# NATIONAL INSTITUTES OF HEALTH

## National Library of Medicine

### Programs and Services

### Fiscal Year 2009

US Department of Health and Human Services
Public Health Service
Bethesda, Maryland

National Library of Medicine Catalog in Publication

# CONTENTS

**Tables**

# PREFACE

Fiscal Year 2009 was another banner year for the National Library of Medicine. We remain focused on our mission of enabling biomedical research, supporting health care and public health, and promoting healthy behavior. We also remain committed to the goals of our 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 the highlights of this year's *NLM Programs and Services:*

- The American Recovery and Reinvestment Act (ARRA) awarded nearly $84 million to the Library for investment in basic and applied research in biomedical informatics. With this unprecedented allocation, NLM substantially increased the number of research grants and supplemented many existing grants to accelerate the pace of research or undertake new research aims.
- The National Center for Biotechnology Information (NCB1) celebrated its 20th anniversary February 5, 2009. NCBI is an invaluable national and international resource for molecular biology information. It introduced a new user interface this year for MEDLINE/PubMed, which experienced significant growth stemming in part from greater compliance with the NIH Public Access policy. That change resulted in the addition of tens of thousands of new full-text articles from NIH-funded authors. A new class of scientific reports, Rapid Research Notes, was added to the NCBI portfolio, allowing swift communication on the timely topic of influenza.
- MedlinePlus released new social media features this year, including search clouds and the popular medlineplus4you Twitter feed. The Spanish/English *NIH MedlinePlus Salud* magazine was launched, bringing NIH research to Spanish-speaking consumers in the US and abroad.
- ClinicalTrials.gov has grown dramatically in scope and size (over, 80,000 trials in 170 countries listed) since its launch in February 2000. The addition of summary results and adverse event information for trials of drugs and devices is already proving beneficial to the public.
- Standardization of medical terminology and interoperability of records are areas where NLM shines. This year, NLM cosponsored a two-day conference on personal electronic health records. In addition, the Lister Hill Center collaborated on a new Web site, the Newborn Screening Coding and Terminology Guide. This marks an important step toward enabling electronic exchange of laboratory test information and preparing newborn screening information for inclusion in electronic health records.
- The Specialized Information Services Division (SIS) has also stepped into the realm of social media, with a Twitter feed to promote its NLM Pillbox pill identification site and other projects. The Division continues its important work with the Disaster Information Management Research Center (DIMRC).

NLM's information services and research programs serve the nation and the world by supporting scientific discovery, clinical research, education, health care delivery, public health response and the empowerment of people to improve their personal health. The Library is committed to the innovative use of computing and communications to enhance effective public access to the results of biomedical research for the citizens of the world. I extend my thanks to the Library staff and to the many advisors and consultants we rely on.

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

# OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT

*Elliot R. Siegel, 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, analysis, and evaluation 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
- Planning, conducting, and evaluating NLM's international programs.

## Planning and Analysis

The NLM Long Range Plan remains at the heart of NLM's planning and budget activities. Its goals form the basis for NLM operating budgets each year. *Charting a Course for the 21st Century: NLM's Long Range Plan 2006–2016* 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. NLM also undertakes 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 plans, develops, and coordinates NLM outreach and consumer health activities. The Committee is chaired and staffed by OHIPD.

One activity, the Physician Information Prescription Project ("Information Rx"), which was initiated with the American College of Physicians in 2003, continues to grow. In FY2009, NLM continued its collaboration with NLM's Specialized Information Services (SIS) and chapters of 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. A new project with the 43,000-member American Academy of Physicians Assistants was initiated in FY2009 in collaboration with Office of Communications and Public Liaison (OCPL) and the National Network of Libraries of Medicine (NN/LM). 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.

OHIPD staff are also pursuing 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, focused on health topics for high school students and took place in selected US metro areas, including New York City and Houston, Texas. This led to an initiative with ExploreHealthCareers.com and Mentoring In Medicine (MIM), which began in FY2008 and resulted in a remarkable and well-attended workshop convened for parents and children in grades three through college that

emphasized the theme "I can be a health care professional!" This program, which pairs students with more than 500 health care professionals and helps them execute a plan for success, encourages and promotes sustainable interest and participation in health careers for African-American and Hispanic students located at schools in New York City's Harlem and South Bronx. In FY2009, the partnership was expanded to include a new MIM Program, Science and Health Career Exploration, which is funded by NLM and co-sponsored by the Friends of the NLM. This initiative is designed to reach six public and charter schools with an after school program to enrich the high school biology curriculum and encourage enrollment in higher education programs leading to degrees in medicine, allied health professions, and medical librarianship. Principals, science teachers and guidance counselors from participating schools will oversee 40 sessions of biology instruction in 12 organ systems, taught by visiting health professionals and mentors over a two-year period. The participating schools constitute an ideal laboratory in which to experiment with new approaches for stimulating and sustaining student interest in health careers, within the larger context of Science, Technology, Engineering, and Mathematics education in the US. Through a combination of personal intervention and online education resources, the program employs an innovative educational curriculum that also strengthens high school students' readiness to pursue health careers.

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.

*Native American Outreach*

In 2009, OHIPD again participated in the NIH American Indian Pow-Wow Initiative to demonstrate the range of NLM information resources for consumer audiences and to enhance awareness of the resources. This included exhibiting at eight pow-wows, mostly in the Mid-Atlantic area. An estimated 4,000 persons visited the NLM booth over the course of these pow-wows. These activities proved to be another viable way to bring NLM's health information to the attention of segments of the Native American community and the general public. OHIPD also 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 FY2009:

- North Dakota—Cankdeska Cikana Community College (via the Greater Midwest RML), Spirit Lake Nation, Ft. Totten, a continuing project to support improvements at the tribal college library and to develop health information-related educational and outreach activities run by the 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 Fort 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 CDROM 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 and Center for Public Service Communications), one Native Hawaiian Community Health Education Projects was 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 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). The installation of a satellite Internet connection completed the technology upgrade, so that community library now has higher bandwidth service.

*Tribal Consultations on Exhibition Concepts*

In FY2009, OHIPD helped coordinate and facilitate a major consultation activity in the Hawaiian Islands. OHIPD arranged for the NLM Director to conduct interviews with Hawaiian physicians, traditional healers, and community leaders on topics relevant to the planned NLM exhibition on "Native Concepts of Health and Illness." Interviews and related site visits took place at various locations on Oahu and Moloka'i, including a visit to the Hansen's Disease Settlement (former Leper Colony) of Kalaupapa. This was the fourth such major consultation activity, with three others having been conducted in FY2007 in Anchorage, Alaska and Santa Fe, New Mexico, and in FY2009 in Seattle, Washington.

*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 FY2009, 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 Grande 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 FY2009, 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.

## Evaluation

### *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. 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 FY2009, the OCHD continued to pursue an integrated approach intended to encourage exchange of information and learning within NLM, and 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; implementation of trans-NLM Web metrics and a data repository Wiki; and access to Internet audience measurement estimates based on Web usage by user panels organized by a private sector company.

During FY2009, 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.

### *Interactive Publications*

Working in collaboration with the journal publisher, Elsevier, NLM evaluated a prototype implementation of a urology paper presented to students in both traditional and interactive electronic formats. The latter employed both presentational and interactive enhancements. Members of the Student National Medical Association served as test subjects. Evaluation research was also undertaken in collaboration with NLM's Lister Hill Center colleagues

and the Optical Society of America. These projects sought to identify those human factors that facilitate improved learning and understanding attributable to interactive publication formats, and user acceptability of visualization fonts that enable data manipulation in novel ways not attainable by standard print with paper publication.

## International Programs

NLM's international partnerships and projects strengthen and expand global access to the world's health literature. 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 medical students, 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.

Targeting one area of opportunity in 1997, NLM played a critical role in the Multilateral Initiative on Malaria. Since that time, NLM's international outreach efforts have focused in particular on Africa where the need is great. All programs are based on making information available to and integrating information from areas where disease is endemic. Building on 12 years of NLM engagement in Sub-Saharan Africa, these programs also often support, connect with, and sustain one another.

### *Information Communication Technology (ICT)*

For the Multilateral Initiative on Malaria (http://www.mimalaria.org/), NLM led the effort to enhance Internet connectivity and access to medical literature for malaria researchers at 27 sites in 14 African countries more than 10 years ago. More recently, NLM has supported the design and implementation of an electronic health management information system (eHMIS – www.ehmis.net) at Tororo Hospital in Eastern Uganda, the first in the country. The first phase of this project has been the implementation of the system in the hospital and training of staff. The second phase of the project will: 1) connect a health center to the system; and 2) make medical information available at point of care.

### *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 as 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, Zambia, Ethiopia) and African research and clinical communities.

The network currently comprises seven librarians from Kenya, Zambia, Mozambique, Mali, Nigeria, Uganda, and Zimbabwe. Since an initial planning meeting hosted by NLM in May 2007, these librarians 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.

The NLM Associates program and the significant role librarians play in other NLM projects have led to creation of Network of African Medical Librarians and Deans. In FY2009, in response to the librarians' interest in creating a course on information retrieval, NLM, in collaboration with Kenyatta University, held a ground-breaking meeting of medical librarians and deans/provosts (or their representatives) from medical schools where the librarians are based. The objective was to develop a course to be adopted at their respective universities as a permanent part of the medical school curriculum. The deans welcomed the initiative, and participants engaged in a lively discussion regarding strategies for expanding the course to encompass basic computer literacy training and the writing of papers for journal publication (and thereby strengthening medical journals) and finally for integrating the course into the medical school curricula. The Web site http://karibouconnections.net/wordpress/mtg_aug09/ was created by NLM's International Programs office for the meeting. Based on the enthusiasm generated at this meeting, the librarians will collaborate online to produce course modules before meeting again in March in Mozambique to assemble the final version. The course is scheduled to be ready in 2010.

At their respective libraries, these librarians continue to be busy training faculty and students as well as engaging in outreach to areas outside of the capital cities. They have carried out workshops for librarians and researchers from around their countries, 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.

NLM's International Programs office also collaborates with the NLM Associates Program in identifying future associates from Africa as well as in structuring the curriculum to support their work when they return to their home libraries.

*MedlinePlus African Tutorials*

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, field tested, and distributed in FY2008 (www.nlm.nih.gov/medlineplus/africa). The current tutorial under development focuses on mental health and is

led by the dean and team of students from Gulu Medical School, located in the epicenter of the Northern Uganda region which was torn by war for 20 years. The text was written by medical school students and faculty while the illustrations are being created by children from around the region, many victims of the conflict. After field testing, the project will publish a final version as well as a training manual for community use.

**Web Sites**

Three Web sites support the international program activities, described below.

Resources for International Librarians, Health Professionals and Researchers in Developing Countries is a continually updated list of NLM training and courses, document delivery, development manuals, NLM databases http://www.nlm.nih.gov/services/international.html of particular interest, and helpful links to local and national organizations.

Malaria Research Resources (www.nlm.nih.gov/mimcom) supports the activities of MIMCom, a project of the Multilateral Initiative on Malaria and the National Library of Medicine to support African scientists and malaria researchers in their ability to connect with one another and sources of information through full access to the Internet and the resources of the World Wide Web. 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.

MIMCom News, a weekly newsletter started by this office, is now a privately run publication known as MalariaWorld (www.malariaworld.org). Malaria World provides the latest information on malaria every week, reaching over 6,000 subscribers. 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.

*African Medical Journal Editors Partnership Program*

This Partnership Program began by focusing on journals associated with MIM sites in Uganda, Ghana, Mali, and Malawi. Currently, comprises editors of *Mali Medical, Ghana Medical Journal*, *African Health Sciences*, *Malawi Medical Journal*, *Ethiopian Journal of Health Sciences*, and *Medical Journal of Zambia*; editors of *JAMA*, *BMJ*, *Lancet*, *Environmental Health Perspectives*, *AJPH*, *Annals of Internal Medicine*, and *New England Journal of Medicine*; and the Council of Scientific Editors. The project's goal is to strengthen the African journals in order that they are accepted into MEDLINE, making their research available to the world. NLM contributes to technical capacity building, providing site visits by experienced IT experts from Africa and helping to

purchase equipment, including computers, printers, scanners and software.

With the support of the Partnership Project, 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.

African journal editors have 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 have been 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. The Web site for members of the project is: http://karibouconnections.net/wordpress/ajpp.

*Visitors*

In FY2009, the Office of Communications and Public Liaison and the History of Medicine Division's Exhibition Program arranged 272 tours—112 regular daily 1:30 tours and 160 specially-arranged group tours and programs. In FY2009 there were 6,600 visitors from the following 67 countries:

Afghanistan, Angola, Austria, Bahamas, Bangladesh, Belarus, Botswana, Bulgaria, Burma, Canada, Cape Verde, China, Colombia, Costa Rica, Cote d'Ivoire, Croatia, Democratic Republic of Congo, Ecuador, Egypt, El Salvador, Estonia, France, Gaza, Georgia, Germany, Ghana, Haiti, Hong Kong, India, Indonesia, Iraq, Israel, Jamaica, Japan, Kazakhstan, Kenya, Latvia, Libya, Malawi, Mali, Mexico, Montenegro, Mozambique, Namibia, Nepal, Nigeria, Oman, Peru, Republic of South Africa, Romania, Sao Tome & Principe, Saudi Arabia, Slovak Republic, South Korea, Spain, Sudan, Sweden, Switzerland, Syria, Tanzania, Thailand, Trinidad & Tobago, Turkey, Uganda, United Arab Emirates, Ukraine, Vietnam.

NLM's International Programs office also participated again in the training of Fogarty International Center Scholars in database searching and access to full text journal articles.

*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.

# LIBRARY OPERATIONS

*Sheldon Kotzin*
*Associate Director*

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

In addition to its basic services, LO develops and mounts exhibitions about the history of science and medicine. It also produces and manages a traveling exhibition program; creates and promotes education and career resources for K-12 and undergraduate students and educators; carries out an active research program in the history of medicine; works with other NLM program areas to enhance NLM products and services; conducts research related to current operations and services; directs and sponsors training programs for health sciences librarians; and contributes to the development of national health data standards policy and to the production and dissemination of clinical vocabulary standards.

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

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

LO also participates with national and international information standards organizations in the development of standards related to preservation, bibliographic control, collection holdings, vocabulary control, and data exchange. In FY2009 more than 22 standards actions were reviewed.

## Program Planning and Management

Priorities for LO programs are based upon the goals and objectives identified in the NLM Long Range Plan 2006-2016 and, where appropriate, in the NLM Strategic Plan to Reduce Racial and Ethnic Disparities. The current NLM Long Range emphasizes connecting and making the results of research from scientific data to published literature to patient and consumer health information readily available. In the next ten years, NLM's programs and services will become even more central to scientific discovery, treatment, prevention, and to reducing health disparities. In FY2009 work continued on the development of a three to five year LO Strategic Plan which will align closely with the NLM Long Range Plan. Three working groups focused on broad areas of Bibliographic Control; Collections, Preservation and Access to Information; and Workforce for the Future submitted their reports including recommendations for priorities in the Strategic Plan. Early in FY2010 the recommendations will be discussed by LO Senior Staff and the LO Strategic Plan will emerge.

In FY2009, LO continued to review and revise policies, procedures, and services to reflect shifting workloads; to use electronic information to enhance basic operations and services; to work with other NLM program areas to ensure permanent access to electronic information; and to develop a succession plan for key LO positions. LO also focused considerable attention on working with other NLM program areas to meet the Library's expanded responsibility for distribution of standard clinical vocabularies within the UMLS Metathesaurus.

An LO-wide Digital Repository Evaluation and Selection Working Group which includes members from OCCS selected Fedora software for the NLM Digital Repository after extensive evaluation and testing. The Digital Repository Implementation Group was established and a pilot project to digitize cholera pamphlets was completed. 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.

Though many of its efforts are directed toward creating and promoting use of electronic information resources and to supporting NLM's high priority outreach initiatives, LO also devotes substantial resources and attention to the care and handling of NLM's extensive collections of physical library materials and to the space and environment in which staff and patrons work and collections are stored. In FY2009, LO continued to make improvements to conditions in the existing NLM library building and develop strategies for handling the projected growth of the collections until a new facility becomes available. These improvements and strategies are discussed elsewhere in this report.

In FY2009, the LO Administrative Office continued to assist managers, supervisors and staff with the transition to a range of new administrative systems and frequently changing administrative requirements. LO continues to encourage its staff to take advantage of telework arrangements as appropriate. More than 161 LO employees have the ability to work at home.

## Collection Development and Management

NLM's comprehensive collection of biomedical literature is the foundation for many of the Library's services. LO ensures that this collection meets the needs of current and future users by updating NLM's literature selection policy; acquiring and processing relevant literature in all languages and formats; organizing and maintaining the collection to facilitate current use; and preserving it for subsequent generations. At the end of FY2009, the NLM collection contained 2,627,858 volumes and 11,745,887 other physical items, including manuscripts, microforms, pictures, audiovisuals, and electronic media.

## Selection

In FY2009, selectors worked on a number of projects to enhance the NLM collections. Several large gift collections were processed which added many titles not widely held in the US. Gifts were received from the National Institute on Drug Abuse, Food and Drug Administration, New York Academy of Medicine, and NASA as well as others. Significant content in the areas of disaster medicine, Native American health, health policy, health technology assessment, health statistics, environmental health and global health issues were added following an extensive review of Internet sites and e-mail communication with US and international agencies.

## Acquisitions

TSD received and processed 152,277 contemporary physical items (books, serial issues, audiovisuals, electronic media), which is slightly below last year's total. The increase in electronic publishing has still not had a significant effect on the number of physical items that NLM acquires. A net total of 33,582 volumes and 2,449,199 other items (including non-print media, manuscripts and pictures acquired by HMD) were added to the NLM collection. In FY2009, 4,913 licensed and 2,924 free electronic journals were available to NLM users, many through the NIH Library licenses to Elsevier's Science Direct and Wiley InterScience for NIH employees. The number of licensed serial titles increased 16 percent over FY2008.

HMD acquired many splendid early printed books, manuscripts, images, and historical films for the NLM collection in FY2009. Important books acquired included: Paracelsus, *Fur Pestilentz…* (Salzburg,1554) an extremely rare first edition of Paracelsus' work on plague; *Aptekarskaia Taksa ili Otensenka Lekarstv* (Saint Petersburg, 1859) a very rare Russian state pharmacopoeia; and Rudbeck, Olaus. *Disputatio Anatomica, De Circulatione Sanguinis…* (Vasteras, 1652), a first edition in which Rudbeck introduces Harvey's theory of the circulation of the blood into Swedish medical literature.

Additions to the archival and manuscript collections included: the papers of Dr. Michael E. DeBakey, pioneering cardiac surgeon and former Chair and member of the NLM Board of Regents; the speeches and presentations of Elias Zerhouni while he was Director, NIH; Rashid Bashir papers (telemedicine); papers of Nobelist Carleton Gajudsek; and the archives of the American Association for the History of Medicine.

## Preservation and Collection Management

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

In FY2009, LO bound 18,866 volumes, repaired 3,174 items in NLM's onsite repair and conservation laboratory, made 303 preservation copies of motion pictures and videos, and conserved 185 rare items. A new multi-year purchase order for binding preparation was awarded in September.

A multiyear inventory of the serials collection which began in FY2006 continued with the completion of *Index Medicus* and *Index Catalogue* titles plus the currently received, non-indexed titles in English. A total of 59,514 titles have been inventoried since the beginning of the project. A new journal donation program was launched in March. The Web-based system enables donating libraries to check their potential donations against NLM's holdings so that only needed issues will be sent. A total of

1,720 missing volumes or single issues were donated by 403 libraries filling in gaps in the NLM collection.

Dwindling space for growth of the NLM collections continues to require significant resources for shifting of collections and contingency planning in the event that NLM does not receive funding for a new facility. Over the course of 174 years, the NLM has repeatedly outgrown its physical facilities as its collection grows. This unparalleled resource of books, journals, and other materials contains much that does not exist elsewhere and thus, as hospital and research libraries face increasing budget and space constraints, the NLM collection will assume even greater importance to the nation. Many health science libraries are reducing the size of their print collections and relying on the online version of collection materials, depending on NLM to have the copy of record.

In FY2009, work continued on the multiyear project to improve building systems in the stacks and expand existing collection space by strengthening the B2 stack level floor and installing compact shelving. The Limited Cataloging and Z (bibliography) collections were moved offsite to an NIH rental space in Rockville where they will stay until all work on B2 is finished. Work in the northwest corner of B2 to strengthen the floor, install new lights and sprinklers and install compact shelving was completed. A new consolidated area for staff on B2 was also constructed and furnished and staff of the Collection Management, Replacement and PMC Backfile Scanning Quality Assurance Units moved in on September 3.

**Permanent Access to Born Digital and Digitized Information**

The preservation of electronic information presents unique challenges that are not yet fully understood. NLM's general approach to addressing these challenges is to use its own electronic services and publications as test-beds and to work with other national libraries, the National Archives and Records Administration, and other interested organizations to develop, test, and implement strategies and standards for ensuring permanent access to electronic information. LO works closely with other NLM program areas on activities related to the preservation of digital materials.

PubMed Central, a digital archive of medical and life sciences journal literature developed by the National Center for Biotechnology Information (NCBI), is NLM's primary test-bed for the development of procedures and methods for ensuring permanent access to electronic journals. LO continues to assist NCBI in expanding current deposits to PubMed Central by soliciting the participation of additional journals, primarily in the fields of clinical medicine, health policy, health services research, and public health. In addition, the Public Services Division also works closely with NCBI to scan and add the backfiles of journals depositing current issues in the digital archive. At the end of FY2009, more than 1,200,000 articles have been scanned and made available. Complete runs of 90 journals have been processed. Each journal is made available in its entirety, including front and back covers, tables of contents, administrative material such as masthead and editorial boards, and advertisements. Views include a full-text summary, HTML view, separate views of TIFF images, and the full PDF. Ten titles were added to PMC during the year including: *British Journal of Cancer, Bulletin of the World Health Organization, Journal of the National Medical Association, Proceedings of the AMIA Symposium,* and *Yale Journal of Biology and Medicine.*

Work has been underway for several years to develop an NLM Digital Repository to preserve and provide access to digitized and born digital content not included in PubMed Central. Significant progress was made in FY2009 in several areas. After the selection of Fedora and Fez software for the repository and submission of the DRESWG final report, the Digital Repository Implementation Group was formed. Their work has focused on ingest, descriptive metadata, and access to repository collections. The pilot phase will involve four different collections: cholera pamphlets; historical motion pictures; images from historical anatomies; and NIH Institute Annual Reports.

**Vocabulary Development and Standards**

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

NLM is the central coordinating body for clinical terminology standards within the Department of Health and Human Services (HHS). LO, in partnership with LHC and OCCS, represents NLM in federal initiatives to select and promote use of standard clinical vocabularies in electronic health records as well as administrative transactions governed by the Health Insurance Portability and Accountability Act of 1996 (HIPAA). NICHSR staff members routinely contribute to the Healthcare Information Technical Standards Panel (HITSP) technical committees and the Federal Health Architecture (FHA) initiative; participate in the Public Health Data Standards Consortium; serve on the Department of Health and Human Services Data Council; and serve as Staff to the National Committee on Vital and Health Statistics

(NCVHS) Standards Subcommittee. With passage of the Health Information Technology for Economic and Clinical Health (HITECH) Act provisions, enacted as part of the American Recovery and Reinvestment Act of 2009 (ARRA), NLM's activities in this area intensified, particularly in relation to the HITSP work on quality and performance measurement.

In a related effort to promote standardized data collection and reporting on quality of care and improvement activities, in FY2009, NICHSR continued to represent NLM and participate on AHRQ Expert Panel for the National Guidelines and the National Quality Measures Clearinghouses, as well as on the AHRQ Expert Panel for the AHRQ Health Care Innovations Exchange. NICHSR also worked with AHRQ to support efforts to promote patient safety and adverse event reporting.

In FY2009, LO staff (with staff from OCCS and LHC) worked closely with the NLM Deputy Director to prepare background materials for, and participate in, the Board of Regents Working Group on Health Data Standards meeting. This working group was established to review NLM's current activities related to standards for electronic health records, identify opportunities to advance the development and deployment of robust standards, and provide advice and recommendations for NLM's future activities in this area. During FY2009, NLM staff began to put many of the Working Group's recommendations into practice. NLM continues working 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 FY2010 and beyond.

In FY2009, NICHSR staff continued to participate (with staff from LHC) on the MLA/NLM Task Force on electronic Personal Health Records.

## Medical Subject Headings (MeSH)

The 2009 edition of MeSH contains 25,588 main headings, 83 subheadings or qualifiers, and more than 186,686 supplementary records for chemicals and other substances. For the 2009 edition, the MeSH Section added 422 new descriptors, replaced 52 descriptors with more up-to-date terminology, and deleted 20 descriptors. A variety of new Category A subtrees were added for the non-animal organism structures originally treed in Category B, following the creation of a new subtree: B5 – Organism Forms: A 18 Plant Structures; A 19 Fungal Structures; A 20 Bacterial Structures; and A 21 Viral Structures. Category B (Organisms Tree) was reorganized to include specific taxonomic descriptors for the third domain of life (EUKARYOTA). The descriptor PROTOZOA was deleted because it has become an outmoded concept from the phylogenetic point of view.

Under an agreement with the Office of Rare Diseases (ORDR), NIH, the list of rare diseases kept by that office has been joined into the MeSH vocabulary. A total of 274 ORDR terms were merged with existing descriptors in MeSH 2010. This increase is in addition to the 440 ORDR terms that were identified as being already in MeSH 2010 descriptors. Additional terms will be merged with MeSH 2010 as descriptors or Supplementary Concept Records. More than 1,300 ORDR terms were deleted because they were abbreviations of fewer than five letters and therefore would have created ambiguity if included in MeSH.

## Clinical Vocabularies

The MeSH Section and its contractors also produce RxNorm, a clinical drug vocabulary that provides standardized names for use in prescribing medications. RxNorm was designated as a US government-wide target clinical vocabulary standard by the Secretary of Health and Human Services. It represents the information that is typically known when a drug is prescribed, rather than the specific product and packaging details that are available at the time a medication is purchased or administered. RxNorm provides a mechanism for connecting information from different commercial drug information services. In FY2009, RxNorm began weekly releases consisting of new material primarily obtained from the labels sent to DailyMed by the FDA. The US Centers for Medicare and Medicaid Services announced on March 26 that Part D sponsors must utilize the National Library of Medicine RxNorm Concept Unique Identifiers contained on the Formulary Reference File for purposes of 2010 Part D formulary submissions which must be made through the Health Plan Management System.

Through LO's NICHSR, NLM supports the continued development and free distribution of LOINC® (Logical Observation Identifiers Names and Codes) by the Regenstrief Institute. LOINC was designated as a US government-wide target clinical vocabulary standard in 2003, and proposed for adoption as standard under the administrative simplification provision of the Health Insurance Portability and Accountability Act (HIPAA) in a notice of proposed rulemaking issued in September 2005. In 2009 LOINC was recommended as a required standard by the Health IT Standards Committee. In 2008, NLM established a new five-year contract to ensure ongoing distribution and maintenance of this important standard.

In FY2009, 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 as one of a suite of standards for use in US Federal Government systems for the electronic exchange of clinical health information. SNOMED CT was subsequently specified as a required standard in US Interoperability Specifications recently released by the Healthcare Information Technology Standards Panel (established under contract to the Office of the National

Coordinator for Health Information Technology, HHS). In 2009 SNOMED CT was recommended as a required standard by the Health IT Standards Committee. 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.

The existence of authoritative electronic mappings from standard clinical vocabularies to administrative code sets is likely to facilitate automated production of bills and statistical reports as a by-product of the capture of detailed patient data. Mapping efforts must involve both vocabulary producers and intended users, undergo technical review and testing within real clinical systems, and establish effective mechanisms for keeping mappings up-to-date and responding to user feedback. NLM has enlisted cooperation from relevant federal agencies and private organizations and initiated and launched projects to map LOINC to Current Procedural Terminology (CPT) and SNOMED CT to CPT and to the International Classification of Diseases, 9th edition. The first draft LOINC to CPT map was distributed for review via the UMLS Knowledge Sources Server in 2006. An expanded and updated version was published in early 2008.

In FY2009, NLM contracted with the World Health Organization and worked in collaboration with the Social Security Administration, National Center for Health Statistics and the Office of the Secretary, to assure that two code sets, the International Classification for Functioning, Disability and Health (ICF) and the International Classification for Functioning, Disability and Health-Children and Youth Version (ICF-CY), are incorporated in and distributed by the NLM UMLS.

With the enactment of the American Recovery and Reinvestment Act, the US 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 will continue to need to be created. NLM plays a role in coordinating and developing official mappings between standard clinical terminologies and HIPAA code sets. Several mappings are in various stages of development and technical validation.

In September 2004, NLM entered into a three-year contract arrangement with Health Level Seven, Inc. (HL7) to: a) align HL7 message standards with CHI standard vocabularies, specifying which subsets of standard vocabularies are valid for particular message segments and replacing HL7-maintained lists of coded values with subsets of standard vocabularies where feasible (NLM-initiated); and b) create implementation guide(s) for transmitting an entire Electronic Health Record (EHR) between disparate systems (on behalf of HHS). As part of this effort, code sets defined for use in version 2 and 3 of HL7 were added to the UMLS Metathesaurus. The EHR work on behalf of HHS was completed in FY2009.

## UMLS Metathesaurus

The MeSH Section manages the content editing of the UMLS Metathesaurus, using systems developed by the Lister Hill Center (LHC) and maintained by OCCS. At the close of FY2009, the Metathesaurus contained more than 2,181,676 million concepts and 9,840,386 million concept names from 153 source vocabularies in different languages. (See further information about UMLS activities in the Information Products section of this chapter.)

## Bibliographic Control

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

## Cataloging

In general, LO adheres to the *Anglo-American Cataloging Rules, 2nd edition (AACR2)*, when creating cataloging and name authority records that reduce the level of cataloging effort required in other libraries. In FY2009, LO continued to review and provide feedback on drafts of *Resource Description and Access,* the anticipated successor to *AACR2*.

In FY2009, TSD's Cataloging Section cataloged 20,615 contemporary books, serial titles, non-print items,

and cataloging-in-publication galleys. Two online training courses were developed and posted on the NLM Web site: "Using Medical Subject Headings (MeSH) in Cataloging" and "Fundamentals of the NLM Classification." Course content was also made available to MLA and the Program for Cooperative Cataloging for their use in developing in-person training. Based on workflows developed by the Bibliographic Control Working Group for the LO Strategic Plan, several efficiencies were implemented immediately including cessation of technicians performing preliminary searching of names for print monographs and decreasing rev. CIP review from 100 percent to 20 percent.

The Cataloging Section released the 2009 edition of the *NLM Classification* with updates to the W (Health Professions), WZ (History of Medicine) and 19th Century schedules.

HMD cataloged 4,675 monographs, 1,437 linear feet of manuscripts, 2,700 historical audiovisuals, and 31,605 pictures.

## Indexing

LO indexed 5,394 biomedical journals for the MEDLINE/PubMed database to assist users in identifying articles on specific biomedical topics. The indexing workload continues to rise, partly due to the selection of new journals for MEDLINE/PubMed, but primarily due to increases in the number of articles published in journals already being indexed. In FY2009, a combination of in-house staff, contractors, and cooperating international institutions indexed 712,675 articles, a 6 percent increase from last year. Previously indexed citations were updated to reflect 249 retractions, and 33,526 comments found in subsequently published notices or articles.

In FY2009, indexers created 82,932 annotated links between newly indexed MEDLINE citations for articles describing gene function in selected organisms and corresponding gene records in the NCBI Entrez Gene database. This represents a 12 percent increase over the previous year. The Index Section began indexing another 1,450 journals from the electronic rather than the print version, resulting in a grand total of 3,311 titles for FY2009 being indexed from the online version, or 61 percent of all currently indexed journals.

Increases in the number of articles and journals that are indexed for MEDLINE require new approaches to indexing to improve effectiveness and efficiency of the critical and high volume indexing process. LO continues to work with other NLM program areas on the Indexing 2015 initiative to identify, test, and implement ways to reduce or eliminate tasks now performed by human indexers. Progress has been made with improving the performance of the Medical Text Indexer (MTI) through implementation of over 200 suggestions made by Indexers this year.

Indexers perform their work after the initial data entry of citations and abstracts is accomplished by one of two means: electronic submission from publishers (the fastest and most economical method), or scanning and optical character recognition (OCR). Work continues to improve the efficiency of data entry. In FY2009, 88 percent of the citations and abstracts were received electronically from publishers and the remaining 12 percent were scanned. A total of 405 publishers are now supplying XML-tagged electronic data for 5441 journals.

NLM selects journals for indexing with the advice of the Literature Selection Technical Review Committee (LSTRC) (**Appendix 6**), an NIH-chartered committee of outside experts. In FY2009, the Committee reviewed 443 journals. A total of 121 journals were recommended, including 14 titles that were recommended provisionally, as they are available only in electronic form and must send citation data in XML and have a permanent archive.

## Information Products

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

## Databases

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

Use of MEDLINE/PubMed was approximately 1.281 billion searches, a 65 percent increase from last year and the first time searches exceeded 1 billion. Page views in PubMed totaled 3.8 billion. Both numbers were affected by a refinement of the data gathering techniques which no longer count Bot activities, such as Web crawlers. MEDLINE/PubMed now includes more than 19 million citations. Citations from 1948 *Current List of Medical Literature* (CLML) were added to PubMed.

BSD staff began using PubMed Alerts to notify subscribers of major Entrez system problems with PubMed, MeSH, Journals or the NLM Catalog. BSD was also proactive in providing users with previews of the new PubMed interface redesign.

MedlinePlus and MedlinePlus en español continues to receive significant traffic. Over 120 million unique visitors viewed over 650 million pages. There are more than 110,000 subscribers to the weekly announcements of new content. Both sites remain in the top 5 of government sites based on the American Customer Satisfaction Index (ACSI). PSD and OCCS continued to expand and improve the basic content and features of MedlinePlus. A Twitter account, medlineplus4you, was launched to provide health news and information using Web 2.0 technology. Thirty four new English health topic pages were added to MedlinePlus to bring the total to 799; 30 were added to the Spanish site for a total of 751. Services to the military and veterans and their families were expanded through the addition of three new health topics on concussion, traumatic brain injury and Veterans and Military Family Health and a new interactive tutorial on Post-Traumatic Stress Disorder was created.

Seven new Go Local sites were released during FY2009, bringing the total to 34 projects in 30 states and the District of Columbia. These sites provide coverage for 46 percent of the US population. Sites released were Tennessee, District of Columbia, Montana, Kansas, Oregon, Oklahoma and Louisiana. Three new proposals for Go Local sites were approved.

NIHSeniorHealth grew by three topics in FY2009 bringing the total to 43 topics. Topics added were Leukemia, Complementary and Alternative Medicine, and Medicare Basics for Caregivers. The Arthritis topic was split into three topics: Gout, Osteoarthritis and Rheumatoid Arthritis.

The DailyMed Web site provides health information providers and the public with a standard, comprehensive, up-to-date look-up and download resource of medication content and labeling as found in medication package inserts. In FY2009, usage grew to 19 million visits.

NLM's History of Medicine Division updated its Images from the History of Medicine (IHM) database of almost 70,000 images. Users can now view search results in a multi-image display, download high resolution copies of images, zoom in on image details, move images into a user-defined workspace for further manipulation, and create media groups for presenting images and sharing them via email or blog postings.

Under the direction of NICHSR, NLM continues to expand and enhance its Web-based resources and databases for health services researchers and public health professionals. The number of serials on topics related to health services research that are indexed in MEDLINE continued to increase during FY2009. NICHSR worked with NCBI to add Centers for Disease Control and Prevention publication *Health US 2008* to the Entrez Bookshelf. Other additions to HSTAT (Health Services

and Technology Assessment Text) on the Bookshelf included 27 new documents produced by the Agency for Healthcare Research and Quality (AHRQ), primarily evidence reports, technology assessments, and evidence syntheses. NICHSR also worked closely with NCBI throughout the year to improve the design, format and content of HSTAT.

In an effort to increase access to research and other relevant information for health services and public health systems researchers, NICHSR redesigned its portal Web site, HSR Information Central, in response to suggestions from users. These changes included improved navigation features, new topic pages and categories including Key Organizations, Health Disparities, Child Health Services and Health Care Reform.

NICHSR continued to work through AcademyHealth and the Sheps Center at the University of North Carolina, Chapel Hill to expand Health Services Research Projects in Progress (HSRProj) and with the Center for Public Health Systems and Services Research (PHSSR) at the University of Kentucky to expand the HSRR database (Health Services and Sciences Research Resources) to include PHSSR datasets in the set. In FY2009, the University of Kentucky, with NICHSR, completed both a description of the field of PHSSR and a webliography of recent important articles in the field, both of which are expected to be updated regularly.

NICHSR was involved in a number of activities related to examining how the NLM's resources could support increasing interest in "comparative effectiveness research." In one such effort, the HSRProj database and ClinicalTrials.gov database were reviewed to determine the volume of comparative effectiveness research being conducted. The study was designed to inform policy makers and the health policy community about the current state of comparative effectiveness research and to inform the discussion about the potential contributions of comparative effectiveness research to health care quality and costs. It also had the effect of highlighting how the database of projects might itself be used to conduct research. (A similar study using the HSRProj database was conducted later in the year concerning research on health reform topics.)

In a related effort, NICHSR developed and tested "comparative effectiveness research" search strategies that can readily query not just the published literature in PubMed but also ongoing research projects in HSRProj and listed clinical trials, from ClinicalTrials.gov. The beta test version of the searches, released at the AcademyHealth Annual Research Meeting in June 2009, can be found at http://www.nlm.nih.gov/nichsr/cer/cerqueries.html. Users are encouraged to use these special queries and submit feedback to NLM using the link on this page.

## Machine-Readable Data

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

Demand for MEDLINE/PubMed data in XML format continues to increase. There are currently 532 MEDLINE licensees, an increase of 19 from last year. A new License Agreement to Lease NLM Data which covers domestic and non-US licensees of bibliographic data was implemented. Of the 563 licensees, 399 report research use. A relatively small number of organizations lease NLM catalog records or one or more of the SIS toxicological or environmental health files in XML format. At the end of FY2009, there were 4,853 UMLS Metathesaurus licensees, an increase of 17 percent from last year.

## Web and Print Publications

NLM's databases and Web sites are its primary method of publication. Demand for the Library's few remaining print publications continues to decline due to increasing electronic access to NLM information and data throughout the world. LO will cease publishing the printed Medical Subject Headings (MeSH) after 2009.

Publications available from the main Web site include recurring newsletters and bulletins, fact sheets, lesson plans and higher education modules, technical reports, and documentation for NLM databases. TSD published the 2009 edition of the *NLM Classification* with two new related online training courses mentioned elsewhere in this report. The 2009 *List of Serials for Online Users* was made available in February in XML.

## Direct User Services

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

## Document Delivery

LO retrieves materials requested by onsite patrons from NLM's closed stacks and also provides interlibrary loan as a backup to document delivery services available from other libraries and information suppliers. In FY2009, PSD's Collection Access Section processed 443,511

requests for documents from the General Collection. HMD handled 7,676 requests for rare books, manuscripts, pictures, and historical audiovisuals.

The number of onsite requests in the NLM Main Reading Room continues to decline due primarily to online access to more journals in the Reading Room and free access to journals in PubMed Central and other sources on the Internet. Main Reading Room users requested 170,919 items from the stacks, a 20 percent decline from last year. Users of the HMD Reading Room requested 6,901 items from historical and special collections, a 35 percent decrease from last year. New contracts for document delivery services and Reading Room Photocopy and Printing were awarded.

In FY2009, PSD's Collection Access Section processed a total of 271,445 interlibrary loan requests, a slight decline from the previous year. A rise in the fill rate, to a new high of 85 percent, occurred for the fifth consecutive year. The one-day processing rate remained at 98 percent.

A total of 3,018 libraries use DOCLINE, NLM's interlibrary loan request and routing system. NLM released DOCLINE 4.0 which includes simplified serials title searching, modified screen designs to improve the display and editing of serial holdings records, search filtering by acquisition status, and a new function called Show All My Holdings which allows authorized users to display a list of all of their library's holdings for efficient review and update. DOCLINE users entered 1.8 million requests in FY2009, a 5 percent decrease from the previous year; 93 percent of the requests were filled. The Library's share of all DOCLINE requests remained steady at 14 percent of the total. Individuals submitted 370,051 requests to DOCLINE libraries via the Loansome Doc feature in MEDLINE/PubMed and the NLM Gateway, a 14 percent decline from the previous year. Document request traffic continues to decline in all Regions of the NN/LM due to expanded availability of electronic full text journals.

NCBI and staff at the Regional Medical Libraries continued to promote the use of PubMed's LinkOut for Libraries and "Outside Tool" as a means for libraries to customize PubMed to display their electronic and print holdings to their users. The number of libraries participating in LinkOut Outside Tool increased to 646, 22 percent more than in the previous year; 2,225 libraries participate in LinkOut, a 10 percent increase over FY2008.

DOCLINE requests are routed to libraries automatically based on holdings data in its serial holdings database. At the end of FY2009, the holdings database contained 1.5 million holdings statements for 57,949 serial titles held by 3,008 libraries.

NLM and the Regional Medical Libraries continued to encourage network libraries to use the Electronic Funds Transfer System (EFTS), operated for the NN/LM by the University of Connecticut, as a mechanism to reduce administrative costs associated with ILL billing. The EFTS staff continued to work toward making EFTS

self sufficient. Software updates occurred and a new blog was announced that replaces the EFTS Newsletter.

**Reference and Customer Service**

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

In FY2009, the Reference and Web Services Section handled a total of 90,340 customer inquiries from onsite and remote patrons, a 3 percent decrease from the previous year. Both onsite and remote requests declined, continuing the downward trend in onsite requests but showing the first decline in remote requests as well.

**Outreach**

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

**National Network of Libraries of Medicine**

The NN/LM works to provide timely, convenient access to biomedical and health information resources for US health professionals, researchers, educators, and the general public, irrespective of their geographical location. It is the core component of NLM's outreach program and its efforts to reduce health disparities and improve health information literacy. The network includes 5,836 full and affiliate members. The full members are libraries with health sciences collections, primarily in hospitals and academic medical centers. Affiliate members include some small hospitals, public libraries, and community organizations that provide health information services, but may have little or no collection of health sciences literature.

FY2009 was the second year of the NN/LM Emergency Preparedness & Response Initiative. Major accomplishments included a Hospital Librarians Summit meeting in Chicago, and the creation and implementation of a new curriculum for training health sciences librarians in emergency preparedness and response. In addition to the continued enhancement of the toolkit and Web site (http://nnlm.gov/ep), new resources were developed to assist libraries in assessing their readiness and making plans for dealing with the effects of pandemic influenza, and promotional materials were developed to help raise awareness and provide contact information about the emergency preparedness resources available from NN/LM.

The NN/LM conducted a Community Day pilot project designed to assist libraries in becoming active partners in their community's emergency preparedness, response and recovery planning. The lead pilot sites were in three different NN/LM regions: the Sarasota County Public Library System in Sarasota, Florida; the University of Oklahoma Health Sciences Center, Robert M. Bird Health Sciences Library partnering with Oklahoma City's Metropolitan Library System; and the Curtis Library in Brunswick, Maine. Each site received a stipend to plan and implement Community Day events. A Community Day Toolkit of emergency preparedness, response and recovery resources was created for the project.

In addition to the basic NN/LM contracts, NLM funds subcontracts for three national centers that serve the entire network. The activities of one of these centers, the National Training Center and Clearinghouse at the New York Academy of Medicine, are described elsewhere in this chapter. The Outreach Evaluation Resource Center (OERC) at the University of Washington provides training and consulting services throughout the NN/LM and assists NLM, the RMLs, and other network members in designing methods for measuring the effectiveness of overall network programs and individual outreach projects. In FY2009, the Center facilitated the gathering of Network member feedback within each region and coordinated the planning evaluation activities related to the NN/LM Emergency Preparedness & Response Initiative and Community Day Pilot Project.

The Web-Services Technology Operations Center (Web-STOC) provides ongoing technical management of the NN/LM Web sites and also investigates, recommends, and directs the implementation of additional Web technology for teleconferencing, Web broadcasting, distance education, online surveys, etc. In FY2009, in addition to an overall improvement of the search engine for nnlm.gov, a major focus of Web-STOC was to implement an improved fail-over service on nnlm.gov which offers major improvements in management of

interactive files in the NN/LM blogs and the NN/LM Moodle, a distance learning platform used by RML staff.

The RMLs and other network members conduct many special projects to reach underserved health professionals and to improve the public's access to high quality health information. Most of these projects involve partnerships between health sciences libraries and other organizations, including public libraries, public health departments, professional associations, schools, churches, and other community based organizations. Some projects are identified by individual RMLs through regional solicitations or ongoing interactions with regional institutions; others are identified by periodic national solicitations for outreach proposals issued simultaneously in all NN/LM regions. In all, the NN/LM issued 280 subcontracts for outreach projects in FY2009. Many of these projects focused on improving access to information for health professionals, consumers and the public health workforce.

With the assistance of other NN/LM members, the RMLs conduct most of the exhibits and demonstrations of NLM products and services at health professional, consumer health, and general library association meetings around the country. LO organizes the exhibits at the Medical Library Association annual meeting, the American Library Association annual meeting, some of the health professional and library meetings in the Washington, DC area, and some distant meetings focused on health services research, public health, and history of medicine. In FY2009, NLM and NN/LM services were exhibited at 303 national, regional, and state meetings 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, provides unified access to public health information resources produced by all members of the Partnership, as well as other reputable organizations. In FY2009, the Web site was expanded significantly, with almost 560 new links added. Unique visitors to the Web site averaged about 18,768 per month, or more than 225,211 for FY2009, an increase of more than 6 percent over FY2008, indicating an audience that is continuing to grow. Work continues on new topic pages; during FY2009, topics pages on Nutrition, Workforce Development and Dental Public Health were added to the site.

One of the most popular resources already on the site is the Healthy People 2010 Information Access Project (HP2010 IAP). For every focus area of Healthy People 2010, the IAP resource includes four or more objective-specific evidence-based PubMed search strategies and links to MedlinePlus topics. The IAP resource will be expanded to support Healthy People 2020, which was under active development during FY2009.

In FY2009, NLM through Partners, provided first-ever support for Internet remote access by members of the Public Health/Health Administration Section of the Medical Library Association to its meeting at the MLA annual meeting in Hawaii, thereby broadening the number of members who could participate in this section's meeting. In addition, in FY2009 NLM through Partners continued to support the Association of Schools of Public Health, which is conducting a national tour of the NLM traveling exhibition on global health *Against the Odds* to schools of public health; the tour began at American Public Health Association annual meeting in October 2009, where the traveling exhibit was awarded "best exhibit" for the conference.

In FY2009, NICHSR continued working in partnership with CDC, the Health Resources and Services Administration (HRSA), Johns Hopkins University School of Public Health, the Public Health Foundation (PHF), the Association of State and Territorial Health Officials (ASTHO), the National Association of County and City Health Officials (NACCHO), and the National Association of Local Boards of Public Health (NALBOH) to develop and refine the Community Health Status Indicators Project (CHSI). CHSI, which became available to the public in FY2009 at www.communityhealth.hhs.gov, provides Web-based comprehensive data and useful information about the health of all counties in the United States. Librarians have a role in helping disseminate CHSI data specific to their local areas, training public and community health workers in using CHSI and related resources, and in developing local collaborations and partnerships to deliver targeted health information interventions.

In a similar activity, NICHSR participated on an Institute of Medicine panel of experts from largely outside of government that met from June through December 2008 and provided recommendations for key national indicators on health to the State of the USA Initiative (SUSA). The initiative will be releasing a national Web site that will target the information needs of both the general public and their information intermediaries on the intersection of health, the economy, the environment and other domains, sometime in FY2010.

## Special NLM Outreach Initiatives

LO participates in many NLM-wide efforts to expand outreach and services to the general public and to address racial and ethnic disparities and participates actively in the Library's Committee on Outreach, Consumer Health, and Health Disparities. 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 when a candidate could be identified. AFP participants from Kenya, Mali, Malawi, Mozambique and Nigeria have become NLM's "ambassadors" working to provide training and outreach on NLM resources to scientists, students, and more recently, journalists within their countries. Although there was no African librarian in the program in FY2009, an informal assessment of the program from the perspective of former Fellows from Kenya, Mali, Mozambique and Nigeria was conducted via in person interviews at a meeting in Nairobi in July. In addition to hearing that it was a life changing experience for most, many excellent suggestions for making the program more effective for these Fellows were received.

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 FY2009: *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 new partnerships were established: *Ethiopian Journal of Health Sciences* with *Annals of Internal Medicine* and *Medical Journal of Zambia* with the *New England Journal of Medicine*. Of the participating journals, *African Health Sciences*, *Malawi Medical Journal* and *Mali Medical* were accepted for indexing in MEDLINE.

LO staff members continue to be involved in NLM's partnership with the SciMaTech Academy at Wilson High School in the District of Columbia. In FY2009, LO provided summer employment and training opportunities for one student.

## 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, *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 one million young people to treat the common ailments of residents of rural China in the 1960s and 1970s, and it 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 Program has developed a series of traveling exhibitions which continue to be wildly popular. Newer projects are booked in a matter of days and older shows remain active for nearly a decade. The very successful exhibition, *Changing the Face of Medicine: Celebrating America's Women Physicians*, continues to reach the public in a traveling version, funded by the NIH Office of Research on Women's Health and NLM. It began touring libraries in the US through collaboration with the American Library Association in FY 2006. It will visit 60 libraries over a five-year period. A Director's special tour was added to reach 22 additional libraries. *Frankenstein: Penetrating the Secrets of Nature*, which originally toured through the ALA during 2002-2006, continues to draw new visitors as it travels to additional sites. *Against the Odds: Making a Difference in Global Health* and *Opening Doors: Contemporary African American Academic Surgeons* are booked through 2012. *Harry Potter's World: Renaissance, Science, Magic and Medicine* received an overwhelming 138 proposals to the American Library Association's announcement. Twelve will be selected for an ALA-managed tour. An additional tour will be managed by NLM for the remaining 122 applicants. Two new traveling exhibitions were added this year: *Rewriting the Book of Nature: Charles Darwin and the Rise of Evolutionary Theory* and *The Literature of Prescription: Charlotte Perkins Gilman and "The Yellow Wall-Paper"*. An online version of *"The Yellow Wall-Paper"* was created that includes digitized manuscript documents in collaboration with the Schlesinger Library, Harvard University. While *"The Yellow Wall-Paper"* was on display at NLM, a special seminar was held featuring Pulitzer-prize-nominated author Helen Lefkowitz Horowitz, who is currently writing a book on Gilman and her physician, S. Weir Mitchell.

Previous NLM exhibitions also live on through heavily used Web sites, printed catalogs, and DVDs. Exhibition Web sites received more than 11,967,676 million page hits in FY2009.

In addition to the major exhibitions in the rotunda, HMD installs "mini-exhibits" generally in cases near the entrance to the HMD Reading Room. *Physician at the Breakfast Table: Oliver Wendell Holmes as Popular Icon* was mounted in connection with the 200th anniversary of Holmes' birth and featured items from local collector Marilyn Barth.

HMD organized a film series and an onsite exhibition to commemorate the 200th anniversary of Charles Darwin's birth and the 150th anniversary of the publication of *On the Origin of Species*. The exhibition, *Rewriting the Book of Nature: Charles Darwin and the Rise of Evolutionary Theory*, features material from the NLM collections including a first edition of the *Origin*. An accompanying Web site was also released. The film series showcased *Inherit the Wind*, *Dr. Jekyll & Mr. Hyde*, *One Million Years B.C.*, and two television episodes from *The Outer Limits* and *Star Trek*.

New Profiles in Science Web sites were released for medical geneticist Victor McKusick and pioneering heart surgeon Adrian Kantrowitz. Kantrowitz, best known for performing the world's second human heart transplant, also developed bioelectronic devices such as cardiac pacemakers, mechanical left heart devices, and the intraaortic balloon pump. Victor McKusick is widely considered to be the founding father of medical genetics. In 1966, McKusick published the first edition of his classic reference work Mendelian Inheritance in Man, a continually updated, annotated catalog of inherited disease genes (now available as the Online Mendelian Inheritance in Man (OMIM) through the National Library of Medicine). McKusick received the 1997 Lasker Special Achievement Award for his lifelong work in medical genetics. The McKusick project is a collaboration with the Alan Mason Chesney Medical Archives of The Johns Hopkins Medical Institutions, the repository of his papers.

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

**Training and Recruitment of Health Sciences Librarians**

LO develops online training programs in the use of MEDLINE/PubMed and other databases for health sciences librarians and other search intermediaries; oversees the activities of the National Online Training Center and Clearinghouse (NTCC) at the New York Academy of Medicine; directs the NLM Associate Fellowship program for post-Masters librarians; and develops and presents continuing education programs for librarians and others in health services research, public health, the UMLS resources, and other topics. LO also collaborates with the Medical Library Association, the Association of Academic Health Sciences Libraries, and

the Association of Research Libraries to increase the diversity of those entering the profession, to provide leadership development opportunities, to promote multi-institution evaluation of library services, and to encourage specialist roles for health sciences librarians.

In FY2009, the MEDLARS Management Section (MMS) and the NTCC trained 1,089 students in 70 classes covering PubMed, the Gateway, ClinicalTrials.gov, TOXNET, and the UMLS. Four new Quick Tours were implemented: under My NCBI, "Custom Filters" and "Sharing Collections," and under LinkOut for Libraries, "Library Submission Utility" and "Retrieving Your Password for the Library Submission Utility." An average 10,333 unique visitors used the PubMed Tutorial for more than 105,000 page views each month.

The UMLS courses are one of a number of NLM training courses useful in preparing librarians for new and expanded roles. A UMLS Basics online tutorial was released in October and six UMLS Webcasts were presented. NLM and the NN/LM are increasing the use of Webcasts to provide current information on NLM's products and services as a cost effective and timely way to reach larger numbers of users. NICHSR continues to make available its suite of courses on health services research, public health, and health policy.

The NLM Associate Fellowship program had nine participants in FY2009, with four 1st year Fellows on site at NLM and five 2nd year Fellows placed at libraries outside NLM. Field placements for the 2nd year fellows were at University of North Carolina, Chapel Hill; Inova Fairfax Hospital; University of Miami Louis Calder Memorial Library; Vanderbilt University; and the National Library of Medicine. In September 2009, a new group of five first-year Fellows began their year at NLM.

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 scholarships for minority students available through the American Library Association, Medical Library Association, and the Association for Research Libraries (ARL). For the first year, NLM participated in an institute of Museum and Library Services grant to ARL, the Career Enhancement Program, designed to attract minority library school students to careers in research libraries. NLM hosted two students during the summer; one student applied for and was selected as an NLM Associate Fellow for FY2009-2010. LO also supports the NLM/AAHSL Leadership Development Program, which provides leadership training, mentorship, and site visits to the mentor's institution for an annual cohort of five mid-career health sciences librarians.

# Table 1

**Growth of Collections**

| Collection | Previous Total (9/30/08) | Added FY2009 | New Total (9/30/09) |
|---|---|---|---|
| *Book Materials* | | | |
| Monographs: | | | |
| Before 1500 | 595 | 0 | 595 |
| 1501 – 1600 | 6,018 | 7 | 6,025 |
| 1601 – 1700 | 10,310 | 4 | 10,314 |
| 1701 - 1800 | 24,820 | 17 | 24,837 |
| 1801 – 1870 | 41,810 | 31 | 41,841 |
| Americana | 2,341 | 0 | 2,341 |
| 1871 – Present | 805,473 | 14,917 | 820,390 |
| Theses (historical) | 288,091 | 0 | 288,091 |
| Pamphlets | 172,021 | 0 | 172,021 |
| Bound serial volumes | 1,367,563 | 21,910 | 1,389,473 |
| Volumes withdrawn | (124,766) | (3,304) | (128,070) |
| Total volumes | 2,595,276 | 33,582 | 2,627,858 |
| | | | |
| *Non-book Materials* | | | |
| Microforms: | | | |
| Reels of microfilm | 148,891 | 130 | 149,021 |
| Number of microfiche | 456,781 | 0 | 456,781 |
| Total microforms | 605,672 | 130 | 605,802 |
| Audiovisuals | 83,837 | 994 | 84,831 |
| Computer software | 2,560 | 0 | 2,560 |
| Pictures | 69,285 | 0 | 69,285 |
| Manuscripts | 8,536,307 | 2,448,075 | 10,984,382 |
| Non-book items added | 9,297,661 | 2,449,199 | 11,746,860 |
| Non-book items withdrawn | (973) | (0) | (973) |
| Total non-book items | 9,296,688 | 2,449,199 | 11,745,887 |
| | | | |
| Total book & non-book items | 11,890,964 | 2,482,781 | 14,373,745 |

# Table 2

**Acquisition Statistics**

| Acquisitions | FY2007 | FY2008 | FY2009 |
|---|---|---|---|
| Serial titles received | 20,165 | 20,901 | 20,096 |
| Publications processed: | | | |
| Serial pieces | 136,543 | 138,893 | 129,053 |
| Other | 21,905 | 21,739 | 23,224 |
| Total | 158,448 | 160,632 | 152,277 |
| Obligations for: | | | |
| Publications | $10,329,902 | $8,778,048 | $7,271,348 |
| (For rare books) | ($402,153) | ($299,764) | ($294,219) |

# Table 3

**Cataloging Statistics**

|  | FY2007 | FY2008 | FY2009 |
|---|---|---|---|
| Completed Cataloging | 22,520 | 21,507 | 20,615 |

# Table 4

**Bibliographic Services**

| Services | FY2007 | FY2008 | FY2009 |
|---|---|---|---|
| Citations published in MEDLINE | 670,900 | 671,904 | 712,675 |
| Journals indexed for MEDLINE | 5,194 | 5,319 | 5,394 |
| Total items archived in PubMed Central | 1,115,778 | 1,683,664 | 1,869,309 |

# Table 5

**Consumer Web Services**

| Services | FY2007 | FY2008 | FY2009 | |
|---|---|---|---|---|
| NLM Web Home Page | | | | |
| Page Views | 63,600,000 | 54,600,000 | 42,700,000 | |
| Unique Visitors | 9,300,000 | 11,700,000 | 7,900,000 | |
| MedlinePlus | | | | |
| Page Views | 906,000,000 | 754,000,000 | 660,000,000 | |
| Unique Visitors | 120,000,000 | 132,300,000 | 123,000,000 | |
| ClinicalTrials.gov | | | | |
| Page Views | 479,926,975 | 563,956,116 | 785,589,163 | |
| Unique Visitors | 6,188,562 | 10,466,607 | 10,794,930 | |
| DailyMed | | | | |
| Page Views | 9,187,248 | 16,436,561 | 66,547,657 | |
| Unique Visitors | 1,298,775 | 1,795,937 | 2,005,636 | |
| Genetics Home Reference | | | | |
| Page Views | 39,829,732 | 27,252,298 | 36,465,215 | |
| Approximate Unique IPs | 1,906,055 | 2,620,080 | 3,005,955 | |
| Household Products Database | | | | |
| Page Views | 12,457,074 | 15,142,059 | 140,340,909 | * |
| Unique Visitors | 1,143,883 | 1,054,447 | 875,237 | |
| Tox Town | | | | |
| Page Views | 5,134,230 | 5,039,964 | 6,352,651 | |
| Unique Visitors | 170,360 | 224,762 | 300,230 | |

*In 2009, Household Products discovered that they had been hit by a "hacking site" that caused repetitive hits of tens of thousands of hits per day.*

# Table 6

**Circulation Statistics**

| Activity | FY2007 | FY2008 | FY2009 |
|---|---|---|---|
| Requests Received | 530,656 | 488,769 | 442,511 |
|     Interlibrary Loan | 298,616 | 283,591 | 271,592 |
|     Onsite | 232,040 | 205,178 | 170,919 |
| | | | |
| Requests Filled | 442,731 | 405,475 | 375,766 |
|     Interlibrary Loan | 246,902 | 234,020 | 229,577 |
|     Onsite | 195,829 | 171,455 | 146,189 |

# 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 | FY2009 |
|---|---|---|
| Page Views | 4,100,051,973 | 3,792,907,905 |
| Interactive Sessions* | 229,571,548 | 349,191,616 |
| Searches | 775,504,557 | 1,281,180,957 |

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

| GateWay Statistics | FY2007 | FY2008 | FY2009 |
|---|---|---|---|
| Page Views | 4,230,910 | 6,391,604 | 7,794,098 |
| Unique Visits | 312,690 | 709,259 | 539,919 |
| Searches | 746,219 | 1,578,208* | 1,128,932 |

*Corrected figure*

# Table 8

**Reference and Customer Services**

| Activity | FY2007 | FY2008 | FY2009 |
|---|---|---|---|
| Offsite requests | 79,502 | 85,457 | 83,140 |
| Onsite requests | 9,919 | 8,142 | 7,200 |
|     Total | 89,421 | 93,599 | 90,340 |

# Table 9

**Preservation Activities**

| Activity | FY2007 | FY2008 | FY2009 |
|---|---|---|---|
| Volumes bound | 18,684 | 18,583 | 18,866 |
| Volumes microfilmed | 1,564 | 1,509 | Ceased in 2008 |
| Volumes repaired onsite | 2,313 | 2,412 | 3,174 |
| Audiovisuals preserved | 318 | 544 | 268 |
| Historical volumes conserved | 163 | 82 | 185 |

# Table 10

**History of Medicine Activities**

| Activity | FY2007 | FY2008 | FY2009 |
|---|---|---|---|
| Acquisitions: | | | |
| Books | 938 | 583 | 592 |
| Modern manuscripts | 372 (in ft) | 820 (in ft) | 2,776 (in ft) |
| Prints and photographs | 839 | 13,262 | 3,914 |
| Historical audiovisuals | 476 | 4 | 430 |
| Processing: | | | |
| Books cataloged | 3,339 | 4,298 | 4,675 |
| Modern manuscripts cataloged | 1,276 (in ft) | 243 (in ft) | 1,437 (in ft) |
| Pictures cataloged | 1,274 | 34,745 | 31,605 |
| Citations indexed | 856 | 5,134 | Now included in Table 4 |
| Public Services: | | | |
| Reference questions answered | 22,237 | 22,506 | 22,414 |
| Onsite requests filled | 12,588 | 13,183 | 13,439 |

# 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, HIV/AIDS disaster health, and other specialized topics including minority health. SIS includes an Outreach and Special Populations Branch, which seeks to improve access to high quality health information by underserved and other targeted populations. The 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 SIS. Over the years, TEHIP has taken advantage of new computer and communication technologies to provide more rapid and effective access to toxicological and environmental health information 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 20 years, and now includes several collaborative efforts in information resource development and deployment, including a focus on the information needs of other special populations. Our outreach program has continued to evolve and reach out to under-served communities through implementation of innovative information access-enabling approaches and dissemination of NLM's resources. This past year has seen the development of new disaster information resource activities as the new DIMRC has become more established.

The SIS Web site provides a view of the full range of the varied programs, activities, and services of the Division. Although users typically approach through one of the specific entry points for the topic of interest (TEHIP, HIV/AIDS, disaster information, or minority health), the Divisional Web site (http://sis.nlm.nih.gov) includes program descriptions and documentation. 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 FY2009 SIS continued to balance efforts to enhance existing information resources with efforts to provide new services in the growing area of disaster information. Highlights for 2009 include the addition of nanomaterials to HSDB, the release of a beta version of PillBox for the identification of prescription pills, research on disaster information tools such as the use of virtual worlds for training and radios for data transmission, program evaluation studies and updates to many existing products and projects. Two SIS resources – Tox Town and REMM - won awards for their quality.

## Toxicology and Environmental Health Resources

**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 Web resources, and in FY2009 customers continued to express their appreciation for the usefulness of these resources. In FY2009, many 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 in FY2009 with records on chemical compounds of interest in monitoring potential terrorist activities. This year, we celebrated the 75th meeting of the Scientific Review Panel of toxicologists that reviews this database. In preparation from this, we arranged for an outside analysis of HSDB's utility in risk assessment development. This led to a project just now underway for a re-engineering of the HSDB, including its content, creation, and review. Central to this project is the inclusion of new important areas of concern to toxicologists, such as the exposure to nanomaterials. Recent expanded scope areas for HSDB have included radioactive chemicals records and biological toxins.
- **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. These side-by-side comparisons of international risk assessments include links to source documentation.

- **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 legacy 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 3.6 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. This database is no longer funded by the multi-agency group that created it, but it is still searchable as a distinct entity.
- **Toxics Release Inventory** (TRI), a series of databases that describe the releases of toxic chemicals into the environment annually for the 1987-2007 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 388,000 chemical records, of which over 295,000 include chemical structures. ChemIDplus includes some toxicity data as well as locators to many important national and international listings of chemicals.
- **Household Products Database**, which provides information on the potential health effects of chemicals contained in more than 9,000 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.

Finally, we have continued to develop the **Dietary Supplements Labels Database**, enhancing it with additional labels from the many brands available in the marketplace, and working with the Office of Dietary Supplements as they developed their prototype label database intended primarily for researchers. Much of the search interface and design and infrastructure has been shared with the ODS group for their project.

In addition to the core TOXNET databases, SIS supports many other databases and resources:

- **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. Enhancements released in FY2009 include updated mortality data, new features such as widget and toolbar, and a new interface for TOXMAP News.
- *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.
- **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. New information resources on cadmium and on chemicals that may act as endocrine disrupters were added this year. Tox Town received the 2009 MERLOT award in the health sciences category and was described as teaching "important concepts to a wide audience in an interesting fashion." Tox Town was demonstrated at several educational conferences and promoted for use by teachers. Tox Town now features changing highlights on its home page on new resources and special events related to environmental health, such as Earth Day and National Radon Month.
- **ToxSeek** is a meta-search tool that enables simultaneous searching of many different information resources and databases on the World Wide Web. The ToxSeek user interface allows selection of resources from a wide range of authoritative sources in environmental health and toxicology. It provides integrated search results from the selected resources and displays related concepts to use in refining searches. Based on user feedback and focus group evaluations, work has continued on enhancements for future releases.
- **ToxMystery,** an interactive Web site for children between the ages of 7-10, was released at the end of FY2006. It provides an animated game-like interface, which includes finding potential chemical hazards in a home and includes fun sound effects. During FY2009 the ToxMystery Kiosk at the Carnegie Science Center

in Pittsburgh, PA was completed and installed near the hands-on Exploration Station for kids and the Building Green exhibit devoted to environmentally sound construction techniques. The kiosk has an adapted version of ToxMystery with all links to the Internet removed. Approximately 46,000 people played the game between March and September. Reports from the Carnegie Science Center show that most users give the kiosk 3 stars. In addition to the work with the museum, a new ToxMystery About page in Spanish and English was added. The About page includes clip art and colorful desktop wallpaper in three versions.

- **Drug Information Portal**, providing current drug information for over 150,000 drugs with links to many credible additional online resources. During FY2009, changes were made to the search interface to allow easier selection of drug name and searching by category of drug.
- At the end of FY2009, we released **Pillbox**, a new drug information resource that focuses on pill images as well as drug names and other physical characteristics information. IT development included creation of a cataloging and quality control system, as well as engagement of the user community during the development process, in part through the employment of social networking tools. This unique pill identification and image exploration application is also an integral part of the patient-safety initiative with the FDA. This division led the coordination of the NLM/FDA patient safety initiative, and worked to engage the pharmaceutical industry in what has become a cooperative effort between government and private industry.

**The Disaster Information Management Research Center**

In 2008, NLM officially created the Office of the Disaster Information Research Center (DIMRC) in SIS to serve as the coordinating office for disaster health information management activities across NLM, as called for in the NLM Long Range Plan 2006-2016. In FY2009, the DIMRC office focused efforts on identifying and providing access to disaster health information resources and conducting informatics research that will be directly beneficial to public health officials, healthcare providers, special populations, and the public.

*Disaster Information and Outreach*

The collection, organization, and dissemination of disaster health information began in FY2009 with DIMRC staff enhancing the DIMRC Web site with new topic pages, adding over 600 state-level disaster organizations to DIRLINE, and initiating the move of the Resource Guide for Public Health Preparedness (http://phpreparedness.info), a grey literature resource to disaster health information, to NLM.

The Disaster Information Specialist pilot program at four institutions continued at four institutions. Each institution has found unique ways in which librarians can assist emergency personnel from serving as informationists for researchers, health care professionals, and policy makers, to providing new resources and training, and serving on disaster preparedness committees at the local or regional level. DIMRC continued to maintain a listserv, now with over 325 librarians and information specialists interested in the disaster information specialist program and began discussions with the Medical Library Association to consider the development of a certification program in Disaster Health Information.

*Research and Development*

The DIMRC office initiated and coordinated a number of informatics and communications research and development and pilot projects on behalf of the Bethesda Hospitals' Emergency Preparedness Partnership (BHEPP). NLM joined in 2008 to help with common issues at hospitals during emergencies: communication, patient management, family reunification, and information access. NLM received approximately $3.5 million to develop a series of prototypes in these areas. Three NLM divisions, SIS, OCCS, and LHNCBC, are working on these projects, coordinated by the DIMRC office.

DIMRC staff worked on several projects including development of a back-up digital communication system via the Military-Affiliated Radio System, development of a virtual world prototype for training hospital personnel in the nationally mandated Hospital Incident Command System, and the development of a research protocol for testing the electronic triage form and digital pen for mass casualty events.

*Tools for Emergency Responders*

Two new versions of the Wireless Information System for Emergency Responders (WISER), a system designed to assist first responders in hazardous materials incidents, were released this year. The new versions include updated data, an improved interface for the Windows version, the addition of images, and the ability to map the safe protective distance from a chemical.

The Radiation Event Medical Management (REMM) tool is a Web-based portal providing guidance to health care providers about the clinical diagnosis and treatment of radiological/nuclear events. This year, NLM, in collaboration with the HHS Office of the Assistant Secretary for Preparedness and Response (ASPR), continued to improve REMM by releasing Mobile REMM for PDAs, changing the Dose Estimator for Exposure algorithm, adding and editing content, images, and animations.

NLM, again in collaboration with HHS ASPR, began development of a new tool for mass casualty chemical incidents, the Chemical Hazards Event Medical Management (CHEMM). This Web-based portal based on the REMM concept, will provide just-in-time medical management guidance to health professionals, as well as assisting in planning and training.

*Partnerships*

NLM continued to work with numerous agencies and organizations on identifying information needs and providing guidance and assistance, as needed. In addition to HHS ASPR, NLM worked with the NCI, CDC, the NIH Biodefense Research Coordinating Committee, the Institute of Medicine's Forum for Medical and Public Health Preparedness for Catastrophic Events and other agencies.

## AIDS Information Services

NLM is the project manager for the multi-agency AIDS*info* 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), in both English and Spanish continues to be used to evaluate *infoSIDA*. The site has scored an 81 on recent quarterly reports. In FY2009 the contract for support of AIDS*info*/*infoSIDA* was recompeted and an award was made to a new contractor resulting in a complex, but seamless transition of databases, Web sites, telephone numbers, and staff.

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 FY2009, NLM made 13 awards.

## Evaluation Activities

In FY2009, SIS conducted a 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. Set-Aside Program funded activities include six focus groups assessing elementary and middle school teachers' needs in locating and using

Web resources for teaching environmental health. The findings of these focus groups served to inform the development of new Envirohealth Kids Portal (currently under construction). Another Set-Aside funded activity involved developing a detailed evaluation plan for the NLM/UNCFSP-HBCU Access Project to promote disease prevention and wellness at Historically Black Colleges and Universities (HBCU). A non-Set Aside-funded evaluation activity, sponsored by the NLM Office of Planning and Evaluation, included a project developing innovative methodology for assessing performance, usage, usability and user feedback of e-government resources in virtual worlds. This project was conducted in the context of SIS-maintained ToxTown in Second Life. In FY2009, SIS also completed usability studies of four special population Web sites: Asian American Health Web portal, American Indian Health Web site and Refugee Health Information Network (RHIN). 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 FY2009 include:

*Outreach to Middle and High School Science Teachers*

SIS staff conducted outreach to teachers during a number of events attended by middle and high school science teachers. The events included National Association of Science Teachers, Maryland Association of Science Teachers and Montgomery County Back-to-School Fair. During these events, SIS staff delivered presentations and demonstrated SIS resources that could be used in science lessons, including ChemIDplus, Household Products Database, ToxTown and ToxMystery.

*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 FY2009 and awarded four HBCUs small grants to develop and implement projects that help to increase the awareness and utilization of NLM resources on campuses and in their communities. The annual June workshop featured a keynote address by the Honorable Louis Sullivan, former Secretary of HHS and additional presentations by Clinton Davis, actor and educator, Las Vegas, Nevada, and Dr. Keith Norris, Acting President of Charles Drew University and Medicine and Science. This was the eighth year for the two day workshop held at NLM Lister Hill Center. In addition, on the following day, an NLM online database training session was conducted at the University of the

District of Columbia (UDC), Washington, DC, for about 30 conference attendees.

*Adopt-a-School program* with Woodrow Wilson Senior High, Washington, DC, encourages students to take an interest in consumer health and promotes interest in science. In addition to providing summer internships and field trips for Wilson students, this year NLM also had summer interns from Charles Flowers High School in Prince Georges County, MD. This summer, four students and two science/computer science teachers had internships at NLM.

The mission of the *Environmental Health Information Outreach Partnership (EnHIOP)* is to enhance the capacity of minority serving academic institutions to reduce health disparities through the access, use and delivery of environmental health information on their campuses and in their communities. The June 2009 meeting was hosted by the University of Texas at El Paso in El Paso, TX. Featured speakers included Chris Lopez, the executive editor of the El Paso Times and Martha Austin-Garrison, Diné College, who spoke on the Navajo Ethno-Medical Encyclopedia (NEME) she has been developing. Dr. and Mrs. Werner, who were instrumental in the initiation of the NEME were present at the meeting as well. 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 ten different clinics in the facility. The topics were chosen by the directors of those clinics. The CHIC manager has done outreach to senior centers and several of the remote clinics. NLM is considering a proposal for a pilot project to implement a CHIC workstation in one of the clinics.

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. ARHC has been very active in using the members-only site to discuss issues of new population groups and to develop and review new materials which have been added to

RHIN. In addition, they have focused extensively on H1N1 materials for refugees and their health providers.

SIS maintains several *special population Web resources* in collaboration with organizations that provide expertise about the particular population groups.

In FY2009, focus groups with health professionals and consumers were held to test the relevancy of the *Asian American Health Web site* for the targeted populations. Strengths of the site included the materials in Asian languages, ease of navigation, and variety of resources available.

Focus testing of the *American Indian Health Web site* identified the look and feel of the site. One of the new additions to the site, "Our Stories," uses the storytelling tradition to convey personal experiences and health information from American Indians. We anticipate that the "Our Stories" section will prove to be one of the most positive features on the site; users are already pointing to its value as a resource for American Indian populations. Additional stories continue to be collected for this section.

*Women's Health Resources*, a collaborative with the Office of Research of Women's Health, NIH, has grown significantly in the past year. The Coordinating Committee for Research on Women's Health at NIH, which has a representative from each of the NIH Institutes, Centers and Offices, provides input and resources to be added to the site. The portal was started as a two page listing of bibliographic online resources, and has since grown to a 10-page resource with several topic areas. Multimedia resources were added to the site including NIH YouTube videos, podcasts and seminars and symposia back to 2005 from the NIH videocast archives.

Paula Maez, currently a second-year NLM Associate Fellow at the University of Texas Health Science Center at San Antonio, carried out a project on the *Arctic Health Web site*. The purpose of the project was to expand and enhance the health topics section and add resources that address health issues of cold climate and indigenous population in the Arctic region. This section has been greatly enhanced with over 400 new links, including many unique resources from groups that work with indigenous populations.

*Central American Network for Disaster and Health Information (CANDHI):* NLM continued its long-term relationship with the Pan American Health Organization, the Regional Disaster Information Center for Latin America and the Caribbean, and six Central American countries to develop and enhance the Central American Network for Disaster and Health Information. CANDHI consists of centers in Honduras, Nicaragua, El Salvador, Guatemala, and Panama and Costa Rica. The CANDHI centers enable health professionals, government agencies, and others in their countries to access vital information previously unavailable. These libraries have acquired the knowledge, skills, and resources that promote delivery of reliable information. This year we completed projects to improve services or technical capabilities. In addition, several Web-based training sessions were held

and the development of a Disaster Health Information Center Toolkit progressed.

## Research and Development Initiatives

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. The pilot project with the state of Delaware, the Widener University School of Law and the Delaware Academy of Medicine was completed at the end of FY2009. A report is being prepared to enable others who wish to pursue this activity to build upon the accomplishments of the project.

NLM and the National Institute of Diabetes and Digestive and Kidney Diseases are developing a Web resource on drug- and herbal medication-induced liver disease. Tentatively named *LiverTox*, the pilot version was shown at a NIH international symposium of experts that was held at NLM on December 1-2, 2008. Feedback from that meeting has proven useful to the system developers. Medical experts, including participants in NIDDK's Drug-Induced Liver Injury Network, are currently being selected to review and evaluate the current pilot system. In the meantime, further content and information technology development is necessary before public release.

SIS conducted several research and development projects in the domain of building advanced disaster information management tools for summarizing disaster literature and monitoring public sentiment on social networking Web sites. One such project, developed by Dr. Tom Rindflesch's group at the Lister Hill National Center for Biomedical Communications, NLM, involved adapting the Semantic MEDLINE tool for analyzing grey literature on preparedness and response to influenza epidemics. The work included developing disaster influenza terminologies and adapting the Semantic MEDLINE search engine and natural language processing algorithm rules to the new domain. Another project, Semantic Twitter, involved using natural language processing to identify Twitter messages about the emergent H1N1 epidemic and using social science methodology to characterize sentiment displayed in these messages.

SIS also conducted research work identifying barriers to consumers' comprehension of health-related texts, identifying text characteristics that make comprehension easier, developing taxonomy of comprehension errors, and making recommendation regarding how librarians can help health consumers avoid these errors.

# 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 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 LHNCBC 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/.

## 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*

The goal of the NLM Personal Health Record (PHR) project is to help individuals manage the care of their elderly parent(s) and/or young children and themselves. The PHR also serves as a test bed for validating and improving NLM clinical vocabularies, for studying consumers' use of PHR systems, for studying the potential of PHR-based educational reminder systems to improve prevention, and as a possible vehicle for gathering information from patients during clinical trials.

As reported last year, the NLM PHR supports the entry and tracking of key measurements and test results, prescriptions, problems and immunizations. It will produce digital and paper copies of its contents in various formats, and users can get access to MedlinePlus information resources about any prescription drug and medical condition or surgery, that they enter into the system by clicking on an icon adjacent to this the name of this drug, condition or surgery It also attaches codes to the medications, observations, and problems it carries. These codes come from terminologies supported by NLM and designated as national standards by HHS. The automatic inclusion of these codes within the NLM PHR enables both the reminders about preventative care and the automatic downloading of clinical information to the PHR.

The LHNCBC is now working with NLM and others to ensure that policies concerning the PHR are developed and in place before we deploy this system at NIH. However, outside organizations have expressed an interest in deploying the software themselves. A nearby hospital will be the likely first user.

In the last year, we have continued to expand and improve the capabilities of the PHR. We have expanded browser support beyond the original Mozilla browser to Internet Explorer 8, Opera, and Apple Safari, and we are finishing a suite that runs all software tests on each of these browsers running on each of the major operating systems. We have also finished the machinery needed for recovering user IDs and passwords via secret questions and have implemented a friendly mechanism for identifying incomplete or erroneous data entry when the user attempts to store the record.

We are finishing a major revision of the decision rules authoring system. In the first version, authors had to use a syntax that was like a programming language and required programming skill. In our new version, such skill is not needed. Instead, users pick from a series of menus and input fields on a web form. The effort is quite similar to what is required to author a Microsoft Access query. To build a base rule, the author identifies the PHR table of interest, e.g., medications, conditions, allergies, etc., and then identifies the subject and the predicate of the rule. A rule author can identify a subject by name, or class. For

example, if the users picked medications as their table of interest, they could identify the subject by name e.g. Simvasatin oral, or by class, e.g. cholesterol lowering drug, and then add criteria such as status = active, or date started was before January 1, 2009. We could define another rule to determine whether a patient's last LDL cholesterol was greater than 130 mg/dl during the last year. These two rules could be combined into a third rule for reminding PHR record holders to ask their doctor about the use of diet and/or a cholesterol lowering agent if they their LDL cholesterol was elevated, *and* their medication list did *not* contain any active cholesterol lowering agents. Of course, what we have described here is for illustrative purposes only and real rules will have to be more complicated.

We have added a number of other capabilities to the PHR including:

1. A special capability for generating pop-up forms for capturing information from almost any questionnaire or laboratory panel. The only prerequisite is that the information that defines the questionnaire (or lab panel) must be in the PHR's master knowledge database and follow the LOINC form structure. Currently the PHR has access to over 200 such "questionnaires."

2. Special controls to manage user corrections and deletions of information they entered about conditions, prescriptions, or other content during a previous data entry session. For example, to change a previously-entered medication record, the users would have to click on the medication of interest and then choose the desired action, e.g. discontinue, revise, or delete.

3. New tables for identifying medication content by ingredient and a system based on this table to warn PHR users when they enter a new prescription with the same ingredient as previously entered one.

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.

*Use of Surescripts Prescription Data in Direct Patient Care*

Surescripts operates the largest electronic prescribing network in the US It handles over 2 billion prescriptions per year and can provide prescription benefit and history information for 65 percent of all patients nationally. 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 studies the feasibility and value of using prescription history information from Surescripts in direct patient care. This external source of medication information that can be obtained automatically could be both time and life saving in disaster circumstances, when the normal route of obtaining patients' medication history is likely to be disrupted due to over-strained medical staff or special patient circumstances (e.g. unconscious patients).

As part of this project, the developers built a system that connects the Emergency Department (ED) of Suburban Hospital to Surescripts data center. The system: (1) receives ADT messages generated by the HL7 engine of Suburban Hospital; (2) filters in five essential key information about the patient (first and last names, gender, date of birth, and ZIP code); (3) sends them to Surescripts; (4) receives prescription history of the patient if the patient's record is found in Surescripts databases; and (5) maps the information to the patient's record in the project database. All messages that are sent and received are in HL7 format. The system was installed at the data center of Suburban Hospital in the first quarter of 2009 and has been in operation since then.

In order to study the value of this additional source of information, Surescripts data was collected in parallel with the usual manual medication history for three months, during which the Surescripts information was not shared with the medical staff (or integrated to the information flow of the ED for patient care). The study collected information on 10,000 Emergency Department patients. Surescripts was able to match 61 percent of these patients in their database and returned some current prescription information for 45 percent. The conventional medical history identified 67 percent of patients on home medications. 39 percent of patients had drug information from both sources. Six percent of patients had drug information only from Surescripts. Preliminary analysis suggests Surescripts information could significantly improve the completeness of the medical history data.

After the first three months of the study period, the Surescripts data was made routinely available to the medical staff. After the triage nurse enter the demographic information of the patient into the hospital system, a one-year-long prescription history of the patient with all necessary details is automatically printed in the ED in a matter of seconds.

The current challenge is to find the best way of presenting this data in a comprehensive, easily understandable but not overly-detailed format. Seamless integration of this additional source of information with the existing workflow is the key to its acceptability and utilization which will ultimately improve patient care.

In the current information flow plan, the raw Surescripts data is reprocessed, clustered by medication names, sorted by date, associated with dispensed drug duration information (when available), and at the end, a one-year prescription history of the patient is plotted on a timeline graph.

*Vocabulary Developments and Related Collaborations (e.g., with Centers for Medicare and Medicaid Service)*

LHNCBC developed and released RxTerms, a drug interface terminology derived from RxNorm to support efficient and user-friendly prescription data entry in e-prescribing systems, and medication histories within electronic medical records and personal health records. The LHNCBC continues to collaborate with the Centers for Medicare and Medicaid Services (CMS) to provide guidance, content and vocabulary lists (for drugs and for common problems) to help them implement their Continuity Assessment Record Evaluation (CARE) data entry form. CMS is using RxTerms in their demonstration project for CARE. LHNCBC is using this same vocabulary for drugs and problem in the PHR we are developing.

We anticipate getting anonymyzed data about the drugs and problems entered into the CARE form from CMS when they complete their pilot project in the early part of 2010. That data will provide a powerful estimate of the adequacy of the entries and the synonymy of RxTerms and the medical problems we provided them.

*Standards for Identifying Clinical Observations, Forms and Panels*

In FY2009, LHNCBC expanded the LOINC database. In collaboration with NLM, HRSA, CMS, and HHS, LHC refined and moved a code set and HL7 implementation guide for clinical genetic test reporting (LOINC codes for the variable and SNOMED CT codes for the answers) through HL7 ballot approval.

Research continues on the development of an approach to the reporting of clinical genetics studies. This collaborative effort with HTSP, Partners of Boston, Intermountain Healthcare, HL7, and NCBI resulted in the creation of 12 LOINC panels and 100 new LOINC terms and a balloted HL7 implementation guide that describes how to report the results of such studies.

**De-identification Tools**

De-identification is a process that enables research on clinical documents by alleviating ethical concerns about protecting personal health information and privacy. The purpose of the project is to produce high-quality software to convert clinical documents into a redacted form that is compliant with the Privacy Rule of the Health Insurance and Accountability Act of 1996. The provisions of the rule dictate removal of 18 individually identifiable health information elements that could be used to identify the individual, the individual's relatives, employers, or household members.

The project consists of three teams: (1) annotators who mark the corpus and build a gold standard set; (2) system developers who design and implement algorithms; and (3) evaluators who compare outcomes of the algorithmic system against the gold standard and suggest improvements.

We obtained narrative reports for more than 70 thousand individuals (under Institutional Review Board exemptions) to improved de-identification. The record mix includes radiology reports, history and physical exam records, occupational therapy notes, discharge summaries, referrals, consult notes, laboratory data and nursing notes. These records are being annotated with the assistance of a software tool that helps the annotators to tag all HIPAA identifiers in the reports. The tool has reduced the annotation time several fold and helped us to reach our goal of 10,000 fully annotated documents.

The de-identification software system operates through a sequence of pipelined processes: (1) Health Level 7 (HL7) message parsing; (2) part of speech tagging; (3) protected personal health information identification; and (4) redaction. The algorithmic approach aims to replicate the gold standard. The preliminary results suggest that the current system performance is measured as greater than 95 percent and close to 100 percent in both sensitivity and specificity. Current efforts are directed toward conserving all clinically important information while ensuring none of the individually identifiable health information slips through the final redacted version of the narrative.

*EMR Database Research and Development*

As part of our medical record research and development, the LHNCBC developed a general purpose longitudinal database structure and populated it with de-identified electronic medical record data from the MIMIC II data base under a restricted use MOU. This most recent version of this EMR data set carries clinical data for over 30,000 ICU encounters and more than 25,000 patients. It includes 11,626,411 laboratory results and over 400,000 radiology reports and discharge summaries. Altogether, it carries nearly 200 million discrete observations. It now includes data about death within the hospital, many sensitive measures of clinical status including Glasgow coma score and arterial–alveolar difference, vital signs, and enough other information to calculate a reasonable Apache score on the day of admission.

We are now testing the feasibility of using this database for clinical research by exploring published relationships between survival and clinical variables such as the relationship between glucose control and survival. We can almost duplicate the results of the early randomized trial and at least one retrospective trial showing the benefits of tight glucose control. Our ultimate goal is to develop general tools for exploring hypotheses via formal statistical methods applied to de-identified longitudinal data bases. The MIMIC EMR data set provides the complexity and size needed to test and tune the database design and explore machinery for hypothesis generation and testing.

*Newborn Screening Coding and Terminology Guide*

Newborn screening is an important part of public health, but use of test results is complicated by wide variations among states in the ways tests are conducted and results recorded and by inefficient, paper-based communications. The current situation can delay rapid attention to a child's health problems, and it creates frustration and extra work for parents, health care providers, and public health authorities. During FY2009, the LHNCBC designed, created, beta tested and deployed a new Web site, the Newborn Screening Coding and Terminology Guide, as a translator to promote more efficient electronic exchange of newborn screening information.

The Web site is designed to help states move toward the use of common terminology and coding and message standards. The ultimate goal is the delivery of newborn screening results to electronic health records and public health database. Developers worked with several other agencies to organize standard definitions and codes (using NLM standards) for more than 100 newborn screening conditions and the test measurements used to detect them. For all the conditions and tests, the Web site indicates the preferred standard terminologies and codes. Users of the Web site can view the information interactively or download electronic datasets of standard names and identifiers for use in their systems. The Web site also includes draft guidance to help newborn screening laboratories create an HL7 version 2.x message using these codes, and an annotated example HL7 message. Links to Genetics Home Reference and other resources provide additional information about newborn screening.

Standardizing coding, terminology, and electronic messaging methods in newborn screening will support quality health care for children. Moreover, public health agencies will be better equipped to observe and compare nationwide trends from newborn screening test results, which will also support efforts of the biomedical research community at NIH and elsewhere to improve newborn screening methods and evaluation. This large project had many partners including the HHS Office of the National Coordinator (ONC) for Health Information Technology, the Health Resources and Services Administration (HRSA), the Centers for Disease Control and Prevention (CDC), the American College of Medical Genetics (ACMG), the Regenstrief Institute and the Federal Advisory Committee on Heritable Disorders in Newborns and Children.

*Biomedical Imaging and Multimedia*

This research area has several objectives: build advanced imaging tools for biomedical research; investigate design principles for, and develop multimedia image/text databases with particular focus on database organization, indexing and retrieval; develop Content-Based Image Retrieval (CBIR) techniques for automated indexing of medical images by image features.

*Imaging Tools for Biomedical Research*

In FY2009, the Center developed and deployed initial versions of imaging tools to enable research in cervical cancer. In collaboration with oncologists and gynecologists at the National Cancer Institute, the team developed a suite of tools: the *Boundary Marking Tool, Multimedia Database Tool, and Virtual Microscope.* Using these tools, our collaborators are conducting research in: developing new biopsy protocols for pre-cervical cancer screening; visual evaluation of the cervix for pre-cancer detection; analyze expert inter-observer agreement; evaluating non-expert health workers to screen for cryotherapy treatment of HPV-positive women; risk factors for Inflammatory Breast Cancer; Phase II clinical trial of treatment of Kaposi sarcoma by nicotine patch; compare efficacy of glass slide diagnosis versus *virtual microscope* diagnosis for lung cancer. Researchers also developed an *HPV Linear Array tool* for the genotyping of HPV types, an important facet of ongoing cancer research work. Our NCI collaborators and their associates are using these tools in the US and in several countries abroad.

As part of the tool building effort, we have also developed the *Teaching Tool*, an image-based knowledge assessment system, for administering, grading and reporting on examinations for medical professionals. The tool is a Web-browser based application using state-of-the-art software technology, including a PHP interface to MySQL and Postgres databases on the server, and a client that uses AJAX and Javascript for efficient response to the user. The American Society for Colposcopy and Cervical Pathology (ASCCP) is using the Teaching Tool to assess the professional knowledge of colposcopists and experts in cervical cancer. Developers created database content for two specific exams of the ASCCP, developed program logic to handle their particular requirements, and created a publicly-accessible version of the Teaching Tool to support two beta tests. These extensive beta tests promise the routine and nationwide use of the tool by the ASCCP. In September 2009, the Board of Scientific Counselors favorably reviewed all the tools developed in this project.

*Interactive Publications Research (IPR)*

This project demonstrates a type of highly interactive multimedia document that serves 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 media objects relevant to the biomedical literature: text, video, audio, bitmapped images, interactive tables and graphs, and clinical DICOM images such as x-rays, CT, MRI, and ultrasound.

In FY2009, researchers developed the Panorama tool for viewing and analyzing video, DICOM clinical images, tables, graphs and animations. Panorama, written in Java, was one of nine semi-finalists out of 70 entrants in Elsevier's Grand Challenge contest.

Ongoing work is in developing the Java-based Forge as an authoring tool for IP. Forge allows authors to develop interactive articles using a set of wizards, while permitting them to write content in their original word processor; developing the capability to allow open community support to add viewers for new media types (e.g., electrocardiograms and other waveform datasets) in an open software plug-in framework; and enhancing Panorama to support a virtual file system layer to >hide= the physical file storage location and format of a self-contained IP instance. This last feature will eliminate the need for time-consuming FTP downloads before readers can view and use a large IP.

Developers from the NLM Office of Computer and Communications Systems (OCCS) are working on another interactive publications project with the Optical Society of America (OSA). The goal is to produce four interactive issues of regularly published OSA journals and to provide a free interactive publication reader similar to the free Adobe PDF Reader. To date, OSA has published two Medline indexed interactive publications on the web. These first two articles contain 24 papers and 287 linked available interactive datasets. The next two planned issues are under development and planned for publication by February 2010. A one-year assessment of readers, focusing on the added value versus the work it takes to have access to the value, is now in progress. The preliminary results are very encouraging.

*Interactive Science Publication*

Data associated with medical research and clinical trials is often sequestered and not available to the greater scientific community for analysis and review. Moreover, convenient tools for viewing and analyzing data are not generally available, source data is often only indirectly linked to journal articles, and digital data is published in non-standard formats with limited ability to search metadata characterizing the associated measurements. NLM is partnering with the Optical Society of America to develop an interactive software and database infrastructure that enables viewing and analysis of curated, supplemental biomedical source data published in conjunction with peer-reviewed manuscripts; to evaluate the educational value of such an infrastructure; and to explore the problems of archiving this medium. In order to accomplish these goals, OSA is publishing four electronic special issues of OSA journals on research topics which lend themselves to interactive publishing. Articles published in these special issues are peer reviewed and fully citable as OSA journal publications indexed in MEDLINE. They will be published on the Web in Acrobat format (PDF) with links to source data, videos, and other media objects. The links allow users to quickly and conveniently download these objects and visualize them using interactive viewing software designed to look like an Acrobat plug-in. The viewing software is freely available as a download for all computer platforms. The journal articles and data sets are open access and the source data and associated metadata are searchable and accessible from outside the publication. The publications are being followed by evaluations of the usability and educational value of this form of publication. The project will also serve as an NLM test bed for archiving this new form of publication.

The first special issue, a survey issue on various biomedical topics of interest involving optics, was published in October 2008, http://www.opticsinfobase.org/oe/virtual_issue.cfm?vid=68. The second special issue focusing on *Optical Coherence Tomography in Ophthalmology* was published in March 2009, http://www.opticsinfobase.org/oe/virtual_issue.cfm?vid=81. The third special issue focusing on *Digital Holography and 3-D Imaging* was published in October 2009, http://www.opticsinfobase.org/ao/virtual_issue.cfm?vid=94. The fourth special issue focusing on *Image Cancer Detection* will be published in March 2010.

The project includes three independent evaluations: a usability evaluation of the interactive interface; a controlled evaluation of the value of the interactivity; and a user evaluation of the perceived value. Dr. Ben Shneiderman of the University of Maryland has completed the evaluation of the user interface. Dr. James L. Mulshine of the Rush University Medical Center is conducting the controlled evaluation. Foresee, Inc. of Ann Arbor, Michigan, with guidance from Dr. Fred Wood and Dr. Elliot Siegel of NLM, is conducting the user evaluation over the Web.

*Computational Photography Project for Pill Identification*

In a national effort to promote patient safety, OHPCC plans to create an authoritative, comprehensive, public digital image inventory of the nation's commercial prescription solid dose medications. Heretofore, we have conducted most of this work in partnership with the NLM Specialized Information Systems Division and the US Veterans Administration to study content-based retrieval methods for medical image databases. Researchers have 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. In FY2009, researchers started a new project, *Computational Photography Project for Pill Identification*, funded under the American Reinvestment and Recovery Act, to create a national collection of digital photographs of prescription tablets and capsules, creating high resolution digital photographs of the front (obverse) and back (reverse) surfaces of pharmacy samples, confirming that the images match the description of the medication, developing and matching the images of the samples to relevant metadata (including size descriptions, dimensions, color, and the provenance of the sample). We believe that there are between 5,000 and 32,000 prescription solid-dose oral medications to be cataloged through this process.

**Multimedia Database R&D**

The Virtual Microscope (VM) and Virtual Slides (VS) are an archive of virtual slides developed from the teaching set of glass slides from the Department of Pathology of the Uniformed Services University 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. Recent additions include annotations and automatic linking to MEDLINE/PubMed. A related collection of images from the Armed Forces Institute of Pathology (AFIP) fascicles allows users to search images and automatically link to MEDLINE citations. Viewing of VS with mobile devices is a recent application. Collaboration with the Massachusetts General Hospital explores its use in distance education and training. We are developing other collaborations to study its use in telemedicine and teleconsultation.

*Virtual Microscope (VM) and Virtual Slides*

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 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 has been enabled. A related collection of images from the AFIP fascicles allows users to search images and automatically link to MEDLINE citations. Viewing of VS with mobile devices is a recent application. A recent collaboration with the Massachusetts General Hospital explores its use in distance education and training. Other collaborations will study its use in telemedicine and teleconsultation.

*Content Based Image Retrieval (CBIR)*

Our research in Content-Based Image Retrieval has several objectives. One is to develop techniques to augment our existing cancer research tools with automated processes. For example, the CervigramFinder automatically indexes and allows retrieval of cervigrams (and spinal x-rays) using shape, color and texture features. This system contains the key elements needed to augment the Boundary Marking Tool with an automated assist for the user in marking boundaries of regions of medical significance. The Cervigram Segmentation Tool combines several image segmentation and shape similarity algorithms for research into selecting the optimal algorithms that enable automated segmentation for image indexing. MOSES is a tool provided as a service for evaluation of segmentations by multiple experts, to reduce

the inter- and intra-observer variability in marking boundaries of significant regions in cervigrams.

In addition, a hybrid system (SPIRS/IRMA) has been developed by linking geographically separated CBIR systems to exploit the characteristics of each. SPIRS/IRMA allows retrieving detailed medical image data in large collections of heterogeneous images of varying modalities, presentations and anatomy. For example, a user may narrow a search to *spinal x-rays* (using IRMA at University of Aachen) and then retrieve specific x-rays containing *osteophytes* at the SPIRS system at NLM.

CBIR is also a component of a process for indexing illustrations in medical journals by using image features, figure captions and in-document text mentions. This research is aimed at enriching the user experience of searching for relevant documents by including the contents of medical images, photographs, graphs and other illustrations found in articles. Techniques developed in this work were evaluated in the international ImageCLEF competition and found to be successful.

Another avenue explored in this research area is distributed computing and use of GPUs for compute-intensive CBIR tasks, with a particular focus on image segmentation.

*The Visible Human Project*

The Visible Human Project image datasets 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 datasets are available through a free license agreement with the NLM. They are distributed to licensees over the Internet at no cost; and on DAT tape for a duplication fee. More than 3,000 licensees in 58 countries are applying the datasets to a wide range of educational, diagnostic, treatment planning, virtual reality, and virtual surgeries, in addition to artistic, mathematical, legal, and industrial uses. The Visible Human Project has been featured in more than 900 newspaper articles, news and science magazines, and radio and television programs worldwide.

FY2009 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.

While designing a database of the parameters and variances defining the normal range of human anatomical structures and the dependencies and covariances between them, an attempt was made to glean the needed statistical data on bone size variation from the existing anatomical literature. Over 1,000 references were scanned. The literature was found to contain a description and images of

the variation in shape that could be expected of each bone, but not the mathematical description of it variances.

In FY2010, the LHNCBC will hold a cross-disciplinary workshop made up of key representatives of the disciplines that are relevant to bone size and shape in an effort to find a source for the needed statistical data. These disciplines will include forensic pathology and other forensic sciences, paleontology, radiology, orthopedics, anatomists, prosthesis designers, clothing designers and furniture designers. By inviting the participation of the interested community combined with using Web tools developed under the Interactive Publication Project, developers expect to collect the needed data.

*3D Informatics*

In FY2009, the 3D Informatics Program (TDI) expanded research efforts concerning problems encountered in the world of three-dimensional and higher-dimensional, time-varying imaging.

The LHNCBC provides continuing support for 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. The archive contains such integrated and multimodal data as virtual colonoscopy matched with recorded video from endoscopic interventions, time-varying 3D cardiac motion, and 4D MRI of a human hand. During 2009, the TDI group installed the necessary software and hardware infrastructure, including a Linux server and a MIDAS software system, to support interactive scientific publication at NLM.

Beyond data collection, we continue to analyze and develop software tools for studying biological problems such as tumor micro-environments and the transmission of HIV/AIDS (from macrophage to T-cell) through 3D optical and electron microscopy. This work combines high performance computing with life sciences research, accelerating and empowering investigators in the detection and prevention of cancer and infectious diseases. Members of the TDI group contributed to research in HIV virus identification in dual-beam electron microscopy of dendritic and macrophage cells. Other work included geometric analysis of mitochondrion in dual-beam data taken of metastatic cancer cells in the human liver. This work was conducted in support of the National Cancer Institute Laboratory for Cell Biology.

Collaborative work in telemedicine has increased the demand for large display technologies, but the cognitive pathways for understanding how the human visual system processes large-scale digital displays are not well understood. Throughout 2009, researchers continued to develop high-end rendering systems for large displays, incorporating multiple GPUs in volume-rendering applications capable of rendering the full VHP Male

dataset at real time rates. We began investigations in the question of presence and visualization, combining forces with the NCRR-funded Biotechnology Resource Center at the Utah Scientific Computing and Imaging (SCI) Institute and their access to the Magnetic Encepholography (MEG) equipment in the University of Utah Department of Radiology. We also began a study of the difference between human-scale versus non-human scale perception, seeking changes in cortical processing of reading and eye-tracking from large wall displays versus hand-held book-sized displays. The result of such a study could have a far-reaching impact on how new digital user interfaces are developed in future generations.

The Center also began a study of the use of rapid prototyping technologies, to develop witness objects or phantoms, reference models as a public engineering and scientific standard for research in image-based computer-aided diagnosis. We have characterized the x-ray attenuation characteristics of some of the 3D-printing materials available at NIH and are presently evaluating the use of contrast agents as printing materials to vary the appearance of the 3D models when viewed with x-ray CT scanners. The goal is to create complex, anatomically-accurate models to test diagnostics systems and evaluate and compare their performance under known conditions. This work is conducted in partnership with the National Institute of Allergy and Infectious Diseases.

*Insight Tool Kit (ITK)*

The Insight Toolkit is approaching its tenth year of development with an official software release of ITK 3.16 in September 2009. Over 845,000 lines of openly available source code comprise ITK, making available a variety of image processing algorithms for computing segmentation and registration of high dimensional medical data on a variety of hardware platforms. ITK can be run on Windows, Macintosh, and Linux platforms, reaching across a broad scientific community that spans over 40 countries and more than 1500 active subscribers to the global software list-serve. A consortium of university and commercial groups, including OHPCC intramural research staff, provide support, development, and maintenance of the software.

ITK is an essential part of the software infrastructure of many projects across and beyond the NIH. The Harvard-led 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. 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.

International software packages that incorporate ITK include *Osirix*, an open-source diagnostic radiological image viewing system available from a research partnership between UCLA and the University of Geneva and the Orfeo Toolbox (OTB) from the Centre Nationale D'Etudes Spatiales, the French National Space Administration.

From 2002 to 2008, the ITK effort has attempted to stimulate scientific development through supplementing 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. In 2009, OHPCC and the ITK project began participation in American Reinvestment and Recovery Act economic stimulus efforts. OHPCC is developing a new acquisition plan for a software revision and development program to upgrade ITK for emerging computational platforms for the upcoming decade. The development of a major new version of ITK will help to continue our international leadership role in medical imaging research.

*Image and Text Indexing for Clinical Decision Support and Education*

As part of the Clinical Information Systems effort, the Image and Text Indexing project seeks to automatically identify illustrations in biomedical articles that could provide multi-media assistance to clinical decision making. We developed an experimental search engine that illustrates retrieval results with images extracted from scientific articles. The retrieval methods implemented in the search engine achieved top performance in the ImageCLEF 2009 case-based retrieval task. This task reflects real-life clinical situations in which a clinician is looking for publications related to a specific patient's case. The quality of our retrieval results in this task allows for evaluation in a teaching hospital that is planned for the next year.

Automated mapping of text from a small teaching collection of electrocardiogram images to UMLS concepts was evaluated for both accuracy and utility in indexing the image. Although a substantial proportion of the mappings were judged as matches, a minority were judged as useful for indexing the image. Most images were judged to need additional descriptors beyond those found in the accompanying text. The additional descriptors suggested were also highly likely to map to an existing UMLS concept or combination of concepts. We propose that sample medical image collections automatically indexed and reviewed by experts might provide a template for more precise automated indexing. The manuscript has been submitted to *JAMIA*.

Observations in published clinical case reports were matched by hand to entries in the Logical Observations Identifiers Names and Codes (LOINC) in a preliminary analysis of the feasibility of automated natural language processing. Coverage of history, physical, laboratory, and imaging findings was nearly complete, and categorization of terms assigned suggests a strategy for automatic coding. This study was presented as a poster (*Proc AMIA Symp*, 2009, 1036).

Researchers plan to further study the influence of augmenting bibliographic references retrieved from a database search with images; new ways to combine text and image retrieval results; 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*

As part of the Clinical Information Systems effort, the **InfoBot** project enables a clinical institution to automatically augment a patient=s emergency medical record (EMR) with pertinent information from NLM and other resources. The InfoBot software runs 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 FY2009, we developed two prototypes and made them available to collaborators from four departments at the NIH Clinical Center (CC). The NLM InfoBot server has handled over 2000 CC requests since the system went live.

The first prototype, a table driven, asynchronous InfoBot system, was evaluated by clinicians at the Clinical Center, and results presented at the 16th National Evidence-Based Practice Conference. This was followed by the development of a second (Web-based) prototype capable of real-time processing of clinical text, generation of database and search-engine queries, summarization and display of the result. This prototype was integrated with CRIS, and made available through the Evidence-Based Practice tab in CRIS July 2009. In FY2010, we plan to complete the evaluation of this prototype at the Clinical Center, implement a module for automatic creation of rulesets by users, replace/expand single-document summarization of retrieval results with multi-document summarization, improve automatic processing of clinical notes, and improve automatic query generation.

*Turning The Pages*

This project allows users to turn and view page images in a photorealistic manner on touch-sensitive monitors in kiosks, as well as "click and turn" in an online version. The goal of the project is to provide the lay public a compelling experience of historically significant and normally inaccessible books in medicine and the life sciences.

In FY2009, we released the kiosk version of the *Edwin Smith Papyrus*, having first created a 3D model for

flat scrolls. In addition, we released both kiosk and online versions of Brunschwig's *Liber de Arte Distillandi*, printed in Strasbourg in 1512, a practical manual on chemical, alchemical, and distillation devices and techniques used to manufacture drug therapies. For offsite demonstrations, we developed standalone projector versions of all eight TTP books. New work includes the conversion of a Japanese medical manuscript by Hanaoka to TTP format.

*Biomedical Image Transmission via Advanced Networks (BITA)*

Research was conducted in the performance of NLM-supported networks such as the Interactive Video Outreach and Distance Learning Network for Minority High School Students, the BHEPP network, 802.11a, 802.11b and higher designation wireless network implementations, and networks exhibiting narrow bandwidths, high latency and high jitter.

We evaluated image quality, usability, and speed performance for multiple image sources, including large image file transfers and high definition videoconferencing over Internet2, through formal demonstrations in collaboration with universities and research centers nationally and internationally.

As part of this effort, research staff represented NLM and NIH at an Internet2 Advisory Council, Joint Engineering Team, Large Scale Networking Team, CSIA (Cybersecurity and Information Assurance) and other forums for high performance/speed networks.

**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 goal of this project is to develop and operate systems to efficiently extract bibliographic data from scanned and Web journal articles to populate MEDLINE. For paper-based journals, the system delivers bibliographic data by combining document scanning, optical character recognition (OCR), and rule-based and 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.

In FY2009, all engineering support for the offsite MARS production facility was provided: installation of upgraded modules, testing, maintenance and operation of all hardware and software for servers, clients and networks, and the necessary system administration. A significant addition was the incorporation of Unicode in the Edit and Reconcile modules that enable operators to view and enter special characters, diacritics and foreign language text in a WYSIWYG manner (exactly as seen in the published article.)

As a key element in helping to achieve goals of NLM's Indexing 2015 Initiative, the Publisher Data Review (PDR) system was developed and currently undergoing field testing. This system is designed to 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. Currently, we are designing a machine learning method for extracting Investigator Names, a labor-intensive effort if done manually since there could be hundreds of such names listed in an article.

Both MARS and PDR rely on the underlying research in image analysis and lexical analysis (within the Analysis of Images for Data Extraction, or AIDE, project), but this research also enables the creation of new initiatives in which these techniques could find application. Examples are the ACORN initiative and MARG database described below.

*ACORN*

The ACORN initiative (*Automatically Creating OldMedline Records for NLM*) 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. In FY2009, we continued our investigation of scanning, image enhancement, OCR, image analysis, pattern matching, and related techniques to extract unique records from the printed indexes. Researchers designed a prototype consisting of three main components: Quality Control, Processing, and Reconcile. The Quality Control module is completed, and work is proceeding toward implementation of the Processing module which will group OCRed text into records for operators to verify using the Reconcile component. In FY2010, a prototype will be delivered to Library Operations as a pilot system.

*Medical Article Records Groundtruth*

As part of the AIDE project, a validated test set for document image analysis has been created for the computer science and informatics communities for research into advanced algorithms for data mining. This Medical Article Records Groundtruth (MARG) database has attracted more than 17,000 visits from 99 countries. It consists of images of journal articles, the corresponding OCR data, zones, labels and verified data obtained from the routine operation of the MARS production system.

*Indexing Initiative*

The Indexing Initiative (II) 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 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 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 development focuses on testing and updating recently added functionality to the 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. In addition a recent effort consisted of successfully modifying MTI for NLM Cataloging; the cataloging version of MTI is now in regular use. System-related objectives for II include completing the final testing of the migration of our systems to a Linux environment.

*Digital Preservation Research (DPR)*

This project 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. Researchers focus on the preservation of digitized documents, and have built and deployed a first version of a prototype System for Preservation of Electronic Resources (SPER). SPER builds on open source systems while incorporating inhouse-developed modules that implement key preservation functions: ingesting, automated metadata extraction and file migration. NLM curators have started using SPER to preserve more than 60,000 court documents from the historic medico-legal FDA collection. In FY2009, approximately 10,000 documents were processed and made Web-accessible to the public.

Research continues into techniques for digital preservation with the goal of improving SPER while processing the remaining FDA documents, as well as investigating ways to accommodate other NLM collections: (a) improve accuracy of the automated metadata extraction (AME) system by enhancing the layout classification model and the pattern recognition module; (b) examine other Support Vector Machine implementations for use in SPER; (c) enhance the extraction model to encrypt metadata patterns and search/extraction rules for each document layout for the FDA collection; and (d) improve the pattern search engine to identify and extract metadata from a broader set of historic documents using this model.

## 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. At the end of FY2009, the site had nearly 80,000 protocol records, over 800 of which display summary results, sponsored by the US Federal government, pharmaceutical industry, academic and international organizations from all 50 States and in 171 countries. Some 36 percent of the trials listed are open to recruitment, and the remaining 64 percent are

closed to recruitment or completed. ClinicalTrials.gov receives over 40 million page views per month and hosts approximately 900,000 unique visitors per month. Data are submitted by over 6,300 study sponsors through a Web-based Protocol Registration System, which allows providers to maintain and validate information about their studies, including summary results data.

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 FY2009 to implement additional 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. Thus, the ClinicalTrials.gov registry was expanded in November 2007 to include additional information about applicable drug and device clinical trials. New links from registration records to the FDA Website (e.g., Drugs@FDA) were added and existing links to NLM's MEDLINE and DailyMed Web sites were enhanced. As a consequence of this law, new registrations in FY2009 have risen to an average of 330 per week, a 32 percent increase from FY2007, the year prior to enactment of the law (average 250 per week). In September 2008, ClinicalTrials.gov launched the "basic results" database, which complements the registry. It includes tables of summary results data on primary and secondary outcomes of registered trials, as well as information on the patient populations studied. In its first year of operation, over 1,400 results records were submitted by 300 study sponsors. The average number of submissions has increased, ending FY2009 with an average of 50 results records submitted per week. To help study sponsors with submitting data to the new results database, a number of documents were developed, including "helpful hints," a summary of common errors, and a guide to results data entry. While an optional module for reporting serious and frequent adverse events observed in a clinical trial was made available in September 2008 for public use and testing, a modified version of the module became mandatory by law on September 27, 2009. The modifications made to the adverse event module were based on feedback from stakeholders and consultation with risk communication experts. As called for by law, NIH held a public meeting on issues related to the expansion of the clinical trial registry and results database on April 20, 2009. In addition to comments presented orally at the meeting, over 60 written comments regarding issues related to the expansion have been submitted to the public docket. 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 the proposed 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, 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 FY2009 in educating the public and the regulated community 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 17,000 new registrations were received over FY2009. ClinicalTrials.gov continues to collaborate with other registries, professional organizations, and regulators in working towards developing global standards of trial registration and reporting to results databases.

*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 GHR Web site currently includes user-friendly summaries of more than 475 genetic conditions, more than 680 genes, all the human chromosomes, and mitochondrial DNA. The Web site also includes a handbook called *Help Me Understand Genetics*, which introduces users to fundamental topics in human genetics including mutations, inheritance, genetic testing, gene therapy, and genomic research. Recent additions to the handbook include information about the validity and reliability of genetic tests.

Genetics Home Reference celebrated its sixth anniversary in 2009. In the past year, the project expanded its genetics content for consumers. Specifically, GHR staff added more than 200 new summaries to the web site in FY2009, an increase of about 17 percent from the previous 12 months. Staff intend to continue this rate of production in FY2010, covering additional Mendelian genetic disorders as well as more complex disorders. The team also plans to continue expanding the Gene Families feature, which currently includes explanations of about 40 families of related genes.

Usage of the GHR Web site continued to increase in FY2009. This year, the site averaged more than 14,100 visitors per day (an increase of about 9 percent over the previous 12 months) and more than 18.8 million hits per month (an increase of 14 percent over the previous 12

months). GHR continues to be recognized as an important health resource.

This year, GHR staff performed outreach activities to increase public awareness of the Web site. The project 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 meetings and conferences. Staff members will continue to educate others about this important resource in FY2010.

## Profiles in Science Digital Library

The Profiles in Science Web site (Profiles) 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. In October 2008, Profiles in Science was awarded the C. Herbert Finch Online Publication award at the Mid-Atlantic Regional Archives Conference.

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 Victor A. McKusick and Adrian Kantrowitz were added this year. Currently 26,347 digital items composed of 139,224 image pages are available on Profiles in Science. Presently the Web site features the archives of 29 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 FY2009, LHNCBC staff performed analyses and made modifications to the Profiles in Science systems. They analyzed the Profiles in Science access log files to compare with ForeSee Results report of the number of American Customer Satisfaction Index (ACSI) surveys completed. During the survey period, 813,647 unique IP addresses accessed Profiles in Science. Of these, 233,811 unique IP addresses were offered the survey at least once. In total, 1,021 respondents completed the ACSI survey at least once, which is the equivalent of 1.25 per 1,000 unique IP addresses, or .0125 percent, that accessed Profiles during the survey period. On an average day during the survey period, 3,670 unique IP addresses accessed Profiles; 60 percent were new and 40 percent had visited previously. Developers also continued to port the Data Entry Program from MS Access Visual Basic for Applications to a Web based application, and continued to port the Solaris version of Profiles in Science to Linux. They also modified, debugged and added reports to the Diagnostic Server, a tool that assists in quality control efforts and provides a read-only view of the Profiles in Science database. Staff debugged the software that logs technical metadata about the digital items, and they rebuilt the software that exports the database to a standard text file. They continued to write software to perform tasks that were done manually or in a semi-automated fashion. Developers debugged, modified and upgraded software underlying the Profiles in Science systems as well as the LHNCBC Web site, the Clinical Questions database, and the LHNCBC Personnel System to adhere to HHS, NIH and NLM security requirements. Staff evaluated software to extract text from PDF files and for making video clips adhere to Section 508 accessibility requirements.

## Nursing Home Screener (NHS)

NHS is a Web 2.0 system using MySQL, Google Maps, and data from CMS to quickly locate and select nursing homes. A user can search for homes by ZIP code or town, and receive a list of homes in order of overall quality (denoted by a one to five star rating), as well as a map showing locations. This list may be filtered by such characteristics as non-profit ownership or accepting Medicare or Medicaid payments. Activities in FY2009 included: developing a beta Web site, with substantially

revised design based on 2008 usability testing; making data updates more automatic, reducing ambiguous location data; conducting further usability testing; reporting the project at AMIA '08 and the Board of Scientific Counselors; and receiving an AMIA Distinguished Paper Award.

FY2010 plans include: launch a public site; implement initial, minimal customized page for each nursing home with tailored information; enhance user interface based on usability testing; run monthly data freshening; add statistics and graphics; pair customized page with per-home blog; and seed blog with news clippings about nursing homes.

### Evidence-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 care via smartphones, wireless PDA's, netbooks 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 need. New features can create a "consensus" opinion of multiple publications. Recent collaborative projects are ongoing to extend its reach to Botswana, Africa and the Pacific Islands.

### User-Focused Portals: NLM Gateway

The NLM Gateway is an ongoing production system that provides results from 24 NLM information resources in response to a single query. Since these resources are frequently updated, improved, and otherwise modified, the Gateway must change with them. Periodic changes to the NLM Document Type Definition (DTD), to MeSH and the MeSH mapping file, and to the UMLS Metathesaurus are accommodated each year. More than 110,000 meeting abstracts are indexed using the tools of the Indexing Initiative. Access to more than 70,000 Images from the History of Medicine was added this year.

Ongoing usability studies have resulted in further improvement of several aspects of the user interface. A user survey sought to determine the degree of user interest in a mobile version of the NLM Gateway for systems like the iPhone.

## 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.

### Videoconferencing and Collaboration

Staff continued to investigate, review, and develop collaboration tools, research their application, and use the tools to support ongoing programs at the NLM.

Work on uncompressed high definition video over IP continued. A second iHDTV node with streams of 1.5 gbps was installed and experiments were also done with UltraGrid uncompressed HD video having similar bit rates. In addition, a compressed version of UltraGrid with bit rates of 250 mbps was tried out successfully with the Rochester Institute of Technology. HD video was tested using new network protocols for dynamic network switching and an upgrade of the OHPCC/Collab subnet to 10 gbps was planned to accommodate HD bandwidth. A demonstration was done at the Internet2 Spring Member Meeting. Two versions of a pan, tilt, zoom (PTZ) remote camera control program were developed for Sony HD cameras (a standalone version and one for use on the AccessGrid) in anticipation of using HD video in telemedicine research. Work continues on additional HD enhancements for telemedicine research. The camera control program was made peer to peer for AccessGrid, while a peering standalone program is under development. Work on a more Windows native version of the University of Washington's iHDTV software and a version allowing bandwidth adjustments have been suspended because ConferenceXP, an open source program originally from Microsoft that was spun off the University of Washington also has HD capability. An uncompressed version was tested locally on the same machine and a compressed version with bit rates from 1 to 5 mbps was tested externally. The plan is to test iHDTV and ConferenceXP over optical connections between machines connected back to back in preparation for external tests when the network upgrade is realized. Staff is working on a manuscript comparing HD and other network video technologies used in the Collab.

Staff continue to look for partners for testing uncompressed HD video in dermatology, since the Puget Sound VA, which help formulate the clinical study, could not participate because it became committed to implementing a regional store and forward teledermatology program. Staff have approached the dermatology departments at the Washington Hospital Center and Medical University of South Carolina. Both have expressed initial interest. Staff will continue exploring other clinical areas where HD research may be appropriate.

A manuscript describing results of clinical research study of video medical interpretation completed at the Medical University of South Carolina is undergoing revision for publication. The study compared conventional standard definition video to phone and in person interpretation. A follow up study using lower quality video

(less than full screen) and cell phone technology has just started. Extensive tests were done with VSee, a low bandwidth video program, in anticipation of the study.

The results of an earlier co-location (presence) study of videoconferencing with the University of Alabama at Birmingham has been accepted for publication. A follow up study using one way video and chat to study learning outcomes and collaborative behaviors when students were co-located and dispersed was conducted, but the results still need to be analyzed.

Work continued with SIS on distance education outreach program for minority high school students and an article was published on expanding the program to Alaska. Farrington High School in Hawaii has been added to the program. The school has a diverse Native Hawaiian, Native Pacific, and Asian student body. It joins Kotzebue High School and King Drew Medical Magnet High School in Los Angeles, with their respective Native Alaskan and African American and Hispanic student populations. Programs continued to be archived and placed on the Collab Web and streaming servers.

An alliance was developed with the NIH Library to continue to offer NCBI database training using the Virtual Computer Lab methodology developed earlier in cooperation with the University of Puerto Rico (UPR). An article describing the Virtual Computer Lab methodology was published and this year programs were conducted with the University of North Carolina at Chapel Hill and the Charles R. Drew University of Medicine and Science. Additional quarterly programs are being planned with UNC and Duke.

PTZ program and other collaboration programs developed by OHPCC continue to be put on the Collab Web server for others to download as they are refined. Documentation for the SIS distance learning program was added to the iHDTV, UltraGrid, and Virtual Computer Lab documentation put online last year.

Staff continued a retrospective review of NLM OHPCC Telemedicine, Next Generation Internet, and Scalable Information Infrastructure Projects with the objective of identifying computing, communication, and health science application themes in this diverse body of work. A draft manuscript was developed and is currently being edited for submission.

*OHPCC Collaboratory for High Performance Computing and Communication*

The "Collab" was originally established 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 now 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, demonstrates 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 multilanguage 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. Through Google, immediate translation of MEDLINE abstracts to French, Italian, Portuguese and Spanish has been added recently.

*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) Team conducted a number of core projects to build, administer, support, and maintain an integrated and secure infrastructure that facilitates the research activities of the LHNCBC and thereby augment the overall effectiveness of research staff members. The integrated secure infrastructure encompasses network management, security management, facility management, and system administration support

for a large number of individual workstations and shared servers.

The network management team plans, implements, tests, deploys, and operates high-speed network connectivity locally as well as over Internet and Internet-2. The core projects include studying and planning the implementation of central network management for effectively responding to network alerts and malfunctions; 10 Giga BPS network to support research projects that require high-speed communication capacity; and an enterprise device management system to update large number of network devices uniformly.

The security management team incorporates security operations into firewall administration, patch management, anti-virus management, intrusion monitoring, security and vulnerability scanning, and vulnerability remediation to ensure a safe working environment from an overall security perspective. The core projects include studying and planning the implementation of a security auditing process, asset management, and configuration management for the consistency and integrity of LHNCBC security profiles.

The facility management team facilitates the deployment of products and servers, including power acquisition, network planning, cabling connection, and space allocation in the B1 computer room as well as co-location in Sterling, Virginia. The core projects include: studying and planning the implementation of redundant LHNCBC infrastructure in the B1 computer room; new network-wiring schemes to the offices in corridor 28 and 30 at B1 level; and Intelligent Platform Management Interface (IPMI) for effective monitoring on the large number of devices in the B1 computer rooms.

The system administration team provides center-wide IT services, such as DNS, NIS, backup, printing, and remote access to ensure an efficient operation across the Center. The core projects include studying and planning the implementation of an enterprise data backup system that utilizes different media at multiple locations for data safety and integrity; unified communication to enhance research collaboration; evaluation of Windows 7, Windows 2008 and Red Hat Linux 5 platforms for LHNCBC desktop and server deployments; and an enterprise search engine. Additionally, the system administration team and other members support Continuity of Operation (COOP) and Federal Information Security Management Act (FISMA) compliance, and providing operation assistance and troubleshooting functions for shared computing resources.

*DocView Project: Tools for Using and Exchanging Library Information*

The goal of this project is to conduct R&D on advanced tools allowing libraries and users to access biomedical information. In FY 2009, research focused on the development of MyDelivery, a new Internet communications system designed to deliver very large files and large numbers of files, especially over potentially unreliable networks such as wireless used by an increasingly mobile population. Health science applications often require the use and exchange of information contained in very large files (e.g., digitized x-ray images, sonographic images, digital video files, MRI, CT scans, PET scans, and scanned document images). 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.

In FY2009, following alpha testing, we conducted an extensive beta test with 200 outside testers. Ongoing work involves completing the MyDelivery API, modifying the system to be fully compliant with FIPS 140-2, integrating it with the Interactive Publications Panorama viewer to demonstrate the capability of the API, and releasing MyDelivery as open source code.

## DocMorph

As part of the DocView project, research and system engineering continues to maintain and improve the operation of DocMorph, a Web-based server providing users remote image and information processing capabilities via the Internet. 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, which uses it mainly in our interlibrary loan service.

## 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 12,000 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.

*Image Storage and Transmission Optimization (ISTO)*

Investigators researched the best approaches for using advanced compression techniques, such as Wavelet transform, for the storage of digitized spine x-rays and uterine cervix images, in addition to implementing an efficient decompression algorithm. Developers incorporated wavelet decompression and multiscale display into a dissemination system for the digitized biomedical images within the WebMIRS system and the Multimedia Database Tool.

Work underway includes completing the development of the HVSQ codec incorporating Wavelet

transform techniques coupled with scalar and vector quantization to achieve maximum compression consistent with image quality for the digitized x-ray images of the lumbar and cervical spine and uterine cervix images.

Other ongoing work includes making these advanced compression techniques compatible with open standards, such as Internet Imaging Protocol (IIP), and developing a standard interface for classification of vertebrae from NHANES x-ray images, and organizing feature vectors for image retrieval.

## Language and Knowledge Processing

### Terminology Research and Services

The Patient Data Management Project (PDM) brings together several activities centered on lexical issues, including development and maintenance of the SPECIALIST lexicon as well as lexical research. The lexicon and lexical tools are distributed to the medical informatics community as free open-source tools and also delivered with the UMLS information sources.

Objectives for FY2010 are:

- Continued expansion and maintenance of the SPECIALIST lexicon with emphasis on clinical vocabulary
- Continued development of the lexical management system
- Continued development of the cross-platform version of the SPECIALIST Lexical Tools
- Continued development of text processing tools (NLP tools)

### Medical Ontology Research (MOR)

The Medical Ontology Research (MOR) project focuses on basic research on biomedical terminologies and ontologies and their applications to natural language processing, clinical decision support, translational medicine, data integration and interoperability.

During FY2009, interoperability issues were investigated among several pairs of biomedical ontologies, including BioTop and the UMLS Semantic Network, SNOMED CT and MedDRA, the Mammalian Phenotype Ontology and OMIM, and between RxNorm and NDF-RT.

RxNav, the standalone browser for RxNorm, NLM's drug terminology integration database, was extended in order to provide better support to our users. The Application Programming Interface (API) has been used for mapping large amounts of NDC codes to RxNorm identifiers. Feasibility of integrating other drug information sources with RxNorm through RxNav was demonstrated.

We kept investigating the benefits of using Semantic Web technologies including RDF - the Resource Description Framework—and triple stores (e.g., Virtuoso) for the integration of biomedical information and in support of translational research activities (e.g., comparing genotype-phenotype associations across species).

This year, our research activities resulted in six journal articles, six papers in conference proceedings, and 13 invited presentations. We continue to collaborate with leading ontology and terminology centers, including the National Center for Biomedical Ontology, the International Health Terminology Standards Development Organization (SNOMED CT) and the World Health Organization (ICD 11).

### Clinical and Translation Science

Using a newly developed database of translational terms, RxNorm and MeSH, the LHNCBC developed a novel search tool to search for innovative, novel and promising translational research. The query is initiated using disease processes and/or interventions or both as search terms. Publications identified as translational in nature are then retrieved with relevant terms highlighted for easy recognition. With interventions, such as drugs in the RxNorm database and disease processes in MeSH that are pertinent clearly identified, the user can quickly find publications to facilitate research, experiment planning and bench-to-bedside applications. An additional benefit from this project is the inclusion of a new term, "translational research," to the MeSH vocabulary suggested by one of the project team members.

### Semantic Knowledge Representation (SKR)

The Semantic Knowledge Representation project conducts basic research in natural language processing (NLP) based on the UMLS knowledge sources. The project focuses on developing SemRep to extract semantic predications from text to support innovative information management applications in biomedicine, including advanced tools for clinical decision support, practice guideline development, and literature-based discovery. Current efforts are adapting this technology to identify selected concepts in clinical narrative.

SKR researchers are developing a SemRep predication database which holds 22 million predications from 6.6 million MEDLINE citations (1999 through October, 2009). The project team is exploiting this database to find publications that support critical questions used during the creation of clinical practice guidelines (with support from NHLBI). Further, the team is collaborating with academic researchers in using the predication database to help interpret the results of microarray experiments, to support literature-based discovery, and to investigate advanced statistical methods for enhanced information retrieval.

**UMLS and Clinical Vocabulary Standards**

This program encompasses multiple projects. The problem list vocabularies project has produced the CORE (Clinical Observations Recording and Coding) subset just released as a UMLS knowledge source. The RxTerms project is an efficient drug interface terminology that links directly to RxNorm. Inter-terminology Mapping activities with IHTSDO (International Health Terminology Standards Development Organization) and direct UMLS-related activities continue. The TREF (Terminology Representation and Exchange Format) specification will be used by NCHS (National Center for Health Statistics) to produce a TREF version of the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM).

*The CORE Problem List Subset of SNOMED CT*

Problem lists in Electronic Health Records (EHR) facilitate the organization and communication of clinical data and reasoning. They also drive additional functions like clinical decision support, generation of reimbursement codes, auditing and research. At present, most institutions use their own problem list vocabularies for problem list entry. The creation and maintenance of these vocabularies are resource and time consuming. The lack of a shared list of terms impairs data interoperability. The UMLS-CORE (short for Clinical Observations Recording and Encoding) Project was started two years ago to study the use of controlled vocabularies in clinical documentation at a summary level, such as in a problem list. This project has two goals:

- To study the problem list vocabularies of large health care institutions and to characterize them in terms of their size, pattern of usage, map-ability to standard terminologies and the extent of overlap
- To identify a CORE subset of terms that have a high frequency of occurrence in problem lists

Problem list terms and their frequency of usage are collected from seven large healthcare providers: Beth Israel Deaconess Medical Center, Kaiser Permanente, Mayo Clinic, Intermountain Healthcare, Regenstrief Institute, Nebraska University Medical Center and the Hong Kong Hospital Authority. The most commonly used terms, covering 95 percent of the usage volume in each institution, are mapped to the UMLS. About 8 percent of these terms cannot be mapped. The overlap among the mapped terms is assessed. The average pairwise overlap among institutions is about 40 percent. The terms that are shared among institutions are used much more frequently than terms that are not shared. Using this data, a subset of 5,000 SNOMED CT concepts, called the CORE Problem List Subset of SNOMED CT is identified and made available to the public. SNOMED CT is the US Consolidated Health Informatics (CHI) designated terminology standard for problem lists. The CORE Subset

represents the SNOMED CT concepts that are most frequently used in problem lists. It can be used as a starter set for the creation of problem list vocabularies. This will facilitate the use of SNOMED CT in EHRs and enhance data interoperability. The CORE Subset was first published in July 2009 and is updated regularly to synchronize with updates in SNOMED CT and the UMLS. It can be downloaded as a standalone file or extracted from the UMLS as a content view subset. Terms that are frequently found in problem lists but not present in SNOMED CT have been submitted to the International Health Terminology Standards Development Organization (IHTSDO) for consideration of addition.

*RxTerms*

RxTerms fills the need for a free, user-friendly and efficient drug interface terminology that links directly to RxNorm, the national terminology standard for clinical drugs. Efficiency of data entry is enhanced by systematic rearrangement of RxNorm drug names and pruning of drugs that are not likely to be useful in prescribing e.g. drugs unavailable in the US, allergenic extracts. Addition of common synonyms and abbreviations further improves data entry efficiency. Other usability enhancing features include 'tall man lettering' of look-alike drug names as recommended by the FDA, and user-friendly concentration units for liquid doses. RxTerms has been evaluated for coverage of commonly prescribed drugs and data entry efficiency. Coverage is over 99 percent for both branded and generic drugs. Data entry efficiency is improved when RxTerms is used compared to the original RxNorm drug names.

RxTerms has been available for free download since November 2008. It is updated every month to synchronize with the monthly full release of RxNorm. There are about 250 registered users to date. RxTerms is currently being used in one of CMS's applications in the post-acute care environment. It is also used in NLM's Personal Health Record. The use of RxTerms will facilitate the inclusion of RxNorm identifiers in Electronic Health Records and promote data interoperability.

**Disaster Information Management**

*Lost Person Finder (LPF)*

The Lost Person Finder (LPF) is a system for family reunification during a mass casualty event. It combines image capture, database and Web technologies.

In FY2009, project activities included requirements gathering, analysis, reporting, test, and demonstration. Developers built a demo version of the LPF site, customizing and enhancing the Sahana open-source disaster management system by focusing on its Missing Person Registry. Customization included a tailored LPF theme and logo, Americanization of words and phrases in

the database and user interface, and incorporation of standard ontologies and localized geographical features.

Other developmental activities included: developed TriagePic software for image capture of triage victims, and integrated Bluetooth camera image transmission and email-base record distribution; built first version of LPF Notification Wall (images of victims on a large display), including Web services for the application to access data from the LPF database, and iconic overlays to represent victim name and status. In October 2009, we participated in a large-scale multi-institutional demonstration (Collaborative Multi-Agency Exercise or CMAX) and demonstrated TriagePic, search capability, and the Notification Wall at collaborating institutions, the Navy and Suburban hospitals in Bethesda.

Plans are under way to: investigate privacy and security issues related to LPF use; add email and notification modules to LPF Web site; refine TriagePic interface in response to feedback at CMAX; customize LPF Web site for individual participating hospitals; develop the LPF Mobile Web site and create search capability for popular mobile platforms; and refine information shown on the Notification Wall for different audiences.

**Video Production, Retrieval, and Reuse Project**

This development area encompasses four projects. The NLM media assets project and the NLM support project contribute to the NLM-wide audio-video support of the NLM Long Range Plan goal of promoting health literacy and increasing understanding. The LHNCBC research support project and the core resources project contribute to ongoing LHNCBC information development projects, working to improve access to high quality biomedical imaging information.

The still image, graphics, and video support staff provide ongoing capability 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. This area of focus includes support to maintain the audio, video, and multimedia capability in the NLM board room, auditorium, and other conference areas.

A number of LHNCBC development projects require videographics, interactive multimedia development, imaging, animation, or video production as part of the overall project objectives. A major effort in this area is the improvement of rendering times for videographics, and 3D visuals and animations for DVD and other interactive multimedia productions. Extensive

development work was directed toward the planning and demonstration of interactive multimedia for the 2010 NLM Exhibition "Native Concepts of Health and Illness."

*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 and participates actively on Center research projects.

In FY2009, the Center provided training to 44 participants from 13 states and five countries. Participants worked on research projects including 3-D informatics, automated indexing, clinical information systems, consumer health information, content-based information retrieval, de-identification of medical records, document processing, preservation of electronic resources, image, text and document processing, information retrieval, interactive publication, medical ontology, medical terminology, mobile computing, natural language processing, personal health record, question/answering and telemedicine research.

The program maintains its focus on diversity through participation in programs supporting minority students, including the Hispanic Association of Colleges and Universities and the National Association for Equal Opportunity in Higher Education summer internship programs.

The Informatics Training Program sponsors a 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 LHNCBC. 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 20 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.

This year, NCBI celebrated its twentieth anniversary with a one-day conference. Leading geneticists, scientists, and bioinformatics leaders attended and spoke in honor of the work that NCBI has done and its contribution to genetics and genomics. In addition, a statement appeared in the Congressional Record on June 26, 2009 from Congressman David Obey applauding NCBI for its achievements over the past 20 years: "NCBI's molecular biology information resources are empowering hundreds of thousands of researchers around the world to identify disease-related genes and develop strategies for treating and preventing disease."

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 incorporating directly submitted data and information from the biomedical literature.
- Performs research into advanced methods of computer-based information processing for analyzing the structure and function of genes and other biologically important molecules and compounds.
- Facilitates the use of genomic databases and software by researchers, students and health professionals.
- Coordinates efforts to gather and disseminate molecular biology 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 NCBI's 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.

Important sources of GenBank data are direct sequence submissions from individual researchers and scientists as well as institutions, such as genome sequencing centers. Sequence records are typically submitted prior to journal publication so that they can be cited by accession number when the publication appears. Sequences 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 (ESTs), Genome Survey Sequence (GSS) records, and the "CoreNucleotide" group. The Transcriptome Shotgun Assembly (TSA), division contains shotgun assemblies of primary (mRNA) sequences deposited in dbEST, the Trace Archive, or the Sequence Read Archive (SRA).

The Third Party Annotation (TPA) database 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, derived by inference, often computationally, where the source molecule or its products have not been the direct result of experimentation. A large amount of data comes from TPA submissions.

The amount of data submitted to GenBank increases each year, resulting in the number of base pairs doubling approximately every 22 months. GenBank's two divisions combined, now contain over 156 million sequence records and over 254 billion basepairs. The traditional nucleotide sequences division increased to 108 million records in FY2009 from 92 million in FY2008. The WGS division grew to over 48 million records and 148 billion basepairs. Approximately 1200 public WGS projects were released by GenBank in FY2009. About 600 complete microbial 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 for updating and submitting large groups of sequences. Sequin version 9.0 was released in FY2009. In this version, feature and qualifier dialogs were updated to comply with the latest version of the INSDC feature, Documentation. A new version of NCBI's Web-based submission tool, BankIt, was released with new modifications. New BankIt features include the ability to save submitter information, submit sets of sequences as well as single sequences, accept source qualifier and sequence feature data as text input, and to pause during a submission without loss of data. 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 enable 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: dbSNP, RefSeq, CCDS, dbGaP, Entrez Gene, Probe, UniGene, HomoloGene, and GEO. Genomic tools include BLAST and Map Viewer. 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 8,835,769 proteins (almost 50 percent more than

last year) representing 9,005 organisms. The RefSeq curation group supports the whole genome annotation process flow for updating existing genomes and processing new submissions.

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 FY2009, several new genomes were annotated and updates were provided to existing genome assemblies for several eukaryotic species, including human, wine grape, zebra finch, Hydra, moss, Ciona intestinalis, and pea aphid. Several bacterial and prokaryotic species were updated as well. Genome resource pages were created for goat and pea aphid.

The RefSeqGene project, a subset of the RefSeq project, provides a set of genomic sequences of well-characterized genes to use as a reference standard. Over the past year, the number of RefSeqGene records tripled from 658 in August 2008 to 2,074 in August 2009. Locus-specific databases are increasingly requesting that sequences be generated for genes they would like to represent.

The Map Viewer is the primary tool for visualization of assembled 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 FY2009, Map Viewer added a feature that allows users to link directly to a chromosomal region. A new clickable icon provides options for chromosome number, assembly, and coordinate range, which then goes directly to the position on a given genome. 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 provides 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. The database continues to be heavily used and manages information for more than five million genes from over 6,500 taxa. During FY2009, most development was focused on internal infrastructure, namely completing the processing so that Gene is delivered only from NCBI's portal system. Gene added a gene-pseudogene cross-referencing system, so users may navigate more readily from functional gene to pseudogene and from pseudogene to functional gene. The phenotype section was enhanced to include links to NHGRI's Genome Wide Association Studies catalog.

As a corollary of establishing RefSeqGene, collaborations with the dbSNP and dbVar groups were activated to establish sequence annotation of clinically important variants in a timely fashion. Descriptions of variants were extracted from OMIM, GeneReviews, and Gene Tests, and processed for submission by dbSNP.

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 coding regions resulting in 3,852 entries, corresponding to approximately 1,300 genes.

The Genome Reference Consortium is an international collaboration that aims to update and improve the mouse and human genome assemblies. NCBI provides informatics support for the project, such as tracking of tiling path files, overlaps between adjacent clones, and curation, as well as generates the final assembly after collaboration and quality assurance. GRC produced a new human genome build in FY2009.

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 114. Two-thirds of interactive traffic on UniGene is for human and mouse entries.

HomoloGene, a comprehensive database of gene homologs, complements 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 2.8 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 92 million submissions of human genome data that have been processed and reduced to a non-redundant set of 17 million refSNP clusters. Fifty-four other organisms are represented in the SNP database, with 64 million submissions curated to 46 million refSNP clusters.

The database of genomic structural variation, dbVar, was introduced in FY2009. dbVar contains data on variant DNA less than or equal to one basepair in size. Submissions are accepted from whole genome comparative studies and locus- and gene-specific data from quantitative studies. This year, dbVar began accepting submissions and a preliminary site was made available to facilitate submissions, obtain documentation, and provide FTP access to loaded studies.

The Probe database, part of the Entrez system, stores molecular probe data, together with information on success or failure of the probes in different experimental contexts. Nucleic acid probes are molecules that complement a specific gene transcript or DNA sequence and are useful in gene silencing, genome mapping, and genome variation analysis. This database contains over 10 million probes as of October 2009. 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.

UniSTS, a database of sequence tagged sites, processed four new maps in FY2009. The resource receives about 2,500 hits per day. Information in UniSTS is being integrated into the Probe database.

## Comparative Genome Data

NCBI currently provides 37 guides for comparing organisms on a genome scale. The Genome Resource Guides provide information on genome-related tools and repositories available through NCBI and various outside centers and institutions. The guides provide 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 FY2009 include the organisms: pea aphid, wine grape, moss, Hydra, Ciona intestinalis, and goat.

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 into 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 3,460 genome sequencing projects completed, in draft form, or in progress. The number of species represented in the database is over 4,900.

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 3,482 reference sequences for 2,360 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 125 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. Fifty-nine plant species are represented in the genome project database.

The Microbial Genome Annotation Pipeline was developed for annotation of prokaryotic genomes. Over 900 genomes have been annotated in-house and NCBI is working with 10 outside groups who submit data.

## Specialized Databases and Tools

The Peptidome resource is a new repository that archives and distributes tandem mass spectrometry peptide and protein identification data. Mass spectrometry results and conclusion-level information are captured, together with sufficient raw data and descriptive information to enable understanding of the experiment and analysis of the underlying data. Results are organized into Studies and Samples. A Sample provides all results common to a biological sample while a Study organizes Samples into datasets that make up an experiment. The database has been public and accepting submissions since March 2009.

The new BioSystems database connects biosystem records with associated literature, molecular, and chemical data via the Entrez system. It also serves to facilitate computation on biosystems data. BioSystems characterizes metabolic pathways and other data collections that together form a meaningful unit with its own name and description. It provides Entrez users with a means to explore functional relationships between genes, proteins, and small molecules. BioSystems currently contains metabolic pathways and similar biological systems imported from KEGG, EcoCyc, and the human Reactome collection.

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. Over 28,600 new influenza virus sequences were entered into NCBI's Influenza Sequence Database in FY2009, of which approximately 13,580 were from the NIAID

Influenza Genome Sequencing Project. This project funded Centers of Excellence for Influenza Research and Surveillance, CDC, and more than 20 institutions worldwide. These sequences were then 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 added to the Influenza Virus Resource include the ability to search sequences by month and date of collection, and to search individual proteins for segments that encode for more than one protein. A special H1N1 Web page was created over the past year to provide updated reports on the pandemic 2009 influenza virus and sequence submissions to GenBank. 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 Sequence Read Archive (SRA) is a repository for data generated from massively parallel sequencing experiments. SRA presents data by sequencer runs rather than by individual traces. Created in June 2007, SRA contains over 1,100 studies and 10 terabytes of data. During FY2009, SRA was added to the Entrez Retrieval system.

The Clone Registry is a database that integrates information about genomic clones and libraries, including sequence data, genomic position, and distributor information. Management of information about clones of interest to users of Gene and UniGene is being consolidated in the Clone Registry. Over the past year, NCBI has converted data from Nucleotide Core and Trace databases into a relational form. The Clone Finder search interface was updated in FY2009, making it easier to search by feature. It also includes the ability to filter clones on results pages and a new graphical view integrates the clone library table view.

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. GEO received over 140,000 new submissions in FY2009, bringing the total to over 447,000 records. This represents a 35 percent increase in submission rate compared to the same period the previous year. About 2,000 new GEO accounts were created with associated data. Seven billion new individual data points were also accepted this year, bringing the total number to over 23 billion. Over 2,500 new manuscripts cited GEO accessions or provided a direct reference to the database this year, bringing the total to over 7,300 manuscripts.

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. Version 5.1, released in FY2009, included a two percent increase in the number of sequences. The related Vector BLAST database was updated to include all GenBank sequences from the current build.

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.

**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. The Web site is visited by more than 50,000 users per day.

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 increased by over 25 million test results in FY2009, for a total of 55 million. The number of bioassays increased from 1,000 in July 2008 to over 1,750 in September 2009.

New types of biological assays are now supported, including multiple-target "panel" assays, siRNA screens, and literature-derived bioactivity summaries. Concomitant changes to the PubChem deposition system and PubChem data analysis tools were made to take advantage of these new assay data models. Annotation of PubChem records was improved by integration of pathway information contained within the BioSystems databases to protein targets and small molecules to existing data presentation and analysis tools.

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 25.5 million compound records, up from 19 million the previous year. The PubChem Substance database contains chemical substance records and associated information. Currently, there are nearly 61 million substance records, up from 40 million.

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. The Entrez interface was revamped to improve discoverability by advertising records with informative annotations. The PubChem homepage was redesigned to improve overall discoverability and usability of available tools and interfaces.

The PubChem 3D system was released in FY2009. This major undertaking contains integrated 3D representations of PubChem Compound records, a pre-computed "similar conformer" relationship to annotate molecules having similar 3D shape and pharmacophore feature orientations, and tools to visualize 3D conformers and their overlays.

Continuous improvements are required to keep up with the steady growth of PubChem. The need to rapidly process data is paramount. Major infrastructure improvements included changes in the way daily public updates are performed and enhancements to how public queries are handled. Memory capacity and data software were also upgraded to handle increased queries and analysis of greater quantities of data.

The PubChem Power User Gateway (PUG) is an interface that allows users to search and download data that is not accessible to the Entrez system, such as structure queries. Programmatic access to BioAssay data was added to PUG in FY2009, which further opened PubChem to data mining. SOAP-based Web services, which complement the existing PUG interfaces by enabling them to be readily used by existing workflow applications as well as many programming and scripting languages, were released in final form and extended.

**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. MMDB now computes and stores molecular interactions observed in three-dimensional structure complexes. In upcoming versions of the MMDB summary pages, these interactions will be displayed prominently to emphasize information about the biological function of those complexes.

An interaction tracking database and corresponding Web service called IBIS (Inferred Biomolecular Interactions Server) have been made accessible to the public. IBIS displays intermolecular interactions observed in three dimensional structures tracked by MMDB. IBIS emphasizes interactions that have been observed multiple times in independent experiments, and clusters such recurring observations so that displays can be made less redundant. IBIS also infers interactions for protein structures that do not come with observed interactions, based on structural and sequence similarities to above mentioned clusters. Molecular interactions and the corresponding sites on proteins can also be inferred for many proteins in the Entrez Protein database that are not directly linked to three-dimensional structure but have significant similarities to structure-linked sequences. IBIS displays and infers interactions between proteins, proteins and nucleic acids, and chemicals.

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 2.17, was released in summer 2009. It mirrors Pfam version 23 and contains about 4,500 models curated by NCBI.

CD-Search, the Web-interface to CDD, 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.

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, and allows a user to browse interactively, viewing superpositions and alignments in Cn3D.

## BLAST Suite of Sequence Comparison Programs

Comparison, whether of morphological features or protein and DNA sequences, lies at the heart of biology. BLAST has made it easier to rapidly scan huge sequence databases for similar sequences and to statistically evaluate the resulting matches. In a matter of seconds, BLAST compares a user's sequence with 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. A new BLAST+ suite of tools was released in FY2009. BLAST+ features improved integration between the Web service and stand-alone package. The Web service improvements include the incorporation of BLAST 2 sequences service on the main BLAST submission forms, changes to output format, and new options for downloading results and search strategies. New standalone options include tasks, search strategies, and custom output format. Also, COBALT multiple-sequence alignments can now be generated from protein BLAST results with a new multiple-alignment link.

SRA transcript sequences are now searchable through a specialized BLAST page. All sequences are from the SRA database.

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, re-rooting at any user-selected node, collapsible subtrees, and sequence grouping.

## Integration of Clinical, Genetic, and Environmental Databases

The NCBI database of Genotypes 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.

## Study Submissions

As of November 2009, dbGaP included 53 studies. Forty studies contain both public summary-level data and individual-level genotype/phenotype data distributed through the authorized access system, three studies contain individual-level sequences (Sequence Read Archive) or references to sequences (Trace Archive) and individual phenotypes, and nine contain public data distributed by the dbGaP FTP site. Four of the latter group are summary-level association results linked to published GWAS studies, and two are GAIN genotype control studies that provide genotype data for HapMap samples and are consented for public distribution.

Studies with individual-level data submitted in 2009 include the disease areas of Multiple Sclerosis, Alcohol Addiction, Lung Cancer, Type II Diabetes, Alzheimer's Disease, Bone Fragility, Glaucoma, Breast Cancer, Pancreatic Cancer, Acute Myeloid Leukemia, and several updates to the Framingham Heart Study. Collectively, the studies released in 2009 include measures for over 54,348 total research participants, and GWAS measurements include:

- Over 22,792 unique phenotype traits.
- 227 XML-based phenotype documents linked to summaries of the measured variables. These documents include descriptions, collection forms, and scientific protocols, and also provide models for future research. Links between searchable documents and variable descriptions provide an unprecedented level of functionality and usability.
- Hundreds of billions of individual genotypes (i.e., single-nucleotide measurements of participant DNA sequence). Analysis of these genotypes can be used to reveal systematic and heritable genetic differences between affected and unaffected individuals.
- 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.

**Authorized Access System Download Activity**

In order to obtain individual-level data, which is de-identified, principal investigators (PIs) must log in to the dbGaP authorized access system and request access. Requests are routed to one or more NIH Institute/Center Data Access Committee (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. After approval, users create and download a password-protected copy of the data to their local secure computing environment.

As of November 2009, 2,726 Data Use Requests (DAR) have been created in the approval system. 1,496 have been approved for access, 406 have been disapproved or have a request for revision, and 520 are in the approval process.

**Data Usability: Tools and Software Development**

The study registration and submission system has been in use since fall 2008 and provides a Web-based system for NIH staff to register, certify, and configure studies for release via the dbGaP authorized access system. For Web-based visualization displays, the dbGaP association results browser was enhanced to show multiple analysis tracks in simultaneous view and show local recombinant rate data. It was also modified to use s-viewer, and support search and navigation functions for dbGaP variables, genes and SNPs.

Many enhancements were performed to the public Web pages, such as inclusion of genotype platform information, dataset-specific pages and statistical summaries broken out by consent group. Manual curation of dbGaP variables to MeSH and other UMLS-controlled vocabularies continues to be performed by dbGaP staff. This effort is being coordinated with international counterparts in order to capture the data in a structured way that can provide a framework for exchange of metadata across resources in the future. FY2009 saw the development of the GaP Plus association results browser, which incorporates published GWAS results obtained through collaboration with the NHGRI GWAS catalog with the most statistically significant results from dbGaP analyses.

**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. Entrez allows users to 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 38 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 established a program to help users better explore and navigate the myriad of data contained in its 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 a medical term in the PubMed database may not be aware that separate databases, on genetics or drugs, for example, contain additional relevant information.

The Discovery Initiative has resulted in many new features being added to the Entrez databases over the past year. The new features are often initially made

available in PubMed and, after a short test period, added to other databases. A "Related Queries" feature provides users with suggested query phrases from the most popular PubMed queries that contain the current search term, possibly providing more precise results. The "Recent Activity" feature shows queries that have been performed within the current search session, including searches and record views in other Entrez databases.

Sensors are a new type of component that detects certain search terms and provides access to relevant results. The Gene Sensor shows a gene symbol when it matches a query, and provides a link to the relevant entry in the Gene database. This feature is available in PubMed as well as the Protein and Nucleotide databases. A Citation Sensor finds terms within a literature citation from the PubMed Citation Matcher. An Accession Sensor provides a direct link to the sequence databases when the search term consists of an NCBI sequence identifier. A Hot Topic Sensor appears for searches relevant to current topical issues such as H1N1 viral sequences.

Various new ads have been included to enhance user discovery as well. A Database Ad displays related information found in other databases that may provide unexpected connections. A PubMed Central (PMC) ad displays articles that have free, full text available in PMC. A Structure ad appears for articles reporting information from structure records. This ad appears in both PubMed and the Structure database. Viral Genome Resource Ads for influenza, dengue, SARS, and retroviruses appear in the sequence databases on the sequence records of viral origin.

The new Entrez Sequence Viewer also has an emphasis on discovery. New features include a related article section that displays article titles that may be relevant to the chosen sequence. Also, links to related mRNA and protein Reference Sequences are available. A link is available to the NCBI primer designing tool, Primer-BLAST. The "All links from this record" component provides a list of resources to which the user can find additional information related to the sequence. On Nucleotide and Protein record pages, a "Change Region Shown" feature allows users to specify the region of the sequence being viewed. Also, a "Customize View" feature provides options for the basic features shown in the record. On results pages, a "Top Organisms" ad provides a list of the organisms for which there are the highest number of results matching the query.

## 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 19 million citations from more than 36,772 journals, some dating back to the 1950s.

PubMed is continually updated and enhanced for better functionality and more precise search results. A Gene Sensor was added to PubMed that recognizes gene symbols found in the Gene database and subsequently reports them on the PubMed results page. New ad features are available on PubMed summary pages. A Structure ad provides a link to the Structure database if that structure is reported in the PubMed article being viewed. Also, a Pubmed Central ad provides links to PMC articles that are cited in the article being viewed. Another new ad shows free articles in addition to links to PMC articles. An "Also try" feature provides a list of search terms that could possibly produce more accurate results based on the original search.

The AbstractPlus display page was enhanced to include a Recent Activity feature that provides a list of previous search terms along with the number of results found. A PMID : PMCID Converter was added to the NCBI site in order to translate ID numbers for articles found in both PubMed and PubMed Central. The converter will convert IDs from PubMed to PubMed Central and vice versa.

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. New features include improved account, navigation, preferences, and filter options and capabilities. Users may now share a collection or bibliography with others.
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 2009, the number of organizations participating in LinkOut increased to over 2,840, representing a twelve percent growth rate over the past year. LinkOut providers include 2,223 libraries, over 320 full-text providers, and 265 providers of non-bibliographic resources, such as biological and chemical databases. Users can now link to 80 million Entrez records, including links to the full text of 54 percent of PubMed records from over 7,300 journals.

Outside Tool is a related service that also links users to outside resources. Participation in this program increased to over 640 institutions. Usage of LinkOut resources reached over 30 million hits per month, about

1.3 million hits per week day.

Training and end-user support of LinkOut libraries have been moved successfully to NLM Library Operations to take advantage of their support infrastructure for libraries. As a result, more libraries can be supported through the Library LinkOut program.

LinkOut infrastructure has been improved in various areas. LinkOut indexing is being re-developed with two aims: to support Unicode and languages in addition to English and to adopt a modular approach to LinkOut indexing to improve efficiency and reliability of LinkOut indexing. A new LinkOut portlet has been developed to display LinkOut resources more prominently in PubMed. This portlet can also be used in other Entrez databases to display LinkOut resources.

*PubMed Central*

PubMed Central (PMC) archives, indexes, and provides free and unrestricted access to full-text articles from life science journals. This repository is integrated with the PubMed biomedical literature database of indexed citations and abstracts. PMC now has more than 1.8 million articles and more than 700 participating journals.

A number of new publishers signed on to PMC in order to support the NIH public access policy. Publishers depositing articles directly in PMC now contribute about twenty-five percent of all NIH-funded articles.

There was a major redesign of PMC software to create a modular infrastructure, which will allow PMC International sties to customize elements of a page. The redesign also supports NCBI's Discovery Initiative, allowing related information to be displayed as "ads" adjacent to specific sections of an article.

In June 2009, NCBI announced support for a new PMCI site, PMC Canada, which is sponsored by the Canadian Institutes for Health Research (CIHR), the Canadian counterpart to the US National Institute of Health.

*Bookshelf*

The NCBI Bookshelf gives users access to the full text of over 120 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. Fifty-three new books were added to the Bookshelf this year, more than any other previous year. Among the new titles are two major collections, Frontiers in Neuroscience and The National Academies Collection: Reports funded by National Institutes of Health. The latter collection also includes 10 prepublications reports, providing rapid access to new information until the final reports are available. Over 200 chapters were added to existing collections. New content is continually added to a number of resources.

These include MICAD (Molecular Imaging and Contrast Agent Database), GeneReviews, the Madame Curie Bioscience database (formerly Eurekah), HSTAT, and the NCBI News newsletter.

**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 over 95 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. Research projects are continuing in support of the PubChem molecular libraries project.

The high caliber of work performed by the CBB is evidenced by the number of peer-reviewed publications

generated—over 60 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 33rd meeting was held in April 2009.

**Bioinformatics Training and Support**

*Outreach and Education*

The outreach and public services component of NCBI is an essential activity to ensure that the research community is aware of all NCBI services and can 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 four scientific conferences, presented at seminars and workshops, provided a number of training courses, and published and distributed various forms of print materials. The public services division also provides user support via e-mail and telephone.

*Outreach: User Guides for NCBI Resources*

NCBI provides 18 "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 three new structure announce lists provide information related to NCBI's structure resources. RSS Web feeds are also available for updates and announcements on many of NCBI's databases.

# EXTRAMURAL PROGRAMS

*Valerie Florance, PhD*
*Acting Director*

The NLM Extramural Programs Division (EP) received Congressional authority for its grant programs from two different authorizing acts: the Medical Library Assistance Act (MLAA) and Public Health Law 301. 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 several categories of grants, all of which pertain to biomedical computing, informatics, and the management and dissemination of biomedical knowledge. Some resource programs, such as Grants for Scholarly Works in Biomedicine and Health, are unique to NLM. Each year, NLM makes new and/or continuing awards in every active grant category.

- Research Project 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 and career development grants for informatics researchers
- 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 American Recovery and Reinvestment Act (ARRA), passed in February 2009, had a dramatic, immediate impact on EP activities in FY2009. NLM's share of ARRA funds was $82 million, designated to be spent largely on research grants in FY2009 and 2010. EP took the lead in planning and implementing NLM's ARRA spending plan. With no additional staff, EP more than doubled the number of new grant awards issued, responded to NIH Director's requests for program statements, used multiple approaches to communicate with potential applicants, and expanded its internal review and reporting activities. Data about ARRA awards are highlighted throughout the report, and a brief overview of EP's 2009 ARRA accomplishments is provided below.

Planned activities relating to the regular appropriations budget in FY2009 continued with a focus on sustaining success rates for NLM applicants and meeting new NIH targets for Early Stage Investigators and new investigators.

NLM'S EP FY2009 base budget was $64,945,000, with $12,137,000 allocated to contracts for the NN/LM. This represents a slight downturn in the grants budget from FY2008, which was $65,294,000.

In 2009, NIH refined target R01 grant commitment levels for new investigators by designating Early Stage Investigators (ESI) (those within 10 years of their terminal degree) as a special focus area. Two NIH targets for new investigator awards were set: (1) the success rate for new investigators obtaining R01 grants should be comparable to that of seasoned investigators, and (2) ESI should represent about 60 percent of the new investigator total. NLM met its targets for both goals through a mix of appropriations budget and ARRA-funded awards.

## Overview of ARRA at NLM

NIH received $10 billion from the American Recovery and Reinvestment Act (ARRA), of which $82 million was assigned to NLM for support of research grants and contracts. All ARRA funds must be committed by NIH no later than September 30, 2010. To meet this ambitious goal, an aggressive program of outreach to NLM's past and present grantees and recent applicants was undertaken, complementing NIH communications about ARRA goals. An initial estimated funding plan was developed, allocating funds to a variety of ARRA funding initiatives. NLM did not issue any ARRA funding announcements of its own, preferring to focus on several of the signature NIH programs, such as Challenge Grants and Summer Research Experiences. The menu of ARRA activities at NLM includes the following:

- NIH Challenge Grants in Health and Science Research
- Funding of reviewed meritorious grants from 2008 and 2009 that had not yet been awarded
- Administrative Supplements for accelerating pace of existing grants
- Competitive Supplements for expanding scope of existing grants
- Summer student & educator research internship supplements
- Restoration of approved but unfunded trainee slots to training programs
- Academic Research Enhancement Award (AREA) (RC4)
- BRDG-SPAN Small Business Pilot Program (RC3)
- Small Business Catalyst Awards (R43)

Some of these programs have completed their cycle of awards, while others are still in process. Additionally, although NLM did not participate in the NIH Grand Opportunity (GO) grant program, three GO grants were assigned to NLM for program oversight, all of them funded by funds from the NIH Director's Office.

**Table 11**

NLM ARRA Awards, 2009

Table 11 summarizes the first year of ARRA awards at NLM.

**Table 11: NLM ARRA Awards, 2009**

| ARRA Activity | Number of awards | Amount (in $1,000) |
|---|---|---|
| Challenge Grants | 16 | 15,021 |
| Grand Opportunity Grants | 3 | 10,937 |
| Fundable Grants not yet Awarded | 31 | 3,234 |
| Supplements to Accelerate/Expand Existing Grants | 58 | 7,602 |
| Training program supplements for Trainees & Equipment | 13 | 5,443 |
| Summer Research Experiences | 10 | 1,081 |
| **Total** | **131** | **$43,318[1]** |
| US States where NLM ARRA awardees are located | 31 | |
| Number of Summer Research Experience trainees | 67 | |
| Number of Informatics trainees, 2-year slots | 44 | |

[1] *Includes nearly $7 million in OD funds.*

As part of the NIH-wide ARRA strategy, each Institute was asked to identify one or more signature areas, where $10 million or more of ARRA funds would be spent. NLM identified two signature areas: Enhancing Electronic Health Records (the primary area) and Training Tomorrow's Informatics Researchers/Health IT Leaders.

**Success Rates of Grant Applicants**

Success rates are computed by dividing the number of awards by the number of applications in a fiscal year. This year, computation of success rates was complicated by the fact that EP gave awards from two different pools of funds and under some special conditions. For example, some previously reviewed grants that were not funded with regular appropriations dollars were funded with ARRA funds. Table 12 shows success rates in 2008 and 2009 for NLM's core grant programs for applications funded with and without ARRA funds. The success rate for R01 research grants funded by EP's base grants budget decreased this year for two reasons: (1) the number of R01 applications increased by 30 percent, from 73 in 2008 to 95 in 2009, without a comparable budget increase; and (2) some RO1 applications were paid with ARRA funds rather than base budget funds and are not counted for the success rate calculation. Success rates for KM/Applied Informatics and Scholarly Works grants were purposely constrained by program changes made to preserve funds for research grant programs. The KM/Applied Informatics and Scholarly Works grant programs moved to a single application deadline each year, with a fixed number of awards.

**Table 12: Success Rate, Core NLM Grant Programs, 2009, with and without ARRA**

| Grant Type | FY2008 | FY2009 w/o ARRA | FY2009 w/ ARRA |
|---|---|---|---|
| Research (R01) | 23% | 17% | 33% |
| KM/Applied Informatics (G08) | 7% | 10% | |
| Scholarly Works (G13) | 19% | 14% | |
| Career Transition (K99) | 38% | 50% | |

**Research Support for Biomedical Informatics and Bioinformatics**

As noted earlier, extramural research support is provided through a variety of grant mechanisms that fund investigator-initiated research. EP's research grants support both basic and applied projects involving the application of computer and information science approaches in clinical medicine, translational science, public health and basic biomedical research. EP's grant program trend and accomplishments are summarized below.

*Research Grant Program*

The R01 research grant program at EP has had two "branches": biomedical informatics (with focus on topics in clinical care, health services research, and public health), and bioinformatics (with focus on basic biomedical research areas such as systems biology, genomics and proteomics). An emerging trend is the rise in research applications that are focused on (1) visualization, integration, mining and sharing of published knowledge or data-banked information or (2) translational science linking biological research findings with clinical phenotypes.

ARRA provided all NIH Institutes with the opportunity to pay meritorious applications that had been received and reviewed but not paid in FY2008 or 2009. EP staff worked with applicants to revise budgets and timelines for selected applications.

At the end of FY2008, EP participated in a trans-NIH RFA on Exceptional, Unconventional Research Enabling Knowledge Acceleration (EUREKA) (R01). The purpose of the RFA was to foster exceptionally innovative research on novel hypotheses or difficult problems germane to the mission of the NLM. Twenty-two applications were received for NLM's two Eureka topics, and eight were funded using ARRA funds. In 2009, NLM is participating in EUREKA again, with a single topic: intelligent search tool for answering clinical and biomedical research questions.

- 95 reviewed R01 applications (73 in FY2008) (non-ARRA applications)
- 16 awarded R01 applications (15 in FY2008) (non-ARRA funds)
- 23 R01 applications awarded using ARRA funds, including 8 EUREKA

*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.

- 20 reviewed R03 applications (33 in FY2008)
- 0 awarded R03 applications (7 in FY2008) (non-ARRA funds)
- 6 R03 applications were awarded with ARRA funds

*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. Five of the FY2009 awards were made with ARRA funds.

- 32 reviewed R21 applications (36 in FY2008)
- 3 awarded R21 applications (7 in FY2008) (non-ARRA funds)
- 5 R21 applications awarded with ARRA funds

*Resource Grants for Biomedical Informatics/Bio-informatics*

The P41 program announcement to support scientific research resources was deactivated in FY2005, but the five existing P41 awardees remain eligible to apply to the program for continuation funding. In FY2009, two of the five existing grants were successfully renewed. The remaining three grantees will submit competitive renewal applications to a limited competition RFA in June 2010.

- 2 reviewed P41 applications (0 in FY2008) (non-ARRA funds)
- 2 awarded P41 applications (0 in FY2008) (non-ARRA funds)

*Conference Grants*

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

- 2 reviewed conference grants applications (3 in FY2008) (non-ARRA funds)
- 2 awarded conference grant application (1 in FY2008) (non-ARRA funds)

*IAIMS Testing & Evaluation Grants*

The IAIMS Testing and Evaluation grant program was closed in FY2008. A single application was received and reviewed in FY2009. No awards were made.

*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 percent of research grants budget) grants and Small Business Technology Transfer Research (STTR) (.5 percent 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 FY2009, 25 applications were "unscored" (out of 37 applications), indicating reviewer assessment that they were not competitive for funding.

- 37 reviewed small business applications (31 in FY2008) (non-ARRA funds)
- 3 awarded small business applications (1 in FY2008) (non-ARRA funds)

**Resource Grants**

Resource Grants support access to information, connect computer and communications systems, and promote collaboration in networking, integrating, and managing health-related information. The Knowledge Management/Applied Informatics series centers on optimizing the management of health-related information. These grants are not research grants and are reviewed with relevant criteria. The Scholarly Works grant program supports the preparation of scholarly manuscripts in health sciences and health public policy areas.

*Knowledge Management (KM) 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 open grant program was suspended, and an RFA for applied informatics grants was issued in FY2009. EP plans to issue annual RFAs in this program.

- 31 reviewed KM Applied Informatics applications (46 in FY2008)
- 3 awarded KM Applied Informatics applications (3 in FY2008)

*Integrated Advanced Information Management Systems (IAIMS) Planning Grant*

The IAIMS Planning grant program was closed in FY2008. Two applications were received and were reviewed in FY2009. No awards were made.

*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 open grant program was suspended, and the program moved to annual RFA issuances. The first of these was issued in FY2009.

- 59 reviewed Scholarly Works applications (52 in FY2008)
- 8 awarded Scholarly Works applications (10 in FY2008)

**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 training support and individual career transition 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 at 18 university-based programs across the country (see Table 4). This program is re-competed every five years. The latest applications were received in March 2006, and five-year awards were made in FY2007. One former NLM training program, 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) and supported two training slots in dental informatics at the University of Pittsburgh.

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. FY2009 was the final year in this four-year collaboration. At the June 2009 NLM Training Conference, the final NLM/RWJF cohort event was held. In FY2009, RWJF funds supported 10 pre- and post-doctoral trainees at the four sites. In summer 2009, RWJF approved use of $159,000 in unexpended funds

from the original grant. This award supports a transition year, during which faculty at the four programs will develop curriculum modules that can be incorporated into the required graduate coursework in schools of public health.

In 2007, NLM restructured its Short-Term Trainee Program (STTP) to focus the program on awards for minority or disadvantaged trainees, with the long-term goal of recruiting more minorities into informatics research careers. Eighteen STTP trainees were supported at seven programs in 2007/08, 18 STTP slots awarded for use in FY2008/09, and 17 slots awarded for use in FY2009/10. Programs may appoint STTP trainees at any time during the training year.

In 2008, with the help of a Congressional Supplement and a change to stipend table allocations, NLM's 18 active training programs were returned to their full awarded level of slots, 171 pre-doctoral and 72 post-doctoral slots, totaling 243. In FY2009, NLM received permission to use ARRA funds to restore slots that were Council-approved in the original applications but not funded in the initial awards. Two-year ARRA slots, with emphasis in clinical and public health informatics, were awarded to 10 programs. ARRA-funded slots will support 28 pre-doctoral and 16 post-doctoral trainees in 2009/10, and 29 pre-doctoral and 20 post-doctoral trainees in 2010/11.

**Table 13: T15 Trainees funded by NLM for FY2009**

| FY2009-2011 AWARDED T15 TRAINEES FUNDED BY NLM | | |
|---|---|---|
| **FISCAL YEAR** | **PRE-DOCTORAL** | **POST-DOCTORAL** |
| **2009-2010** | **28** | **16** |
| **2010-2011** | **29** | **20** |

**Table 14:**
**In FY 2009, NLM Sponsored Trainees at 18 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 Missouri-Columbia (Columbia, MO)
10. Columbia University Health Sciences (New York, NY)
11. Oregon Health & Science University (Portland, OR)
12. University of Pittsburgh at Pittsburgh (Pittsburgh, PA)
13. Vanderbilt University (Nashville, TN)
14. Rice University (Houston, TX)
15. University of Utah (Salt Lake City, UT)
16. University of Virginia Charlottesville (Charlottesville, VA)
17. University of Washington (Seattle, WA)
18. University of Wisconsin Madison (Madison, WI)

Every summer, all NLM-supported trainees attend a national informatics training conference. On June 23-24, 2009, the two-day meeting took place at the Oregon Health and Sciences University, Portland, Oregon. Research projects were presented in plenary and semi-plenary sessions by 39 informatics trainees. An additional 43 trainees presented posters at the meeting. There were 314 attendees, including directors, faculty, and staff from all NLM-funded informatics training programs; faculty and trainees from the Veteran's Administration informatics training sites; and NLM staff and guests. Two trainees received awards for the best poster, selected by their peers: Yalini Senathirajah of Columbia, for "Development and Evaluation of a Widget-based Web 2.0 Electronic Health Record" and Ben Compani of UC-Irvine, for "Sigmoid: an Integrative System for Pathway Bioinformatics and Systems Biology."

### Career Support

### K99/R00 Pathway to Independence

In January 2006, NIH announced a new career transition program, the NIH Pathway 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 program are not restricted to NLM's informatics trainees, they are preferred applicants.
- 6 reviewed K99/R00 applications (8 in FY2008)
- 4 awarded K99/R00 applications (3 in FY2008)

### 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 FY2009, EP funded 4 of 12 applications.
- 12 reviewed Loan Repayment Program applications (15 in FY2008)
- 4 awarded Loan Repayment Program applications (4 in FY2008)

### Pan-NIH Projects

### ARRA

The NIH strategy for ARRA involved EP participation in several new grant program initiatives. General programs included the NIH Challenge Grants, ARRA-related AREA and SBIR/STTR issuances, and grant supplements for research enhancement and Summer Research Experience awards. New topical foci included Comparative Evaluation Research, Community-based Research, and a new NIH initiative called OPP Net, designed to expand NIH's presence in social and behavioral science basic research.

### NCBC

The National Centers for Biomedical Computing (NCBC) program is funded through NIH Roadmap funds under renewable cooperative agreements. NLM provided an additional $800,000 per year to the NCBC program for the first five years, which ended in FY2008. An RFA for an open competition for the second five-year period of the NCBC was released on November 10, 2009. The NIH Roadmap funds for this program will diminish over years 6 through 10 of the program. Participating ICs will provide 100 percent of the funding for NCBCs by the final year of the program. Eleven NIH ICs are participating in the renewal RFA, including NLM. 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. EP participation is confined to topical programs which do not duplicate its existing grant programs. The active multi-institute programs NLM participates in are listed in **Table 17**. One new addition worth noting is the NIH Transformative T-R01 grant. Like the Eureka program in which NLM also participates, the T-R01 grants emphasize high-risk/high-reward research. The applications for multi-institute programs are reviewed by CSR, and participating institutes select grants for full or shared funding. Applications assigned to NLM that receive fundable merit scores 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

In FY2009, EP contributed approximately $1,043,660 million in collaborative co-funding agreements to other NIH Institutes and Centers. The listing below outlines the projects and amount of support NLM contributed to other ICs.

- Co-funding to the National Institute of General Medical Sciences (NIGMS) for the continued support of a cooperative agreement grant entitled "The PharmGKB: Catalyzing Research in Pharmacogenetics" ($310,664).
- Co-funding to the National Institute on Aging (NIA) for SBIR/STTR grant entitled "Anti-Aging

Mitochondrial Protein Therapy for Alzheimer's Disease and Sarcopenia" ($86,000).

- Co-funding to the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) for SBIR/STTR grant entitled "Clinical Validation of BCT – Phase II" ($216,555).
- Co-funding to the Fogarty International Center (TW) for International Informatics capacity building, for an award to OHSU entitled "Transitional Informatics for Global Health" ($224,541).

In FY2009, the NLM grants which received co-funding from other NIH agencies included the following:

- 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).
- Support from the NIMH for a grant supplement to Stanford University's Biomedical Research Resource P41 grant entitled, "A Resource for Biomedical Ontologies and Knowledge Bases", $160,000. This supplement support was for the development of autism ontologies.
- NLM received funding from NCMHD for two small research grants that are linked awards as part of the Partners In Research Program, $63,625.
- The OD/NIH provided full funding support for the National Center for Biocomputing (NCBC) entitled, "Informatics for Integrating Biology and the Bedside (i2b2), $3,931,786. This included grant supplemental additions in support of the i2b2 Natural Language Processing challenge conference and for postdoctoral support to train young researchers in Utah to install and test i2b2 tools.
- NLM received funding for support of two trainees in dental informatics from the National Institute of Dental and Craniofacial Research (NIDCR), $135,170. Funds were awarded as a supplement to the University of Pittsburgh's biomedical informatics training program.
- The Office of Research for 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 full funding support from the NIH OD for one linked research proposal at the UCLA Consortium for Neuropsychiatric Phenomics, an NIH Roadmap Interdisciplinary Center awardee. The research grant is entitled "Hypothesis Web Development for Neuropsychiatric Phenomics," $303,721.
- The NIH Office of the Director provided $999,457 (2 awards) for ARRA Challenge Grants and $5,866,079 (3 awards) for ARRA Grand Opportunities grants.

*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, $631,285. 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 project's specific aim it to improve medical guidelines for clinicians and to operationalize their use in a clinical setting.
- The NLM included, at end-of-fiscal year, two agreements in support of the Assistant Secretary's Office for Public Affairs, OS, DHHS and the Social Security Administration:
  - $55,179 co-funding from the Assistant Secretary for Public Affairs (ASPA), OS, HHS for NLM grant G08-LM-9026-01A2, Healthmap: Knowledge Management for Emerging Infectious Disease Intelligence. The funds are provided to integrate the Healthmap platform into FLU.GOV.
  - $153,012 from the Social Security Administration to R01 LM009723-01A1, "Capturing Patient-Provider Encounter through Text Speech and Dialogue Processing." The SSA is interested in the use of medical natural language processing (MNLP) technology in data mining for the adjudication of disability claims.

**Extramural Programs Web Site**

An entirely new section was created on the EP Web page for ARRA-related funding opportunities and other ARRA information, incorporating NIH graphic elements. A link on the NLM home page brings interested parties to the EP Web site for more information. The EP Web site (http://www.nlm.nih.gov/ep/funded.html) lists grants awarded since 1997. In 2009, the EP awards pages were enhanced by replacing links to the de-commissioned NIH CRISP system with a search 'widget' for access to abstracts, grant-related publications and other project information provided via the new NIH RePORTER system (http://projectreporter.nih.gov/reporter.cfm).ARRA-funded awards are listed in a separate section on the EP awards pages.

During FY2009, the EP Web site received approximately 186,000 page views, 82,000 visits, and 60,000 visitors, across roughly 127 pages.

**EP Operating Units**

*Program Office*

*Grant Program Development*: Program activities in FY2009 were focused on implementing EP's ARRA activities without neglecting the implementation of program plans for the regular grant programs. Participation in ARRA required new or expanded processes for communicating with potential applicants, selecting previously reviewed applications for awards, reviewing supplement requests, negotiating budgets, and answering hundreds of 'cold calls' from people who wanted to know more about NLM's 13 challenge grant topics. ARRA awards affected all aspects of program work, and each program staff member took on new responsibilities, to assure that awards were made in a timely way. Standard NIH processes also had to be employed for the new ARRA awards. For example, each ARRA award had to go through validity checks for the Research Condition and Disease Categorization (RCDC) fingerprinting. Our program unit's hard work related to ARRA activities was recognized with a Group Award at the NLM Honors Awards Ceremony.

One of our core resource grant programs, the G08 Knowledge Management/Applied Informatics grants, moved from three deadlines per year to one. EP's planned Challenge Grant initiative was launched by EP's participation in the GM's EUREKA program calling for high-risk, high-reward innovative applications. The NLM component requested applications in the areas of computational discovery and hypothesis testing. NLM participated in the 2010 EUREKA program as well, this time with an updated version of one of the ARRA challenge topics relating to intelligent information retrieval and summarization. SBIR/STTR grant programs for FY2009 focused on areas of particular interest from small business including Semantic Web technologies and intelligent search agent to improve information retrieval,

knowledge discovery, and management model for biological discovery and clinical decision support. The NIH timetable for transition to electronic applications has been completed for all of NLM's active grant programs except P41, and T15. Until the transition is completed, these applications continue to be submitted on paper.

*Program Staff Activities*: EP program staff represented NLM on various NIH and NLM standing committees, including Extramural Programs Management Committee, Program Leadership Committee, Training Advisory 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, BISTI, IMAG/Multi-Scale Modeling; NCBC Management Team, Trans-NIH Genomic Working Group, and a number of RCDC fingerprint "expert" committees. EP program officers attended and completed the NIH Senior Leadership Program.

*Program Oversight, Management & Evaluation*: EP completed its two-year contract with Humanitas, Inc. for two program assessment activities related to the university-based training programs.

Implementation of EP's eGrants system for managing electronic grant files is fully completed. The program analyst transferred responsibility for quality control of this new resource to DEAS staff, and completed work on the retention policy that determines when hardcopy grant files should be destroyed. In line with NIH-wide initiative, EP made the transition to eAdditions, which allows additional documents to be added to the grant folder in IMPACII. EP additions to the eGrants system have ceased, though eGrants remains an archive for scanned copies of older grants.

*Dissemination and Staff Activities*: Presentations were made to the NLM Board of Regents on the following topics: EP 2009 budget; Approval of Operating Procedures; 2009 Biennial Report on Compliance with Inclusion Guidelines; Ethics Policies of NLM Grantee Institutions; NIH Loan Repayment Program; American Recovery & Reinvestment Act and NLM Grant programs; Concept Review and Approval for Insight Toolkit Projects for ARRA Funding; Reporter – Expanded Information on NIH Grants; Concept Review of Research Contract for Artificial Intelligence. 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: Mr. Jonathan Chen, NLM Informatics Fellow, on Artificial Intelligence in Chemistry: "An Organic Computer for Predicting Organic Chemistry Reactions"; Dr. Matthew Scotch on "Zoonotic Disease Surveillance:

Combining Animal and Human Data"; Dr. Charles Caldwell on "What is Epigenetics?"; Dr. Lawrence Hunter on NLM University-based Training in Biomedical Informatics – Colorado Computational Bioscience Program; Dr. Stephanie Guerlain on University of Virginia Medical Informatics Training Program. Dr. John Pani, BLIRC Member, presented "I Thought Neuroanatomy was Difficult: Development of New Computer-Based Methods for Learning the Anatomy of the Brain" in the NLM Informatics Lecture Series. Dr. Dominik Aronsky, BLIRC Member, presented "Computer-based Decision Support in the Emergency Department" in the NLM Informatics Lecture Series.

## Grants Management Office

The NLM issued 127 non-competing grant awards and 47 competing grant awards in FY2009. In addition, we issued nine supplement awards and funded five other NIH – ICs through co-funding. The total base grants budget was $49,588,180. Other extramural awards supported the NIH Loan Repayment Program (LRP), four awards, one interagency agreement and the grants office provided the end-of-year accounting for the National Networks of Libraries of Medicine, nine awards as extramural contracts. These other extramural awards totaled $12,532,820 for a total of 14 awards.

The EP/NLM grants office provides the daily and end-of- fiscal year accounting for the EP budget including all awarding mechanisms. FY2009 required additional accounting of awards for tracking grant expenditures under the ARRA funding initiative. The grants office provides full fiscal support of the NIH LRP and approval for funding of the annual LRP contracts. Grants staff also provided support for the final year of the Public Health Informatics Training Program, four supplement awards, $668,826. NLM grants staff are responsible for the annual maintenance of the Conflict of Interest (COI) database and the *Freedom of Information (FOI) database.*

## NLM Grants Management Service Center

The NLM continued its support in FY2009 as a GMSC to the Office of the National Coordinator for Health Information Technology (ONCHIT), OS, DHHS. NLM staff were required to process all noncompeting renewal cooperative agreements for awards originally issued in FY2008, seven awards, $4,596,305. A new award was issued in support of the State-Level Health Information Exchange project, $1,294,919. In addition, a supplement for the Annual-American Hospital Association (AHA) Survey was issued in support of improving the survey framework, $54,840. Total Program FY2009: $6,095,057.

***The total GMSC support for the program includes the following:***

| | |
|---|---|
| National *e*-Health Collaborative (N*e*HC) & NLM Staffing Costs under GMSC | $4,047,020 |
| National Health Information Network (NHIN) | $599,427 |
| Information Technology Survey (AHA) | $98,751 |
| State Level Health Information Exchange Consensus Project | $1,294,919 |
| Information Technology Survey (AHA) | $54,840 |

*Scientific Review Office*

*Grant Review Activities*: Overall, 706 applications were reviewed for which NLM was the primary assignment. Of those, 244 were reviewed by NLM. The remaining 462 applications, including all ARRA Challenge grants, were reviewed by CSR. Most of NLM's non-ARRA grants reviewed by CSR were SBIR/STTR grant applications. Of the applications reviewed by NLM, 61 percent were in one of the four research grant mechanisms (100 R01s, including 21 Eureka grants, 24 R21s, and 2 R03s, and 3 R13s). Knowledge management grants represented 16 percent of the applications reviewed (40 apps.), Career Transitions Awards – 2 percent (5 apps.), and Scholarly Works at 22 percent (55 apps.) of the total applications reviewed. Other mechanisms reviewed included Competitive Revision applications (type 3s, 13) and Loan Repayment Program applications (LRPs, 12).

*BLIRC:* NLM's standing review group, the Biomedical Library and Informatics Review Committee (BLIRC), evaluates grant applications assigned to NLM for possible funding. BLIRC met three times in FY2009 and reviewed 127 applications (as compared to 158 in 2008). 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):* Nine Special Emphasis Panels were held during FY2009 compared to eight in 2008. These panels are convened on a one-time basis to review applications for which the BLIRC 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 117 applications during FY2009, compared to 100 in 2008.

*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, conducts early concurrence reviews electronically on lists 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. In 2009, the subcommittee held three early concurrence panels, including one covering ARRA Challenge Grants in the following areas: Behavior, Behavioral Change and Prevention; Comparative Effectiveness Research; Enabling Technologies; Enhancing Clinical Trials; Information Technology for Processing Health Care Data; Translational Science. The Board of Regents conducts an *en bloc* vote for all other applications that meet criteria for further consideration for funding. The EP subcommittee also reviews rebuttal requests from applicants who challenge the review of their grants. There were no rebuttals received in 2009. 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

*Human Resources Planning:* 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. FY2009 was a year of great change for EP, because after 20 years of service to EP, Dr. Milton Corn accepted a Senior Management position at NLM, thus creating a vacancy at the Associate Director position. In April 2009, EP Deputy Director Dr. Valerie Florance was appointed Acting Director of EP. In FY2009, EP began the recruitment process for two new personnel (one Administrative Officer, and one Program Analyst). We are happy to report that both vacancies were filled as of September 27, 2009. However, as EP gained 1 new program analyst we also lost another, who accepted a position at another NIH institute. EP planned recruitments for FY2010 will include hiring a Program Officer, a new Program Analyst and a Junior Committee Management Officer. Additionally, the permanent position for Associate Director, EP will be filled.

*COOP Planning:* In 2009, there was an increased emphasis on Emergency Preparedness Planning, also known as Continuity of Operations Planning (COOP). For the past couple of years, with assistance from NIH, NLM has been making incremental progress with internal efforts to develop a comprehensive plan. Two individuals from each NLM Division were selected as members of the newly-formed NLM COOP Committee. The

representatives from EP were Dr. Valerie Florance, Acting Associate Director, and Annmarie Carr, Chief Administrative Officer. The individuals posted to this committee had the authority to speak for their Program Director regarding program mission essential functions and could facilitate the collection of existing emergency response documents and/or development (as needed) of required, program specific, emergency response documentation. EP's representatives were specifically tasked with assessing the current status of emergency response preparedness for EP.

*Grants Administration Support:* Support for grants administration continues to be provided by five staff members from the NIH/OER/DEAS Division. However, in 2009, EP welcomed the addition of two temporary DEAS staff brought on board as a result of the American Recovery and Reinvestment Act (ARRA), which provided a small pool of funds for expanding staff to handle the increased workload. The DEAS staff also continue to assist the Review staff with the preparation and facilitation of grant review meetings; the Grants staff in tracking the NBS Acquisitions CAN errors and grants closeouts; the Admin staff in preparing and setting up of Video and Telephone Conferences; and the Program staff in administrative review of grant applications, mailings to grantees and applicants, and special projects.

*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). 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. EP continued with its newly revamped telework policy, and the program has an 80 percent participation rate.

*Facilities & Equipment:* In FY2009, the EP facilities had many improvements, including the upgrade of the EP Video Teleconference center to 100 percent High Definition, the much needed painting of office walls, the installation of new digital desk phones for the entire staff, the encryption of all staff desktop computers, and the installation of new carpeting throughout the entire office suite is scheduled for November 2009. Another big change to the EP landscape in 2009 was the redesign of the old EP Grants File room. The renovation of this space was completed in early October 2009. It will provide workspace for two staff members and, at the same time, continues to hold the dwindling collection of hardcopy grant applications. The office "facelifts" of this past year were welcomed by the EP staff, as the office has not seen upgrades to its basic facilities in 12 years.

## Table 15:
## Extramural Programs Grants Budget by NIH Mechanism Groupings, and by Activity Code, FY2009

(dollars in thousands)

| FY 2009 operating budget request by NIH mechanism groupings | FY 2009 Budget Total | |
|---|---|---|
| | No. | Amount |
| **Research Project Grants** (R01, R03, R21, R00, RL1, U01) | 87 | 26,097,666 |
| **SBIR/STTR** (R41, R42, R43, R44) | 6 | 805,563 |
| **Research Centers - Specialized/Comprehensive** (U54) | - | - |
| **Other Research - Research Careers** (K99) | 6 | 628,693 |
| **Other Research - Biomedical Research Support** (P41) | 5 | 2,782,625 |
| **Other Research - Other** (G08, G13, R13, R24, D43) | 45 | 3,994,449 |
| **Training - Individual** (F37) | 2 | 125,864 |
| **Training - Institutional** (T15) | 23 | 15,153,320 |
| **R&D Contracts** (L30, L40, N01, Y01) | 15 | 12,532,820 |
| | | |
| **EP budget excluding TAPS and RMS/Direct Operations** | **189** | **62,121,000** |
| | | |

(dollars in thousands)

| FY 2009 operating budget request by activity code | FY 2009 Budget Total | |
|---|---|---|
| | No. | Amount |
| **D43:** International Training Grants in Epidemiology (cofund) | 1 | 224,541 |
| **F37:** Individual Informationist Fellowships | 2 | 125,864 |
| **G08:** Knowledge Management & Applied Informatics; Planning Grant for IAIMS | 16 | 2,070,089 |
| **G13:** Scholarly Works in Biomedicine and Health | 23 | 1,600,155 |
| **K99:** Pathway to Independence | 6 | 628,693 |
| **L30:** Extramural Loan Repayment Program | 3 | 137,050 |
| **L40:** Extramural Loan Repayment Program | 1 | 58,770 |
| **N01:** NN/LM Contracts | 9 | 12,137,000 |
| **P41:** Biomedical Resource Grant | 5 | 2,782,625 |
| **R00:** Pathway to Independence | 2 | 495,368 |
| **R01:** Research Project Grants | 67 | 22,998,867 |
| **R03:** Small Project Grants | 7 | 444,552 |
| **R13:** Conference Grants | 5 | 99,664 |
| **R15:** Academic Research Enhancement Award (AREA) | - | - |
| **R21:** Exploratory/Developmental Grants | 10 | 1,848,315 |
| **R24:** IAIMS Test & Evaluation | - | - |
| **R41:** Small Business Technology Transfer (STTR) | - | - |
| **R42:** Small Business Technology Transfer (STTR) | 1 | 86,000 |
| **R43:** Small Business Innovation Research (SBIR) | 4 | 503,008 |
| **R44:** Small Business Innovation Research (SBIR) | 1 | 216,555 |
| **R56:** Director's Bridge Award | - | - |
| **RL1:** Linked Research Project Grant | - | - |
| **T15:** University Biomedical Informatics Research Training Programs | 23 | 15,153,320 |
| **U01:** Cooperative Agreement | 1 | 310,564 |
| **U54:** NCBC Roadmap Center | - | - |
| **Y01:** Inter-Agency Agreement | 2 | 200,000 |
| | | |
| **EP budget excluding TAPS and RMS/Direct Operations** | **189** | **62,121,000** |
| | | |

**Table 16:**

**FY2009 EP Budget, Dollars in Thousands**



■ Training - Institutional
$15,154

■ Research Project Grants
$26,098

Other Research - Other
$3,995

■ Other Research -
Biomedical Research
Support
$2,783

■ Training - Individual
$126

■ Other Research -
Research Careers
$629

■ R&D Contracts
$12,533

■ Small Business
$806
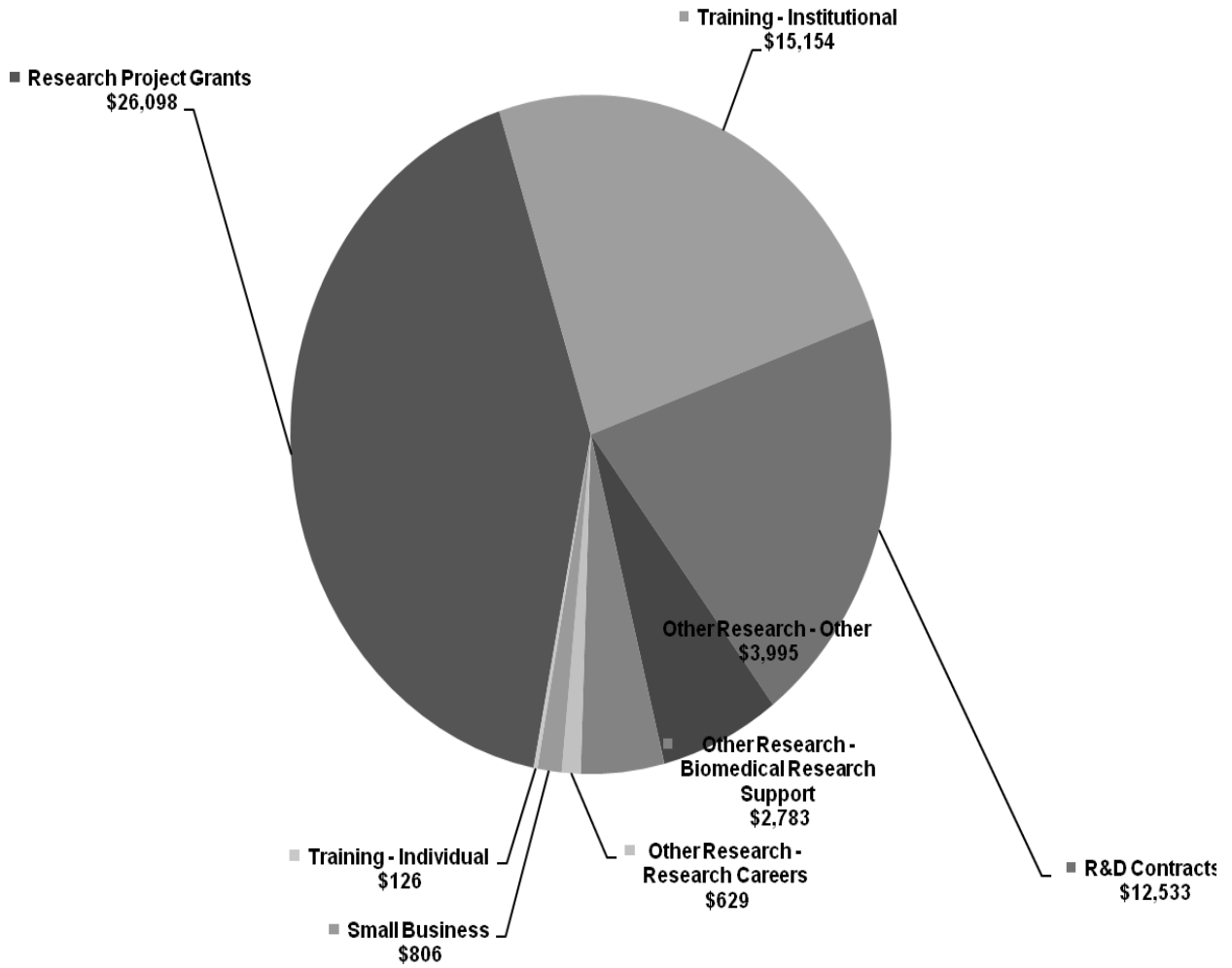
# Table 17:
# RFA/PA Actions in FY2009

| NLM'S Core Active Grant Programs | | |
|---|---|---|
| **Announcement** | **Title** | **Expiration** |
| PA-07-070 | Research Project Grant (Parent R01) | 1/8/2010 |
| PAR-08-080 | NLM Express Research Grants in Biomedical Informatics (R01) | 5/8/2011 |
| PA-09-164 | NIH Exploratory/Developmental Research Grant Program (Parent R21) | 5/8/2012 |
| PA-08-149 | NIH Support for Conferences and Scientific Meetings (Parent R13/U13) | 5/8/2011 |
| PA-09-036 | NIH Pathway to Independence Award (K99/R00) | 1/8/2012 |
| PA-08-191 | Research Supplements to Promote Re-Entry into Biomedical and Behavioral | 9/30/2011 |
| RFA-LM-09-001 | NLM Applied Informatics Grants (G08) | 7/2/2009 |
| PAR-09-030 | NLM Grants for Scholarly Works in Biomedicine and Health (G13) | 2/3/2010 |
| PA-06-042 | Academic Research Enhancement Award (Parent R15) | 1/8/2010 |
| PA-09-080 | PHS 2009-02 Omnibus Solicitation of the NIH, CDC, FDA and ACF for Small Business Innovation Research Grant Applications (Parent SBIR [R43/R44]) | 1/8/2010 |
| PA-09-081 | PHS 2009-02 Omnibus Solicitation of the NIH for Small Business Technology Transfer Grant Applications (Parent STTR [R41/R42]) | 1/8/2010 |
| NOT-OD-08-083 | NIH Extramural Loan Repayment Programs (LRP) | annual in December |
| | | |
| **NIH Programs in which NLM Participates** | | |
| **Announcement** | **Title** | **Expiration** |
| PAR-07-019 | Understanding and Promoting Health Literacy (R03) | 1/26/2010 |
| PAR-07-020 | Understanding and Promoting Health Literacy (R01) | 1/26/2010 |
| PAR-08-023 | Predictive Multiscale Models of the Physiome in Health and Disease (R01) | 9/16/2010 |
| PAR-09-129 | Solicitation of Assays for High Throughput Screening (HTS) in the Molecular Libraries Probe Production Centers Network (MLPCN) (R03) | 1/5/2012 |
| PAR-09-220 | Innovations in Biomedical Computational Science and Technology Initiative | 9/8/2012 |
| PAR-09-221 | Innovations in Biomedical Computational Science and Technology Initiative | 9/8/2012 |
| PAS-07-381 | Advancing Novel Science in Womens Health Research (ANSWHR) [R21] | 1/8/2010 |
| PAS-07-382 | Advancing Novel Science in Women's Health Research (ANSWHR) (R03) | 1/8/2010 |
| RFA-GM-10-009 | Exceptional, Unconventional Research Enabling Knowledge Acceleration | 11/25/2009 |
| RFA-OD-09-010 | Recovery Act Limited Competition: Building Sustainable Community-Linked Infrastructure to Enable Health Science Research (RC4) | 12/12/2009 |

# Table 18: NLM New Grants Awarded in FY2009

(*Sorted by PI name, within each grant category*)

**Appropriations Budget Awards**

**RESEARCH GRANTS**

ALTMAN, RUSS BIAGIO
2-R01-LM005652-14
STANFORD UNIVERSITY
Annotating Functional Sites in 3D Biological Structures

ASSELBERGS, FOLKERT WOUTER; MOORE,
JASON H. (CONTACT); WILLIAMS, SCOTT
MATTHEW (New Investigators)
1-R01-LM010098-01
DARTMOUTH COLLEGE
Bioinformatics Strategies for Genome-Wide Association
Studies

CHANG, CHUNG-CHE ; ZHOU, XIAOBO
(CONTACT) (New Investigators)
1-R01-LM010185-01
METHODIST HOSPITAL RESEARCH INSTITUTE
System Biology Approach for Signaling Transduction
Study of Complex Phenotypes

HAUSKRECHT, MILOS (New Investigator)
1-R01-LM010019-01A1
UNIVERSITY OF PITTSBURGH AT PITTSBURGH
Using medical records repositories to improve the alert
system design

HRIPCSAK, GEORGE M.
2-R01-LM006910-10
COLUMBIA UNIVERSITY HEALTH SCIENCES
Discovering and Applying Knowledge in Clinical
Databases

HUNTER, LAWRENCE E
2-R01-LM009254-04
UNIVERSITY OF COLORADO DENVER
Biomedical Language Processing Writ Large: Scaling to
all of PubMedCentral

KRAUTHAMMER, MICHAEL (New Investigator)
1-R01-LM009956-01A1
YALE UNIVERSITY
Advancing Literature Mining through Image Processing
and Analysis

LU, XINGHUA
1-R01-LM010144-01
MEDICAL UNIVERSITY OF SOUTH CAROLINA
Statistical methods for integromics discoveries

MALIN, BRADLEY A (New Investigator)
1-R01-LM009989-01A1
VANDERBILT UNIVERSITY
Technologies to Enable Privacy in Biomedical Databanks

MALIN, BRADLEY A (New Investigator)
1-R01-LM010207-01
VANDERBILT UNIVERSITY
Automated Detection of Anomalous Accesses to
Electronic Health Records

SCHADOW, GUNTHER
1-R01-LM009897-01A1
INDIANA UNIV-PURDUE UNIV AT INDIANAPOLIS
Clinical Knowledge Hub - Conceptual Integration of
Rules, Data Sets, and Queries

SCHNEEWEISS, SEBASTIAN G.
1-R01-LM010213-01
BRIGHAM AND WOMEN'S HOSPITAL
Analyzing Complex Healthcare Data to Determine
Causality of Observed Drug Effects

WENG, CHUNHUA
1-R01-LM009886-01A1
COLUMBIA UNIVERSITY HEALTH SCIENCES
Bridging the Semantic Gap Between Research Eligibility
Criteria and Clinical Data

ZENG, QING
1-R01-LM009966-01A1
BRIGHAM AND WOMEN'S HOSPITAL
Improve Discharge Instruction
Through Automated Pictograph Enhancement

**KNOWLEDGE
MANAGEMENT/APPLIED
INFORMATICS GRANTS**

BROWNSTEIN, JOHN S. (New Investigator)
1-G08-LM009776-01A2
CHILDREN'S HOSPITAL BOSTON
HealthMap: Knowledge Management for Emerging
Infectious Disease Intelligence

WALJI, MUHAMMAD (New Investigator)
1-G08-LM010075-01
UNIVERSITY OF TEXAS HLTH SCI CTR HOUSTON
Development of an Inter-University Oral Health Research
Database

## SCHOLARLY WORKS

BU, LIPING
1-G13-LM009601-01A1
ALMA COLLEGE
Public Health Education and Campaigns in China, 1910-1990

DE VOS, PAULA SUSAN (New Investigator)
1-G13-LM010089-01
PAULA DE VOS
The Craft of Medicine in Colonial Mexico: The Art and Science of the Apothecary

HALPERN, SYDNEY A. (New Investigator)
1-G13-LM010000-01
UNIVERSITY OF ILLINOIS AT CHICAGO
Human Hepatitis Experiments in the United States, 1942-1972

KAHN, JONATHAN
1-G13-LM010073-01
HAMLINE UNIVERSITY
Race in a Bottle: Law, Commerce and the Production of Racial Categories in Biomedicine

MATLIN, KARL S
1-G13-LM009860-01A1
UNIVERSITY OF CHICAGO
A History of Cell Biology from the Perspective of the Signal Hypothesis

MCCUSKER, KRISTINE MARIE (New Investigator)
1-G13-LM010074-01
MIDDLE TENNESSEE STATE UNIVERSITY
"Just Enough to Put Him Away Decent": The Management of Death and the Evolution of Public Health Policy in the South, 1918-1945

VOEKS, ROBERT ALLEN (New Investigator)
1-G13-LM009716-01A1
CALIFORNIA STATE UNIVERSITY FULLERTON
Roots of Rainforest Medicine

## CAREER DEVELOPMENT

KALPATHY-CRAMER, JAYASHREE (New Investigator)

1-K99-LM009889-01A1
OREGON HEALTH AND SCIENCE UNIVERSITY
Clinical Image Retrieval: User Needs Assessment, Toolbox Development & Evaluation

NOVAK, LAURIE L (New Investigator)
1-K99-LM010038-01A1
VANDERBILT UNIVERSITY
Patient Safety Through Implementation Science: Clinical Work Culture & IT Design

SOLTI, IMRE (New Investigator)
1-K99-LM010227-01
UNIVERSITY OF WASHINGTON
Increasing Clinical Trial Enrollment: A Semi-Automated Patient Centered Approach

## SMALL BUSINESS GRANTS

BROMBERG, YANA (New Investigator)
1-R43-LM010156-01
BIOSOF, LLC
Improved Manuscript Search Through PubSeq

CARNEY, DONALD P (New Investigator)
1-R43-LM010746-01A1
FLUIDITY SOFTWARE, INC.
Adapting Mathematical Pen-Computing for Classroom Use on Interactive Whiteboards

REININGER, DANIEL J
3-R44-LM008474-03S2
SEMANDEX NETWORKS, INC.
Medical Emergency Disaster Response Network

TURNER, WES (CONTACT); WILLIAMS, MICHELLE A (New Investigators)
1-R43-LM010245-01
KITWARE, INC.
Flexible Information Visualization and Analysis Platform for Biomedical Research

ZHU, YUERONG
5-R43-LM009913-02
BIOINFORX, INC.
Development of a Highly-Automated Microarray Data Analysis System That Allows Re-analyze Deposited Microarray Data with New Algorithms

# Table 19:  NLM New Grants Awarded in FY2009
(*Sorted by PI name, within each grant category*)

## American Recovery and Reinvestment Act Awards

**CHALLENGE GRANTS**

BARRETT, JEFFREY SCOTT
1-RC1-LM010367-01
CHILDREN'S HOSPITAL OF PHILADELPHIA
Decision Support System to Guide
Pediatric Pharmacotheraphy

BATES, DAVID W
1-RC1-LM010526-01
BRIGHAM AND WOMEN'S HOSPITAL
Refining IT Support for Genetics in Medicine

BROWN, JEFFREY STUART; KULLDORFF,
MARTIN (contact)
1-RC1-LM010371-01
HARVARD PILGRIM HEALTH CARE, INC.
Data Mining Electronic Health Records for Drug
Adverse Events

COBB, LOREN (New Investigator)
1-RC1-LM010641-01
UNIVERSITY OF COLORADO DENVER
Improved Tracking for Emerging Diseases from Climate
Change

DAS, AMARENDRA K; WALDINGER, RICHARD J.
(contact)
1-RC1-LM010583-01
SRI INTERNATIONAL
Intelligent Question Answering in the HIV Domain (SRI
Proposal ECU 09-630)

HOLMES, JOHN H (New Investigator)
1-RC1-LM010342-01
UNIVERSITY OF PENNSYLVANIA
Mining Internet Conversation For Evidence Of Herbal
Association

KEMPE, ALLISON (New Investigator)
1-RC1-LM010513-01
UNIVERSITY OF COLORADO DENVER
Population versus Practice-Based Interventions to
Increase Immunization Rates

KOHANE, ISAAC S.; MANDL, KENNETH D (contact)
1-RC1-LM010470-01
CHILDREN'S HOSPITAL BOSTON
Presenting Genome Information in Patient Electronic

Records: The Informed Cohort

LOUWERSE, MAX (New Investigator)
1-RC1-LM010442-01
UNIVERSITY OF MEMPHIS
The Importance of Language Characteristics in
Documenting Clinical Encounters

MILLER, MARLENE ROSEMARY
(New Investigator)
1-RC1-LM010512-01
JOHNS HOPKINS UNIVERSITY
Improving Childhood Immunization Compliance Using
Electronic Health Records

RICHESSON, RACHEL L (New Investigator)
1-RC1-LM010455-01
UNIVERSITY OF SOUTH FLORIDA
Library of Standardized Patient Registry Questions for
Rare Diseases

SAVOVA, GUERGANA (New Investigator)
1-RC1-LM010608-01
MAYO CLINIC COLL OF MEDICINE, ROCHESTER
Multi-source clinical Question Answering system

SCHNEEWEISS, SEBASTIAN G.
1-RC1-LM010351-01
BRIGHAM AND WOMEN'S HOSPITAL
"Stopping Rules" for Drug Safety Monitoring Using
Healthcare Databases

TURCHIN, ALEXANDER (New Investigator)
1-RC1-LM010460-01
BRIGHAM AND WOMEN'S HOSPITAL
Natural Language Processing to Study Epidemiology of
Statin Side Effects

WETZEL, RANDALL C (New Investigator)
1-RC1-LM010639-01
CHILDREN'S HOSPITAL LOS ANGELES
Advanced Computational Framework for Decision
Support in Critically Ill Children

WHEELER, ARTHUR P (New Investigator)
1-RC1-LM010310-01
VANDERBILT UNIVERSITY
A Model-integrated, Guideline-driven Process
Management System for Sepsis

**GRAND OPPORTUNITIES GRANTS**

MOSKOWITZ, JAY
1-RC2-LM010796-01
UNIVERSITY OF SOUTH CAROLINA AT
COLUMBIA
An Open Source Research Permissions Framework for
South Carolina

NARUS, SCOTT P
(New Investigator)
1-RC2-LM010798-01
UNIVERSITY OF UTAH
Development of a Statewide Master Person Index

TERDIMAN, JOSEPH (New Investigator)
1-RC2-LM010797-01
KAISER FOUNDATION RESEARCH INSTITUTE
The Kaiser Permanente National Research Database

**Other ARRA Research Grants (R01, R03, R21)**

ALEXOV, EMIL GEORGIEV (New Investigator)
1-R03-LM009748-01A2
CLEMSON UNIVERSITY
The Effect of Single Nucleotide Polymorphisms on
Protein Structure, Function and Interactions

ASH, JOAN S.
2-R01-LM006942-08
OREGON HEALTH AND SCIENCE UNIVERSITY
Clinical Decision Support in Community Hospitals:
Barriers & Facilitators

BLALOCK, TRAVIS N (New Investigator)
1-R21-LM009763-01
UNIVERSITY OF VIRGINIA CHARLOTTESVILLE
Wireless Real-Time Monitoring of Research Animals

BRY, LYNN (New Investigator)
1-R01-LM010100-01
BRIGHAM AND WOMEN'S HOSPITAL
Crimson - i2b2 integration for high-throughput, scalable
sample collection

CARROLL, AARON E (New Investigator)
1-R01-LM010031-01
INDIANA UNIV-PURDUE UNIV AT INDIANAPOLIS
Computer Decision Aid for ADHD Management
(CDAAM)

CLERMONT, GILLES
1-R21-LM009936-01A1
UNIVERSITY OF PITTSBURGH AT PITTSBURGH
Quantitative, Model-based Medical Decision support by
Bayesian Inference

COOPER, GREGORY F
1-R01-LM010020-01
UNIVERSITY OF PITTSBURGH AT PITTSBURGH
Predicting Patient Outcomes from Clinical and Genome-
Wide Data

CRANGLE, COLLEEN ELIZABETH (New Investigator)
1-R03-LM009752-01
CONVERSPEECH, LLC
Beyond information extraction: Identifying Gene Ontology
concepts in text

CROWLEY, REBECCA S
2-R01-LM007891-05
UNIVERSITY OF PITTSBURGH AT PITTSBURGH
Computational Methods for Personalized and Adaptive
Cognitive Training

DAS, AMARENDRA K.
1-R01-LM009607-01A2
STANFORD UNIVERSITY
Open-Source toolkit for Knowledge-Based Querying of
Time-Oriented Data

ELHADAD, NOEMIE (New Investigator)
1-R01-LM010027-01
COLUMBIA UNIVERSITY HEALTH SCIENCES
An NLP Approach to Generating Patient Record
Summaries

EVANS, JAMES ALLEN; FOSTER, IAN ; RZHETSKY,
ANDREY (contact)
1-R01-LM010132-01
UNIVERSITY OF CHICAGO
Large-scale Discovery of Scientific Hypotheses;
Computation Over Expert Opinions

GRAUR, DAN
1-R01-LM010009-01
UNIVERSITY OF HOUSTON
Error Correction in Multiple Sequence Alignments

GRISWOLD, WILLIAM G (New Investigator)
1-R01-LM009522-01A1
UNIVERSITY OF CALIFORNIA SAN DIEGO
WIISARD
SAGE: Self-Scaling Systems for Mass Casualty
Management

GURCAN, METIN NAFI (New Investigator)
1-R01-LM010119-01
OHIO STATE UNIVERSITY
OAMiner: Integrative Knowledge Anchored Hypothesis
Discovery

JAGADISH, H V
1-R01-LM010138-01
UNIVERSITY OF MICHIGAN AT ANN ARBOR

Literature and Data Driven Hypothesis Generation for High Throughput Experiments

KOHANE, ISAAC S.
1-R01-LM010125-01
HARVARD UNIVERSITY (MEDICAL SCHOOL)
Preventing the Incidentalome

LI, JING
2-R01-LM008991-04
CASE WESTERN RESERVE UNIVERSITY
Multi-point and Multi-locus Analysis of Genomic Association Data

LIU, HONGFANG (New Investigator)
1-R01-LM009959-01A1
GEORGETOWN UNIVERSITY
Onto-BioThesaurus: Ontological Representation of Gene/Protein names for Biomedical Literature

LOGAN, JUDITH R (New Investigator)
1-R21-LM009550-01A1
OREGON HEALTH AND SCIENCE UNIVERSITY
Exploiting the User Interface for Data Integration in Effectiveness Research

LU, GUOQING (New Investigator)
1-R01-LM009985-01A1
UNIVERSITY OF NEBRASKA OMAHA
A Computational Genotyping System for Improved Influenza Surveillance

MA, SHUANGGE (New Investigator)
1-R03-LM009754-01A1
YALE UNIVERSITY
Effective Clustering Penalized Methods for Genomic Biomarker Selection

QUACKENBUSH, JOHN
1-R01-LM010129-01
DANA-FARBER CANCER INSTITUTE
Integrated Approaches to Deriving Predictive Networks from Public Data Sources

RABADAN, RAUL (New Investigator)
1-R01-LM010140-01
COLUMBIA UNIVERSITY HEALTH SCIENCES
Integrated Discovery and Hypothesis Testing of New Associations in Rare Diseases

RITCHIE, MARYLYN D. (New Investigator)
1-R01-LM010040-01

VANDERBILT UNIVERSITY
Analysis Tool for Heritable and Environmental Network Associations

ANDREI (New Investigator)
1-R03-LM009738-01A1
UNIVERSITY OF TEXAS HLTH SCI CTR HOUSTON
Multivariate Analysis of Candidate Blood Pressure Response Genes in Hypertensives

ROSENBERG, MICHAEL S
1-R01-LM009505-01A1
ARIZONA STATE UNIVERSITY-TEMPE CAMPUS
Sequence Alignment Fidelity in Genomics and Bioinformatics

ROST, BURKHARD
1-R03-LM009739-01
COLUMBIA UNIVERSITY HEALTH SCIENCES
Novel Method to Identify Competing Protein-protein Binders

SZALLASI, ZOLTAN (New Investigator)
1-R03-LM009979-01
CHILDREN'S HOSPITAL BOSTON
Increasing the Reliability of Clinical Microarray Data Analysis by Systematic Bias Correction

TAIRA, RICKY K.
1-R01-LM009961-01
UNIVERSITY OF CALIFORNIA LOS ANGELES
Data Structuring and Visualization System for Neuro-oncology

TONELLATO, PETER J.
1-R01-LM010130-01
HARVARD UNIVERSITY (MEDICAL SCHOOL)
Method for Prediction of Efficacy of Genetic-Based Prediction Models of Personalized Medicine with Clinical Avatars

VERSPOOR, KARIN MARIA (New Investigator)
1-R01-LM010120-01
UNIVERSITY OF COLORADO DENVER
Automated Literature Mining for Validation of High-Throughput Function Prediction

WANG, YU-PING (New Investigator)
1-R21-LM010042-01
UNIVERSITY OF MISSOURI KANSAS CITY
A New Paradigm for Integrated Analysis of Multiscale Genomic Imaging Datasets

# 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.

The following describes OCCS accomplishments in FY2009:

## Consumer Health

*MedlinePlus*: From a systems design perspective, MedlinePlus' greatest achievement for FY2009 was the new HealthDay news video feature. In the first month of release, 44 videos were added, one per day from Monday to Saturday. Links to the videos will appear on the Health Topic pages under Latest News, News-by-Date, and News-by-Topic pages. The News-by-Date page will also feature an embedded player displaying the most recent video. A new video input system module was added to accommodate the HealthDay TV videos. The input system module is designed to allow Library Operations to upload videos, generate SWF files and public HTML pages, edit video URLs, create approved press release site records for the videos, and select a current pick for the News-by-Date page.

There were three versions of MedlinePlus released, version 21.2 in October, version 21.3 in January and 22.0 in September. These releases included enhancements to the MedlinePlus search functions with improved export module, and to include the HealthDay news video.

MedlinePlus Go Local implemented three releases, version 4.2 in October, version 4.5 in December and 5.0 in May. These releases included replacing the MapQuest capability with Google Map, and adding the ability to delete incomplete, approved, returned and audit records queues. The ability to import vocabulary, service areas, and display status fields was also implemented. MedlinePlus added Go Local sites serving Tennessee, DC, Montana, Kansas, Oregon, Oklahoma and Louisiana. Forty-six percent of the US population is covered by Go Local.

A new, NLM-developed MedlinePlus tutorial was released in March, to teach laypersons about medical terminology and acronyms.

A proof of concept Web service was created. The service provides a way for public entities to request the URL of MedlinePlus Topics pages that are a best match to the ICD9 code they supply. This pilot is being conducted in conjunction with The New York City based Institute for Family Health and their MyChart® Electronic Health Record and developed/implemented by Epic. To date there are 3,600 ICD9 codes mapped to the 800 MedlinePlus topics. Also added capability to return Spanish Medical Topics and expanded pilot to create an HL7-compliant Infobutton variant of the service which will serve as a resource node for Electronic Health Records that follow the HL7 standard. This service returns the links as well as summaries and will work in both English and Spanish.

MedlinePlus recognized over 643 million page views and over 121 million unique visitors. Additionally, MedlinePlus and Spanish MedlinePlus each received a score of 85 on the ACSI E-Government Satisfaction Index, a survey that tracks trends in customer satisfaction.

*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. In FY2009, the DailyMed system:

- Added over 1,000 new labels/package inserts.
- Added/updated 894 SPL labels.
- Published the Product Identification System.
- Published more than 5,167 prescription drug labels for public access.
- Completed extracting the Drug Delivery Devices (DDD) imprint data.
- Modified DailyMed processing system to receive & process Animal drug labels.
- Recognized more than two million unique visitors.
- Increased usage by 287 percent from 16 million to over 62 million page views.

*NIHSeniorHealth Project*: NIHSeniorHealth is a joint NLM and National Institute on Aging (NIA) project that

provides health information on the Web using modes of delivery, such as 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 43 topics now available in SeniorHealth, many new topics were added including Complementary and Alternative Medicine (CAM), Medicare Basics for Caregivers and Leukemia. Audio review and text-to-speech pronunciation adjustments were completed for several topics. Telling Show was implemented, which is an application embodying speech, data mapping, tactile components, 3D, information retrieval/extraction, rich media interface and Web services technologies. NIHSeniorHealth recognized over 21 million page views and over 865,000 unique visitors with 19.47 percent being International visitors.

The Accent module received numerous enhancements including Java Database Connectivity (JDBC) libraries configured with Oracle 10g, and auto-restart services and scripts. Application upgrade included renewed annual license for Loquendo voice font (Susan/Dave), and Loquendo license (PIK) patches.

## 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.

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

Federal regulations mandate that systems be re-accredited every three years. A System Testing and Evaluation (ST&E) was conducted by an independent vendor to validate a number of random management, operational and technical controls for the Certification and Accreditation (C&A) of NLM's MEDLARS and TOXNET systems. The analysis process was tailored to meet the requirements of the Office of Management and Budget (OMB) Circular A-130, National Institute of Standards and Technology (NIST) guidance, and DHHS/NIH security processes and procedures including HHS C&A Checklist. The evaluation was completed in August 2009 with certifications issued authorizing continued operations of MEDLARS and TOXNET.

In support of the NLM Continuity of Operations Plan (COOP), refinement and improvement of the living document continued through lessons learned to ensure that essential NLM functions will be maintained under all circumstances.

NLM continued to support the NIH Enterprise Risk Management Program. The NIH Risk Management Program is an ongoing effort to perform standardized repeatable activities that promote the overall efficiency, effectiveness, accountability and integrity of the organization's work.

Due to the increase in new vulnerabilities and the rapid emergence of associated threats, the deployment of software patches is more frequent and more urgent than ever before. In FY2009 NLM's automated patch management program applied over 83,000 patches on commodity desktops fixing known vulnerabilities to software.

Anti-virus screening was implemented for all outgoing Web (HTTP) connections to the Internet. As a result of this new initiative, the number of virus threats at desktop computers decreased by 93 percent from 672 in FY2008 to 48 in FY2009. This Web anti-virus service has blocked over 1,500 security violations, minimizing the burden on desktop security software to fend off threats to desktop computers.

More than 950 desktop computers were updated with over 4.5 million signatures, a 275 percent increase from FY2008; and over 10.8 million intrusion detection system alerts were handled, an increase of the more than 3.3 million alerts in FY2008.

In order to support the NLM-wide vulnerability management program, including compliance monitoring, centralized logging and security event management, Tenable Security Center was implemented as a new vulnerability management solution tool with significant additional vulnerability information being provided from credentialed scans.

To comply with DHHS' policy on encryption and data security, NLM initiated a policy to have a 'full-disk' encryption if the desktop stores sensitive information. NLM has completed 93 percent encryption of required desktops. In addition, all Apple Macintosh laptops will be encrypted with Apple FileVault.

The OMB requires that HHS computer users complete annual IT security awareness training. NLM has completed 100 percent of the mandatory FY2009 Security Awareness Training for employees, contractors and fellows. A NLM policy requires completion of the training within five days for new hires.

## High Speed Communication Network

Reliable LAN and Internet communications services continued, 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. Steps were taken to increase the capabilities and reliability of network services and storage by providing for the following:
- NIH Consolidated Collocation Site (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 Level 3. Internet connectivity was upgraded from an OC3 (155 Mbps) circuit to a 1 Gigabit Ethernet (GigE) circuit in May of 2008 then to a full Gigabit connection in September of 2009. 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 from 1 Gigabit 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 NLM perimeter network provides a 10 Gigabit security boundary to aggregate connections to NIHnet, Internet, Internet-2, and the NCCS. It also provides for 10 Gigabit connections to interconnect OCCS, LHC, and NCBI divisional networks. A 1 Gigabit private, dark fiber connection has been deployed from the NLM main campus site to NLM satellite sites EP, SIS and OAM.

State of the art firewalls guard the network at the perimeter and lower layers within the local network. The NLM perimeter and NCCS firewalls were upgraded from the legacy 2 Gigabit platform to modern 10 Gigabit appliances. A technology refresh was also conducted on public (second layer) firewalls.

NLM's Citrix remote access capability increased by 16 percent. Registration for Citrix services increased from 525 users to 625 users.

Network availability was provided at more than 99.95 percent and network services availability at 99.98 percent or higher.

Implementation of the High Availability Computing Solution continued to ensure that critical applications and resources remain available to NLM users. Work continued on clustered Oracle server systems and clustered storage systems as NLM's high availability computing resources. This initiative will increase the NLM mission critical database storage capacity eight times and will increase non-database storage capacity four times. Applications ported to the system were DOCLINE, DCMS, BizFlow, CSMS, Newfiles and Bioethics applications.

SharePoint is supported as a production service for information sharing and document collaboration at NLM.

**Data Center Reengineering**

The NLM data center has tripled its use of electrical power, cooling and data transmission capacity over the last seven years due to the rapid growth in IT systems. Recognizing this growth will continue in the years ahead, a detailed process continued for evaluating the safety, reliability and performance requirements of the data center. Reengineering activities included:
- Expanded data center space by an additional 2,800 square feet.
- Increased electrical power capacity by 87.5 percent from 480 KW to 900 KW.
- Increased air conditioning capacity by 33 percent from 200 tons to 266 tons.
- 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. Thirty-six of the 40 IRCs have been installed and configured in groups of eight, for a total of five groups. Each group is configured for N+1 redundancy.
- The installation of new cable trays for backbone blown-fiber optic cabling throughout Buildings 38 and 38A continued during the year. 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 incidents. The environmental system provides real-time status of temperature, humidity, UPS power usage per phase as well as Computer Room Air Conditioner (CRAC) status.
- The B1 data center's UPS system was reconfigured. The facility is now supported by three 300 KW UPS, two parallel and one stand alone.
- To keep up with growing electrical power demands, increased L5-20 circuits from 336 to 353.
- Increased UPS battery run time from 12 minutes to 20 minutes.

**Controlled Medical Vocabularies**

*Unified Medical Language System (UMLS) Project:* The Unified Medical Language System (UMLS) Metathesaurus is a large multi-purpose, multi-lingual vocabulary database that contains information about biomedical and health-related concepts. In the 2009AA edition of the UMLS Metathesaurus, there are 152 source vocabularies in 19 languages. The Metathesaurus contains over 2.1 million concepts, a 36 percent increase from last year (with nearly

9.7 million names for those concepts, a 24 percent increase) along with over 46 million relationships between and among these concepts. This represents a nearly 10 percent growth in concepts since the 2008AB edition a year ago. In the past two editing cycles (2009AA and 2009AB), over 59 sources were successfully inverted and inserted. Six new sources were added, improving the coverage and richness of the Metathesaurus content. The SPECIALIST tools and the lexicon were updated in 2009AA.

Development was completed on a re-architected MetamorphoSys using a plug-in framework. This major upgrade will be included in the 2009AB release. Other accomplishments include work on cataloging, requirements gathering, specification, and initial reference implementation of a common technology platform for supporting UMLS applications, including MetamorphoSys, KSS, and Metathesarus Editing and Maintenance Environment (MEME). Release Four of the UMLS licensing system went into production during this fiscal year.

*RxNorm Project:* Thirteen major versions and 51 weekly releases were produced from the RxNorm Editing System. Versions 7.1 and 7.2 were released in October, 7.3 and 7.4 in December, 7.5 in January, 7.6 in February, 7.7 and 7.8 in March, 7.9 in April, 8.0, 8.1, 8.2 in June and 8.3 in September. This reflects a 29 percent increase in the number of source insertions performed, a 6.8 percent increase in the number of NDC codes, a 4.9 percent increase in the number of new ingredients and a 2.2 percent increase in the number of drugs. Implementation of these releases included the addition of specific categories of information to NDC code conflicts page, the ability to create multiple synonyms for BPCK, GPCK, SCD and SBD, and the ability to change NDC Code Conflicts Page to add five new categories of conflicts in addition to the source related conflicts.

The monthly inversion and insertions were completed for eight regular sources, 152 new quality assurance checks were added to the database, 421 concepts were processed for data quality purposes, and 154 sources were inserted. Two resynchronizations were completed with UMLS data, and resynchronization codes were updated for handling merges between base and precise ingredients, prevention of brand name merges with disparate ingredients sets and support for multiple synonyms.

RxNorm currently contains 18,843 active generic drugs, 15,626 active branded drugs, 4,330 active ingredients, 9,896 active brand names and 278,975 distinct NDC codes for RxNorm forms.

*Medical Subject Headings (MeSH) and Related Systems*:
MeSH is the National Library of Medicine's controlled vocabulary used for indexing articles for MEDLINE/PubMed. MeSH terminology provides a consistent way to retrieve information that may use different terminology for the same concepts.

Several systems support the ongoing maintenance of MeSH and its translations. A list of enhancements to those systems is shown below:

- The M2000 maintenance system was enhanced to accommodate the increased tree parent/child node levels for descriptor records from 11 to 13. This change also had impact on the following MeSH modules: Cutover Process, XML Distribution files, and MeSH MARC files.

- A new User Interface, M2000Plus, was developed to view changes in the new version of MeSH and to view new statistical information for M2000 activities from 2005-2009.

- MeSH Translation Maintenance System (MTMS) is an inter-lingual database of translations and a system for extending and maintaining foreign MeSH terms. A new information page was developed in MTMS to view and compare the MTMS statistical information in various years. The use of MTMS has increased MeSH record translation to other languages by seven percent, from 97K to 103K. The cutovers for the French, German, Italian, Czech and Croatian languages were completed. New 2010 MeSH XML files were generated and distributed for these translations. Polish and Norwegian translations were loaded into the system for the first time

- Global Citation Management System (GCMS) provides XML data of annual changes which are made by NLM in the MeSH indexing of citations in PubMed and distributed MEDLINE. In 2009, NLM developed a new module to handle tasks related to Publication Types Conversion for both Auto Year-end Processing (YEP) and the current MeSH year.

- A new supporting system was developed and successfully implemented to manage Journal Descriptors records (Journal Descriptors Maintenance System - JDMS). A separate section was included for the JDMS User Manual in the GCMS user interface.

- The Office of Rare Diseases (ORD) terminology was integrated with the existing MeSH vocabulary system by merging over 3,500 records and nearly 5,000 synonyms. A new browser was also developed for ORD to help the editors of Rare Diseases.

- As in past years, NLM performed additional maintenance to MEDLINE data in the fall time frame. This maintenance is known as Year-end Processing (YEP). The volume for the year-end processing increased by 4.1% from 197,000 to 206,000 records.

**Research and Development**

*NLM Digital Repository Project*:  The hands-on testing of candidate repository software tools and the selecting of the best tool for NLM's needs was completed.  A final report was prepared that summarized the results of the analysis and testing activities,  and recommended that the open source Fedora Commons software be used as the core component for the NLM repository.

Also, OCCS assisted in designing and implementing the cholera repository collection which is the first of four planned NLM digital repository pilot collections.  The cholera collection consists of 545 English-language monographs related to cholera, dated 1830 to 1890.  The digital repository cholera collection pilot provides a means to preserve these materials, enable users to search the collection, and provide improved user viewing using page-turner software.  The repository was implemented using several open source software components, including the Fedora Commons repository core, open source book viewer software from Northwestern University, and open source JPEG2000 server software from Los Alamos National Labs.

Tool research was conducted, and a prototype was developed of an advanced video player that can be used to add innovative video search capabilities to video collections in the NLM digital repository.  The prototype is planned to be used as a key component of the second pilot collection in the NLM Digital Repository, a collection of HMD historical videos, which is scheduled for implementation in FY2010.

In the past year, the NLM Digital Repository has grown from only a requirements statement, to an implemented repository containing a substantial pilot collection of historical NLM materials as follows:
- 545 books in NLM Repository (cholera collection)
- Approximately 55,000 scanned pages in NLM Repository
- Approximately 330,000 total content files in NLM Repository

*Database*:  Migrated the MedlinePlus input system and DOCLINE to the production 10g Real Application Clusters (RAC) database.  Also, implemented Oracle Dataguard Standby database for the 10g RAC database, and implemented database security changes based on recommendations from the security team.  Of particular significance was the support of year-end processing by the M2000 and DCMS application staff and support of the move of Knowledge Source Server (KSS) data from LHC to OCCS.  Also, completed  migration of the remaining few schemas in the primary 9i standalone database  (other than the KSS related schemas) to 10g RAC database and moved the remaining database with the KSS related schemas to an Medical Language Branch (MLB) database server.

*Search Engine*:  Upgraded Vivisimo to version 6.1-4, evaluated the implementation of some of the new version 6 features in an internal enterprise search application and continued testing the beta release of Vivisimo 7.5 in the NLM environment.

*ReportNet Migration*:  Migration from Impromptu and Impromptu Web Reports continued to the COGNOS product ReportNET.  ReportNET will provide decision-support reporting capability across the spectrum of NLM activities.   The Voyager application was migrated to ReportNet, the use of COGNOS8 by various Library Operations staff was supported, and modifications were made to the Serials Extract File COGNOS model to account for the new data from Meridian.

*Comparative Effectiveness Research (CER)*:  Developed a new NLM Web resource to inform investigations of comparative effectiveness.  Comparative effectiveness research (CER) focuses on studies which examine the relative effect of interventions that address diseases and conditions in "real world" practice.  CER is an evolving discipline, so there are currently few studies conducted that have wide applicability beyond the settings in which they are based. The purpose of the search strategies developed by NLM is to help *inform* discussions on or relating to comparative effectiveness by providing retrieval of a full-range of studies, both those already published and those still in progress, that address efficacy/effectiveness in a "real world" context.

*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, a new branch and workflow was created to support Go Local contributor documentation as well as the testing of TeamSite 6.7.

*Web Analytics*:   NLM uses the Web Trends 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.   An architectural improvement was made by adding a fifth Web Trends server in June.  The new server was added to the pool of existing servers in order to increase the processing and analysis resources for the application.

**Bethesda Hospital Partnership**

The Bethesda Hospitals' Emergency Preparedness Partnership (BHEPP) consists of the National Naval Medical Center, National Institutes of Health Clinical Center, Suburban Hospital Healthcare System, and the National Library of Medicine. BHEPP's main goals include respond rapidly and successfully to any emergency situation; integrate the collaborative response with other community, regional, and national responses; and create

collaborative structures and processes to serve as an exportable model for other similar centers in the Nation.

OCCS led BHEPP research in the following areas:

- Data & voice communication infrastructure – to facilitate cooperation and collaboration among BHEPP member sites by providing a laser communications network and linking the disparate voice communications systems into a unified interoperable system.
- Disaster patient data management – to investigate the practicality and effectiveness of using the digital pen technology to capture patient assessment data during triage and exchanging a minimum core patient data among BHEPP member sites to support disaster response operations.
- Disaster patient tracking – to explore the effectiveness of using Radio Frequency Identification (RFID) and Real Time Location System (RTLS) technologies to track disaster patients, medical responders, and critical disaster medical equipments.

**Green Computing Initiatives**

Green computing initiatives were advanced to more efficiently use available electrical power, cooling and computing resources. Major accomplishments include:
- Increased computer virtualization by 28 percent by implementing more virtual servers than physical servers. The increase in virtual servers leads to maximizing the utilization of available computing resources. In FY2008, the ratio of virtual servers to physical servers was 91:309, a 29 percent virtualization. In FY2009, the ratio of virtual servers to physical servers changed to 123:214, a 57 percent virtualization.
- Consolidated computing racks in the NLM Data Center and reduced used rack space by 19 percent. The freed space was made available to support other growth areas within NLM. In FY2009, 58 racks were used in the Data Center, a reduction from 72 racks in FY2008.
- Upgraded older equipment to more energy efficient hardware.
- Created hot and cold aisles to better utilize cooling resources.
- Plugged gaps within cabinets to redirect cold air for equipment cooling.
    Desktop virtualization was introduced with virtual machines (VM) being integrated into desktop operations. Use of a VM allows for flexibility allowing the user to access the VM from their primary computer without requiring them to maintain a second physical computer. In addition, several software applications were virtualized such as Oxygen XML Editor 10.3, SQL developer 1.5 and MS Exp Web 3. Software virtualization benefits include using fewer resources, accelerated application deployment,

and implementation of security by removing the requirement for end-users to have administrator privileges.

**Medical Literature Support & Document Delivery Services**

*Data Creation and Maintenance System (DCMS)*: The major event for DCMS was the baseline extraction, which is a re-release of all DCMS citations that follows the MeSH Year-end Processing (YEP). The 2009 MEDLINE/PubMed baseline database contains 17.7 million, a 16 percent increase from FY2008. In addition, OCCS loaded and processed over 70,000 "new" OLDMEDLINE records for the publisher year 1947 and 1948, 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:
- Reduced the elapsed time for extraction of the baseline to 10 hours and 49 minutes, a 36 percent reduction over the time required in the prior year. The number of records exported per hour increased by 67 percent.
- Journal Article publication type was removed from 13,000+ Meeting abstracts and the records were re-exported.
- 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 2,600 domestic and international libraries in processing more than 1.4 million interlibrary loan transactions and 275,000 Loansome Doc requests. Three versions of DOCLINE were released, version 3.4.5 in October, 3.4.3 in December and 4.0 in June. In FY2009, the Serials Holdings Search and Update modules were redesigned for more efficient editing of holdings. New features include an easy way to filter and list library holdings and more intuitive title searching. More efficient identification of electronic holdings was incorporated. There were 165 DOCLINE and Loansome Doc enhancement requests implemented in addition to internal security and server configuration changes. The system was upgraded to support 22,190 journals with articles, 16,005 monographs, 903,717 articles, and 1,703,826 ILL requests. Also, there were over 30 million page views and over 1.1 million visitors to the site.

*Voyager Integrated Library System (ILS)*: The year-end processing of the Voyager Integrated Library System database was completed, which included altering numerous fields and upgrading the MeSH terms to the 2009 version. Also, a new release of Voyager, version 2007.1, was promoted into production use.

*Relais*: NLM uses the commercial off-the-shelf Relais system for electronic document delivery and Interlibrary

Loan (ILL) management. Documents requested from NLM via DOCLINE are scanned and automatically delivered using the borrower's requested delivery method. The download procedure was updated to read the Meridian values from the SEF and write them to the Relais download file. The ultimate goal is for these Meridian fields to display on the Relais pick-slip.

*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 557 journals (55,447 issues) in various stages of scanning for entry into PMC.

*Print Journal Donation Tracking System*: An automatic system to track the print journal donations was implemented. This system would facilitate donation offers from libraries, etc. using DOCLINE holdings if available. One hundred and four donations have been submitted since the application launched, 353 donors and 330 institutions registered the system. There are 689 total titles offered, 1,709 total volumes offered and 234 titles received.

*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. The Medline Review Application Form is required for all journals initially reviewed by the LSTRC for inclusion in MEDLINE. The ability to keep deleted records was also added.

*Serials Extract File (SEF)*: Among numerous upgrades and fixes, there was a new addition of data from Meridian Electronic Resource Management System. This data contains information about NLM licenses for electronic serial publications. More than 4,000 Meridian records now reside in SEF. Generated the List of Serials Indexed for Online Users (LSIOU). The 2009 LSIOU expanded with 396 additional serials. A request to transfer URL data was completed in the Managed Serials form when there is a title change. Over 250 titles were changed.

*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. The major Classification re-engineering accomplishments include:
- Removing the need for manual HTML coding by cataloger.
- Validating of Publication Types used exclusively by Cataloging.
- Examining the feasibility of connecting Editor directly to M2000 rather than DCMS MeSH.

- Refining the searching capabilities to allow the use of punctuation marks.

## 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. VPN and Citrix remote access services were implemented at the NCCS as well as the install of Oracle's Dataguard to synchronize databases between NLM and NCCS. Various other upgrades were performed to the storage systems and servers located at this site.

## Office Automation & Customer Support

A "Store" model was adopted to serve hardware and software to its constituent users. In past years, equipment was procured as requested by the user. The procurement process was lengthy and introduced delays in order fulfillment. By introducing the Store concept, users are now able to order items from the Store with wait time for fulfillment decreasing from greater than four months to an average of three weeks. In addition, Standard Operating Procedures and Work Instructions were established to manage the Store. This effort enhanced the Order Fulfillment team's ability to better serve the customer and greatly reduced the turn-around time from when a user's request is submitted to the deployment of the IT asset.

Altiris software metering agent was implemented which allows for discovery of applications on all workstations via the metering process. This project started with the monitoring of MS Office 2007 and Adobe Acrobat 9 Pro with additional "application monitors" created for the remaining limited seat license products. This feature will be used to monitor and manage the usage of limited seat licenses.

Eighty-one new Microsoft operating system security patches that were released were applied to the 950 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.

A 100 percent electronic enforcement capability for managing the inventory of remote computers was achieved. Prior to this implementation, the capability to electronically audit remote computers did not exist.

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. Over 10,000 NLM ticket requests for IT support were entered and tracked. NLM IT Staff resolved 63 percent of the calls (6,300 tickets) with 37 percent of support calls being completed by NIH staff.

Five courses were coordinated including "MS Visio," "MS Excel," "Altiris," "Remedy" and "MS OneNote. Focused training was provided in support of the Administrative Account procedures, as well as for the Stay-In-School and NLM Associates' programs. An OCS training SharePoint site was also implemented.

## Public Health

*Health Services Research Projects in Progress (HSRProj)*: There are currently 8,167 projects in the Ongoing and Completed HSRPROJ file, as well as 10,959 archived projects. The HSRProj Web site was updated with 1,608 new records. Additionally, 1,133 records with the Final Date between 1/2004 and 12/2004 were relocated to the archive file. Version four was released in June that included many enhancements such as making the Award Type and Study Design fields searchable, adding addition information to Search Tips, improving search retrieval results display and printing options and improving the Advanced Search display. Also, increased monthly page views from an average of 22,000 to 40,000, added 10 percent more health services/science research resource records, increased HSR information central page views by 12 percent and increased public health partnership Web site page views by 11 percent.

*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., NLM, the Centers for Disease Control and Prevention, Agency for Healthcare Reseach and Quality), public health organizations (e.g. American Public Health Association, National Association of County and City Health Officials) and health sciences libraries (e.g., National Network of Libraries of Medicine).

Version six was released for the input system, which includes modification of the home page display and the addition of new pages. The following were added: display sequence field for categories and sub-categories, a spotlight selection function, and the ability to allow users to edit display start and end dates of all links. Also, a new functionality was created to Add/Edit the link form that allows new organization names to be inserted if not currently in the list of names.

## Outreach and Customer Services

*Against the Odds Web Exhibit & Web interactive activities*: A new Podcast feature was released. The podcast format was chosen to publish a five-part series of speakers who addressed an audience of high school students from Maryland, Virginia and the District of Columbia at the opening program for the interactive exhibition.

*Exhibit Asset Manager*: The Exhibit Asset Manager database is a new database 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 initial release of the core features were completed in January which included contacts input, asset input and basic reporting features.

*Images from the History of Medicine (IHM)*: A new system, LUNA, was implemented for the IHM project in April. Over 70,000 Voyager IHM records were processed and numerous fields were enhanced. A complete reload and index of the IHM data in the LUNA system was implemented.

*HMD Directory*: The HMD Directory is an online version of the History of Medicine component of DIRLINE (Directory of Information Resources Online), a National Library of Medicine (NLM) database, which contains location and descriptive information about a wide variety of health and biomedical resources. The DIRLINE History of Medicine component aims to assist scholars and researchers in identifying useful medical history collections throughout the world. Some of the accomplishments included:
- Changing structure, data and relationship of tables.
- Adding validation to all required fields.
- Implementing the ability to save the form in Microsoft Word.
- Adding administrator's name that approves the form.

*Customer Service Support System (Siebel)*: Multiple defect resolutions, enhancement releases, and system upgrades were delivered for Customer Service, Change Request, and Firewall Service Request Management applications. Key accomplishments included:
- Migrated database from Oracle 9.2.0.7 to Oracle 10.2.0.3.
- Moved the system gateway servers to Virtual host and removed physical box from the network as part of a green computing strategy.
- Enhanced Web form submission performance by adjusting Siebel configuration settings. As a result, Web form SR submission has been dramatically reduced from 12-15 seconds to 1-2 seconds.

- Fixed contact purge performance issue by adding hint/update false script in configuration file. The contact purge process took about 22 minutes for 1,000 records then further reduced to 13 seconds after adding the script.

- Supported customer service to process 86,369 service requests, 46,000 inbound e-mail messages, 51,176 outbound email messages, 549 firewall change requests, and 824 application change requests.

# ADMINISTRATION

*Todd D. Danielson*
*Executive Officer*

**Table 20: Financial Resources and Allocations, FY2009**
*(Dollars in Thousands)*

Budget Allocation:

| | |
|---|---|
| Extramural Programs | $66,070 |
| Intramural Programs | 260,162 |
|    Library Operations | (78,567) |
|    Computer & Communications Systems | (30,445) |
|    Lister Hill National Center for Biomedical Communications | (49,846) |
|    National Center for Biotechnology Information | (83,551) |
|    Specialized Information Services | (17,753) |
| Research Management and Support | 13,459 |
| Total Appropriation | 339,691 |
| Plus: Reimbursements | 35,935 |
| **Total Resources** | **$375,626** |

## Personnel

In October 2008, **Stacey Arnesen** became head of the Office of the Disaster Information Management Research Center, Division of Specialized Information Services. Stacey came to NLM in 1986 as a MeSH indexer in the Bibliographic Services Division of Library Operations. In 1987 she joined the Specialized Information Services Division to coordinate the EMIC and DART databases. Subsequently she worked on HIV/AIDS information services and headed the TOXNET search team. Since 2002, Ms. Arnesen has served as the Advisor for Special Projects, SIS. In this capacity, she heads the SIS Web team, coordinates the Central American Network for Disaster Health Information, working with the Pan American Health Organization and the Regional Disaster Information Center in Costa Rica, and is a member of the Bethesda Hospitals' Emergency Preparedness Partnership. Ms. Arnesen received her AB degree from Smith College and an MS in Neurobiology and Behavior from Cornell University.

In October 2008, **Evgueni Belyi** was appointed to a Staff Scientist position after having been a TAJ Technologies contractor since 2003 at NCBI. Mr. Belyi earned his MS in Electronics Engineering from the Moscow Institute of Physics and Technology, then worked as a physicist and software designer. He joined TAJ Technologies in May 2003 as a contractor for the National Center for Biotechnology Information (NCBI) working as a software developer with Information Engineering Branch and has been involved in several projects related to development and support of ID databases, loading and retrieving software, quality assurance system. As a Staff Scientist, he will continue to work on support of ID system and its ongoing and future developments.

In October 2008, **John Kilbourne, MD**, rejoined NLM as a Medical Officer. He received his medical degree from the University of Illinois College of Medicine. After completing his Family Medicine residency and practicing near Chicago, he joined the SNOMED division of the College of American Pathologists (CAP) as a clinical editor. His role there expanded to include implementation training, and in 2005 he joined the NLM. He left in 2007 to work in consulting, then rejoined the CAP. John returned to the NLM, with primary responsibility for monitoring standards developments as they relate to terminology, and to assist users of RxNorm and the Unified Medical language System.

In October 2008, **Byungsoo Kim, PhD**, was appointed to a Staff Scientist after having been a Computercraft contractor at NCBI/IEB since May, 2007. Dr. Kim will continue in his role of senior quality assurance engineer where he is responsible for implementing software quality assurance processes for major NCBI systems including NCBI Portal and Entrez. Byungsoo received a PhD in Mechanical Engineering from Pennsylvania State University in University Park and has been working as a software quality assurance engineer since 2000. During his career, Byungsoo has worked on a variety of projects where he was responsible for a range of application components including user interfaces, generic search APIs, backend services, and a video search engine.

In October 2008, **Evgeny Kireev** was appointed to a Staff Scientist position at NCBI Information Engineering Branch (IEB) after having been a Lockheed Martin Information Systems contractor since 2006 at NCBI. Mr. Kireev received an MS in computer science in 1991 from the Moscow Aviation Institute at the Russian National Space Mission Control Center. After receiving his graduate degree, Mr. Kireev was a software engineer at the Russian National Space Mission Control Center in Moscow. He later joined Maris Multimedia, LTD., in the UK as a senior software engineer and a team lead. In 1998, Mr. Kireev joined the US corporation, Invitrogen, initially serving as senior software engineer and promoted later to a senior software development manager. Over the course of his eight year tenure, he oversaw the software teams developing desktop and Web-based bioinformatics enterprise applications and databases. Since joining NCBI in 2006 as a Senior Technical Project Manager, Mr. Kireev managed Entrez Portal Conversion project and served as the acting Program Manager for NCBI Discovery Initiative. As a Staff Scientist, he will continue to manage development of new releases of the NCBI Portal System and Entrez projects in Public Services Section of IEB.

In October 2008, **William Klimke, PhD**, was appointed to a Staff Scientist position after having been a ComputerCraft contractor since 2002 at NCBI. In 2002, he received his PhD degree in Molecular Biology and

Genetics from the University of Alberta (Edmonton, Alberta, Canada) based on studies of horizontal gene transfer and F plasmid biology. He joined Computercraft in Sept 2002 as a contractor for the NCBI working as an indexer for GenBank and then switched in 2003 to the Reference Sequence (RefSeq) Microbial Genomes Annotation Project which included processing submissions to the Prokaryotic Genome Automatic Annotation Pipeline, curation of prokaryotic genomic RefSeq and Genome Project records, and development, curation, and organization of the Protein Clusters database (ProtClustDB). As a staff scientist (IEB), Dr. Klimke will continue to work on prokaryotic genomic RefSeq records, Protein Clusters, and other tools and databases which will be used for the analyses of prokaryotic genomes, the human microbiome, and other metagenomic datasets.

In October 2008, **Anastasia Nikolskaya, PhD**, was appointed to a Staff Scientist position after having been a Computercraft contractor since July 2008. Anastasia earned her PhD in Genetics from Michigan State University. She has worked on protein classification and annotation as a Computercraft contractor for NCBI in 2000 - 2002. This was followed by six years as a Research Assistant Professor at Georgetown University (Protein Information Resource) where she worked on the PIRSF protein classification system. Anastasia will continue her work at the CBB, joining the ongoing research on the evolution of influenza virus.

In November 2008, **Dianna Adams** was appointed Head, Desktop Services Section, Systems Technology Branch. Ms. Adams brings with her extensive experience in IT management in both the Military and the private sector. Most recently, Ms. Adams worked for Lockheed Martin as the Infrastructure Operations Manager of a 170-member team servicing the Department of Homeland Security. Ms. Adams has a Bachelors in Information Systems Management, and a Masters in Computer Systems Management.

In November 2008, **Christopher Miller, MD,** joined NLM as a Visiting Fellow. Dr. Miller has a medical degree from University of Iowa. He did an internship year in internal medicine at University Hospitals of Cleveland. Prior to coming to NLM he was a postdoctoral fellow in Sleep Medicine, Neurobiology and Epidemiology at Case Western Reserve. At NIH, Dr. Miller is working on informatics projects involving large clinical databases with the Clinical Center and with NLM.

In November 2008, **Mahmudur Rahman, PhD,** joined NLM as a Visiting Fellow. Dr. Rahman received his doctorate degree from the Computer Science Department of Concordia U, Montreal, Canada in January 2008. For his dissertation, he developed approaches to perform semantic-based searches of photographs in medical image databases. At NLM, Dr. Rahman is working on research to develop an interactive classification and retrieval framework to search images in heterogeneous medical image collections.

In December 2008, **Svetlana Dracheva, PhD**, was appointed to a Staff Scientist position after having been a ComputerCraft and later Lockheed Martin (MSD) contractor since 2000 at NCBI. Dr. Dracheva earned a PhD in Physics and Mathematics (Biophysics) from the Moscow State University for work on the study of the photocycle of 13-cis retinal analogs of bacteriorhodopsin. She then did a postdoctoral internship at NIH with Dr. Richard Hendler at the National Heart, Lung and Blood Institute (NHLBI) working on studying the role of Halobacterium halobium lipids on bacterhiorhodopsin photocycle. This was followed by a postdoctoral fellowship at NIH with Dr. Ronald Wilder in the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) working on rat animal models of collagen-induced arthritis. She joined ComputerCraft in May 2000 (later switched to MSD - Lockheed Martin) as a contractor for the National Center for Biotechnology Information (NCBI) working as Data Analyst for LocusLink/Gene and then switched to the new PubChem project with Computational Biology Branch (CBB) of NCBI in August of 2005. As a Staff Scientist, she will continue to work on the PubChem BioAssay and Substance databases in the Structure Group of CBB.

In December 2008, **Jessica Fong, PhD**, was appointed to a Staff Scientist position. Dr. Fong received a PhD in Computer Science from Princeton University in 2005. For her thesis research, she developed methods to predict coiled-coil protein-protein interactions from protein sequence data. After graduating, she joined Dr. Stephen Bryant's group as a postdoctoral fellow. At the NCBI, she created computational methods for improving domain classification using the Conserved Domain Database, including new heuristics for assigning specific domains to protein sequences and an algorithm for profile-profile alignment. She has also researched the evolution of domain architectures and studied protein disorder. As a Staff Scientist, she will continue working with the Structure group on computational tools and methods for performing protein classification and for studying protein interactions.

In December 2008, **Karen Karapetyan, PhD**, was appointed to a Staff Scientist position after having been a TAJ Technologies contractor at NCBI. Dr. Karapetyan earned a PhD in Organic Chemistry from the Institute of Fine Organic Chemistry of Armenia for synthesis and characterization of novel benzoquinazolines with anticancer properties. He then did a postdoctoral internship with Dr. Kenneth Jordan at University of Pittsburgh working on quantum-chemical investigation of water clusters and carbon nanotubes. In 2002 Dr. Karapetyan joined Chemical Abstracts Service as Scientific Information Analyst and curated journal articles and patents. He then joined TAJ Technologies as a contractor for the National Center for Biotechnology Information (NCBI) working as a curator of PubChem data. As a Staff Scientist, he will continue curation of

PubChem data and will be involved in other projects in the Structure Group of CBB.

In December 2008, **Alla Keselman, PhD**, joined the Specialized Information Services (SIS) as a Social Scientist. She holds a PhD in human cognition and learning (2001) and an MA in biomedical informatics (2005) from Columbia University. Prior to joining the SIS, she served as an Aquilent, Inc. Senior User Experience Consultant to the NLM, and before that, as a postdoctoral researcher in the Department of Biomedical Informatics, Columbia University. In her previous positions, Dr. Keselman led research efforts in the areas of consumer health informatics, health cognition and evaluation of health information technology. Her research interests include evaluation of consumer health Web sites, lay understanding of complex health concepts (particularly, among children and adolescents), health literacy, readability of health texts and machine algorithms for advanced health information management and summarization. She has established an extensive scholarly publication record in these areas. Dr. Keselman's duties as the SIS will include leading a K-12 team that focuses on the development of environmental health resources for school-age children and educational research, evaluation of SIS Web sites and outreach programs, and research and development of advanced tools for disaster information management.

In December 2008, **Yoo-Ah Kim, PhD**, joined NCBI as a Research Fellow. She will be working in Teresa Przytycka's research group on algorithmic methods in computational and systems Biology (AlgoCSB). She obtained her PhD degree in Computer Science at the University of Maryland in 2005. Since then she has been an assistant professor in the Computer Science and Engineering Department at the University of Connecticut. While her previous research interest has been focused on combinatorial optimization and algorithms, she became increasingly interested in applications of computational techniques to biology research, working on problems such as analysis of gene regulation pathways using microarray data. In NCBI, she will work on computational modeling of gene regulation based on expression data and other projects in AlgoCSB group.

In January 2009, **Mark Johnson** was appointed to a Staff Scientist position after over five years as an MSD/Lockheed-Martin contractor. Mr. Johnson came to NCBI with numerous years of experience as a software engineer, book and magazine author, and speaker. Shortly after coming to NCBI in 2003, he became the technical lead of IEB's usability group. He has worked on dozens of NCBI projects, including key roles in user interface and interaction design of several Entrez databases, including PubMed, Gene, and the Entrez core itself; the user interface for Variation Viewer; the redesign of the BLAST Web user interface; the client-side framework for the NCBI Portal system; and an MVC application framework for Portal applications. He will continue as the lead of the IEB Core Web team, providing design and development

support and reusable components for NCBI Web applications.

In January 2009, **Melanie Modlin** was appointed Deputy Director of the Office of Communications and Public Liaison (OCPL). She has more than 12 years' experience at NLM, having served as a Public Affairs Specialist since 1996. During her service with NLM's OCPL, she received the NLM Director's Award for her writing and editing, and for managing NLM's tour operation. Active in promoting health issues and health policy for nearly three decades, she worked for Congressman Claude Pepper for eight years, on the staff of his House Select Committee on Aging, where, among other accomplishments, she helped coordinate hearings resulting in the creation of the National Center for Biotechnology Information. She also held communications positions at the Foundation for Hospice and Home Care and the Alliance for Aging Research before joining the NLM staff. Ms. Modlin attended Smith College and graduated from the University of North Carolina at Chapel Hill, where she also did graduate studies in journalism.

In January 2009, **Conrad Schoch** joined NCBI as a Staff Scientist. He will be working with Scott Federhen in the Taxonomy group where he will be responsible for the curation of fungal records in GenBank. Prior to this appointment he was a postdoctoral researcher in the laboratory of Joey Spatafora at Oregon State University. He was responsible for analyzing and collecting non-lichenized Ascomycota for the collaborative 'Assembling the Fungal Tree of Life' project, funded by NSF. Whilst at Oregon State, he was also lead researcher on another NSF funded project concerning evolution and phylogenetics of the fungal class Dothideomycetes. Earlier post doctoral experience included research on motor proteins in Nectria haematocca at Cornell University. Dr Schoch obtained his PhD in Plant Pathology from Stellenbosch University, South Africa in 2000 with a study on the systematics of the plant pathogen *Calonectria*.

In January 2009, **Chunlin Xiao, PhD**, was appointed to a Staff Scientist position. He earned his doctoral degree from University of Illinois at Champaign-Urbana. Prior to his appointment, he spent several years on genome research and bioinformatics application development in both academia and industrial sectors including Protein Information Resources (PIR) of Georgetown University Medical Center, Celera Genomics, and Applied Biosystems. As a Staff Scientist, he will be working on genetics/variation-related projects at Information Engineering Branch of NCBI.

In February 2009, **Benjamin G. Slade** was appointed to a Staff Scientist position after having been an MSD/Lockheed Martin contractor since 2006 at NCBI. Mr. Slade received a BS in Electrical Engineering from the University of Maryland and an MS in Computer Science from Johns Hopkins University. He brings to NCBI an extensive, diverse background in large database administration and systems analysis including his

experience as a Systems Architect for the AOL Email Operations Division. He joined NCBI to work on the design, implementation, and operations of Sybase database systems running the RefTrack, Id, and Trace applications. As a Staff Scientist, he will continue this work along with other projects.

In February 2009, **Constantin Vasilyev** was appointed to a Staff Scientist position after having been TAJ Technologies contractor since 2000 at NCBI. He earned a Masters degree in Applied Mathematics and Computational Physics at the Moscow Engineering and Physics Institute in 1996. He then was employed by the State Center "Interphysica" in Moscow, where his work involved data analysis and development of applications relating to nuclear physics. Later, at Russia's State Agency of International Cooperation in Science and Education, he was employed as a systems administrator, network administrator, and Web developer. Mr. Vasilyev joined NCBI to work as a database administrator. During his tenure, he has designed, deployed, and maintained the Microsoft SQL server infrastructure, which is a backend for NCBI's major resources. He also participated in design and performance tuning of many scientific databases developed at NCBI. He will continue this work as a Staff Scientist.

In March 2009, **Lianyi Han, PhD**, was appointed to a Staff Scientist position. Dr. Han earned a doctorate in Bioinformatics from the National University of Singapore, Department of Pharmacy, for working in bioinformatics and computer aided drug design. He has worked on research projects involving computational analysis of protein interactions, prediction of novel protein functions, drug like compounds virtual screening, and the exploration of druggable proteins. He also worked on Bioinformatics databases such as therapeutically relevant multiple pathways database, Therapeutic target database and Drug adverse reaction target database. He joined NCBI in 2006 as a Postdoctoral Research Fellow for PubChem bioassays data analysis and research on virtual hits selection. As a Staff Scientist, he will work with the PubChem Database and the BioSystems Database production service along with other research projects in the Structure Group of CBB.

In April 2009, **Vincent C. Calhoun, PhD**, was appointed to a Staff Scientist position after having been a Computercraft contractor since 2007 at NCBI. Dr. Calhoun earned a PhD in Molecular and Cell Biology from the University of California, Berkeley for work on transcription regulation and enhancer-promoter specificity in Drosophila melanogaster. He then completed a postdoc with Dr. Gary Felsenfeld at the National Institute of Diabetes and Digestive and Kidney Diseases working on nucleosome occupancy and chromosome condensation at the chicken beta-globin locus. He joined Computercraft in April 2007 as a contractor for NCBI working as an indexer for GenBank. As a Staff Scientist, he will continue working as a GenBank indexer in addition to other projects.

In April 2009, **Milton Corn, MD**, was appointed NLM Deputy Director for Research and Education. Prior to joining the NLM, Dr. Corn served with distinction in numerous top leadership positions at Georgetown University including Dean of the School of Medicine. Since 1990, as NLM Associate Director for Extramural Programs, Dr. Corn has led the Library's Division of Extramural Programs in an exemplary manner, establishing numerous collaborations with other agencies and organizations that have expanded informatics research and training opportunities. Dr. Corn creatively modified NLM's Training Program in Biomedical Informatics to keep pace with developments in biomedical computing. Dr. Corn will serve as principal medical advisor on research and development and medical education priorities for the NLM. Dr. Corn will provide direction and advice on the full range of NLM's intramural and extramural research portfolio.

In April 2009, **Anton Golikov, PhD**, was appointed to a Staff Scientist position, after having been a TAJ contractor since 1999 at NCBI. Dr. Golikov earned a PhD in IT sciences from Moscow State University of Aircraft Technology in Russia. For the past 10 years, he has worked on the Entrez PubMed Web site and participated in building an advanced search engine for medical data. In 2002, he created the new MyNCBI Web portal. As a Staff Scientist, he will work on Short Read Archive (SRA) along with other projects.

In April 2009, **Aleksey Grichenko, PhD**, was appointed to a Staff Scientist after having been a MSD/Lockheed Martin contractor since 2001 at NCBI. Dr. Grichenko earned a doctorate in Biophysics from the Institute of Theoretical and experimental Biophysics, Russian Academy of Sciences for investigating and modeling the regulation of the Ca2+ channels in cardiac cells of hibernating animals. He learned biophysics at the college level in the Moscow Institute of Physics and Technology. In addition to his research work, he has developed software and hardware for recording and analyzing the Ca2+ currents through cell membrane. He also assisted in the development of other scientific and medical tools including long-term radioactivity analyzer, a tool for the express blood analysis etc. He participated in the development of the NCBI C++ Toolkit (core library, Biological Object Manager, CGI framework) as a contractor at NCBI. As a Staff Scientist, he will continue this work along with other projects.

In April 2009, **Anne Kiang, PhD**, was appointed to a Staff Scientist position after having been a Computercraft contractor since 2007 at NCBI. Dr. Kiang earned a doctorate in the University Program of Genetics and Genomics from Duke University for work in adenoviral gene transfer and associated complement-dependent inflammatory responses. She joined dbGaP at NCBI in March 2007 as a scientific curator working on ensuring phenotypic data quality control, curation, annotation and integration of study documents and data,

and communication with external investigators. As a Staff Scientist, she will continue her work with dbGaP.

In April 2009, **Terence D. Murphy, PhD**, was appointed to a Staff Scientist position after having been a Computercraft contractor since 2006 at NCBI. Dr. Murphy earned a doctorate in Molecular Genetics from the University of California, San Diego for work on chromosome inheritance. He subsequently developed his own research program as a Staff Associate at the Carnegie Institution of Washington in Baltimore, MD, focusing on regulatory mechanisms of centrosome duplication. He joined NCBI in March 2006 as a contractor on the Reference Sequence project and as the Insect Genome Champion. As a Staff Scientist, he will continue this work along with other projects.

In April 2009, **Ronald P. Shaw** filled the newly created position of Junior Administrative Officer (AO). Mr. Shaw's primary responsibility will be the administration of NLM's Scientific Review and Evaluation Award Program (SREA). As our Junior AO he will also support our Chief, Administrative Officer in coordinating and performing administrative activities for the Division. Mr. Shaw recently completed two years as an intern in the NIH STRIDE program, where he gained invaluable administrative experience through several rotations with NHLBI, NIAID, and NCI, in both their Intramural and Extramural Administrative Branches. Prior to accepting his current position here at NLM, Mr. Shaw served as an Administrative Officer with the National Center for Complementary and Alternative Medicine, NCCAM.

In April 2009, **Françoise Thibaud-Nissen, PhD**, was appointed to a Staff Scientist position after having been a Computercraft contractor since 2008 at NCBI. Dr. Thibaud-Nissen earned a PhD in Plant Molecular Biology from the University of Illinois for her work on embryogenesis and the flavonoid pathway in soybean. This was followed by a postdoc focusing on the development of whole-genome arrays for ChIP-chip analysis in Arabidopsis at The Institute for Genomic Research with Dr. Chris Town. She then became a Staff Scientist with Dr. Robin Buell where she contributed to the annotation of the rice genome. She joined Computercraft in January 2008 as a contractor for NCBI, working as a genome pipeline assistant coordinator. As a Staff Scientist, she will continue to work on the genome pipeline.

In April 2009, **Kamen Todorov** was appointed to a Staff Scientist position after having been an MSD/Lockheed Martin contractor since 2000 at NCBI. Mr. Todorov earned a MSc in Computer Science from the University of Veliko Turnovo in Bulgaria. In addition to his academic work in Mathematics and Computer Science, he developed software and firmware at Systems for Teleprocessing and Networks, Plc. and later Tremol Ltd. where he focused on networks and application software. Mr. Todorov later worked on development of ERP and B2B software at Datastream Systems, Inc. He joined NCBI in December 2000 and has been involved in a variety of projects: systems for submissions of biological data, toolkit software for dealing with sequence alignments, development of internal systems and databases. As a Staff Scientist, he will continue to expand the scope of his work at NCBI.

In April 2009, **Guo-Yun Yu** was appointed to a Staff Scientist position after having been a Computercraft contractor since 2006 at NCBI. Dr. Yu received a PhD in Plant Pathology from the University of Illinois at Urbana-Champaign in 1998 for work on characterization of antifungal peptides under James B. Sinclair. He also completed a two-year postdoc in the Department of Agronomy and Plant genetics at the University of Minnesota studying the gene expression profile of chemical induced plant resistance to fungal infection. Meanwhile, he earned a Master's degree in Software Engineering from the University of St. Thomas. He then worked as a Research Associate for six years with Christopher M. Gomez in the Department of Neurology of the University of Minnesota on genetic analyses and molecular characterization of spinocerebellar ataxias. He joined Computercraft in April 2006 as a contractor dbSNP curator for NCBI. As a Staff Scientist, he will continue his dbSNP work.

In April 2009, **Zhigang Zhou, PhD**, was appointed to a Staff Scientist position. Dr. Zhou earned a doctorate in Computational Chemistry from Duquesne University for work on the small molecule binding and interaction with bio-molecular receptors. He then worked at Purdue University as a postdoctoral research associate working on computational screening and design of antivirals against dengue and yellow fever viral infections. He joined NCBI in January 2008 as a visiting scholar working on the PubChem project. As a Staff Scientist, he will continue to work on the PubChem project to develop Web service for public users in the Structure Group of the Computational Biology Branch in NCBI.

In May 2009, **Robert D. Sanders** was appointed to a Staff Scientist position after having been an MSD/Lockheed Martin contractor since 2007 at NCBI. Mr. Sanders earned his MS degree in Computer Science from Johns Hopkins University in Baltimore, Maryland. Prior to joining Lockheed Martin, Mr. Sanders worked at Celera Genomics and Applied Biosystems where he integrated bioinformatics data into the assay development and annotation processes, generated and staged data for incorporation into the Celera Discovery System, and assisted in gene discovery efforts. He joined MSD in December 2007 as a contractor for NCBI to help in the evolution of the Trace and Short Read Archives, including providing data management and software development support. As a Staff Scientist, he will continue to work in support of these Archives and associated projects such as the 1000 Genomes Project.

In May 2009, **Anjana R. Vatsan, PhD**, was appointed to a Staff Scientist position after having been a Computercraft contractor since 2004 at NCBI. Dr. Vatsan earned a doctorate in Molecular Biology from the Jawaharlal Nehru University, Delhi, India for work on the

nutritionally important seed specific proteins. She then did a postdoc with Dr. Roy Curtiss at Washington University, St. Louis on transgenic plants. She subsequently joined the laboratory of Dr. Ben Matthews at the US Department of Agriculture to work on the Bioinformatics of soybean genes involved in soybean cyst nematode and other pests, pathogens, and stresses. She joined Computercraft in January 2004 as a contractor for NCBI, working as a plant genome biologist. As a Staff Scientist, she will continue to work on the various projects associated with plant genomes.

In May 2009, **Lana Yeganova, PhD**, was appointed to a Staff Scientist position after having been an MSD/Lockheed Martin contractor since 2001 at NCBI. Dr. Yeganova earned a doctorate in Operations Research from the George Washington University for work in mathematical optimization. She joined MSD in July 2001 as a contractor for NCBI to work on the problems of Text Processing and Information Retrieval. As a Staff Scientist, she will continue this work along with other projects.

In June 2009, **Vadim Miller** was appointed to a Staff Scientist position after having been a TAJ Technologies contractor since 2002 at NCBI. Mr. Miller earned a Master's degree in Electronics Engineering from the Moscow Institute of Physics and Technology. He worked as a Senior System Analyst in Informax, Inc. developing Sequence Analysis and Gene Expression Analysis software. He joined TAJ Technologies in September, 2002 as a contractor for NCBI to work on the Bookshelf project. Later he joined the Portal project group and now works on Portal core engine and Search Backend service. As a Staff Scientist, he will continue this work along with other projects.

In July 2009, **Catherine M. Farrell, PhD**, was appointed to a Staff Scientist position after having been a Computercraft contractor since 2006 at NCBI. Dr. Farrell earned a PhD in Molecular Biology and Biochemistry from Wesleyan University, CT, for work on collagen gene expression and antisense RNA. She then became a postdoctoral fellow in the laboratory of Dr. Gary Felsenfeld at the National Institute of Diabetes and Digestive and Kidney Diseases, working in the field of chromatin and gene expression. This was followed by a Contract Worker Specialist position in the laboratory of Dr. Keji Zhao at the National Heart, Lung, and Blood Institute, with continued research on chromatin, gene expression and epigenomics. Dr. Farrell joined Computercraft in December 2006 as a contractor for NCBI working as a curator in the RefSeq group, and becoming a lead curator for the Consensus Coding Sequence (CCDS) project. As a Staff Scientist, she will continue to work on the RefSeq and CCDS projects.

In July 2009, **Lei He, PhD** joined NLM as a Visiting Fellow. Dr. He is an associate professor in the Department of Information Technology at Armstrong Atlantic State University in Savannah, GA. Dr. He earned his PhD degree in Electrical Engineering from the University of Cincinnati in 2003 and has over 10 years of research and industry experience in image processing, computer vision and pattern recognition. At NLM, Dr. He's research is focused on biomedical image understanding, specifically the use of Content Based Image Retrieval (CBIR) techniques involving shape matching and learning-based segmentation. These techniques are applied to extraction of pre-cancer evidence from cervigrams and HPV linear array genotyping.

In July 2009, **Kimberly Marshall** was appointed to a Staff Scientist position after having been a Computercraft contractor since October 2006 at NCBI. Ms. Marshall earned a MS in Forest Genetics from the University of Vermont. Prior to coming to NCBI, she worked as a biologist for the US Forest Service in Berkeley, CA where she focused on molecular genetics and comparative genomics of phytochrome genes in western conifers and as a Web developer and curator of Dendrome and the TreeGenes database. She then worked for MMI Genomics in Davis, CA as a bioinformatics scientist in SNP discovery and whole genome analysis of bovine, chicken and swine. As a staff scientist, she will continue to work as a curator in the Gene Expression Omnibus group.

In July 2009, **Katherine H. Phillippy** was appointed to a Staff Scientist position after having been a Computercraft contractor since 2007 at NCBI. Ms. Phillippy earned a MS in Biotechnology in 2003 from Johns Hopkins University with a concentration in Bioinformatics. Before joining NCBI, she worked for 9 years at the Institute for Genomic Research, and has a total of eleven years of experience in the fields of molecular biology, genome sequencing and closure, high-throughput functional genomics, microarray-based experiments, annotation, data curation, and bioinformatics. She joined Computercraft in September 2007 as a contractor for NCBI, working as a curator for Gene Expression Omnibus (GEO), the leading fully public repository for high-throughput functional genomic data. As a Staff Scientist, she will continue to process and curate GEO submissions.

In July 2009, **Jane E. Rall** was appointed to a Staff Scientist position after having been an MSD/Lockheed Martin contractor since 2007 at NCBI. Ms. Rall earned a Master of Science degree from the University of Tennessee for work in numerical analysis. Over the course of her career, she has worked as a software designer, software developer and software engineering manager at Manager Software Products, Inc., HFS, Inc. and ACE*COMM Corporation. She joined Lockheed Martin in August 2007 as a contractor for NCBI to provide configuration management and process development support for the newly created Quality Assurance group in the Information Engineering Branch. As a Staff Scientist, she will continue this work along with developing monitoring and reporting tools to alert project teams of problems with the NCBI Web site and provide NCBI management with production system metrics and error rates.

In July 2009, **Han Zhang, MD** joined NLM as a Visiting Fellow. Dr. Zhang is an associate professor in the Department of Medical Informatics at China Medical University. She has an MD degree from China Medical University and a Masters degree in epidemiology and health statistics. Her research involves data mining of a biomedical literature database, extracting semantic relations for knowledge discovery and construction of domain ontology. At NLM, Professor Zhang will be exploring statistical and graph theoretic methods to exploit MEDLINE citations for text mining, literature-based discovery, and enhanced information retrieval.

*NLM Associate Fellows Program for 2009 – 2010*

The National Library of Medicine Associate Fellowship Program is a one-year postgraduate training fellowship at the NLM in Bethesda, Maryland, with an optional second year program component. The program is designed to provide a broad foundation in health sciences information services, and to prepare librarians for future leadership roles in health sciences libraries and in health services research. The Associate Fellows are introduced to a wide range of technologies and skills used in managing information at a national library. Five new Associate Fellows begin their year at NLM on August 31, 2009.

**Ada Cornell** – Syracuse University, May 2009. A double major in Environmental Studies and Spanish, Ada also has a certificate in international studies from Binghamton University (NY). She most recently was a reference librarian, youth services assistant, and clerk at the Fayetteville Free Library (Fayetteville, NY). Ada has completed internships at F. Franklin Moon Library at SUNY College of Environmental Science and Forestry, a conservation internship at Green Mountain National Forest and a legislative internship at the Sierra Club (Albany, NY).

**MaShana Davis** – University of Maryland, College Park, December 2007. MaShana received a BS in Computer Science from the University of Maryland, Eastern Shore. MaShana is currently working as a Technical Communications Liaison for the Statistics and Measurement Program at the Association of Research Libraries (ARL) in Washington, DC. She facilitates communications for and provides training to participants and other interested parties, and performs Web development and planning activities for ARL New Measures Initiatives: LibQUAL+®, ClimateQUAL, and MINES for Libraries.

**Sarah Westphal** – University of Wisconsin – Milwaukee, May 2009. Sarah received a BS in Agricultural Sciences, with a major in Food Science from the University of Wisconsin – Madison. She was previously a regulatory specialist with The Pillsbury Company and Target Corporation. Sarah is currently working as a cataloging intern at Marquette University's Raynor Memorial Libraries in Milwaukee, WI.

**Yani Yancey** – University of Maryland, College Park, August 2009. Yani received a BS in Speech Language Pathology and Audiology from Ithaca College in Ithaca, NY. Yani has worked in medical reimbursement services and patient assistance. Currently she is working as a Graduate Assistant at the libraries in University of Maryland, College Park, working in the Information Technology Division to configure and maintain Metalib and SFX for the University and affiliated institutions. She is a 2008-2009 ALA Spectrum Scholar and also a fellow in the Association of Research Libraries Career Enhancement Program for 2009.

**Holly Zerbe** – MSLIS, Drexel University, March 2009. Holly received a BS in Biology from Houghton College, Houghton, NY. She has previously worked as a curriculum developer for an online home school company. She is currently working at two internships, one in the print and photograph department of The Library Company of Philadelphia, PA; and as a licensing intern in the electronic acquisitions department at the Van Pelt Library at the University of Pennsylvania.

*Retirements and Separations*

**Donald W. King, MD** retired as Deputy Director for Research and Education in November 2008. Shortly after his arrival in 2002, he played a key role in negotiations that resulted in the US-wide license for the SNOMED CT clinical terminology, an important milestone in Federal health IT policy. As Acting Director of the Lister Hill Center for more than a year, Dr. King supported and enhanced its high-caliber research programs. He was also a pioneering force in the formation and evolution of the NIH Medline*Plus* magazine, which was created in 2006 and currently enjoys a readership of nearly 5 million. Highly supportive of NIH programs for recruitment, retention and leadership development, Dr. King was instrumental in initiating new avenues for developing the nation's next cadre of scientists. He worked to develop a series of nationwide seminars, to excite young people about possible careers in science. Throughout his tenure, Dr. King was a public health ambassador and advocate for NLM informational products and services, nurturing relationships with a diverse range of scientific and public health organizations and the library community, and orchestrating many memorable events to highlight NLM's mission and achievements.

In December 2008, **Philip Teigen, PhD**, retired as Deputy Chief of the History of Medicine Division after 24 years of service to NLM. Dr. Teigen received his BS from the University of Minnesota, and his MS and PhD from the University of Wisconsin. He worked for ten years at the Osler Library of the History of Medicine, McGill University, Montreal, Canada before coming to HMD in 1984. His long list of publications include work on bibliography (in particular but not limited to William Osler), Tudor-Stuart medicine, and veterinary medicine.

Dr. Teigen was an active member of the American Association for the History of Medicine, most recently serving as Chair of the Program Committee for the Annual Meeting, in 2007. He was also an active member of the Archivists and Librarians in the History of the Health Sciences, and took a special interest in the study of rhetoric as applied to the writing of history.

*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 recipients of the 2009 award were **Alan R. Aronson, PhD** for the development of two natural language processing tools, MetaMap and the NLM Medical Text Indexer which provides in-house assistance to NLM indexers; and

**James R. Marcetich,** in recognition of significant contributions in the ongoing effort to automate and enhance the indexing of the biomedical journal literature for the MEDLINE Database.

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 2009 award were **Stacey J. Arnesen** in recognition of innovative and exceptional contributions to NLM's disaster information resources and activities; and **Terry M. Wittig,** in recognition of her excellent leadership in developing an active program for licensing access to e-Journals and other electronic resources for NLM.

The NLM Director's Honor Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to **Valerie Florance, PhD** in recognition of sustained excellence in the administration of NLM's Informatics Research Training Programs and outstanding leadership in the development of a creative plan of action in response to the American Recovery and Reinvestment Act of 2009; and **Deborah A. Zarin, MD,** in recognition of her inventiveness, vision and continuous enthusiasm in leading ClinicalTrials.gov through a period of extraordinary growth in volume and scope.

The NIH Merit Award was presented to four individuals. The individual recipients were **Sameer K. Antani, PhD** for significant contributions to Content-Based Image Retrieval by integrating two complementary, geographically-separated systems, enabling biomedical image retrieval by exploiting the strengths of each; **Kin Wah Fung, MD** for sustained excellence in multiple initiatives involving applied medical terminology research; **Paula M. Kitendaugh** for exceptional customer service and management of many NLM products for consumers and professionals including MedlinePlus, MedlinePlus en español, and Go Local; and **Quang Le** for his outstanding information technology support for the Cognitive Science

Branch and the Lister Hill Center Computing Resources Team.

The NIH Director's Award was presented to two groups. The award recipients were **Jill L. Newmark** and **Patricia N. Tuohy** for the development of the highly successful Traveling Exhibition Program in support of NLM's and NIH's goals for outreach and education; and **Elliot R. Siegel, PhD**, as a member of the National Center on Minority Health and Health Disparities, NIH-GROUP, NIH Health Disparities Summit, in recognition of contributions made in planning and coordinating the first NIH Summit: The Science of Eliminating Health Disparities, December 16-18, 2008.

The Phillip C. Coleman Award was presented to **Margaret V. Slovikosky** for continued support of and participation in the NLM Mentoring TEAM (Together Everyone Achieves More) program.

### Table 21: FY2009 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 | 54 |
| Office of Computer and Communications Systems | 49 |
| Extramural Programs | 15 |
| Lister Hill National Center for Biomedical Communications | 64 |
| National Center for Biotechnology Information | 224 |
| Specialized Information Services | 40 |
| Library Operations | 284 |
| **TOTAL FTEs** | **754** |

### American Recovery and Reinvestment Act

On February 17, 2009, H.R.1 was signed into law as the American Recovery and Reinvestment Act (ARRA) in an unprecedented effort to jumpstart the economy, create or save millions of jobs, and put a down payment on addressing long-neglected challenges facing the nation. NIH directly received $10 billion of two-year funds, of which $7.4 billion was transferred directly to the ICs and Common Fund in proportion to FY2009 Appropriations. NLM's portion of this distribution was $83,643,000, of which $37,070,349 was obligated in FY2009. The balance will be obligated in FY2010.

### NLM Diversity Council

*NLM Director's Employee Education Fund*

The Director's Employee Education Fund was established in 1998 by the Director, Donald A.B. Lindberg, MD, to help enhance and further careers of NLM employees. In 2009, the NLM Diversity Council continued coordination of the Director's Employee Education Fund. During FY 2009, the fund made continuing education possible for 68 employees to take 75 courses. Undergraduate classes made up of a majority of the classes supported. The University of Maryland had the highest enrollment (22), with Montgomery College running second (11), and Strayer University took third place (five). Other institutions attended included: Catholic University, Johns Hopkins University, George Mason University, American University, Montgomery College, Walden University, Bowie State University, University District of Columbia, the USDA Graduate School, College of Notre Dame of Maryland, Wesley Theological Seminary, and Frederick Community College. Course disciplines included: Strategic Management of Technologies; Information Technology Project Management; Principles of College Chemistry; Human Anatomy and Physiology I; Elementary Logic and Semantics; Introduction to Computer Applications; Fundamentals of Epidemiology; Introduction to Networking; SQL Programming; Advanced Public Relations; Basic Indexing; Elementary Japanese II; and Nursing-Health Policy, among others. The Diversity Council continues to publicize the availability of the fund.

**NLM 2009 Health Expo**

On April 23, 2009, the NLM Diversity Council once again sponsored the "Healthy Lifestyles for You and Your Family" expo in conjunction with NIH's "Take Your Child to Work Day." This was the second of what most likely will be an annual event and this year NLM was joined by the NIH Diversity Council who sponsored a table and coloring event for the children attending the Expo. The products of this event, which depicted the children's ideas of how diversity looks, were later displayed in building 10 during the NIH Faces of Diversity Celebration. The NLM Expo featured a full day's worth of educational, family-oriented activities focusing on healthy living styles for both adults and children and this year was also flavored with a multi cultural theme. The Emmy Award-winning Food-Play Productions which uses both live theatre and interactive media to promote healthy eating and exercise activities were joined this year by multiple yoga classes, Tae Kuk Martial Arts demonstrations and a traditional Chinese Lion Dance all promoting healthy exercise as a way of maintaining good health. In addition, an inflatable obstacle course was erected for the children to use and exercise during this event proved to be a great hit. This year's Expo was attended by more than 1,500 NIH employees and their families and is proving to be one of the most popular events held here at the NIH. With this overwhelming response, NLM will be planning to continue this event for many   years.

# 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.

HARRIS, C. Martin, MD **(Chair)**
Chief Information Officer and Chairman
Information Technology Division
The Cleveland Clinic Foundation
Cleveland, OH 44195

**Appointed Members**

COHEN, Jordan J., MD
Professor of Medicine
George Washington University
Washington, DC

CONNOLLY, John E., MD, Honorable
Professor of Surgery
University of California, Irvine
Orange, CA

JAMES, Bruce R., Honorable
President and CEO
Nevada New-Tech Inc.
Incline Village, NV

MITCHELL, Joyce A., PhD
Professor and Chair and Associate Vice
President
Department of Biomedical Informatics
University of Utah School of Medicine
Salt Lake City, UT

FRIEDMAN, Carol, PhD
Professor and Vice Chair
Department of Biomedical Informatics
Columbia University
New York, NY

ROSSITER, Louis F., PhD
Research Professor
The Thomas Jefferson Program in Public Policy
Williamsburg, VA

ISOM, O. Wayne, MD
Terry Allen Kramer Professor of Cardiothoracic Surgery
New York Presbyterian-Weill Cornell Medical School
New York, NY

STANLEY, Eileen H., MLS
Roseville, MN

TANJI, Virginia, MSLS, MED
Director
Health Science Library
University of Hawaii at Manoa
School of Medicine
Honolulu, HI

**EX OFFICIO MEMBERS**

FRIERSON, Eleanor G., MSLS
Acting Director
National Agricultural Library
US Department of Agriculture
Beltsville, MD

BILLINGTON, James H., D. Phil.
Librarian of Congress
Library of Congress
10 First Street, S.E.
Washington, DC

ROUDEBUSH, James G., Lt. Gen., USAF, MC
Surgeon General
United States Air Force
1780 Air Force Pentagon
Washington, DC

COLLINS, James, PhD
Asst. Director, Biological Sciences
National Science Foundation
Arlington, VA

KUSSMAN, Michael J., MD, MS, MACP
Acting Under Secretary for Health
Veterans Health Administration
Washington, DC

MARCUM, Deanna, PhD
Associate Librarian for Library Services
Library of Congress
Washington, DC

HEPBURN, Byron, MD
Deputy Surgeon General
United States Air Force
Bolling AFB, DC

HIRSH, Haym, PhD
Division Director, Information and Intelligence Systems
National Science Foundation
Arlington, VA

GRAHAM, Gail
Director, Health Data and Informatics
Veterans Health Administration
Washington, DC

ROBINSON, Adam M., Vice Admiral
Surgeon General of the Navy
Chief, Bureau of Medicine and Surgery
Department of the Navy
Washington, DC

CORRIERE, Michael A., PhD
CDR, MSC, USN
Dean of Academics
Director, Officer Programs & Plans
Naval Medical Manpower, Training and Education
Command
Bethesda, MD

SCHOOMAKER, Eric B., MD, PhD
The Surgeon General/Commander
US Army Medical Command
Falls Church, VA

BENJAMIN, Regina., MD, MBA
Surgeon General
US Public Health Service
Rockville, MD

RICE, Charles L., MD
President
Uniformed Services University of the Health Sciences
Bethesda, MD

POWERS, John, COL
Director, Medical Education
Office of the Surgeon General
Department of the Army
Falls Church, VA

MISHOE, Helena O., PhD
Director, Office of Minority Health Affairs
National Heart, Lung and Blood Institute
National Institutes of Health
Bethesda, MD

SMITH, Dale, PhD
Senior Vice President
Uniformed Services University of the Health Sciences
Bethesda, MD

# 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
Associate Director
Department of Medical Informatics and Clinical Epidemiology
Oregon Health Sciences University
Portland, OR

BAKKEN, Suzanne, DNSc, RN, FAAN
Professor
Departments of Nursing and Biomedical Informatics
Columbia University
New York, NY

CHUEH, Henry C., MD
Director & Chief
Laboratory of Computer Science
Division of Biomedical Informatics
Massachusetts General Hospital
Boston, MA

HUFF, Stanley M., MD
Chief Medical Informatics Officer
Intermountain Health Care
Information Systems
Salt Lake City, UT

LUMPKIN, John R., MD, MPH
Senior Vice President & Director
Health Care Group
The Robert Wood Johnson Foundation
Princeton, NJ

SHNEIDERMAN, Ben, PhD
Professor
Department of Computer Science
University of Maryland, College Park
College Park, MD

SILVERSTEIN, Jonathan C., MD
Assistant Professor
Department of Surgery
University of Chicago
Computation Institute
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.

GINSBURG, David, MD **(Chair)**
James V. Neel Distinguished University Professor
Internal Medicine and Human Genetics
Life Sciences Institute
University of Michigan
Ann Arbor, MI

**Appointed Members**

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

BABBITT, Patricia C., PhD
Professor
Department of Bioengineering & Therapeutic Sciences
University of California, San Francisco
San Francisco, CA

BENHAM, Craig J., PhD
Professor
Department of Biomedical Engineering
UC Davis Genome Center

University of California, Davis
Davis, CA

LEE, Christopher J., PhD
Professor
Department of Chemistry & Biochemistry
Molecular Biology Institute
University of California, Los Angeles
Los Angeles, CA

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

WENG, Zhiping, PhD
Professor and Director
Program in Bioinformatics and Integrative Biology
University of Massachusetts Medical School
Worchester, 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.

MARCHIONINI, Gary J., PhD **(Chair)**
Cary C. Boshamer Professor
School of Information and Library Science
University of North Carolina at Chapel Hill
Chapel Hill, NC

**Appointed Members**

CONSALES, Judith C., MLS
Director
Louis M. Darling Biomedical Library
University of California, Los Angeles
Los Angeles, CA

COOPER, Gregory F., MD, PhD
Associate Professor
Department of Biomedical Informatics
University of Pittsburgh
Pittsburgh, PA

GONZALEZ, Graciela H., PhD
Assistant Professor
Department of Biomedical Informatics
Arizona State University
Phoenix, AZ

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

KAHN, Michael G., MD, PhD
Associate Professor
Department of Pediatrics
University of Colorado
The Children's Hospital
Aurora, CO

LU, Xinghua, MD, PhD
Associate Professor
Department of Biochemistry
Medical University of South Carolina
Charleston, SC

MANDL, Kenneth D., MD, MPH
Associate Professor

Children's Hospital Informatics Program
Children's Hospital Boston
Boston, MA

MENDONCA, Eneida A., MD, PhD
Associate Professor
Department of Pediatrics, Section of
Hematology/Oncology
The University of Chicago
Chicago, IL

MOORE, Jason H., PhD
Professor of Genetics
Dartmouth-Hitchcock Medical Center
Dartmouth College
Lebanon, NH

MURPHY, Beverly, MLS
Assistant Director
Marketing and Publications
Duke University Medical Center Library
Durham, NC

PANI, John R., PhD
Associate Professor
University of Louisville
Louisville, KY

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

SHATKAY, Hagit, PhD
Associate Professor
Head of the Computational Biology & Machine Learning
Lab
School of Computing
Queen's University
Kingston, Ontario, Canada

SITTIG, Dean F., PhD
Associate Professor
UT Memorial Hermann Center for Healthcare Quality and
Safety

The University of Texas School of Health Information
Sciences at Houston
Houston, TX

SRINIVASAN, Padmini, PhD
Professor
Computer Science, Management Sciences, School of
Library & Information Sciences
The University of Iowa
Iowa City, IA

STARREN, Justin, MD, PhD
Director
Biomedical Informatics Research Center
Marshfield Clinic Research Foundation
Marshfield, WI

STATES, David J., MD, PhD
Professor of Health Information Science
School of Health Information Sciences
Brown Foundation Institute of Molecular Medicine

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.

FLEMING, David A., MD **(Chair)**
Professor and Chairman,
Department of Internal Medicine
Director, MU Center for Health Ethics
University of Missouri School of Medicine
Columbia, MO

**Appointed Members**

CHRISTOPHER, Mary M., PhD
Professor of Pathology
Dept. of Pathology, Microbiology & Immunology
School of Veterinary Medicine
University of California
Davis, CA

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

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

DU, Chunying, PhD
Associate Professor
Cancer and Cell Biology
University of Cincinnati
School of Medicine
Cincinnati, OH

ELPERN, David J., MD
Dermatologist
The Skin Clinic
Williamstown, MA

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

JACKSON, Gretchen P., MD, PhD

Assistant Professor
Surgery and Biomedical Informatics
Vanderbilt Children's Hospital
Nashville, TN

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

NORTON, Catherine N., MLS
Director, Information Technology
Marine Biological Laboratory
Woods Hole, MA

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

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

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

ZHANG, Ge, MD, PhD
Assistant Professor
Department of Biomedical Engineering
University of Akron
Akron, OH

**Scientific Review Administrator**

KOTZIN, Sheldon, MLS
Associate Director, Library Operations
National Library of Medicine, NIH
Bethesda, MD

# 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.

WARD, Gary E., PhD **(Chair)**
Professor
Department of Microbiology & Molecular Genetics
University of Vermont
Burlington, VT

**Appointed Members**

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

ALIRE, Camila, EDD
Dean Emeritus
University Libraries
University of New Mexico &
Colorado State University
Sedalia, CO

BIRD, Christopher J., BA
Solicitor
Legal Department
Wellcome Trust
London, England
United Kingdom

BLANTON, Ronald E., MD
Professor of Medicine
Center for Global Health and Diseases
Case Western Reserve University
Cleveland, OH

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

KANN, Maricel G., PhD
Assistant Professor
Department of Biological Sciences

University of Maryland
Baltimore, MD

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, Chapel Hill
Chapel Hill, NC

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

TANNER, R. Michael., PhD
Provost and Vice Chancellor for Academic Affairs
University of Illinois at Chicago
601 S. Morgan Street
Chicago, IL

VELTEROP, Johannes, PhD
Chief Executive Officer
Knewco, Inc.
Rockville, Maryland

WEINTRAUB, Susan T., PhD
Professor, Department of Biochemistry
Director, Mass Spectrometry Laboratory
The University of Texas Health Science Center
at San Antonio
San Antonio, TX

WILBANKS, John T., BA
Boston, MA

**Executive Secretary**

LIPMAN, David J., MD
Director
National Center for Biotechnology Information
National Library of Medicine
Bethesda, MD

102

# 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
ARRA      American Recovery and Reinvestment Act
ASCCP     American Society for Cervical Pathology and Colposcopy
ASPR      Assistant Secretary for Preparedness and Response, HHS Office of the
BAC       Bacterial Artificial Chromosome
BarSTool  Barcode Submission Tool
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
CAM       Complementary and Alternative Medicine
C&A       Certification & Accreditation (audit)
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
CEL       Affymetrix Cell intensity (file)
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
CRAC      Computer Room Air Conditioner
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 | | GENSAT | Gene Expression Nervous System Atlas |
| dbSNP | Database of Single Nucleotide Polymorphism | | GENE-TOX | Genetic Toxicology |
| DDBJ | DNA Data Bank of Japan | | GHR | Genetics Home Reference |
| DDD | Drug Delivery Devices | | GIS | Geographic Information System |
| DCMS | Data Creation and Maintenance System | | GO | Grand Opportunity grant |
| DEAS | Division of Extramural Administrative Support | | GMAC | Grants Management Advisory Committee |
| DHHS | Department of Health and Human Services | | GPS | Global Positioning System |
| DICOM | Digital Imaging and Communications in Medicine | | GRMS | Global Record Maintenance System |
| | | | GSA | General Services Administration |
| DIMRC | Disaster Information Management Research Center | | GSS | Genome Survey Sequences |
| | | | GUI | Graphic User Interface |
| DIRLINE | Directory of Information Resources Online | | GWAS | Genome Wide Association Studies |
| DNA | Deoxyribonucleic Acid | | HapMap | Haplotype Map |
| DPR | Digital Preservation Research | | HAVnet | Haptic Audio Video Network for Education Technology |
| DRAGON | Dynamic Resource Allocation in GMPLS Optical Networks | | HBCU | Historically Black Colleges and Universities |
| DRESWG | Digital Repository Evaluation and Selection Working Group | | HD | High Definition |
| | | | HHS | Health and Human Services |
| DRIG | Digital Repository Implementation Group | | HIPAA | Health Insurance Portability and Accounting Act |
| DTD | Document Type Definition | | HITSP | Healthcare Information Technology Standards Panel |
| DVTS | Digital Video Transport System | | | |
| EBI | European Bioinformatics Institute | | HLA | Human Leukocyte Antigen |
| EBP | Evidence-Based Practice | | HL7 | Health Level Seven, Inc. |
| ECHO | European Community Humanitarian Office | | HMD | History of Medicine Division |
| | | | HSDB | Hazardous Substances Data Bank |
| Educollab | Educational Collaborators | | HPCC | High Performance Computing and Communications |
| EEO | Equal Employment Opportunity | | | |
| EFTS | Electronic Funds Transfer Service | | HPV | Human Papillomavirus |
| EHR | Electronic Health Record | | HRSA | Health Resources and Services Administration |
| EMBL | European Molecular Biology Laboratory | | | |
| EMR | Electronic Medical Record | | HSRProj | Health Services Research Projects |
| EMS | Emergency Medical Services | | HSRInfo | Health Services Research Information |
| EnHIP | Environmental Health Information Partnership | | HSRR | Health Services and Sciences Research Resources |
| EnHIOP | Environmental Health Information Outreach Program | | HSTAT | Health Services and Technology Assessment Text |
| EP | Extramural Programs | | HuGENet | Human Genome Epidemiology Network |
| EPA | Environmental Protection Agency | | I3 | Image Indexing Initiative |
| eRA | Electronic Research Administration | | IAIMS | Integrated Advanced Information Management Systems |
| ESI | Early Stage Investigators | | | |
| EST | Expressed Sequence Tag | | IBIS | Inferred Biomolecular Interactions Server |
| ETIC | Environmental Teratology Information Center | | ICMJE | International Committee of Medical Journal Editors |
| EUREKA | Exceptional, Unconventional Research Enabling Knowledge Acceleration | | ICs | Institutes and Centers (of NIH) |
| | | | ICT | Information and Communication Technologies |
| FDA | Food and Drug Administration | | | |
| FDCC | Federal Desktop Core Configuration | | IEB | Information Engineering Branch |
| FHA | Federal Health Architecture | | IGSTK | Image Guided Surgery Toolkit |
| FIC | Fogarty International Center | | IHTSDO | International Health Terminology Standards Development Organization |
| FNLM | Friends of the National Library of Medicine | | | |
| | | | IHM | Images from the History of Medicine |
| FTE | Full Time Employee | | ILL | Interlibrary Loan |
| FTP | File Transfer Protocol | | ILS | Integrated Library System |
| GAIN | Genetic Association Information Network | | IMPAC | Information Management Planning Analysis And Coordination |
| Gbps | Gigabits per Second | | | |
| GDS | GEO DataSet | | InCHIs | IUPAC International Identifiers |
| GEO | Gene Expression Omnibus (database) | | | |

| | | | | |
|---|---|---|---|
| INDSC | International Nucleotide Sequence Database Collaboration (formerly DDBJ/EMBL/GenBank) | MLAA | Medical Library Assistance Act |
| *info*SIDA | *info*Síndrome de Inmunodeficiencia Adquirida (database) | MLB | Medical Language Branch (database server) |
| IP | Interactive Publications | MMDB | Molecular Modeling DataBase |
| IPv6 | Next Generation Internet, Version 6 | MMS | MEDLARS Management Section |
| IRB | Institutional Review Board | MMTx | MetaMap Technology Transfer |
| IRIS | Integrated Risk Information System | MOR | Medical Ontology Research |
| IRMA | Image Retrieval for Medical Applications | MOU | Memorandum of Understanding |
| ISTO | Image Storage and Transmission Optimization | MTHSPL | Metathesaurus Structured Product Labels |
| IT | Information Technology | MTI | Medical Text Indexer |
| ITER | International Toxicity Estimates for Risk | MTMS | MeSH Translation Management System |
| ITK | Insight Toolkit | NA-MIC | National Alliance of Medical Image Computing |
| ITP | Informatics Training Program | NAS | National Academy of Sciences |
| IUPAC | International Union of Pure and Applied Chemistry | NCBC | National Centers for Biomedical Computing |
| JDBC | Java Database Connectivity | NCBI | National Center for Biotechnology Information |
| JDI | Journal Descriptor Indexing | NCCS | NIH Consolidated Collocation Site |
| JDMS | Journal Descriptor Maintenance System | NCHS | National Center for Health Statistics |
| JRE | Java Runtime Environment | NCMHD | National Center for Minority Health and Health Disparities |
| KSS | Knowledge Source Server (data) | NCI | National Cancer Institute |
| LactMed | Drugs and Lactation (database) | NCRR | National Center for Research Resources |
| LAN | Local Area Network | NCVHS | National Committee on Vital and Health Statistics |
| LHC | Lister Hill Center | N*e*HC | National *e*-Health Collaborative |
| LHNCBC | Lister Hill National Center for Biomedical Communications | NEI | National Eye Institute |
| LJI | List of Journals Indexed | NGI | Next Generation Internet |
| LO | Library Operations | NHANES | National Health and Nutrition Examination Surveys |
| LOINC | Logical Observations Identifiers, Names, Codes | NHGRI | National Human Genome Research Institute |
| LPF | Lost Person Finder | NHIN | National Health Information Network |
| LRP | Long Range Plan (NLM) | NHLBI | National Heart, Lung, and Blood Institute |
| LSI | List of Serials Indexed | NIA | National Institute on Aging |
| LSTRC | Literature Selection Technical Review Committee | NIAID | National Institute of Allergy and Infectious Diseases |
| LVG | Lexical Variant Generator | NIBIB | National Institute of Biomedical Imaging and Bioengineering |
| MARG | Medical Article Records Groundtruth | NICHD | National Institute of Child Health and Human Development |
| MARS | Medical Article Records System | NICHSR | National Information Center on Health Services Research and Health Care Technology |
| MAX | Mid Atlantic Exchange, U. of Maryland | | |
| MDoT | MEDLINE Database on Tap | NIDCD | National Institute on Deafness and other Communication Disorders |
| MDT | Multimedia Database Tool | | |
| MEDLARS | Medical Literature Analysis and Retrieval System | NIDCR | National Institute of Dental and Cranio-facial Research |
| MEDLINE | MEDLARS Online | NIDDK | National Institute of Diabetes, Digestive, and Kidney Diseases |
| MegaBLAST | Basic Local Alignment Search Tool | | |
| MEME | Metathesaurus Editing and Maintenance Environment | NIEHS | National Institute of Environmental Health Sciences |
| MEO | Medical Education and Outreach | NIGMS | National Institute of General Medical Sciences |
| MeSH | Medical Subject Headings | | |
| MHC | Major Histocompatibility Complex | NIH | National Institutes of Health |
| MID | Manuscript Identifiers | NIHMS | NIH Manuscript Submission |
| MICAD | Molecular Imaging and Contrast Database | NIH PI | NIH Pathways to Independence Award |
| MIM | Multilateral Initiative on Malaria | | |
| MIMCom | MIM Communications Working Group | | |
| MIRS | Medical Information Retrieval System | | |
| MLA | Medical Library Association | | |

| | | | | |
|---|---|---|---|---|
| NIMH | National Institute of Mental Health | | PHLIP | Public Health Law Information Project |
| NINDS | National Institute of Neurological Disorders and Stroke | | PHII | Public Health Informatics Institute |
| | | | PHP | Public Health Partners |
| NIOSH | National Institute for Occupational Safety and Health | | PHR | Personal Health Record |
| | | | PHS | Public Health Service |
| NIST | National Institute of Standards and Technology | | PI | Pathway to Independence award |
| | | | PICO | Patient/Population, Intervention, Comparison, and Outcome |
| NLM | National Library of Medicine | | | |
| NLP | Natural Language Processing System | | PLAWARe | Programmable Layered Architecture With Artistic Rendering |
| NN/LM | National Network of Libraries of Medicine | | | |
| NNMC | National Naval Medical Center | | PMC | PubMed Central |
| NNO | National Network Office | | PMCI | PubMed Central International |
| NOSC | Network Operations and Security Center | | PMC ID | PubMed Central Identification (number) |
| NOVA | National Online Volumetric Archive | | PRS | Protocol Registration System |
| NRCBL | National Reference Center for Bioethics Literature | | PSD | Public Services Division |
| | | | PUG | PubChem Power User Gateway |
| NSF | National Science Foundation | | QCIM | Quarterly Cumulative Index Medicus |
| NTCC | National Online Training Center and Clearinghouse | | RAC | Real Application Clusters |
| | | | RCDC | Research Condition and Disease Categorization |
| OACF | Onsite Alternate Computing Facility | | | |
| OAM | Office of Administrative Management | | RCSB | Research Collaboratory for Structural Bioinformatics |
| OCCS | Office of Computer and Communications Systems | | | |
| | | | RDMS | Rare Disease Maintenance System |
| OCHD | Coordinating Committee on Outreach, Consumer Health and Health Disparities | | RefSeq | Reference Sequence (database) |
| | | | REMM | Radiation Event Medical Management |
| OCPL | Office of Communications and Public Liaison | | RFA | Request for Applications |
| | | | RFID | Radio Frequency Identification |
| OCR | Optical Character Recognition | | RFP | Request for Proposals |
| OD | Office of the Director | | RHIN | Refugee Health Information Network |
| ODIMRC | Office of the Disaster Information Management Research Center | | RIDeM | Repository for Informed Decision Making |
| | | | RML | Regional Medical Library |
| OERC | Outreach Evaluation Resource Center | | RNA | Ribonucleic Acid |
| ORF | Original Release Format | | RNAi | RNA Interference |
| OHIPD | Office of Health Information Programs Development | | RPS-BLAST | Reversed Position Specific BLAST |
| | | | RRF | Rich Release Format |
| OMB | Office of Management and Budget | | RSS | Really Simple Syndication |
| OMIA | Online Inheritance in Animals (database) | | RTECS | Registry of Toxic Effects of Chemical Substances |
| OMIM | Online Mendelian Inheritance in Man (database) | | | |
| | | | RTLS | Real Time Location System |
| OMSSA | Open Mass Spectrometry Search Algorithm | | RWJF | Robert Wood Johnson Foundation |
| | | | SAB | Source Abbreviations |
| ONC | Office of the National Coordinator (for Health Information Technology) | | SBIR | Small Business Innovation Research |
| | | | SCR | (MeSH) Supplemental Chemical Records |
| OPASI | Office of Portfolio Analysis and Strategic Initiatives | | SDK | Software Development Kit |
| | | | SEF | Serials Extract File |
| OSA | Optical Society of America | | SEP | Special Emphasis Panel |
| ORWH | Office of Research on Women's Health | | SIDA | Swedish International Development Agency |
| OSIRIS | Open Source Independent Review and Interpretation System | | | |
| | | | SII | Scalable Information Infrastructure |
| PAHO | Pan American Health Organization | | SIS | Specialized Information Services |
| PBM | Pharmacy Benefit Manager | | SKR | Semantic Knowledge Representation |
| PCA | Personal Computer Advisory Committee | | SMART | Scalable Medical Alert and Response Technology |
| PCR | Polymerase Chain Reaction | | | |
| PDA | Personal Digital Assistant | | SNOMEDCT | Systematized Nomenclature of Medicine Clinical Terms |
| PDR | Publisher Data Review | | | |
| PDB | Protein Data Bank | | SOAP | Simple Object Oriented Protocol (formerly Simple Object Access Protocol) |
| PDF | Portable Document Format | | | |
| PDL | Personal Digital Library | | 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 |
| TBL | The bottom line |
| 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*
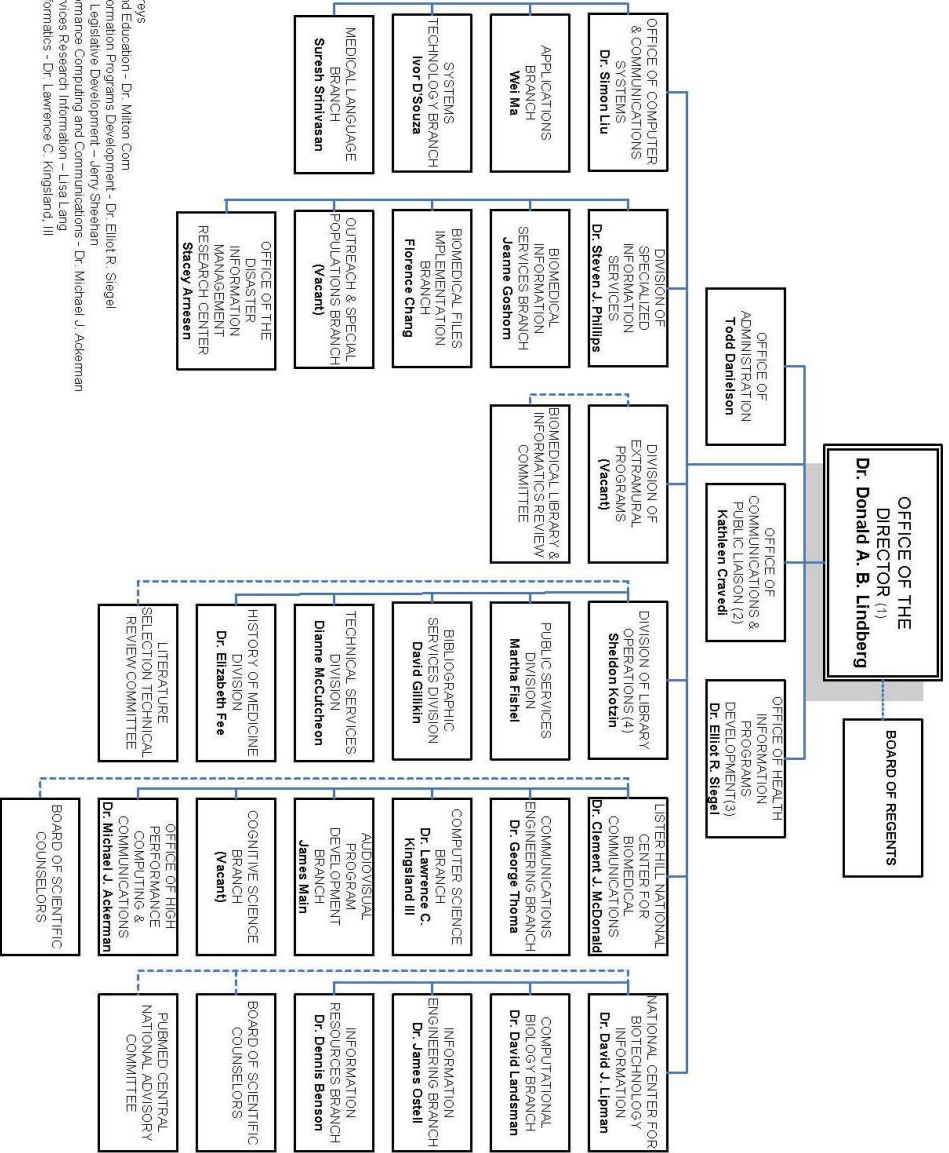
*Bethesda, MD 20894*

*301-496-6308*

*E-mail: publicinfo@nlm.nih.gov*

*Web: www.nlm.nih.gov*

**Cover**:  Poster by NIH Medical Arts commemorating the Twentieth Anniversary of the National Center for Biotechnology Information (1989-2009).

# National Library of Medicine

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

BOARD OF REGENTS

OFFICE OF ADMINISTRATION
**Todd Danielson**

OFFICE OF COMMUNICATIONS & PUBLIC LIAISON (2)
**Kathleen Cravedi**

OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT(3)
**Dr. Elliot R. Siegel**

## OFFICE OF COMPUTER & COMMUNICATIONS SYSTEMS
**Dr. Simon Liu**

- APPLICATIONS BRANCH — **Wei Ma**
- SYSTEMS TECHNOLOGY BRANCH — **Ivor D'Souza**
- MEDICAL LANGUAGE BRANCH — **Suresh Srinivasan**

## DIVISION OF SPECIALIZED INFORMATION SERVICES
**Dr. Steven J. Phillips**

- BIOMEDICAL INFORMATION SERVICES BRANCH — **Jeanne Goshorn**
- BIOMEDICAL FILES IMPLEMENTATION BRANCH — **Florence Chang**
- OUTREACH & SPECIAL POPULATIONS BRANCH — **(Vacant)**
- OFFICE OF THE DISASTER INFORMATION MANAGEMENT RESEARCH CENTER — **Stacey Arnesen**

## DIVISION OF EXTRAMURAL PROGRAMS
**(Vacant)**

BIOMEDICAL LIBRARY & INFORMATICS REVIEW COMMITTEE

## DIVISION OF LIBRARY OPERATIONS (4)
**Sheldon Kotzin**

- PUBLIC SERVICES DIVISION — **Martha Fishel**
- BIBLIOGRAPHIC SERVICES DIVISION — **David Gillikin**
- TECHNICAL SERVICES DIVISION — **Dianne McCutcheon**
- HISTORY OF MEDICINE DIVISION — **Dr. Elizabeth Fee**

LITERATURE SELECTION TECHNICAL REVIEW COMMITTEE

BOARD OF SCIENTIFIC COUNSELORS

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

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

BOARD OF SCIENTIFIC COUNSELORS

## NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION
**Dr. David J. Lipman**

- COMPUTATIONAL BIOLOGY BRANCH — **Dr. David Landsman**
- INFORMATION ENGINEERING BRANCH — **Dr. James Ostell**
- INFORMATION RESOURCES BRANCH — **Dr. Dennis Benson**

PUBMED CENTRAL NATIONAL ADVISORY COMMITTEE

---

1. Deputy Director – Betsy Humphreys
   Deputy Director for Research and Education - Dr. Milton Corn
   Associate Director for Health Information Programs Development - Dr. Elliot R. Siegel
   Assistant Director for Policy and Legislative Development – Jerry Sheehan
   Assistant Director for High Performance Computing and Communications - Dr. Michael J. Ackerman
   Assistant Director for Health Services Research Information – Lisa Lang
   Assistant Director for Applied Informatics - Dr. Lawrence C. Kingsland, III

2. Includes Public Liaison Officer

3. Includes International Programs

4. Includes:
   National Network of Libraries of Medicine, Head – Dr. Angela Ruffin
   Medical Subject Headings Section, Chief - Dr. Stuart Nelson
   National Information Center on Health Services Research & Health Care Technology, Head – Lisa Lang

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

FY2009