutcher was involved in NLM's early efforts to provide AIDS information services, working with the National Institute of Allergy and Infectious Diseases and the Food and Drug Administration to develop the AIDSTRIALS and AIDSDRUGS databases in 1989. She currently coordinates AIDSinfo, which NLM manages on behalf of the Department of Health and Human Services. Ms Dutcher also helps to coordinate the Library's overall AIDS programs. As the Chief of the Outreach and Special Populations Branch, Ms. Dutcher is a leader of the Library's efforts in information outreach to minority populations. She has developed programs to increase the capacity of underserved populations to access electronic health resources, develop health literacy, promote awareness of resources and their use, provide training, develop specialized Web sites and other electronic resources, and connect users to local libraries that can serve their needs. She received a BS degree in biology from Stony Brook University, an MS in zoology from the College of Environmental Science and Forestry, and an MLS from CW Post College, Long Island University.

NLM Associate Fellows Program for 2008 - 2009

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, 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. Four new Associate Fellows begin their year at NLM on September 2, 2008.

Kathleen Amos received her MLIS in May 2008, from Dalhousie University in Halifax, Nova Scotia and also holds a BA in Sociology and Social Anthropology from Dalhousie. She has experience in cataloging and database management as a student assistant in an academic library as well as varied practicum experience in the Health Sciences Library and Family Resource Library of the Izaak Walton Killam Health Centre. She has been involved as a volunteer with a patient record database at the Maritime Medical Genetics Service and served as the Secretary for the Dalhousie University Student Chapter of the Canadian Library Association.

Amy Donahue received her MLIS from the University of Washington in June 2007. She has worked as a content indexer for MSN Shopping and as a library

volunteer for the Richard Hugo House in Seattle, a center for the literary arts. Amy has also worked part time as a processing assistant in the Henry M. Jackson special collections at the University of Washington and as the library intern for the biotech company Amgen. She has a dual bachelor of arts in Mathematics and Russian Language and Literature from Grinnell College.

Paula Maez received her MLS in May 2008 from the University of Arizona, Tucson and has worked as a graduate assistant for the Knowledge River School of Information Resources and Library Sciences Wellness Education Search Project. She has also worked as a student assistant in the Arizona Health Sciences Library. She is interested in the role of information access in overcoming health disparities among underserved populations. She has a BA degree in Education and a BS in Speech and Health Sciences, both from the University of Arizona.

Patrick McLaughlin received his MLIS in May 2008 from the University of South Carolina. He has experience as an intern at the Centers for Disease Control and Prevention Public Health Library and Information Center and the Rodale Library. Mr. McLaughlin also has experience from an assistantship at the University of South Carolina School of Medicine Library. He has an undergraduate degree from Clemson in Biological Sciences.

Retirements and Separations

In January 2008, **Susan Buyer**, Chief, Planning and Analysis, Office of Health Information Programs Development (OHIPD), retired from the National Library of Medicine after more than 36 years of Federal government service. Ms. Buyer received her MA in comprehensive health planning from the George Washington University. Following her selection by the NIH Management Intern Program she came to NLM in 1972 where she began work in planning and evaluation, which remained the primary focus of her distinguished contributions to the Library that culminated with the recent publication of the latest long range plan of the Board of Regents, "Charting a Course for the 21st Century: 2006-2016". Susan's immediate supervisors changed periodically over the years as did her increasing scope of responsibility; her high standards and pursuit of excellence remained a constant and salutary presence. Susan was recognized by presentations of the NLM Director's Award, the NIH Director's Award, and the NIH Merit Award.

Awards

The NLM Board of Regents Award for Scholarship or Technical Achievement is awarded to recognize and stimulate independent creativity leading to scholarly and/or technical achievements that enrich biomedicine. The 2008 recipient was **Michael J. Ackerman, PhD** in recognition of his energetic and visionary leadership during the development and implementation of the Visible Human Project.

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 2008 award were **Teressa Wittig** in recognition of her excellent leadership in developing an active program for licensing access to ejournals and other resources for NLM; and **Susan Von Braunsberg** for dedication, technical proficiency, and outstanding organizational skills related to creation of electronic versions of citations published before 1966, furthering the NLM mission to provide access to the world's biomedical journal literature.

The NLM Director's Honor Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to **Lucy D. Ozarin, MD** for her outstanding volunteer efforts that have improved access to NLM's resources in the history of medicine and uncovered hidden gems in the historical collections.

The NIH Merit Award was presented to 6 individuals. The individual recipients were **Marie E. Gallagher** for her unwavering commitment and contributions to NLM's Profiles in Science; **Lori J. Klein** for exceptional contributions to many NLM products for consumer and professionals including MedlinePlus, *Citing Medicine*, Go Local, and customer service; **John M. Rozier** for his leadership and dedicated efforts enhancing and optimizing the Data Citation Management System and the NLM Year-End-Process to improve the efficiency of NLM business processes; **Margaret V. Slovikosky** for her outstanding administrative and management skills for maintaining a very high and effective operational level of the Lister Hill Center; **Janice M. Ward** for successfully expanding the scope of gene indexing to include clinical genetics articles; and **Deborah A. Zarin, MD** for her exceptional leadership and unwavering commitment to the ClinicalTrials.gov Project.

The NIH Director's Award was presented to two individuals and 6 groups. The individual award recipients were **Dianne E. McCutcheon** for exceptional initiative and leadership in carrying out activities to improve NLM programs and services; and **Nadgy P. Roey** in recognition of her leadership and skill in the daily administration and coordination of an outstanding ethics program. The group recipients were **Phillip D. Osborne** as a member of the Acquisition Offices group in recognition of unwavering leadership of staff and commitment to mission; **James M. Ostell, PhD** as a member of the Roadmap Recognition group for exceptional service and teamwork in implementation of the second cohort of Roadmap

programs; **Susan Buyer** and **Jerrard R. Sheehan** as members of the NIH Biennial Report Leadership Team for conceptualizing and producing the first of the new NIH Biennial Reports, mandated by the NIH Reform Act of 2006; **David J. Lipman, MD**, **Edwin P. Sequeira** and **Barton W. Trawick, PhD** as members of the Public Access group for outstanding management of the voluntary Public Access Policy and proactive development of a policy for mandatory Public Access; **David J. Lipman, MD**, **James M. Ostell, PhD**, **Alan S. Graeff** and **Stephen T. Sherry, PhD** as members of the Genome-Wide Association Studies Policy Development Team in recognition of the outstanding teamwork, skill, and dedication in the development of the landmark NIH GWAS data sharing policy; and **Alejandro A. Schaffer, PhD** as a member of the Job's Syndrome Group in recognition of a team effort to elucidate the mechanisms of pathogenesis of Job's Syndrome.

The EEO Special Achievement Award was presented to **Cassandra R. Allen** for her active participation in providing critical support to the Student National Medical Association's (SNMA) Brotherhood Alliance for Science Education (BASE) project.

The Phillip C. Coleman Award was presented to **George W. Franklin** for his outstanding contributions in recruitment, training, and organizing the cadre of NLM staff that conduct the NLM Powwow outreach activities.

Table 21

FY2008 Full-Time Equivalents (Actual)

Office of the Director	9
Office of Health Information Programs Development	6
Office of Communication and Public Liaison	9
Office of Administration	44
Office of Computer and Communications Systems	52
Extramural Programs	15
Lister Hill National Center for Biomedical Communications	68
National Center for Biotechnology Information	196
Specialized Information Services	38
Library Operations	265
TOTAL FTEs	**702**

NLM Diversity Council

NLM Director's Employee Education Fund

NLM's Director, Dr. Donald A.B. Lindberg, established the Director's Employee Education Fund in 1998 because of his strong belief in continuing education for Library employees. This year, the NLM Diversity Council continued its coordination of the Director's Employee Education Fund. In FY2008, the Fund enabled 50 staff members to take 72 courses. Undergraduate classes made up the majority of the classes supported. The school with the largest number of NLM enrollees was the University of Maryland (21), with Strayer University coming in second (18), and Montgomery College third (8). Other institutions attended include: Walden University, American University, Catholic University, Howard University, Johns Hopkins University, the Foundation for Advanced Education in the Sciences (FAES), Virginia Tech, University of the District of Columbia, IMPAC Government Services, Bowie University, George Mason University, USDA graduate school, University of Notre Dame, Frederick Community College, and Wesley Theological Seminary. Course disciplines enrolled in include: computer technology, sociology, psychology, business administration, communications, and mathematics, among others. In addition to traditional classroom instruction, some students took courses on the internet. The Diversity Council continues its efforts to publicize the availability of the Fund. This year witnessed an increase in staff participation; ten more employees signed up for courses versus last year's total.

NLM Diversity Council Healthy Lifestyles Expo

On April 24, the Diversity Council sponsored the "Healthy Lifestyles for You and Your Family Expo" in conjunction with NIH's "Take Your Child to Work Day." The Expo featured a day's worth of educational, family-oriented activities focusing on healthy living styles for both children and adults. The Emmy Award-winning FoodPlay Productions, a performance group that uses both live theatre and interactive media to promote healthy eating and exercise activities, performed a morning and an afternoon show. The community at large also contributed to help make the Expo an outstanding success. Whole Foods gave a healthy cooking demonstration, and exercise professionals instructed children and their families on everything from yoga to jump rope. Water balloon fights, raffles, and a moon bounce helped round out the afternoon activities. The Expo was attended by more than 1,000 NIH employees and their families. Due to this overwhelming enthusiasm, the Diversity Council plans to make the Healthy Lifestyles Expo an annual event on the NIH campus.

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 MD 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
US 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
Roseville, MN

Ex Officio Members:

Director
National Agricultural Library

Librarian of Congress

Surgeon General
United States Air Force

Surgeon General
Department of the Navy

Surgeon General
Department of the Army

Acting Surgeon General
US 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 (BSC) provides advice on NLM's intramural research and development programs for the Lister Hill Center.

Appointed Members:

ASH, Joan S., PhD (Chair)
Associate Professor
Department of Medical Informatics and Epidemiology
Oregon Health and 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 Healthcare
Information Systems
Salt Lake City, UT

LUMPKIN, John R., MD
Senior Vice President and Director
Healthcare 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
Associate Professor
Department of Surgery
University of Chicago Research Institutes
Chicago, IL

Appendix 4: Board of Scientific Counselors, National Center for Biotechnology Information

The Board of Scientific Counselors (BSC) provides advice on NLM's intramural research and development programs for the National Center for Biotechnology Information.

Appointed Members:

GINSBURG, David, MD (Chair)
Professor
Internal Medicine and Human Genetics
Life Sciences Institute
University of Michigan
Ann Arbor, MI

ALLEWELL, Norma M., PhD
Dean
Department of Chemistry and Biochemistry
University of Maryland
College Park, MD

LEVINE, Arthur S, MD
Senior Vice Chancellor for Health Sciences
Dean, School of Medicine
University of Pittsburgh
Pittsburgh, PA

LYNCH, Michael R., PhD
Distinguished Professor
Department of Biology
Indiana University
Bloomington, IN

NICKERSON, Deborah A., PhD
Professor
Department of Genome Sciences
University of Washington School of Medicine
Seattle, WA

SALEMME, F. Raymond, PhD
President
Imiplex, LLC
Yardley, PA

THOMAS, Annette C., PhD
Managing Director and President
Nature Publishing Group
MacMillan Publishers LTD
London, UK

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 is the initial review group for the NLM, and provides the first level of review for NLM's grant programs.

Appointed Members:

WALKER, James M., MD (Chair)
Chief Health Information Officer
Geisinger Health System
Danville, PA

ARONSKY, Dominik, MD, PhD
Assistant Professor
Department of Biomedical Informatics
Eskind Biomedical Library
Vanderbilt University Medical Center
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, Informatics
Center for Computational Biology & Bioinformatics
Indiana University Schools of Medicine and Informatics
Indianapolis, IN

HURDLE, John F., MD, PhD
Associate Professor
Department of Biomedical Informatics
University of Utah School of Medicine
Salt Lake City, UT

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

LIDDY, Elizabeth D., PhD
Dean and Trustee Professor
School of Information Studies
Syracuse University
Syracuse, NY

LU, Xinghua, MD, PhD
Associate Professor
Department of Biostatistics, Bioinformatics, and Epidemiology
Medical University of South Carolina
Charleston, SC

MANDL, Kenneth D., MD, MPH
Associate Professor
Childrens Hospital Informatics Program
Childrens 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
Associate Professor
Department of Pediatrics
Section of Hematology/Oncology
Computation Institute
The University of Chicago
Chicago, IL

NADKARNI, Prakash M., MD
Senior Research Investigator
Geisinger Center for Healthcare Research
Danville, PA

PANI, John R., PhD
Associate Professor
Department of Psychological and Brain Sciences
University of Louisville
Louisville, KY

SALTZ, Joel H., MD, PhD
Director
Center for Comprehensive Informatics
Emory University
Atlanta, GA

SHATKAY, Hagit, PhD
Head, Computational Biology and Machine Learning Lab

School of Computing
Queen's University
Kingston, Ontario

SHEDLOCK, James, AMLS
Director
Galter Health Sciences Library
Feinberg School of Medicine
Northwestern University
Chicago, IL

SITTIG, Dean F., PhD
Associate Professor
UT Memorial Herman Center
Healthcare Quality and Safety
University of Texas School of Health Information Sciences (At Houston)
Houston, TX

STATES, David J., MD, PhD
Professor of Health Information Science
Institute for Molecular Medicine
School for Health Information Sciences
University of Texas Health Science Center at Houston
Houston, TX

TENNANT, Michele R., PhD
Bioinformatics Librarian
Health Science Center Libraries
And UF Genetics Institute
University of Florida
Gainesville, FL

TONELLATO, Peter J., PhD
Senior Research Scientist
Center for Biomedical Informatics
Harvard Medical School
Boston, MA

WARD, Deborah, MA, MSLS
Director
Health Sciences Libraries
University of Missouri-Columbia
Columbia, MO

Appendix 6: Literature Selection Technical Review Committee

The Literature Selection Technical Review Committee advises the NLM on matters of policy related to the evaluation and recommendations of biomedical publications to be considered for indexing and inclusion in Medline.

Appointed Members:

BAUCHNER, Howard, MD (Chair)
Professor of Pediatrics and Public Health
Department of Pediatrics
Boston University School of Medicine
Boston, MA

CHRISTOPHER, Mary M., PhD
Professer of Pathology
Department of Pathology, Microbiology & Immunology
School of Veterianary Medicine
University of California
Davis, CA

COPELAND, Robert L., PhD
Associate Professor
Department of Pharmacology
Howard University
College of Medicine
Washington, DC

DOSWELL, Willa A., PhD
Associate Professor
School of Nursing
University of Pittsburgh
Pittsburgh, PA

FLEMING, David A., PhD
Professor of Medicine
Director, Center for Health Ethics
University of Missouri School of Medicine
Columbia, MO

HASHIMOTO, Frederick, MD
Distinguished Professor of Medicine
Division of General Internal Medicine
University of New Mexico
School of Medicine
Albuquerque, NM

MORENO, Carlos A., MD
Professor and Chair
Department of Family and Community Medicine
University of Texas Health Science Center
Houston, TX

NORTON, Catherine N., MLS
Director, Information Technology
MBL/WHOI Library
Marine Biological Laboratory
Woods Hole, MA

PHILLPOT, Caroline C., MD
Chief
Genetics and Metabolism Section
Liver Diseases Branch, NIDDK
National Institutes of Health
Bethesda, MD

RACZ, Gabor B., MD
Grover Murray Professor and Director of Pain Services
Department of Anesthesiology
Texas Tech University Health Science Center
Lubbock, TX

SMITH, Paul D., MD
Associate Professor
Department of Family Medicine
University of Wisconsin Medical School
Madison, WI

SPANN, Melvin, PhD
Associate Director, Retired
Silver Spring, MD

VAN PEENEN, Hubert J., MD
Professor of Medicine, Retired
Eugene, OR

WALTON, Linda J., MLS
Associate University Libriarian and Director
Hardin Library for the Health Sciences
University of Iowa Libraries
University of Iowa
Iowa City, IA

Appendix 7: PubMed Central National Advisory Committee

The PubMed Central National Advisory Committee establishes criteria for groups submitting materials to the PubMed system, monitoring its operation, and ensuring that as PubMed Central evolves it remains responsive to the needs of researchers, publishers, librarians, and the general public.

Appointed Members:

KILEY, Robert J., MS (Chair)
Head
Systems Strategy
Wellcome Library
Wellcome Trust
London

ADLER, Prue S., MS, MA
Associate Executive Director
Association of Research Libraries
Washington, DC

ALIRE, Camila A., EDD
Dean of Library Services
University Libriaries
University of New Mexico
Albuquerque, NM

GREENSTEIN, Daniel, PhD
University Librarian
California Digital Library
Oakland, CA

HAWLEY, John, BA
Executive Director
American Society for Clinical Investigation
Ann Arbor, MI

HENDERSON, Cynthia L., MLS.
Director
Health Sciences Library
Morehouse School of Medicine
Atlanta, GA

KOHANE, Isaac S., MD, PhD
Director
Informatics Program
Department of Medicine
Children's Hospital Boston
Boston, MA

MICHALAK, Sarah, MLS.
Associate Provost for University Libraries
University of North Carolina at Chapel Hill
Chapel Hill, NC

PARTHASARATHY, Hemai B., PhD
Vice President of Life Sciences
Feinstein Kean Healthcare
San Francisco, CA

SOBEL, Mark E., MD, PhD
Executive Officer
American Society for Investigative Pathology
Bethesda, MD

VELTEROP, Jan, PhD
Director of Open Access
and Journal Publishing Facilitation
Springer Publishing
Asbourne House
United Kingdom

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

WEINTRAUB, Susan T., PhD
Professor
Department of Biochemistry
University of Texas
Health Science Center
San Antonio, TX

WILBANKS, John, PhD
Executive Director, Retired
Boston, MA

Appendix 8: Organizational Acronyms and Initialisms Used In This Report

AAHSL	Association of Academic Health Sciences Libraries
ABC	Advanced Biomedical Tele-Collaboration Test Bed)
ACORN	Automatically Creating OLDMEDLINE Records for NLM
ACP	American College of Physicians
ACSI	American Customer Satisfaction Index
AFIP	Armed Forces Institute of Pathology
AG	Access Grid
AHIC	American Health Information Community
AHRQ	Agency for Healthcare Research and Quality
AIDS*info*	Acquired Immune Deficiency Syndrome*info* (database)
ALTBIB	Alternatives to Animal Testing
AME	Automated Metadata Extraction
AMIA	American Medical Informatics Association
AMPA	American Medical Publishers Association
AMWA	American Medical Women's Association
APDB	Audiovisual Program Development Branch
APIRE	American Psychiatric Institute for Research and Education
ASCCP	American Society for Cervical Pathology and Colposcopy
BAC	Bacterial Artificial Chromosome
BarSTool	Barcode Submission Tool r
BGMUT	Blood Group Antigen Gene Mutation Database
BHEPP	Bethesda Hospitals' Emergency Preparedness Partnership
BISTI	Biomedical Information Science and Technology Initiative
BITA	Biomedical Image Transmission via Advanced Networks
BLAST	Basic Local Alignment Search Tool
BLIRC	Biomedical Library and Informatics Review Committee
BMT	Boundary Marking Tool
BOR	Board of Regents
BSAT	BMT Study Administration Tool
BoSC	Board of Scientific Counselors
BSD	Bibliographic Services Division
BSN	Bioinformatics Support Network
CANDHI	Central American Network for Disaster and Health Information
CAS	Collection Access Section
CBB	Computational Biology Branch
CBIR	Content-Based Image Retrieval
CCB	Configuration Control Board
CCDS	Consensus CoDing Sequence
CCHIT	Commission for Healthcare Information Technology
CCR	Central Contractor Registration
CCRIS	Chemical Carcinogenesis Research Information System
CDD	Conserved Domain Database
cDNA	Complementary DNA
CEB	Communications Engineering Branch
CgSB	Cognitive Science Branch
ChemIDplus	Chemical Identification File
CHEMM	Chemical Hazard Event Medical Management
CHRIS	Consumer Health Resource Information Service
CHIC	Chickasaw Health Information Center
CIT	Center for Information Technology
CLML	Current List of Medical Literature
CMS	Centers for Medicare and Medicaid Services
COOP	(NIH Pandemic Flu) Continuity of Operations Plan
CORE	Clinical Observations Recording and Encoding
CoreBio	Core Bioinformatics Facility
CPSC	Center for Public Service Communication
CPT	Current Procedural Terminology
CRISP	Computer Retrieval of Information on Scientific Projects
CSB	Computer Science Branch
CSI	Commission on Systemic Interoperability
CSR	Center for Scientific Review
CT	Computer Tomography
CTSA	(NIH Roadmap) Clinical Translational Science Award Centers
CUIs	Concept Unique Identifiers
DAC	Data Access Committees
DART/ETIC	Developmental and Reproductive Toxicology/Environmental Teratology Information
dbEST	Database of Expressed Sequence Tags Center
dbGaP	Database of Genotypes and Phenotypes
dbMHC	Database for the Major Histocompatibility Complex

dbRBC	Database of Red Blood Cells
dbSNP	Database of Single Nucleotide Polymorphism
DDBJ	DNA Data Bank of Japan
DCMS	Data Creation and Maintenance System
DEAS	Division of Extramural Administrative Support
DHHS	Department of Health and Human Services
DIMRC	Disaster Information Management Research Center
DIRLINE	Directory of Information Resources Online
DNA	Deoxyribonucleic Acid
DPR	Digital Preservation Research
DRAGON	Dynamic Resource Allocation in GMPLS Optical Networks
DRESWG	Digital Repository Evaluation and Selection Working Group
DTD	Document Type Definition
DVTS	Digital Video Transport System
EBI	European Bioinformatics Institute
EBP	Evidence-Based Practice
ECHO	European Community Humanitarian Office
Educollab	Educational Collaborators
EEO	Equal Employment Opportunity
EFTS	Electronic Funds Transfer Service
EHR	Electronic Health Record
EMBL	European Molecular Biology Laboratory
EMR	Electronic Medical Record
EMS	Emergency Medical Services
EnHIP	Environmental Health Information Partnership
EnHIOP	Environmental Health Information Outreach Program
EP	Extramural Programs
EPA	Environmental Protection Agency
eRA	Electronic Research Administration
EST	Expressed Sequence Tag
ETIC	Environmental Teratology Information Center
FDA	Food and Drug Administration
FDCC	Federal Desktop Core Configuration
FHA	Federal Health Architecture
FIC	Fogarty International Center
FNLM	Friends of the National Library of Medicine
FTE	Full Time Employee
FTP	File Transfer Protocol
Gbps	Gigabits per Second
GDS	GEO DataSet
GEO	Gene Expression Omnibus (database)
GENSAT	Gene Expression Nervous System Atlas
GENE-TOX	Genetic Toxicology
GHR	Genetics Home Reference
GIS	Geographic Information System
GMAC	Grants Management Advisory Committee
GPS	Global Positioning System
GRMS	Global Record Maintenance System
GSA	General Services Administration
GSS	Genome Survey Sequences
GUI	Graphic User Interface
GWAS	Genome Wide Association Studies
HapMap	Haplotype Map
HAVnet	Haptic Audio Video Network for Education Technology
HBCU	Historically Black Colleges and Universities
HHS	Health and Human Services
HIPAA	Health Insurance Portability and Accounting Act
HITSP	Healthcare Information Technology Standards Panel
HLA	Human Leukocyte Antigen
HL7	Health Level Seven, Inc.
HMD	History of Medicine Division
HSDB	Hazardous Substances Data Bank
HPCC	High Performance Computing and Communications
HPV	Human Papillomavirus
HRSA	Health Resources and Services Administration
HSRProj	Health Services Research Projects
HSRInfo	Health Services Research Information
HSRR	Health Services and Sciences Research Resources
HSTAT	Health Services and Technology Assessment Text
HuGENet	Human Genome Epidemiology Network
I3	Image Indexing Initiative
IAIMS	Integrated Advanced Information Management Systems
ICMJE	International Committee of Medical Journal Editors
ICs	Institutes and Centers (of NIH)
IEB	Information Engineering Branch
IGSTK	Image Guided Surgery Toolkit
IHTSDO	International Health Terminology Standards Development Organisation
IHM	Images from the History of Medicine
ILL	Interlibrary Loan
ILS	Integrated Library System
IMPAC	Information Management Planning Analysis And Coordination
InCHIs	IUPAC International Identifiers
*info*SIDA	*info*Síndrome de Inmunodeficiencia Adquirida (database)
IP	Interactive Publications
IRIS	Integrated Risk Information System
IRMA	Image Retrieval for Medical Applications
IT	Information Technology
ITER	International Toxicity Estimates for Risk
ITK	Insight Toolkit
ITP	Informatics Training Program
IUPAC	International Union of Pure and Applied Chemistry
JDI	Journal Descriptor Indexing
JDMS	Journal Descriptor Maintenance System

JRE	Java Runtime Environment
LactMed	Drugs and Lactation (database)
LAN	Local Area Network
LHC	Lister Hill Center
LHNCBC	Lister Hill National Center for Biomedical Communications
LJI	List of Journals Indexed
LO	Library Operations
LOINC	Logical Observations Identifiers, Names, Codes
LPF	Lost Person Finder
LRP	Long Range Plan (NLM)
LSI	List of Serials Indexed
LSTRC	Literature Selection Technical Review Committee
LVG	Lexical Variant Generator
MARG	Medical Article Records Groundtruth
MARS	Medical Article Records System
MAX	Mid Atlantic Exchange, U. of Maryland
MDoT	MEDLINE Database on Tap
MDT	Multimedia Database Tool
MEDLARS	Medical Literature Analysis and Retrieval System
MEDLINE	MEDLARS Online
MegaBLAST	Basic Local Alignment Search Tool
MEME	Metathesaurus Editing and Maintenance Environment
MEO	Medical Education and Outreach
MeSH	Medical Subject Headings
MHC	Major Histocompatibility Complex
MID	Manuscript Identifiers
MICAD	Molecular Imaging and Contrast Database
MIM	Multilateral Initiative on Malaria
MIMCom	MIM Communications Working Group
MIRS	Medical Information Retrieval System
MLA	Medical Library Association
MLAA	Medical Library Assistance Act
MMDB	Molecular Modeling DataBase
MMS	MEDLARS Management Section
MMTx	MetaMap Technology Transfer
MOR	Medical Ontology Research
MOU	Memorandum of Understanding
MTHSPL	Metathesaurus Structured Product Labels
MTI	Medical Text Indexer
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
NCCS	NIH Consolidated Collocation Site
NCHS	National Center for Health Statistics
NCMHD	National Center for Minority Health and Health Disparities
NCI	National Cancer Institute
NCRR	National Center for Research Resources
NCVHS	National Committee on Vital and Health Statistics
NEI	National Eye Institute
NGI	Next Generation Internet
NHANES	National Health and Nutrition Examination Surveys
NHGRI	National Human Genome Research Institute
NHIN	National Health Information Network
NHLBI	National Heart, Lung, and Blood Institute
NIA	National Institute on Aging
NIAID	National Institute of Allergy and Infectious Diseases
NIBIB	National Institute of Biomedical Imaging and Bioengineering
NICHD	National Institute of Child Health and Human Development
NICHSR	National Information Center on Health Services Research and Health Care Technology
NIDCD	National Institute on Deafness and other Communication Disorders
NIDCR	National Institute of Dental and Craniofacial Research
NIDDK	National Institute of Diabetes, Digestive, and Kidney Diseases
NIEHS	National Institute of Environmental Health Sciences
NIGMS	National Institute of General Medical Sciences
NIH	National Institutes of Health
NIHMS	NIH Manuscript Submission
NIH PI	NIH Pathways to Independence Award
NIMH	National Institute of Mental Health
NINDS	National Institute of Neurological Disorders and Stroke
NIOSH	National Institute for Occupational Safety and Health
NIST	National Institute of Standards and Technology
NLM	National Library of Medicine
NLP	Natural Language Processing System
NN/LM	National Network of Libraries of Medicine
NNMC	National Naval Medical Center
NNO	National Network Office
NOSC	Network Operations and Security Center
NOVA	National Online Volumetric Archive
NRCBL	National Reference Center for Bioethics Literature
NSF	National Science Foundation
NTCC	National Online Training Center and Clearinghouse
OACF	Onsite Alternate Computing Facility
OAM	Office of Administrative Management
OCCS	Office of Computer and Communications Systems

OCHD	Coordinating Committee on Outreach, Consumer Health and Health Disparities
OCPL	Office of Communications and Public Liaison
OCR	Optical Character Recognition
OD	Office of the Director
ODIMRC	Office of the Disaster Information Management Research Center
OERC	Outreach Evaluation Resource Center
ORF	Original Release Format
OHIPD	Office of Health Information Programs Development
OMB	Office of Management and Budget
OMIA	Online Inheritance in Animals (database)
OMIM	Online Mendelian Inheritance in Man (database)
OMSSA	Open Mass Spectrometry Search Algorithm
ONC	Office of the National Coordinator (for Health Information Technology)
OPASI	Office of Portfolio Analysis and Strategic Initiatives
OSA	Optical Society of America
ORWH	Office of Research on Women's Health
OSIRIS	Open Source Independent Review and Interpretation System
PAHO	Pan American Health Organization
PBM	Pharmacy Benefit Manager
PCA	Personal Computer Advisory Committee
PCR	Polymerase Chain Reaction
PDA	Personal Digital Assistant
PDR	Publisher Data Review
PDB	Protein Data Bank
PDF	Portable Document Format
PDL	Personal Digital Library
PHLIP	Public Health Law Information Project
PHII	Public Health Informatics Institute
PHP	Public Health Partners
PHR	Personal Health Record
PHS	Public Health Service
PICO	Patient/Population, Intervention, Comparison, and Outcome
PLAWARe	Programmable Layered Architecture With Artistic Rendering
PMC	PubMed Central
PMCI	PubMed Central International
PMC ID	PubMed Central Identification (number)
PRS	Protocol Registration System
PSD	Public Services Division
PUG	PubChem Power User Gateway
QCIM	Quarterly Cumulative Index Medicus
RCDC	Research Condition and Disease Categorization
RCSB	Research Collaboratory for Structural Bioinformatics
RDMS	Rare Disease Maintenance System
RefSeq	Reference Sequence (database)
REMM	Radiation Event Medical Management
RFA	Request for Applications
RFP	Request for Proposals
RHIN	Refugee Health Information Network
RIDeM	Repository for Informed Decision Making
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
SBIR	Small Business Innovation Research
SCR	(MeSH) Supplemental Chemical Records
SDK	Software Development Kit
SEF	Serials Extract File
SEP	Special Emphasis Panel
SIDA	Swedish International Development Agency
SII	Scalable Information Infrastructure
SIS	Specialized Information Services
SKR	Semantic Knowledge Representation
SMART	Scalable Medical Alert and Response Technology
SNOMEDCT	Systematized Nomenclature of Medicine Clinical Terms
SOAP	Simple Object Oriented Protocol (formerly Simple Object Access Protocol)
SPER	System for the Preservation of Electronic Resources
SPIN	Shared Pathology Informatics Network
SPIRS	Spine Pathology Image Retrieval System
SRA	Short Read Archive
STB	Systems Technology Branch
STTP	Short-Term Trainee Program
STTR	Small Business Technology Transfer Research
STS	Sequence Tagged Site
SVM	Support Vector Machine
TDI	3D Informatics (Group)
TEHIP	Toxicology and Environmental Health Information Program
TERA	Toxicology Excellence for Risk Assessment
TIE	Telemedicine Information Exchange
TIFF	Tagged Image File Format
TILE	Text to Image Linking Engine
TIOP	Toxicology Information Outreach Project
TOXLINE	Toxicology Information Online
TOXNET	Toxicology Data Network
TPA	Third Party Annotation (database)
TREF	Terminology Representation and Exchange Format
TRI	The Toxics Release Inventory
TSA	Transcriptome Shotgun Assembly
TSD	Technical Services Division

TT	Teaching Tool
TTP	Turning the Pages
UKPMC	United Kingdom PubMed Central
UMLS	Unified Medical Language System
UMLSKS	UMLS Knowledge Source Server
UNCFSP	United Negro College Fund Special Programs Corporation
UPS	Uninterrupted Power Supply
VAST	Vector Alignment Search Tool
VHP	Visible Human Project
VM	Virtual Microscope
VPN	Virtual Private Network
VS	Virtual Slides
VTS	Visual Triage Study (NCI)
WAI	WebMARS Assisted Indexing
WashCAS	Washington Area Computer Assisted Surgery
WebMARS	Web-based Medical Article Records System
WebMIRS	Web-based Medical Information Retrieval System
Web-STOC	Web-Services Technology Operations Center
WGS	Whole Genome Shotgun
WIISARD	Wireless Internet Information System for Medical Response in Disasters
WISER	Wireless Information System for Emergency Responders
WSD	Word Sense Disambiguation
WTEC	World Technology Evaluation Center
XML	Extensible Markup Language
XSLT	eXtensible Stylesheet Language Transformations
YEP	(MeSH) Year-end Processing

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: A sampling of front covers of the *NIH Medline*Plus Magazine, produced by the Friends of the National Library of Medicine (FNLM). The magazines feature the research and latest findings of the National Institutes of Health. This free publication, recommended by the Congress, is distributed quarterly to the public via physicians' offices nationwide.