NATIONAL INSTITUTES OF HEALTH

National Library of Medicine

Programs and Services

Fiscal Year 2012

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

National Library of Medicine Catalog in Publication

CONTENTS

Appendixes

Figures

Tables

PREFACE

Fiscal Year 2012 was a time of innovation and accomplishment at the National Library of Medicine. Notable achievements included the following:

- The Library kept its HHS*innovates* award streak alive when the Bethesda Hospitals' Emergency Preparedness Partnership won distinction as a "Secretary's Pick" for developing the Patient Tracking and Locating System, to help hospitals handle the surge of patients during a disaster. NLM is a BHEPP member.

- We launched the Genetic Testing Registry to address the need for information about genetic tests for healthcare providers, patients, researchers and others. The number of laboratory tests is rapidly growing; currently tests are available for over 2,500 diseases.

- NLM became a partner in a special arrangement to make the world's largest set of data on human genetic variation—produced by the international 1000 Genomes Project—publicly available on the Amazon Web Services cloud.

- We conducted our first *API Showcase*, a free webinar showing software developers how to tap into free NLM data covering drugs, medical literature, clinical trials, consumer health information, molecular biology and more.

- ClinicalTrials.gov, with information on over 134,000 clinical studies in all US states and 180 nations, was refreshed with an easy-to-navigate design and many new tables, charts, maps and other features.

- NLM joined leading Internet companies in the World IPv6 Launch event, June 6, 2012. NLM is successfully migrating to this updated and improved communications protocol, which succeeds the 30-year-old IPv4.

- This year, NLM forged new partnerships, private and public, collaborating with the Folger Shakespeare Library on an exhibition, *"And there's the humor of it" — Shakespeare and the four humors*, and with the National Endowment for the Humanities on future initiatives of common interest.

- With an eye toward improving the public health, NCBI developed a pathogen detection pipeline in conjunction with an FDA project that aims to rapidly identify pathogenic isolates in order to combat outbreaks of foodborne illnesses and antibiotic resistant hospital infections.

- An impressive milestone was reached when NLM scanned its one millionth page for the Medical Heritage Library Project. This cooperative venture is working to digitize historical materials from the collections of some of the nation's most prominent libraries.

- In conjunction with the National Institute on Aging (NIA), the Library released a redesigned version NIHSeniorHealth. The redesign improves usability and modernizes the site's appearance.

- The *Native Voices* exhibition continued to welcome visitors from the US and many nations. The release of a new iPad app allowed those who can't come to NLM to experience its rich contents digitally.

Proudly and with regularity, we continue to introduce new programs and services, improve those we have, and make all of them more readily available to users around the globe, for the betterment of the public health. I commend our talented and committed staff, advisors and consultants for continuing to do so, brilliantly.

Donald A.B. Lindberg, MD
Director

OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT

Michael F. Huerta, PhD
Associate Director

The Office of Health Information Programs Development (OHIPD) is responsible for four 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;
- Planning, conducting, and evaluating NLM's international programs; and
- Contributing to trans-NIH collaborations in digital biomedicine activities.

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 is organized around four key goals:

> **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

OHIPD also has overall responsibility for developing and coordinating the NLM Health Disparities Strategic Plan, outlining NLM strategies and activities undertaken in support of NIH efforts to understand and eliminate health disparities between minority and majority populations. This plan addresses priorities related to improving access to health information as a means to address health disparities, with special emphasis on rural, minority, and other underserved populations. Areas of emphasis include programs in research and research capacity building as well as information dissemination, community outreach, and public health education. NLM's Health Disparities Strategic Plan for 2009-2013 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 and the public. In FY2012, OHIPD staff continued outreach initiatives intended to encourage underrepresented minority high school students to utilize NLM's health information resources and to pursue careers in medicine and the health sciences, carried out in collaboration with other divisions of NLM. An NLM-wide Coordinating Committee on Outreach, Consumer Health and Health Disparities (OCHD) plans, develops, and coordinates NLM outreach and consumer health activities. The OCHD is chaired and staffed by OHIPD.

An initiative with ExploreHealthCareers.com and Mentoring In Medicine (MIM) continues to support an annual workshop entitled "Yes, I can be a Healthcare Professional," an inspiring and well-attended workshop convened for parents and children in grades 3 through college. 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 underrepresented minority African-American and Hispanic students located at schools in New York City's Harlem and South Bronx neighborhoods. An additional MIM Program, Science and Health Career Exploration, continues to be funded by NLM and co-sponsored by the Friends of the NLM. This initiative reaches 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 oversee 40 sessions of biology instruction in 12 organ systems, taught by visiting health professionals/ mentors over a two-year period. During its four-year operation, the program has exposed over 700 minority students to health care career instruction. In the process, the program has developed four after-school curricula, nine knowledge tests and a series of pre- and post-course evaluation surveys and instructor feedback forms. Program evaluations demonstrate continued impressive gains in health care knowledge. 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.

NLM also continues its focused efforts to meet the health information needs of the Hispanic population in Texas and elsewhere. In FY2012, NLM continued its support for the South Texas High School for the Health Professions, known as MedHigh, a magnet health high school in the Lower Rio Grand Valley of Texas. The MedHigh iVIVA! 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. MedlinePlus en español is being emphasized where applicable. The project also includes a health careers tutoring component that involves the students, teachers, and guidance counselors. In FY12, a health literacy video challenge was conducted to encourage students in grades 6-12 to develop videos promoting use of Medlineplus.gov. Over 80 entries were received. Peer tutoring has been extended to other magnet high schools in the Lower Rio Grande Valley, where it continues to be an effective outreach program. A practical "how to" implementation guide was drafted, to assist other high schools in considering and implementing a peer tutoring program. The guide will be finalized in FY2013.

Native American Outreach

In FY2012 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 a dozen pow-wows, mostly in the Mid-Atlantic area. An estimated 10,000 persons visited the NLM booth over the course of these pow-wows. These activities are yet another way to increase awareness of NLM's health information resources among segments of the Native American community and the general public.
OHIPD also supported a project in North Dakota that resulted largely from the Native American Listening Circles conducted in prior years and continued into FY2012. At Cankdeska Cikana Community College (via the Greater Midwest RML), Spirit Lake Nation, Ft. Totten, ND, a continuing project supports improvements at the tribal college library and development of health information-related educational and community outreach activities run by the library.

Native Voices Exhibition

OHIPD staff also contributed to the outreach, follow-up and enhancement of the NLM exhibition, *Native Voices: Native Peoples' Concepts of Health and Illness*.
In FY2012, OHIPD staff participated with other NLM divisions in coordinating and facilitating visits to

Native Voices by the Hawaiian Civic Club of Washington DC, South Central Foundation of Anchorage, American Public Health Association (Native Caucus), American Indian Higher Education Consortium, National Association of Community Health Centers, and National Indian Health Board, among others. OHIPD staff also participated in *Native Voices* displays at various conference or event locales, such as the Hawaii Healing Weekend at the National Museum of the American Indian on the National Mall, and the annual Tribal Consumer Health Conference held by the NIHB in Denver, CO.

OHIPD assisted with exhibition enhancement projects including: upgrades to the Healing Ways Galleries and totem pole displays; *Native Voices* Web site construction and expansion; and traveling exhibit planning.

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 and intensely competitive. These two factors continue to support the need for a 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, measures of customer perception, and use of NLM Web sites. During FY2012, the OCHD continued to pursue an integrated approach intended to encourage exchange of information and learning within NLM, and help better inform NLM management decision-making on Web site research, development, and implementation. The year's evaluation activities included: analysis of NLM Web site log data; continuation a of trans-NLM Web metrics program; access to Internet audience measurement estimates based on Web usage by user panels organized by a private sector company; and access to the results of syndicated nationwide surveys of consumer and physician use of online health information. During FY2012, 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 managers.

OHIPD staff participated in FY2012 in completing the work of the Trans-NIH Web Analytics Project Team, which was charged with conducting an NIH-wide survey and assessment of NIH Web analytics opportunities and challenges.

International Programs

NLM's international partnerships and projects strengthen and expand global access to the world's health literature. Programs are intended to strengthen all phases of the research process, with particular emphasis by OHIPD on outreach to researchers, physicians, medical

students, health workers, journal editors, and librarians in developing countries. Initiatives include demonstration projects that relate to NLM's major programs and databases, including information dissemination strategies and training opportunities. NLM also contributes to the Fogarty International Center's Medical Education Partnership Initiative in Africa (MEPI) program, which provides grants to institutions in sub-Saharan African countries to strengthen the medical education systems and increase the quality, quantity and retention of health care workers and the faculty needed to train them.

Information and Communication Technology (ICT)

The African Medical Journal Editors Partnership Program (AJPP), developed in collaboration with Fogarty International Center in 2003, aims to strengthen the quality and viability of African medical journals through capacity building and partnerships with major medical journals in the US and UK. This program makes important research being carried out in endemic countries available to the world. Since the program began, six participating African journals have qualified for selection for indexing in MEDLINE. NLM is the principal funder of the project, and OHIPD staff have participated with Library Operations staff in its development and implementation since its inception. In FY2012, responsibility for management of the project transitioned to OHIPD.

In FY2012, NLM continued support for the implementation of an electronic health management information system at Tororo Hospital in Eastern Uganda. In the first phase, a team of medical students from Uganda's Makerere University Faculty of Medicine used a digital pen featuring a tiny camera and "smart paper" to collect data for an observational survey of bed net use in a remote village in the eastern district of Tororo. In the second phase, the data was integrated in to the Electronic Health Management System being pioneered by Simon Ndira at Tororo Hospital.

Network of African Medical Librarians (NAML)

NLM continues its commitment to utilizing and expanding the leadership of a growing Network of African Medical Librarians (NAML) who received training as NLM Associate Fellows or as Medical Library Association Cunningham Fellows. The objective is to help African librarians, who already have links to NLM, build library capacity through outreach and training in Africa. The current network consists of eight librarians, from Kenya, Zambia, Mozambique, Mali, Nigeria, Morocco, Uganda, and Zimbabwe, respectively.

Building on the involvement with medical school curricula that started with an FY2009 meeting of NAML members with leaders of medical schools where they were based, the librarians have developed a course for the medical school curriculum of their respective universities as well as a training manual for use in Schools of Health

Sciences in Africa called *Finding, Organizing, and Using Medical Information: A Training Manual for Students, Researchers and Health Workers in Africa.* Available free of charge in print and electronic formats, including YouTube, the manual comprises seven modules: Information Sources, Searching Tools, Electronic Information Searching Techniques, Intellectual Property Rights, Management of Information, Evaluating Electronic Resources, and Scholarly Communications. The training course is also being used outside of the original institutions involved. In FY2012, OHIPD supported health information resource training provided by NAML librarians at the Association for Health Information and Libraries in Africa (AHILA) 2012 international congress in Cape Verde, at the MEPI Symposium 2012 in Ethiopia, and at a Bioinformatics and Medical Informatics Workshop in Morocco.

At their respective libraries, in FY2012, NAML librarians continued to train faculty and students and engage in outreach to areas outside of the capital cities through workshops, newsletters, and lunchtime training sessions for staff. Several have developed institutional repositories that can be accessed online from anywhere. OHIPD provided infrastructure support for the acquisition of scanning technology to facilitate the implementation and content-development of such repositories.

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, including malaria, diarrhea, Burkitt's lymphoma, and mental health. The tutorials were developed with medical faculty, hospital staff, and students in Uganda and translated into six local languages. The community education materials for Burkitt's lymphoma also includes a video, a pictorial informed consent form, and flyers and posters to promote Burkitt's lymphoma awareness. In FY2012, NLM continued to make these materials available on its Web site.

Web Sites

NLM hosts a Malaria Research Resources Web site which 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. It also maintains a Web site that lists resources for International Librarians, Health Professionals and Researchers in Developing Countries.

Trans-NIH Collaboration

NLM participates in the NIH Global Health Research Working Group and the NIH mHealth Working Group and mHealth Summit.

NLM's International Programs Office, in collaboration with HMD, is assisting the Enteric and Hepatic Diseases Branch at NIAID to scan, archive and make available rare historical documents on the origination of oral rehydration therapy in the US-Japan Cooperative Medical Sciences Program/Cholera and Other Bacterial Enteric Infections.

International Visitors

In FY2012, the Office of Communications and Public Liaison and the History of Medicine Division's Exhibition Program arranged tours and special programs for visitors from the following 76 countries:

Algeria, Argentina, Armenia, Australia, Austria, Bangladesh, Bhutan, Brazil, Botswana, Burma, Burundi, Cambodia, Canada, Chile, China, Czech Republic, Egypt, Eritrea, Estonia, Fiji, France, Germany, Ghana, Greece, Guatemala, Guyana, Hong Kong, Hungary, India, Indonesia, Iran, Iraq, Israel, Japan, Kazakhstan, Kenya, Korea (South), Lebanon, Libya, Lithuania, Malaysia, Malta, Mexico, Nairobi, Netherlands, Niger, Nigeria, Norway, Oman, Pakistan, Palestinian Territories, Peru, Philippines, Russia, Saudi Arabia, Serbia, South Africa, Spain, Sri Lanka, Sudan, Swaziland, Sweden, Switzerland, Taiwan, Thailand, Tunisia, Uganda, Ukraine, United Arab Emirates, United States, Uruguay, Venezuela, Vietnam, Yemen, Zambia, Zimbabwe.

Trans-NIH Collaborations in Digital Biomedicine

The use of sophisticated computational approaches to explore and analyze vast amounts of complex, diverse data has transformed the modern world, and is increasingly being used in biomedical research. While the NLM has been a pioneer and leader in digital biomedicine through resources such as PubMed and GenBank, numerous major initiatives are currently underway across all of NIH. In FY2012, OHIPD contributed leadership and expertise to several of these efforts, most notably to the planning of the trans-NIH Big Data to Knowledge (BD2K) initiative. Currently, this initiative is planned as a seven year endeavor, funded at approximately $100 million per year. Importantly, BD2K will produce not only a set of research funding initiatives, but will also change NIH policies and practices that will help transform the biomedical research enterprise.

Planned BD2K funding initiatives will support centers of excellence to advance the science and technology of data and big data, research project grants to support software development, and a host of training activities. OHIPD is providing leadership to a group that is developing policy changes that will increase data sharing across the biomedical research community. That group is also planning the development of NIH resources that will make information about NIH-supported data sets discoverable, citable and linked to the scientific literature. In FY2013, this group will organize and hold three workshops related to these efforts.

These funding, policy, and practice initiatives will be coordinated with each other to raise the prominence of data in the biomedical research enterprise, and to enable the use of such data by the broad research community. The intended structural changes in the enterprise are expected to add value to the NIH research investment and accelerate the pace of progress.

LIBRARY OPERATIONS

Joyce E. B. Backus
Associate Director for Library Operations

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

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

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

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 (LHNCBC). LO staff work closely with these program areas on the design, development, and testing of new systems and system features.

Program Planning and Management

LO sets priorities based on the goals and objectives in the NLM Long Range Plan 2006-2016, and the closely related NLM Strategic Plan to Reduce Racial and Ethnic Disparities. In FY2012 LO continued its work in implementing its own Strategic Plan for 2010-2015, within this broader framework. The LO Strategic Plan has three major directions: Transforming Access to the Collection, Redesigning Systems and Workflows, and Developing a 21st Century Workforce. Significant progress on the first two directions is reported throughout this chapter. In addition, an LO Administrative Efficiencies group captured best business practices on administrative topics and developed and published on a Sharepoint site checklists, policies, and other resources to aid LO divisions in their work.

In the area of Developing a 21st Century Workforce, an LO Workforce Development group held four quarterly all LO staff meetings in which new staff are recognized and updates from every area of LO are shared with staff, including those listening in from off-site. LO continued its first full year of an Emerging Leaders program, a career enrichment program for selected LO staff to learn about the big picture within LO and NLM and to work on a project of institutional significance. LO also is holding ongoing supervisory discussion groups for 22 supervisors, and held customized group training for 80 LO staff in project management and retirement planning.

Collection Development and Management

NLM's comprehensive collection of biomedical literature is the foundation for many of the Library's services. LO ensures that this collection meets the needs of current and future users by updating NLM's literature selection policy; acquiring and processing relevant literature in all languages and formats; organizing and maintaining the collection to facilitate current use; and preserving the content for subsequent generations. At the end of FY2012, the NLM collection contained 2,728,662 volumes and 18,140,582 other physical items, including manuscripts, microforms, pictures, audiovisuals, and electronic media.

The Office of the Associate Director (ADLO), the Technical Services Division (TSD), and the History of Medicine Division (HMD) together established overall NLM cataloging priorities for current and historical material. To ensure these priorities were met, almost all TSD catalogers and technicians, as part of their regular duties, worked on HMD-related cataloging tasks.

Thousands of formerly "hidden collection" historical materials are now discoverable from NLM. Many TSD staff members have also undertaken work to thoroughly document selection, acquisitions, licensing and cataloging workflows. This documentation and subsequent analysis will inform division workflow changes and reorganization.

Selection

In FY2012, selectors worked on a number of projects to enhance the breadth and depth of NLM collections. Staff worked with book dealers and the Library of Congress Overseas Offices to improve the supply of works from the US, UK, Australia, New Zealand, China, Russia and the former Soviet Union, Eastern Europe, Latin America, India, Taiwan, Hong Kong, Iran, Pakistan, India, and East Africa. Selectors carried out a review of new video titles, and acquired a number of new documentaries for the collection. Selectors also worked on individual projects in subjects such as military medicine, veteran's health and rehabilitation, health effects of radiation, disaster management, emergency medicine, and the history of medicine and health care, and acquired commercially published works and grey literature in these and other subject areas. Staff reviewed numerous donations, enabling NLM to enrich its collections with many titles not widely held in the US. Gifts were received from the National Cancer Institute, Army Research Laboratory Library, Navy Bureau of Medicine Office of the Medical Historian, US Fire Administrative Library, Social Security Administration Library, Carl R. Darnall Army Medical Center Library, Environmental Protection Agency libraries, and other institutional and individual donors.

LO uses subscription agents and book vendors to acquire current literature published around the world. In FY2012, a new five-year subscription services contract was awarded for acquisition of periodicals and serials from North America, Western Europe, the UK, Australia, New Zealand, and Japan.

Acquisitions

The Technical Services Division received and processed approximately 119,000 contemporary physical items (books, serial issues, audiovisuals, and electronic media) which is below last year's total. The number of electronic-only serials continues to increase and now stands at 2,336. This represents approximately 12 percent of the serial titles acquired. Efforts continued to reduce serial expenditures by cancelling some subscriptions for titles that were duplicates, office copies or in peripheral or out of scope subject areas for the general and reference collections. More than 500 subscriptions were canceled for 2012 with an expected savings of approximately $300,000. Net totals of 30,059 volumes and 1,717,881 other items, including nonprint media, manuscripts, and pictures acquired by the HMD, were added to the NLM collection.

HMD acquired a wide variety of important printed books, manuscripts and modern archives, images, and historical films during FY2012, including two early manuscripts: *Commentary on the Ars parva of Galen*, written in Latin in Italy c. 1480-1500 by Lucas Maiorius, and *Medical Miscellany*, an extremely rare Polish manuscript written in Yiddish and Hebrew dated 1596. Probably written by a physician, the latter manuscript contains a collection of medical cures arranged in the order of the ailment or the organ affected.

Among early printed books acquired was *De locutione et eius instrumentis liber* (Venice, 1601) by Hieronymus Fabricius ab Aquapendente, the very rare first edition of an early monograph on the organ of speech. Fabricius studied medicine at Padua, became professor of anatomy and surgery, and later was William Harvey's teacher at Padua. Fabricius is perhaps best known for his description of the valves in veins in his *De Venarum Ostiolis* (1603) which influenced William Harvey in his own studies of the circulation of blood. The Library also acquired Alfonso Ferri's *De sclopetorvm sive archibvsorvm vvlneribvs libri tres* (Lyon, 1553), a rare second edition of one of the earliest works on the treatment of gunshot wounds. Ferri had developed a special instrument for removing bullets called an 'alphonsinum' which is illustrated.

NLM also acquired a collection of over 200 items related to the literary achievements of the prominent American Civil War surgeon Silas Weir Mitchell (1829-1914). Mitchell was among the foremost physicians of his time and a prolific writer of both scientific and literary works. His service as a surgeon during the Civil War led to his interest in neurology and neuro-psychiatry. He published over 100 scientific articles and monographs, but was also a successful poet and novelist. He published more than 25 literary titles. As described in *The Literature of Prescription: Charlotte Perkins Gilman and "The Yellow Wall-Paper"*, a traveling exhibition released by NLM in 2009, a proposed treatment regimen for one of Dr. Mitchell's patients also prompted the writing of short story that became a feminist classic. The collection includes many volumes signed by Mitchell and various editions which document the changes in styles and tastes of American publishing of the time. The Library also received two major transfers of books and journals, the Stitt collection from the US Navy Bureau of Medicine and Surgery's Office of Medical History and a second collection from the current Armed Forces Medical Library.

Recently acquired archives and modern manuscript collections included the papers of Dr. Bernadine Healy, former Director of the National Institutes of Health; the papers of Dr. Alexis I. Shelokov, a virologist who worked on the polio vaccine while at NIAID; the papers of Dr. Charles F. Whitten, pioneer in sickle-cell disease; and the records of the American College of Cardiology. The prints and photographs collections received two World War I–era photo albums, one by Roy Sheets, a soldier in the war, and the other showing the

efforts of Base Hospital No. 21, in Rouen, France, a field X-ray unit assembled from Washington University in St. Louis. Films and videos on Navy medicine from the history office of BUMED, the Navy Bureau of Medicine and Surgery, enriched the historical audiovisuals collections.

Preservation and Collection Management

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

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

Collection Space and Maintenance

In FY2012, LO bound 15,000 volumes, repaired 2,346 items, made 534 preservation copies of films and audiovisuals, and conserved 997 items. A total of 429,923 items were shelved. This figure is an 11.5 percent decrease from FY2011 and reflects the decline in interlibrary loan and Main Reading Room requests.

Dwindling space for growth of NLM collections required a major effort to plan for additional storage capacity in the existing NLM facility, award contracts to strengthen the B-2 floor, purchase compact shelving, install new sprinkler systems, and shift materials into available space or temporarily offsite as new space is prepared. The multi-year collection space expansion project proceeded smoothly in FY2012. By the end of FY2012, a total of 74,665 linear feet of new compact shelving had been installed on the B-2 and B-3 floors.

National Cooperative Preservation

Following the September 2011 launch of MedPrint, a cooperative project to preserve biomedical journals in print, much headway was gained in FY2012. An early 2012 survey of the 120 resource libraries in the regions revealed that 41 percent of them plan to participate in MedPrint, committing to hold titles for 25 years or to 2036. The program is open to all US libraries who participate in DOCLINE®, NLM's interlibrary loan system that stores journal holdings information for over 2,700 libraries. Print retention commitments are also stored in DOCLINE. At year's end, NLM had received 15 signed agreements from institutions representing all eight regions. Of the 250 titles identified in the MedPrint program, libraries across the country had made commitments to all but 10 titles. Many more libraries made commitments to titles not officially in the program, as over 700 titles now show print retention commitments. The goal of the program is to get commitments for 12 copies of each MedPrint title, with NLM's copy being the 13th. Thirty-nine of the 250 titles now have 12 or more commitments.

Digitization Program

The NLM digital repository, Digital Collections, now holds approximately 7,000 scanned texts and 70 historical films. A number of enhancements were made throughout the year, including the ability to handle multi-volume resources with intuitive search and browse, the addition of faceted browsing and browse by language, and increased automation of the processing tasks involved in taking in content from the in-house book digitization workflow. Preservation services were also enhanced, including the versioning of all datastreams, attachment of audit histories to objects, and the establishment of redundant storage of all files at the NLM offsite data center.

HMD, in cooperation with the Public Services Division(PSD), TSD and OCCS, continued to select, conserve when necessary, and put forward for digitization monographs that became part of the Library's contribution to the Medical Heritage Library Project, a cooperative venture to digitize historical materials from the collections of the NLM, the Countway Library at Harvard, the Harvey Cushing/John Hay Whitney Library at Yale, the Augustus C. Long Health Sciences Library at Columbia University, and the New York Public Library. The goal of the project, launched in January of 2010, is to digitize books and journals that document the evolution of American medicine from 17th century colonial medicine to 20th century research hospitals. The whole of NLM's contribution of over 6,000 books is available through NLM's Digital Collections repository, and the entire content of the Medical Heritage Library will be available through the Internet Archive. In FY2012, the Library scanned 1.5 million pages (8,131 volumes) for the Medical Heritage Library Project.

PubMed Central (PMC), a digital archive of medical and life sciences journal literature developed by NCBI, is the NLM vehicle for ensuring permanent access to electronic journals and digitized back files. LO assists NCBI in soliciting participation of additional journals, particularly in the fields of clinical medicine, health policy, health services research, and public health. LO provides support for the PMC efforts ranging from review of potential journals for appropriateness for the NLM collection to cataloging and authority data creation for the PMC system. By the end of FY2012, 252 new journals had been added to PMC, and 263,203 new articles had

been added. The total number of articles in the database is 2,540,952.

Permanent Access to Electronic Information

LO selected Web content as part of its mission to collect, preserve, and make accessible the scholarly biomedical literature as well as resources that illustrate a diversity of philosophical and cultural perspectives not found in the technical literature. New forms of publication on the Web, such as blogs authored by doctors and patients, illuminate health care thought and practice in the 21st century. In launching this initiative, the Library is capturing and providing a unique resource for future scholarship.

The inaugural collection of Web content is "Health and Medicine Blogs," presenting the perspectives of physicians, nurses, hospital administrators and other individuals in health care fields. The collection also includes patients chronicling their experiences with conditions such as cancer, diabetes and arthritis. The site currently contains 12 blogs, including KevinMD.com, "social media's leading physician voice"; Not Running a Hospital, a blog by a former CEO of a large Boston hospital; e-patient Dave, a cancer survivor and leader in the participatory medicine movement; and Wheelchair Kamikaze, who writes about his personal experience living with multiple sclerosis (MS).

Vocabulary Development and Standards

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

Medical Subject Headings (MeSH)

In FY2012 work was completed on the 2013 edition of MeSH which contains 26,853 descriptors and more than 214,000 supplementary records for chemicals and other substances. The MeSH Section added 302 new descriptors, replaced 26 descriptors with more up-to-date terminology and deleted 30 descriptors.

Work continued on enhancing the rare disease coverage in the MeSH vocabulary via two separate projects: ensuring that the list of rare disease terms maintained by the Office of Rare Diseases Research (ORDR) is available in the MeSH vocabulary and adding the disease portion of the Online Mendelian Inheritance in Man (OMIM) database that is available in the UMLS. The ORDR rare disease terms that matched existing MeSH descriptors were merged with those descriptors. The remaining terms were introduced as Supplementary Concept Records (SCRS) in MeSH 2012. OMIM is a database that catalogues human diseases with genetic components. Although OMIM disease names are available for searches in PubMed, it is often difficult to index and search for the articles on rare diseases with genetic components because of multiple synonyms used by different scientists that often do not overlap. MeSH created 3,774 new disease SCRs, and identified and tagged 1,498 existing ORDR SCRs as rare diseases with genetic components during the OMIM load. All OMIM disease names therefore will be available starting with MeSH 2013 for indexers and searchers. Both of these projects should lead to more consistent indexing and effective retrieval for rare diseases.

In FY2012 TSD developed an online tracking system for new MeSH requests suggested by the Cataloging Section. Subsequently BSD began using the system to track requests from Indexers and then from HMD and PSD. This SharePoint-based system allows everyone in LO to see all internal MeSH requests.

In partnership with the US Food and Drug Administration (FDA) Unique Ingredient Identifiers (UNIIs) have been added to a subset of 2013 MeSH SCRs found to match FDA Substance Registration System (SRS) terminology. The UNIIs were inserted into the RN (Registry Number) field of MeSH chemical SCRs and serve as a new system of unique identifiers that will enhance indexing and searching for chemicals and enzymes.

UMLS Metathesaurus

The MeSH Section and its contractors are responsible for content editing of the UMLS Metathesaurus using systems developed originally by the LHNCBC and now managed by OCCS. Responsibility for the production of the Metathesaurus rests with LO/OCCS. The MEDLARS Management Section (MMS) plays a major role in Quality Assurance and Documentation, and MeSH continues its supervision and training for Metathesaurus editing. The MeSH staff also assumes responsibility for monitoring vocabulary updates, the Metathesaurus production schedule, vocabulary licenses, and other agreements. Working with OCCS, a Metathesaurus production coordination group began meeting regularly to coordinate the production efforts, including regular review of inversions and insertions of updated and new vocabularies to the Metathesaurus. The Metathesaurus has shifted to twice yearly releases in response to user requests. The releases in 2012 took place in May and November.

Clinical Vocabularies

NLM is the central coordinating body for clinical terminology standards within the Department of Health and Human Services (HHS). LO, in partnership with LHNCBC 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). With enactment of the Health Information Technology for Economic and Clinical Health (HITECH) Act, included as part of the American Recovery and Reinvestment Act of 2009 (ARRA), NLM's activities in this area intensified, particularly in on the areas of quality and performance measurement, lab services, and newborn screening.

In FY2012, NLM established a follow-on interagency agreement with the Office of the National Coordinator for Health Information Technology (ONC) to collaborate in all areas of terminology standards enhancement and related projects that support meaningful use of electronic health records; effective health information exchange for health care, public health, and research; and the goals of the Standards and Interoperability framework. NLM also established an interagency agreement with the Centers for Medicare and Medicaid Services (CMS) for the validation of quality measurement value sets associated with Meaningful Use Stage Two criteria, for the development of tools to support value set creation and validation by original value set authors, for robust public dissemination of value sets, and for additions and maintenance to clinical terminology standards that may be required for quality measurement value sets.

Finally, NLM established an interagency agreement with the HHS Office of the Assistant Secretary for Planning and Evaluation (ASPE), as a mechanism for ASPE, ONC, and NLM to collaborate on a project to develop infrastructure to facilitate the use of electronic health records (EHRs) in comparative effectiveness research (CER). Launching of the NLM's new Value Set Authority Center (VSAC) was a direct outcome of these interagency agreements. Initially the VSAC provides downloadable access to all official versions of vocabulary value sets contained in the 2014 Clinical Quality Measures. The value sets provide lists of the numerical values and individual names from standard vocabularies used to define the clinical concepts (e.g. diabetes, clinical visit) used in the quality measures. In the future the content of the VSAC will gradually expand to incorporate value sets for other use cases, as well as for new measures and updates to existing measures.

RxNorm

The MeSH Section and its contractors also produce RxNorm, a clinical drug vocabulary that provides standardized names for use in prescribing. It is released through the UMLS and also separately. The final US rule for 2014 EHR certification criteria published in September 2012 requires the use of RxNorm. It represents the information that is typically known when a drug is prescribed, rather than the specific product and packaging details that are available at the time a medication is purchased or administered, and provides a mechanism for connecting information from different commercial drug information services. Inactive ingredient content from DailyMed was added to RxNorm to allow users to better record allergy content using RxNorm. RxNorm editors prepare and release monthly updates to the clinical drug vocabulary.

NDF-RT is a terminology used to code clinical drug properties, including mechanism of action, physiologic effect, and therapeutic category. It is an extension of the VHA National Drug File (VANDF), another RxNorm source vocabulary. NLM, the Department of Veterans Affairs (VA), Veterans Health Administration VHA, and FDA met at NLM to discuss the future of the NDF-RT which is included as a source vocabulary in RxNorm. As a consequence, the NDF-RT content appearing in RxNorm will be refined and enhanced to better meet the needs those attempting to achieve "meaningful use" of electronic health records, as defined by CMS.

LOINC

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 is a clinical terminology important for laboratory test orders and results. The final rule for 2014 EHR certification criteria published in September 2012 expanded the requirements for use of LOINC. In FY2012, NLM continued its support for LOINC and NLM staff, primarily in LHCNBC, also made substantial contributions to enhancing its content.

SNOMED CT

In FY2012, NLM also 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 owned by the International Health Terminology Standards Development Organisation (IHTSDO). NLM, on behalf of HHS, is the US member of the IHTSDO responsible for distribution of SNOMED CT within the US and for representing US interests in the continued development of SNOMED CT. The final rule for 2014 EHR certification criteria published in September 2012 expanded the requirements for use of SNOMED CT.

FY2012 saw several important advances that will improve the use of SNOMED CT in the US and elsewhere. NLM launched multiple updates to the US SNOMED CT

Content Request Submission System and established a twice annual release cycle for the US Extension to SNOMED CT. The IHTSDO, after evaluating options available from IHTSDO Members, decided to adopt the SNOMED CT Content Request Submission System developed by NLM as the system for all SNOMED CT requests. In May 2012 the IHTSDO and the American Dental Association (ADA) entered into an agreement permitting the IHTSDO to continue to distribute the ADA-developed dental diagnostic concepts within SNOMED CT and enabling active ADA involvement in the development of new dental content for SNOMED CT. In July 2012 the IHTSDO and the Regenstrief Institute began a long-term collaborative relationship to develop coded content to support order entry and result reporting critical to the computerized transmission of medical information. They expect to sign a formal agreement in FY 2013.

NLM continues to be involved in activities related to the development of mappings from important standard vocabularies to administrative code sets (e.g. SNOMED CT to ICD-10-CM), to support implementation of these standards by tying them to billing systems, with the ultimate goal of helping to improve the overall quality and efficiency of health care services. In FY2012, NLM (including LHNCBC, NICHSR, and MeSH staff) continued to participate in several mapping projects, both nationally and internationally. In FY2012 the IHTSDO published an initial map from SNOMED CT to ICD-10 based on NLM's CORE Problem List Subset of SNOMED CT as well as the SNOMED CT concepts that correspond to the top ranking ICD-10 codes used in five IHTSDO Member countries. Building on the IHTSDO's map from SNOMED CT to ICD-10 and leveraging contributions from Kaiser Permanente's Convergent Medical Terminology, NLM produced a map from SNOMED CT to ICD-10-CM in FY2012. To facilitate migration to SNOMED CT, NLM also produced a map from SNOMED CT to ICD-9-CM in FY2012. In 2012, Kaiser contributions included problem list subsets for endocrinology and gastrointestinal diseases as well as the first installment of Kaiser's Enterprise Terminology Tool used for the creation, management, and quality assurance of Kaiser's terminology products such as the CMT subsets. Where appropriate, CMT content is being incorporated into the International Release of SNOMED CT for international use or the US Extension to SNOMED CT when relevant for use in the US.

BSD launched a new Web page, NLM Tools for EHR Certification and Meaningful Use, providing a centralized location for users to access vocabulary standards, applications, and related tools that can be used to meet US EHR certification criteria and to achieve Meaningful Use of EHRs.

Bibliographic Control

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

Cataloging

LO catalogs the biomedical literature acquired by NLM to document what is available in the Library's collection or identify high-quality, relevant medical resources on the Web. NLM also works with the Library of Congress (LC) to provide cataloging-in-publication (CIP) data for medical books prior to their publication in the US. The cataloging and authority data created by NLM is used by libraries world-wide and minimizes the cost and effort of creating cataloging data for these health science libraries. Cataloging is performed by the TSD Cataloging Section, staff in HMD, and contractors. Additionally, the Cataloging Section is responsible for maintenance and updating of the NLM Classification, the maintenance and updating of the NLM Catalog DTD, transformation of MARC data into other metadata formats, and name authority control for selected NLM Web Services.

In FY2012, the Cataloging Section cataloged a total of 19,656 books, serials, electronic resources, and audiovisuals. These figures include 2,186 CIP records created from publisher electronic galleys and 1,926 titles cataloged for the HMD collection, including material from online exhibits and special collections. This represents a decline of eight percent from the previous year's total.

The 2012 edition of the NLM Classification was released in online and PDF formats with updates to the QV (Pharmacology) schedule.

NLM loaded 338,594 records from Library of Congress Name Authority File for headings that matched existing records in the NLM Name Authority File. This has simplified and streamlined name authority workflows for the Section and allows better and more accurate sharing of NLM data with the library community.

Three TSD staff served on the US RDA Test Coordinating Committee, which reviewed the progress on tasks and action items that needed to be completed before implementation of the new cataloging code RDA: Resource Description and Access. In general, NLM and the US RDA Test Coordinating Committee found the action items (reword the instructions in clear, understandable English; have the RDA elements sets and vocabularies completed and published on the Web; improve the functionality of the online RDA Toolkit; integrate complete RDA record examples in MARC and other encoding schema into the Toolkit; and demonstrate credible progress towards a replacement for MARC) were either completed or demonstrated good progress. The three

US national libraries agreed to implement RDA in March 2013.

As part of these implementation plans, the Cataloging Section has done a great deal of preparatory training for NLM staff, both within and outside of TSD. The Section developed training sessions tailored to specific audiences on an introduction to FRBR (Functional Requirements for Bibliographic Records) concepts, which are the backbone of RDA, and on the differences between AACR2 and RDA.

A member of the Cataloging Section is representing NLM in the Bibliographic Framework Initiative (BFI), a major project sponsored by the Library of Congress. The BFI will develop a new means for capturing and sharing bibliographic data that will replace the MARC format as the common exchange currency for bibliographic data.

Indexing

LO indexes 5,633 biomedical journals for the MEDLINE/PubMed database to assist users in identifying articles on specific biomedical topics. A combination of Index Section staff, contractors, and cooperating US and international institutions indexed 760,903 articles in FY2012, bringing the total number of MEDLINE citations to over 20 million. Indexed citations were updated to reflect 375 retracted articles, 8,142 published corrections, and 37,158 comments found in subsequently published notices or articles.

In FY2012, indexers created 88,944 annotated links between newly indexed MEDLINE citations for articles describing gene function in selected organisms and corresponding gene records in the NCBI Entrez Gene database. This was an 11 percent increase from last year.

LO continues to work with other NLM program areas to identify, test, and implement ways to reduce or eliminate tasks now performed by human indexers. The Index Section continues to increase the number of articles indexed from the online version of journals. At the end of FY2012, 4,157 journals (74 percent of all indexed journals) were indexed from an online version, including online-only journals and those with a print version.

The Medical Text Indexer First Line (MTIFL) pilot project became a regular production process. During FY2012, 4,729 articles from 24 journals were indexed using MTIFL, which is a combination of machine-suggested indexing further reviewed and approved by senior indexers. The Index Section also developed a methodology to identify appropriate journals for use with MTIFL in order to continue to expand this program.

Indexers perform their work after the initial data entry of citations and abstracts has been accomplished. Over the past 10 years, great strides have been made in improving the efficiency of data entry. By the end of FY2012, 93 percent of all citation data entry consisted of XML-submitted data from publishers, a one percent increase over FY2011. The remaining citations were created by scanning and optical character recognition (OCR). A total of 29,652 more citations were received from publishers compared with the previous year for a grand total of 690,161 XML citations.

NLM selects journals for indexing with the advice of the Literature Selection Technical Review Committee (LSTRC) (Appendix 6), an NIH-chartered committee of outside experts. In FY2012, LSTRC reviewed 559 journals and rated 97 of them highly enough for NLM to begin indexing them.

Information Products

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

Databases

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

Use of MEDLINE/PubMed, which now includes over 22 million citations, registered 2.2 billion searches in FY2012, a 22 percent increase from the previous year. BSD staff assisted NCBI with the testing and implementation of many enhancements to PubMed, notably: replacing the Limits page with a Filter Sidebar; modifying the Advanced Search page; enhancing the My NCBI and My Bibliography Award View display; and introducing a search feature to limit to citations that have a structured abstract.

Use of MedlinePlus increased to 213 million unique visitors in FY2012, a 30 percent increase from FY2011; MedlinePlus en español saw 36 percent increase, with 96.9 million unique visitors in FY2012. Use of MedlinePlus Mobile increased 26 percent to 820,000 unique visitors. Requests to MedlinePlus Connect increased to 7.5 million in FY2012 from 1.1 million in FY2011. MedlinePlus, MedlinePlus en español and MedlinePlus Mobile continue to receive high ratings from customers in the American Customer Satisfaction Index

(ASCI), ranking among the top government news/information sites.

PSD and OCCS continued to expand and improve the suite of MedlinePlus-branded products and services. At the end of FY2012, MedlinePlus featured 931 topics in English and 918 in Spanish. An auto-complete feature for the search boxes on MedlinePlus and MedlinePlus en español was released to facilitate more effective searching by site visitors. Several enhancements, including expanded mappings between MedlinePlus health topics and electronic health record coding systems and improvements to the precision of drug information responses, were released for MedlinePlus Connect. MedlinePlus Connect now responds to requests for diagnosis, lab test and medication information in either English or Spanish. The MedlinePlus health topics Web service and XML files were also enhanced to open up a wealth of MedlinePlus data to the public. The Web service and the XML files are both now available in English and Spanish, and they both provide health topic name and group data, summaries, and all links that appear on the health topic pages. The health topics Web service now also supports both keyword searching and fielded searching. A Spanish twitter feed, @MedlinePlusEsp, was launched to accompany the English @MedlinePlus feed. At the end of FY2012 the English feed had 40,000 followers, up from 22,000 followers at the end of FY2011. Subscriptions to the MedlinePlus and MedlinePlus en español GovDelivery email update service grew to 430,000 subscribers with 46.9 million subscriptions, up from 270,000 subscribers with 34.3 million subscriptions at the end of FY2011.

Under the direction of NICHSR, NLM continues to expand and enhance its databases for health services researchers and public health professionals. The number of serials on topics related to health services research that are indexed in MEDLINE continued to increase during FY2012.

NICHSR collaborates with NCBI to maintain and expand HSR, public health resources, and the set of clinical guidelines-related materials available on Bookshelf.

In FY2012, NICHSR continued to improve the depth and breadth of content available on its information portal for the health services research community, HSR Info Central. As a result of these efforts, and increased promotion to the HSR community, unique visitors increased more than 35 percent through FY2012.

NICHSR contributes to the field of health services research through its support of NLM databases containing information about newly-funded research and accessible data collection tools and resources. With the assistance of AcademyHealth and the Sheps Center at the University of North Carolina, Chapel Hill, the content of HSRProj (Health Services Research Projects in Progress) continued to expand, incorporating work funded by additional foundations, states, and other organizations. In FY2012, the database grew to more than 9,500 active records representing more than 340 funders and was increasingly recognized as a valuable resources and used to monitor projects related to comparative effectiveness, and other topics.

The Health Services Research Resources (HSRR) database also expanded to cover additional datasets, surveys, other research instruments, and software packages used with datasets.

DailyMed, a Web site that presents high quality information about drugs, including the FDA approved packaging information (labels) for drugs, grew tremendously throughout the year to approximately 20,000 labels; 75-100 updates are received each day. Over the past year unapproved prescription drugs and the over-the-counter product labels were also added. Use of DailyMed increased to 8,812,640 visitors in FY2012 from 5,166,000 in FY2011.

Machine Readable Data

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

Demand for MEDLINE/PubMed data in XML format continues to increase. At the end of FY2012, there were 782 licensees of MEDLINE data, a seven percent increase over the previous year. The majority use the data for research and data-mining.

At the end of FY2012, there were 5,640 UMLS licensees, a decrease of 14 percent from the previous year. NLM services and support to licensees of UMLS were enhanced by new infrastructure for the UMLS Annual Report requirement.

Web and Print Publications

NLM databases and Web sites are its primary publication media. Publications available on the main Web site include recurring newsletters and bulletins, fact sheets, technical reports, and documentation for NLM databases. The BSD MEDLARS Management Section edits and produces the *NLM Technical Bulletin*, which provides timely, detailed information about changes and additions to the NLM databases and related policies, primarily for librarians and other information professionals. Published since 1969, the *Technical Bulletin* also serves as the historical record of the evolution of NLM's online systems and databases. A redesign of the *Technical Bulletin* was launched in January 2012, providing users with an updated design and improved functionality

NLM added one new profile to the Profiles in Science Web site during FY2012, bringing the total to 36. The new site focuses on the papers of Henry Swan II (1913-1996), the American surgeon who pioneered the use of hypothermia to make possible the first open-heart surgeries. Launched in September 1998, Profiles in Science promotes the use of the Internet for research and teaching in the history of biomedical science by making widely available archival collections of leaders in biomedical research and public health. Published and unpublished materials appear on the site, including books, journal volumes, pamphlets, diaries, letters, manuscripts, photographs, audiotapes, and video clips. Altogether the program has digitized more than 190,000 pages of material, 20 video clips, and 2,421 still images.

In FY2012, LO staff continued to be involved in the two publications designed for patients, families, and the public. The Director's Comments podcasts bring current health news to listeners. Four issues of the *NIH MedlinePlus Magazine* and one issue of *NIH MedlinePlus Salud* were published in print and online in FY2012.

Direct User Services

In addition to producing heavily used electronic resources, LO is responsible for document delivery, reference, and customer service for both onsite users and remote users as a back-up to services available from other members of the National Network of Libraries of Medicine (NN/LM).

Reference and Customer Services

LO provides reference and research assistance to onsite and remote users as a backup to services available from other health sciences and public libraries. LO also has primary responsibility for responding to inquiries about NLM's products and services and how to use them effectively. LO's Reference and Web Services Section responds to initial inquiries and also handles the majority of questions requiring second-level attention. Staff from throughout LO and NLM assist with second-level service when their special expertise is required. Customer Service received 100,023 inquiries in FY2012, a slight increase from FY2011. The number of onsite inquires (6,549) declined seven percent from FY2011. The number of remote inquiries (93,899) increased one percent.

Document Delivery

LO retrieves documents requested by onsite patrons from NLM's closed stacks and an off-site storage facility and also provides interlibrary loan as a backup to document delivery services available from other libraries and information suppliers. In FY2012, PSD's Collection Access Section (CAS) processed 320,014 requests for contemporary documents, a 12 percent decrease from FY2011. The History of Medicine Division (HMD) handled 7,217 requests for rare books, manuscripts, pictures, and historical audiovisuals, a nine percent decrease from FY2011.

Onsite

The number of onsite users registering to use the collection declined by 18 percent from last year and use of NLM's collection by users in the Main Reading Room declined by 28 percent, marking a new low of only 85,352 requests. Users of the HMD Reading Room requested 6,692 items from the historical and special collections, an increase of 34.7 percent from last year, due to high-volume requests by users working over sustained periods in the HMD Reading Room and to intensive internal collection use for special scanning projects. Paid printing at Main Reading Room workstations decreased 30 percent, from 872,216 pages in FY2011 to 613,815 pages in FY2012, a trend that corresponds with the decrease in retrieval of print materials and reflects the increased use of flash drives to download electronic journal content.

OffSite

The Collection Access Section (CAS) received 234,662 interlibrary loan requests, a four percent decline from FY2011, and filled 83 percent of them. NLM ILL staff keyboarded a total of 34,435 requests, a four percent increase from FY2011. The number of requests processed in 12 hours increased slightly to 98 percent, with 99 percent processed within one day of receipt. NLM continues to deliver 96 percent of interlibrary loan requests electronically.

A total of 2,731 libraries use DOCLINE, NLM's interlibrary loan request and routing system, a decrease of three percent from FY2011. DOCLINE users entered 1,555,154 requests in FY2012, a five percent decline from last year. The number of DOCLINE requests has decreased 49 percent since the transaction volume peaked in 2002. There were 273,872 Loansome Doc requests entered into the system by end-users of the DOCLINE libraries, down eight percent from FY2011.

CAS and OCCS worked together to create an online form for non-DOCLINE libraries to submit ILL requests to NLM. The form automatically creates service requests in the Siebel customer service software, thus eliminating manual work by the customer service contract, and standardizes information submission in a standard, easy to read format for ILL staff.

NCBI, BSD and the staff at the Regional Medical Libraries continued to support and promote the use of the PubMed LinkOut for Libraries and Outside Tool, the open-URL services that allow libraries to link directly from PubMed to a wide range of resources beyond the Entrez system. Using these tools, libraries can create custom displays of their electronic and print holdings for their primary clientele. The number of libraries participating in LinkOut increased by six percent in FY2012 to 2,796;

there are 961 libraries participating in the Outside Tool option, an increase of 15 percent over last year.

NLM and the Regional Medical Libraries continued to encourage network libraries to use the Electronic Funds Transfer System (EFTS), operated for the NN/LM by the University of Connecticut, as a mechanism to reduce administrative costs associated with interlibrary loan service billing. At the end of FY2010, there were 1,349 libraries participating in EFTS. Participants receive either a single net consolidated bill or a net consolidated payment each month.

During FY2012, the LO Strategic Planning work group on DOCLINE and Resource Sharing in the National Network of Libraries of Medicine (NN/LM), held three focus group sessions with network librarians to solicit opinions and gather information. Data from focus groups was used to draft a survey that will gather quantitative data from network libraries on the current state of resource sharing and their expectations for the future and help to guide decisions regarding NLM and network services related to data sharing.

Outreach and Training

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

National Network of Libraries of Medicine

The NN/LM works to provide timely, convenient access to biomedical and health information for US health professionals, researchers, and the general public irrespective of their geographic location. With more than 6,000 full and affiliate members, the Network is the core component of NLM's outreach program and its efforts to reduce health disparities and to improve health information literacy. Full members are libraries with health sciences collections, primarily in hospitals and academic medical centers. Affiliate members include some smaller hospitals, public libraries, and community organizations that provide health information service, but have little or no collection of health sciences literature. LO's NN/LM Office (NNO)

oversees network programs that are administered by eight Regional Medical Libraries (RMLs) under contract to NLM. In FY2012, the NN/LM identified initiatives related to promoting MedlinePlus Connect and ClinicalTrials.gov Results and outreach to community colleges as major initiatives for the five year contract period.

RMLs and other network members conduct many special projects to reach underserved health professionals and to improve the public's access to high quality health information. Virtually all of these projects involve partnerships between health sciences libraries and other organizations, including public libraries, public health departments, professional associations, schools, churches, and other community-based groups. In FY2012, the NN/LM initiated 232 outreach projects which target rural and inner city communities and special populations in 37 states, the District of Columbia, US Virgin Islands, Puerto Rico and Guam.

With the assistance of other NN/LM members, the RMLs do most of the exhibits and demonstrations of NLM products and services at health professional, consumer health, and general library association meetings around the country. LO organizes the exhibits at the Medical Library Association annual meeting, the American Library Association annual meeting, some of the health professional and library meetings in the Washington, DC area, and some distant meetings focused on health services research, public health, and history of medicine. In FY2012, NLM and NN/LM services were exhibited at 29 national and 281 regional, state, and local conferences across the US. These exhibits highlight all NLM services relevant to attendees.

Partners in Information Access for the Public Health Workforce

The NN/LM is a key member of the Partners in Information Access for the Public Health Workforce, a 14-member public-private agency collaboration initiated by NLM, the Centers for Disease Control and Prevention, and the NN/LM in 1997 to help the public health workforce make effective use of electronic information sources and to equip health sciences librarians to provide better service to the public health community. The NICHSR coordinates the Partners for NLM; staff members from the National Network Office, SIS, and the Office of the Associate Director for Library Operations serve on the Steering Committee, as do representatives from several RMLs.

The Partners Web site (PHPartners.org), managed by NLM with assistance from the New England RML at the University of Massachusetts, provides unified access to public health information resources produced by all members of the Partnership, as well as other reputable organizations. In FY2012, the Web site was expanded with more than 510 new links.

One of the most popular resources on the PHPartners site has been pre-formulated PubMed search queries in support of the HHS Healthy People 2010

initiative. NICHSR has worked with the HHS Office of Disease Prevention and Health Promotion, responsible for coordinating the development and management of the Healthy People initiative, to arrange for the development of comparable structured searches for Healthy People 2020, which launched at the end of 2010. In FY2012, NICHSR worked with librarians and public health subject matter experts to develop "structured evidence queries" or "SEQs" (pronounced "seeks") for each of the Healthy People 2020 objectives and sub-objectives. By the end of the year, SEQs for 25 HP2020 topics had been posted, with work underway for release of the remaining SEQs in FY2013. The SEQs are accessible both from the PHPartners.org site and from the HealthyPeople 2020 site, at www.healthypeople.gov.

Special NLM Outreach Initiatives

LO participates actively in the Library's Committee on Outreach, Consumer Health, and Health Disparities and in many NLM-wide outreach efforts designed to expand outreach and services to the public as well as to address racial and ethnic disparities.

For several years, LO has worked in collaboration with NLM's Chief of International Programs to improve health information capacity in sub-Saharan Africa by devoting one position in the NLM Associate Fellowship Program (AFP) to an African librarian. Former AFP participants from Kenya, Mali, Malawi, Mozambique, Nigeria, and Morocco form an African Network of Medical Librarians. These African librarians have become NLM's "ambassadors" providing information services, training and outreach bringing NLM resources to health professionals, scientists, students and other librarians within their countries. In addition, LO has worked with NLM's Chief of International Programs on a project to build journal capacity and enhance the quality of African medical journals through partnerships between an established Western medical journal and a sub-Saharan Africa medical journal.

LO also continued to provide support to NLM's joint fellowship program with the Association of Health Care Journalists (AHCJ) now in its fourth year. Staff from NICHSR and PSD presented to the journalists and AHJC staff on health services research and public health resources available through NLM and its partner agencies, Administration's OpenGov Initiative, and the use of MedlinePlus.gov to quickly find pharmaceutical, medical and public health information.

Historical Exhibitions and Programs

HMD supports the development and installation of major exhibitions in the NLM rotunda, under direction of the Office of the Director and with assistance from offices across the NLM. Designed to appeal to the interested public as well as the specialist, these exhibitions highlight the Library's resources and are an important part of

NLM's outreach program. The current exhibition on display, *Native Voices: Native Peoples' Concepts of Health and Illness*, explores the interconnectedness of wellness, illness, and cultural life for Native Americans, Alaska Natives, and Native Hawaiians. Through the extensive use of views and media, visitors will discover how Native concepts of health and illness are closely tied to the concepts of community, spirit, and the land.

HMD promoted Native Voices through email blasts to K-12 educators who previously visited exhibitions. HMD renewed contacts with pre-collegiate programs sponsoring tours for young people; submitted notices to the Bethesda Urban partnership, and participated in a Hawaiian cultural celebration at the National Museum of the American Indian. HMD sent emails to staff at the American Indian Higher Education Consortium, the Washington Internship for Native Students, the Udall Native American Congressional Internship, the Native American Political Leadership Program, the National Congress of American Indians, the National Indian Education Legislative Summit, the National American Indian Housing Council, and Barbara Jordan Health Policy scholars, followed up by phone calls, inviting all to view the Native Voices exhibition.

HMD traveled 25 copies of 12 exhibitions to 156 host venues in 43 states and five international locations during FY 2012. Host venues and their audiences had positive and favorable responses to the traveling exhibition program.

In addition to the flagship exhibitions mounted in the Rotunda Gallery, and travelling exhibitions, HMD featured special displays in the History of Medicine Division Reading Room. *"And there's the humor of it" Shakespeare and the four humors* opened on January 30, 2012. The project was developed in collaboration with the former director of the Folger Shakespeare Library, Gail Kern Paster, and Theodore Brown, a noted historian of medicine. The special display featured numerous works borrowed from the Folger in addition to rare books, incunables, and manuscripts from the History of Medicine Division.

On August 27, 2012, a new special display opened titled *Pick Your Poison: Intoxicating Pleasures and Medical Prescriptions*. The project was developed in cooperation with the National Museum of American History.

HMD develops resources to support K-12 educators and student engagement with online exhibitions and introduces those audiences to NLM collections. In February 2012, a new middle-school Chinese language lesson plan titled: China's Hygiene Education for Children during 1950s was launched. Using several online Chinese Public Health posters as primary sources, the lesson plan helps students studying the Chinese language acquire vocabulary and knowledge of sentence structure. In addition HMD launched a higher education module featuring online Chinese public health posters. Disseminating Health Knowledge: Public Health

Campaigns in 20th-Century China expands HMD educational resources. Accompanying both online adaptations of this year's special displays—*"And there's the humor of it,"* and Pick Your Poison, HMD produced two K-12 lesson plans and a higher education module for each show.

HMD launched a new education initiative this year hosting the first NLM Teacher Institute from June 19 to June 21, 2012. Twelve K-12 educators from five area school systems (District of Columbia, Montgomery, Prince George's, Frederick, and Loudoun counties) with subject area focus in history/social studies, health education, science, and library media worked with the Exhibition Educator to identify and develop ideas and products for education resources in support of online exhibitions and emphasizing Library collections and resources.

Collaborating with the Office of the Director, HMD conceived and developed *Hidden Treasure: The National Library of Medicine*, a richly-illustrated coffee-table book about the history of the Library and its collections. The book was reviewed positively by several major media outlets and professional journals, including the *Lancet*, *BMJ*, and *The New York Times*. Complementing the appearance of Hidden Treasure were numerous staff publications in peer-reviewed books and journals, contributions to the popular press, and interviews with media including CBS Sunday Morning, *Emory Magazine*, *Maryland Life Magazine*, NIH Research Radio, *NIH Catalyst*, and *NIH Record*, and *Rochester Review*.

Training and Recruitment of Health Sciences Librarians

LO staff members develop online training programs and online tutorials to teach the use of MEDLINE/PubMed and other NLM databases to health sciences librarians and other information professionals; oversee the activities of the National Training Center (NTC) at the University of Utah; direct the NLM Associate Fellowship program for post-masters librarians; and present continuing education programs for librarians and others in health services research, public health, the UMLS resources, and other topics. LO also collaborates with the Medical Library Association, the American Library Association, the Association of Academic Health Sciences Libraries (AAHSL), and the Association of Research Libraries to increase the diversity of those entering the profession, to provide leadership development opportunities, and to encourage specialist roles for health sciences librarians.

In FY2012, the MEDLARS Management Section (MMS) and the NTC launched a new "train-the-trainer" PubMed course, consisting of three distance education sessions and one on-site session. Overall, 408 students were trained in 26 classes covering the subjects of PubMed, ClinicalTrials.gov, "Teaching with Technology", and TOXNET. In addition, use of NLM distance education offerings continues to grow with 2,426,246 page views of the 57 online tutorials and Quick Tours available.

The NLM Associate Fellowship program had nine participants in FY2012: five 2nd year fellows at sites across the country (UCLA, Oregon Health & Science University, University of Cincinnati, Johns Hopkins University, and the National Library of Medicine) and four 1st year fellows, who completed their year at NLM in August 2012. Second year placements were arranged for three of these fellows at the National Institutes of Health Library, Bethesda, MD; Inova Fairfax Hospital, Fairfax, VA; and the University of Utah, Salt Lake City, UT. One fellow took a position at the University of California, Irvine. Four new fellows began the 2012/13 Associate fellowship year at NLM. Efforts to recruit fellows from underrepresented groups have been successful in attracting diverse groups of fellows to the program, including Hispanic representation in three of the last six cohorts and African-American representation in two of the last six cohorts.

NLM works with several organizations on librarian recruitment, diversity, and leadership development initiatives. For librarian recruitment and diversity, NLM works collaboratively with the American Library Association and the Medical Library Association to provide funding for the ALA Spectrum Scholarship Program. NLM has supported the Spectrum Scholarship Program for 12 years, with a commitment to continue support for another eight years. The ALA Spectrum Scholarship program recruits library science students from underrepresented populations and supports their academic and leadership development through tuition support and attendance at the ALA's annual meeting. NLM is also a partner library in an Institute of Museum and Library Services (IMLS)-funded initiative with the Association of Research Libraries (ARL), for the ARL Career Enhancement Program (CEP). The ARL CEP supports library science students from underrepresented populations with paid internships at an ARL library and attendance at an annual Leadership Symposium. LO hosted two ARL CEP fellows in spring 2012. LO also supports the NLM/AAHSL Leadership Development Program which provides leadership training, mentorship, and site visits to a mentor's institution for an annual cohort of five mid-career health sciences librarians. Recruitment efforts have emphasized and been successful in attracting minority candidates.

Table 1: Collections

Physical	Total[1]	FY2012	FY2011	FY2010
Monographs[2]				
Before 1500	597	0	1	1
1501-1600	6,053	0	7	21
1601-1700	10,343	0	5	24
1701-1800	272,731	4	21	54
1801-1870	256,690	37	66	108
1871-Present	861,182	13,287	14,650	12,855
Bound Serial Volumes[3]	1,451,840	17,025	21,924	23,418
Microforms[4]	606,076	10	15	249
Audiovisuals and Computer Software	91,260	1,310	1,224	1,335
Prints and Photographs	69,652	336	20	11
Manuscripts[5]	17,374,682	1,716,225	1,143,275	3,530,800
Withdrawn Items	(131,862)	(294)	(1,272)	(1,253)
Total items	20,869,244	1,747,940	1,179,936	3,567,623

Digital	Total[1]	FY2012	FY2011	FY2010
PubMed Central Articles	2,540,952	263,203	223,404	179,153
PubMed Central Titles[6]	1,149	252	203	158
Bookshelf Titles[7]	1,396	323	295	501
Digital Collections Repository[8]				
Texts[9]	6,979	5,075	1,406	518
Audiovisuals[10]	70	41	18	11

---

[1] Total: Numbers are cumulative as of the end of the fiscal year.

[2] Monographs: A bibliographic resource complete in one part or finite number of separate parts. Includes Americana, theses and pamphlets. Starting in FY11 numbers for these materials are reported under monographs by publication year.

[3] Bound serial volumes: A serial is a continuing resource issued in separate parts with no predetermined conclusion. Bound serial volumes include serials bound, serials pamphlet bound and bound serial gifts.

[4] Microforms: Reduced size reproductions of monographs and serials including microfilm and microfiche.

[5] Manuscripts: Total manuscripts equivalent to 8,948 linear feet of material, multiplied by a common factor to provide an item number estimate.

[6] PMC Titles: Only fully deposited titles.

[7] Bookshelf Titles: Titles of books, reports, databases, documentation, and collections. Spike in 2010 due to restructuring of HSTAT database in 2010; creation of new titles for existing content (2004-2010).

[8] Digital Collections Repository: Digitized content in the public domain. In the future will contain born digital items as well as reformatted items.

[9] Includes monographs and serials such as annual reports. Referred to as "Print Materials" on Digital Collections Website.

[10] Referred to as "Films and Videos" on Digital Collections Website.

Table 2: Collection Activities

Acquisitions and Processing	FY2012	FY2011	FY2010
Active Serial Subscriptions	19,184	19,731	20,645*
Items Processed[11]			
Serial Pieces	101,294	117,091	119,067
Monographs (pre-1914)[12]	715	886	1,041*
Monographs (1914-)	17,012	20,639	21,194
Audiovisuals[13]	1,335		
Prints and Photographs	47,982	897	1,845
Total	168,338	159,244	163,486

Archival Materials Acquired			
Modern Manuscripts (in linear feet)	497 (In ft)	866 (In ft)	681 (In ft)*

Expenditures	FY2012	FY2011	FY2010
Publications	$10,207,330	$10,216,505	$10,922,470
Rare Books, Manuscripts, and other Historical Materials	$299,584	$299,794	$299,822
Total[14]	$10,506,914	$10,516,299	$11,222,292

Preservation	FY2012	FY2011	FY2010
Volumes Bound	15,000	18,997	19,629
Volumes Repaired Onsite[15]	2,346	2,304	2,727
Audiovisuals Preserved	534	594	414*
Historical Volumes Conserved	997	340	4,125
Pages Digitized[16]	643,372	787,865	72,622

*These figures are corrected from those in the FY2010 report.

---

[11] Items processed: Serial issues, monographs and nonprint receipts processed.

[12] Monographs (pre-1914) includes historical manuscripts (those written prior to the year 1600).

[13] Audiovisuals became a separate tracking category in FY2012. For prior year reports, Audiovisuals were grouped with Monographs (1914-).

[14] Used to be reported in "Publications" prior to FY2012 and "Rare Books" was a portion of the amount.

[15] Volumes repaired onsite: General Collection monographs and serials only

[16] Number excludes digitization projects not associated with the Digital Collections Repository, e.g. Profiles in Science.

Table 3: Cataloging and Indexing

Cataloging	*FY2012*	*FY2011*	*FY2010*
General Collection Items[17]	19,656	21,268	18,820
Historical Monographs (pre-1914)	4,238	1,128	4,328
Modern Manuscripts (In ft)[18]	1,050 (in ft)	843 (in ft)	2,018 (in ft)
Prints and Photographs[19]	2,281	704*	14,784*
Historical Audiovisuals	992	105	792

Indexing	*FY2012*	*FY2011*	*FY2010*
Citations Indexed for MEDLINE	760,903	724,831	699,420
Journals Indexed for MEDLINE	5,633	5,559	5,484

*These figures are corrected from those in the previous reports.

---

[17] Items: Includes monographs, serials, nonprint and integrating resources.

[18] Number reflects manuscripts that are fully processed and have a catalog record.

[19] Number includes accessioned prints and photographs that are described by finding aids.

Table 4: Services to the Public

Document Delivery	*FY2012*	*FY2011*	*FY2010*
Interlibrary Loan Requests Received	234,662	244,701	256,459
Interlibrary Loan Requests Filled	194,255	200,581	213,330
General Reading Room Requests Received	85,352	119,209	145,240
General Reading Room Requests Filled	74,993	102,054	124,815
History of Medicine Reading Room Requests Filled	6,989	5,346*	5,954*
Customer Service Inquiries	*FY2012*	*FY2011*	*FY2010*
Offsite Inquiries[20]			
General	100,023	97,796	91,540
History of Medicine	10,857	8,667	8,811
Onsite Inquiries[21]			
General	6,124	6,531	7,076
History of Medicine	14,763	15,057	14,635
Data Licensees	*FY2012*	*FY2011*	*FY2010*
MEDLINE	782	688	621
UMLS	5,640	6,566	6,116
Tours and Visitors	*FY2012*	*FY2011*	*FY2010*
Exhibitions			
Visitors	2,769	2,522	7,183
Daily Tours			
Tours	122	100	115
Visitors	668	548	622
Special Tours			
Tours	63	71	69
Visitors	1,047	1,392	1,558

*These figures are updated due to a change in tracking more consistent with other LO areas.

---

[20] Offsite Inquiries: Inquiries via telephone, fax, US mail, and e-mail. Includes BSD interactions with data licensees.
[21] Onsite Inquiries: In person

Table 5: Selected Web Resources

Resource	FY2012	FY2011	FY2010
ClinicalTrials.gov			
Number of Trials	133,291	118,682	100,919
Page Views[22]	1,474,668,417	1,347,564,098	857,270,086
Visitors[23]	11,201,854	9,518,503	10,583,773
DailyMed			
Number of Labels	39,232	31,000	11,600
Page Views	85,025,368	118,163,000	70,195,000
Visitors	8,812,640	5,166,000	3,035,000
Genetics Home Reference			
Summaries	1,907	1,622	1,408
Page Views	145,647,998	142,967,488	118,267,551
Visitors	6,382,520	4,819,287	3,641,936
Household Products Database			
Number of Products	11,000	11,000	10,000
Page Views	17,994,336	10,254,423	10,984,725
Visitors	786,346	694,042	850,332
MEDLINE/PubMed			
PubMed Citations	22,174,097	21,200,952	20,209,357
Page Views	6,113,756,836	4,783,804,040	4,104,309,671
Unique Visits[24]	482,175,373	539,184,717	366,396,717
Searches[25]	2,204,710,177	1,834,537,403	1,578,714,477
MedlinePlus			
Number of Topics (English/Spanish)	931/918	903/881	859/826
Page Views	745,000,000	728,000,000	716,000,000
Visitors	213,000,000	164,000,000	153,000,000
NLM Main Web Site			
Page Views	44,7000,000	39,300,000	44,500,000
Visitors	8,900,000	8,300,000	7,700,000
ToxTown			
Page Views	6,757,011	6,164,031	6,096,928
Visitors	240,277	192,737	286,004

---

[22] Page Views: Number of times that a single page is viewed or downloaded.
[23] Visitors: Number of people visiting a Website in a defined period of time.
[24] Unique Visits: Total number of times that all users visit a Website, regardless of the number of individual pages viewed.
[25] Searches: Number of searches performed.

SPECIALIZED INFORMATION SERVICES

Steven Phillips, MD
Associate Director

The Division of Specialized Information Services (SIS) at the National Library of Medicine (NLM) offers a diverse collection of publicly accessible free online information resources and services in toxicology, environmental health, chemistry, HIV/AIDS, disaster management, minority health, and other specialized topics. SIS includes an Outreach and Special Populations Branch, which actively seeks to improve access to high quality health information by underserved and other targeted populations. SIS also manages NLM's Disaster Information Management Research Center (DIMRC). This year, we created a Clinical Toxicology Office within the Division to facilitate providing relevant resources to health professionals. Working closely with other parts of the Division, the new office focuses on the identification of the needs of clinical toxicologists and other potential users, development of new clinical toxicology resources or enhancements to existing SIS resources, and interaction with NIH, government agencies, professional societies, and educational institutions on clinical toxicology-related information needs and collaborations. At the same time, the Office of Clinical Toxicology was created, DIMRC was moved from being part of the SIS Office of the Director, to branch status.

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 as a result of recommendations of the President's Science Advisory Committee. From its inception, TEHIP has strived to use the most up-to-date and relevant information together with up-to-date information technologies to provide rapid and effective access to the latest toxicological and environmental health information. We not only create databases ourselves, but direct users to relevant sources of toxicological and environmental health information wherever these sources may reside as part of the library's role in collecting and organizing health and medical information. Identifying high quality resources is significant facet of the work done within SIS, and this has resulted in a robust collection of guides to specific toxicology and environmental health topics including Epigenomics, Water Pollution, and many resources related to all aspects of disaster preparedness and response. This year as part of our involvement in the trans-NIH Tobacco and Nicotine Research Interest Group, we prepared a links page on Tobacco, Smoking, and Health.

SIS has been involved in the development of HIV/AIDS information resources since 1989, and we now have responsibility for several collaborative efforts in information resource development as well as a major outreach effort to the organizations serving those affected by HIV/AIDS.

The Division's outreach programs have continued to innovate and reach out to new populations that could benefit from using NLM's resources. In addition to minority populations, which have always been an important focus of these programs, SIS also reaches out to the public health workforce, women, disaster planners and, most recently, the K-12 educational community.

This year, DIMRC continued to expand its activities as well as respond to national and international crises. Released in FY2011, the Chemical Hazards Emergency Medical Management (CHEMM) resource, a companion to REMM, has continued to be developed. This tool assists emergency managers, planners, responders and other healthcare professionals in responding to mass casualty chemical events.

In FY2012, SIS completed its fourth year of co-directing the Public Health Informatics course at the Uniformed Services University of the Health Sciences (USUHS). This course is jointly sponsored by the USUSHS Biomedical Informatics Department and the National Library of Medicine providing students with a conceptual framework for understanding the emerging field of public health informatics, highlighting the latest approaches and technologies used by NLM and other organizations. New for 2012 was a speaker from the National Institute for Environmental Health Sciences (NIEHS). SIS was also invited to lecture several times a year in the toxicology courses offered by the Preventive Medicine & Biometrics Department at USUHS, covering topics as diverse as toxicokinetics, risk assessment, risk management, snake and other venoms, plant toxins, and NLM's toxicology databases and other resources.

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. In addition to the Web sites and databases, SIS added Twitter and Facebook pages to its repertoire of communication tools, developed apps for smartphones, and used crowd-sourcing for resource development.

Toxicology and Environmental Health Resources

Figure 1: Stay Connected
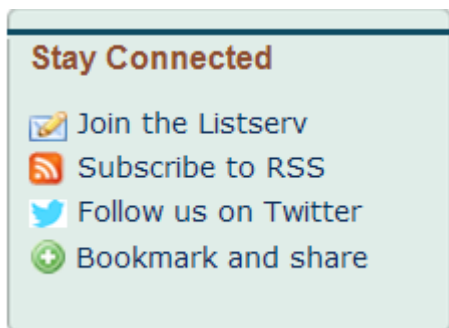
We continue to embrace all avenues of communication to reach our diverse user communities. Here is a box that appears on our newly redesigned Environmental Health and Toxicology home page, offering choices for users to connect with us. We have moved beyond the Web resources we provide in the area of toxicology and environmental health information to social media methods that allow us to rapidly deploy new important relevant information on our own resources or other ones as well. We use a toxicology and environmental health listserv, Facebook accounts, Twitter accounts, and RSS feeds to supplement our traditional Web site and searchable databases. This helps us reach the general public as well as scientists, researchers, and health workers. In addition, in response to the ongoing trend in information access to move towards more mobile resources, we continue to develop and enhance applications for mobile access.

Our **TOXNET** (TOXicology Data NETwork) databases are the core content in our information resources. These databases are varied in what they cover, and how they are created and maintained. Some are legacy ones, evolving from their earlier forms as we included information on a wider range of toxic chemicals, such as adding Superfund cited compounds, focusing on worker exposures, taking on the possible safety issues with nanomaterials, and considering household products safety. Others are databases originally created elsewhere and put up on the TOXNET system for greater visibility and to provide continued access when the project that created them was stopped. Enhancements to TOXNET continue, based on user feedback and upgrades or additions of data and capabilities. These databases are currently part of the TOXNET system:

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. Complementary and alternatives medicines (CAM) are now included in LactMed,   with a focus on supplements that are frequently used by nursing mothers. A LactMed app for the iPhone, iPod, iPad, and Android released last year continues to be very popular.

HSDB (Hazardous Substances Data Bank), is a peer reviewed database focusing on the toxicology of over 5,500 potentially hazardous chemicals. This flagship database continues to be enhanced with records on nanomaterials and includes structures from ChemIDplus displayed as images at the start of the record. This year records were added on EPA's green chemicals and pesticides, plant derived toxins, and animal venoms. We completed an HSDB Needs Assessment project for the re-engineering of the HSDB, and began work to better incorporate new resources important to today's toxicologists. We have several concrete steps that are being taken immediately.

ITER (International Toxicity Estimates for Risk), is a database containing data in support of human health risk assessments. It is compiled by Toxicology Excellence for Risk Assessment (TERA) and contains 690 chemical records. These side-by-side comparisons of international risk assessments include links to source documentation.

TOXLINE, is a comprehensive database providing journal and report literature coverage of the biochemical, pharmacological, physiological, and toxicological effects of drugs and other chemicals from 1965 to the present. TOXLINE contains nearly 4.6 million citations, almost all with abstracts and/or index terms and CAS Registry Numbers. This year we added updated records on toxicology research projects from the CRISP database, and began the process of extracting additional project information from NIH's new system RePorter.

DART/ETIC (Development and Reproductive Toxicology /Environmental Teratology Information Center), is 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 as well as a part of TOXLINE.

Toxics Release Inventory (TRI), is a series of databases that describe the releases of toxic chemicals into the environment annually for the 1987-2010 reporting years.

ChemIDplus, is a database providing access to structure and nomenclature authority databases used for the identification of chemical substances cited in NLM databases. ChemIDplus contains over 387,000 chemical records, of which over 300,000 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 12,000 common household products used inside and around the home. New categories of products were added this year, and we now include nine categories. Work was begun with the Consumer Product Safety Commission (CPSC) to discuss possible types of nanomaterials and additional categories of products to add to the Household Product Database.

Haz-Map, is 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. Additional jobs and hazardous tasks with occupational diseases and their symptoms were added this year. We continue to collaborate with the Department of Labor to include tasks and chemicals associated with work at the Department of Energy hazardous sites . This year a mobile version of Haz-Map was released, and the Web interface was changed to allow both versions to be updated together.

ALTBIB is a Web portal providing information on resources about alternatives to the use of live vertebrates in biomedical research and testing, developed as part of NLM's participation in the Interagency Coordinating Committee on the Validation of Alternate Methods (ICCVAM). Work was finished this year for the first update and expansion of the content of the ALTBIB resource in over five years. The new version of ALTBIB was released in September 2012, and it now includes access to PubMed searches for citations covering 17 topics such as ocular toxicity and numerous subtopics aligned with current approaches. This includes information on *in silico*, *in vitro*, improved (refined) testing methods, and testing strategies and other approaches. ALTBIB now also provides easy access to key organizations, the latest news from some of those organizations, and to US and international compilations of validated methods and methods undergoing validation.

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

TOXMAP is a Geographic Information System (GIS) 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 FY2012 include EPA Toxics Releases (2010 data) and EPA Superfund NPL data. Work was continued on a new version of TOXMAP which will more closely align with the look and feel of current mapping tools.

The **Drug Information Portal** provides current drug information for over 28,000 drugs, with links to many credible additional online resources. During FY2012, changes were made to the search interface to allow easier selection of drug names and searching by category of drug. This year, we continued to work with FDA on the increasing the availability of pill images for inclusion in Pillbox and other NLM services and research projects. In collaboration with LHCNBC, a standardized methodology was developed for creating high-resolution images of medications suitable for applications in health IT, as well as image processing and informatics research.

The **Dietary Supplements Labels Database** was enhanced with additional labels from the many brands available in the marketplace. SIS is also collaborating with the NIH Office of Dietary Supplements (ODS) in developing a comprehensive labels database intended for researchers.

The **Comparative Toxicogenomics Database (CTD)** was developed at the Mount Desert Island Biological Laboratory and is included in TOXNET searching. CTD receives funding from the National Institute of Environmental Health Sciences (NIEHS) and NLM. This database focuses on molecular mechanisms by which environmental chemicals affect human disease, providing insight into complex chemical-gene and protein interaction networks.

ToxLearn is a tutorial providing an introduction to concepts and principles of toxicology, it is under development as a joint project with the Society of Toxicology. Module 1, Introduction to Toxicology and Dose-Response, was released this year. Future modules will be released as they are developed.

LIVERTOX, a Web site that provides up-to-date, accurate, and easily accessed information on the diagnosis, cause, frequency, patterns, and management of liver injury attributable to prescription and nonprescription medications, herbals and dietary supplements, was released this year. LIVERTOX also includes a case registry that will enable scientific analysis and better characterization of the clinical patterns of liver injury. The LIVERTOX Website provides a comprehensive resource for physicians and their patients, and for clinical academicians and researchers who specialize in idiosyncratic drug induced hepatotoxicity. Its content is produced by the National Institute of Diabetes and Digestive and Kidney Diseases and NLM.

Disaster Information

SIS acquired additional space in FY2012, allowing for a larger and enclosed emergency information center. The new emergency information center includes several computers, TVs, smartboard, radios, and improved collaboration facilities.

The NLM Board of Regents ad hoc Working Group on Disaster Health Information Management held its second and final meeting in September 2012. The

group was formed to review the current activities of NLM's Disaster Information Management Research Center (DIMRC) and provide guidance on future authorization, support and direction for DIMRC. The Working Group made recommendations for the inclusion of NLM/DIMRC in the national agenda for disaster preparedness and response. The report includes recommendations regarding several major areas: disaster medicine and public health information (collection, organization, and dissemination); information technology tools and research; outreach and training; evaluation; and authorization.

Disaster Information Outreach

The Disaster Information Specialist Program is a collaborative effort to explore and promote the role of information specialists in the provision of disaster-related information resources to the disaster preparedness and response workforce and to communities. In FY 2012, NLM collaborated with the Medical Library Association (MLA) to offer a continuing education program that leads to a basic or advanced certificate. The first course was taught in FY2011; this year it was joined by five additional courses, taught both in person and online. The online courses are also archived for viewing at any time. Courses included: US Response to Disaster and Public Health Emergencies; Information Roles in Disaster Management; Ethical and Legal Aspects of Disaster Response; A Seat at the Table: Working with Local Responders; and Health and Disasters: Understanding the International Context.

At the end of FY2011, NLM awarded purchase orders to seven organizations for disaster information outreach projects. The projects required an organization working in disasters (public health department, emergency management department, fire/rescue service, academia, etc.) to partner with a library to improve access and knowledge of NLM and other disaster health resources and to identify roles for librarians in disaster management. These seven projects were successfully completed in 2012 and included training of the disaster health workforce in accessing disaster health information from NLM and other organizations, developing Web and mobile access to local disaster health resources, and working with healthcare coalitions to improve access to information. At the end of FY2012, NLM awarded another seven purchase orders. These projects include: assisting and training the Medical Reserve Corps of Arkansas; and teaching emergency responders and public health officials to use mobile apps related to disaster health, develop situational awareness briefs, and assist with information management issues related to disasters including family reunification, situation awareness in the hospital command center, etc.

DIMRC continued to promote awareness and information sharing via the Disaster Outreach Listserv an the NLM_DIMRC Twitter feed. The listserv membership grew from over 650 subscribers at the end of FY2011 to over 800 at the end of FY 2012. During FY2012, DIMRC expanded the use of Twitter substantially and engaged with the Social Media for Emergency Management community. The NLM_DIMRC Twitter account expanded from 100 subscribers at the end of FY2011 to over 500 subscribers at the end of FY2012. Twitter allowed us to quickly and easily promote NLM resources, especially after Hurricane Sandy, and to collaborate with other agencies and organizations involved in disaster medicine and public health.

The Disaster Information Specialist monthly meetings/conference calls continued throughout FY2012. Topics this past year included: needs of public libraries following disasters; use of information for situation awareness in fusion centers; the American Red Cross digital operations center; disaster apps and mobile Web sites; exercises and drills 101; and updates on the disaster outreach purchase orders.

DIMRC continued to work with the Pan American Health Organization (PAHO) and the Regional Disaster Information for Latin America and the Caribbean (CRID) on the network of disaster information centers in the region. Several virtual meetings of RELACIGER (Spanish acronym for the Latin American Network of Disaster Health Information Centers) were held throughout the year. Small projects were started by a number of the centers that would benefit the entire network, such as: a Web site for RELACIGER; a GIS application mapping disaster organizations in the network countries; development of promotional materials to be used by the centers, and the development of a training Web site and a training course on disaster health information management. In addition, PAHO, CRID, and NLM continued to update and promote the availability of a Haiti Disaster Information Center Web site that contains approximately 1,000 health-related documents generated following the 2010 earthquake and subsequent cholera outbreak.

Health Information Technology Research and Development for Disaster Management

DIMRC continued to coordinate and conduct a number of informatics and communications research and development projects on behalf of the Bethesda Hospitals' Emergency Preparedness Partnership (BHEPP). NLM joined BHEPP in 2008 to help with common issues at hospitals during emergencies: communication, patient management, family reunification, and information access. Three NLM divisions, SIS, OCCS, and LHNCBC, are working on these projects.

DIMRC continued to enhance the emergency email communication system to be used when the Internet is not available. This system uses the Military-Affiliated Radio System (MARS) to assist in relaying email when the Internet is not available. A station at the NIH Radio Amateur Club and a portable station are now available and the system was tested at the National Capital Region Capital Shield Exercise in October 2011. The new,

enhanced version is expected to be demonstrated at the disaster exercise to be held at the Walter Reed National Military Medical Center in October 2012.

NLM and Suburban Hospital/Johns Hopkins Medicine conducted the first disaster exercise in a virtual world in June. This exercise introduced Suburban Hospital staff to the virtual world platform (Second Life) to determine if this platform could enhance the Hospital Incident Command training experience. Participants include healthcare workers and administrators who may be called to staff the incident command center during a disaster. The virtual world provided a more detailed level of engagement than a lecture or table-top exhibit yet was not as time-consuming or disruptive to the hospital as a full-scale exercise. An after action report was drafted and will assist in the planning of a second exercise scheduled for FY2013.

Figure 2: Virtual Drill 2012

Virtual Drill 2012. The virtual hospital incident command center was developed by NLM as a training method for disaster and emergency response.

Disaster Health Information Resources and Tools

The Resource Guide for Disaster Medicine and Public Health, a database of free online documents and resources related to disaster medicine and public health, was completely restructured in FY2012. Both the database input system and the search engine were revised and enhanced. The scope and coverage manual and selection criteria were also revised. Over 1,000 new resources were added to the database and the backlog of materials to be included was eliminated. In addition, links to PubMed and MedlinePlus searches were added to the results pages for easy access to additional sources of information.

Several new disaster health topic pages were developed including one on disaster health apps and mobile Web sites. This page proved to be extremely popular, especially after Craig Fugate, FEMA Administrator, tweeted about it. A large increase in access to that page what noted in the statistics.

SIS also began working with the HHS Assistant Secretary for Preparedness and Response (ASPR) on the development of an interactive, searchable All-Hazards

Playbook (or Plan). ASPR staff wrote the new plan to be used by HHS staff at headquarters, in the field offices, and at CDC. NLM turned the document into a database-driven resource that can easily be searched and provide quick and easy access to specific portions of the resource. A beta version of the Playbook was used in the Secretary's Operation Center (SOC) for Hurricane Isaac and an enhanced version was used for Hurricane Sandy. This new tool allowed staff in the SOC to quickly identify what needed to be done at each phase of the incident (before, during, and after) and who needed to do it.

SIS signed an Inter-Agency Agreement with the Food and Drug Administration (FDA) for mobile app development of data capture for adverse effects of drugs approved via Emergency Use Authorization in Public Health Emergencies. Initial discussions were held in FY2012 and the development of the apps is expected in FY2013.

WISER

Several enhancements were made to the mobile versions of the Wireless Information System for Emergency Responders (WISER), a tool for hazardous materials (hazmat) and chemical, biological, radiological, and nuclear (CBRN) incidents. A universal iOS app was developed that will work on the iPhone, iPod Touch, or iPad. This new universal version includes important features including the Help Identify an unknown chemical and the protective distance mapping. In addition, the protective distance mapping feature was added to the Web version of WISER (WebWISER). The protective distance mapping feature takes the safe protective distance as identified in the Department of Transportation's (DOT) Emergency Response Guidebook (ERG) and allows users to enter a location and wind direction to quickly and easily map the location potentially affected by an incident. At the end of FY2012, WISER had been downloaded approximately 450,000 times.

CHEMM

The Chemical Hazards Emergency Medical Management tool continued to be enhanced in FY2012. NLM began working on integrating the CHEMM content into WISER to create a single app. Funding from the trans-NIH CounterACT (Countermeasures Against Chemical Threats) program is being used for the development and enhancement of easy to access information on: 1) the toxic effects and other information such as patient management for fifty possible mass casualty-related chemicals; and 2) over 20 medical countermeasures (including efficacy, relevant publications, research in progress, and FDA and other global regulatory status information). In addition, it includes the development of a prototype tool to capture feedback on possible mass casualty chemical-related events, including on-site and post-incident information to

develop lessons learned. This work will continue into FY2013.

An Interagency Agreement between the Department of Homeland Security (DHS) Office of Health Affairs (OHA) and the NLM resulted in a technical workshop on May 8-9, 2012 to discuss and develop a consistent lexicon to describe toxic chemical syndromes, or toxidromes for CHEMM. During this workshop, that included over 40 practitioners and experts in emergency response, emergency medicine, and medical toxicology, names and definitions for twelve unique toxidromes that describe and differentiate the clinical signs and symptoms from exposures to chemicals were developed and agreed upon. These toxidromes lay the foundation for a consistent lexicon for use in CHEMM and for other uses that could improve response to chemical mass exposure incidents.

Outreach Initiatives

The **Environmental Health Information Partnership** (EnHIP) enhances the capacity of minority serving academic institutions to reduce health disparities through the access, use and delivery of environmental health information on the campuses of HBCUs, tribal colleges and minority serving institutions and in their communities. EnHIP held a successful meeting of its 22 partnering institutions including 15 HBCUs, three tribal colleges, three Hispanic-serving institutions and one Alaskan university. The program focused on *Children's Health and the Environment,* with a presentation by Dr. Yvonne Maddox, deputy director of the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). Representatives enjoyed an overview and tour of the *Native Voices: Native People's Concepts of Health and Illness* exhibition. Seven small outreach projects were funded with EnHIP partner institutions in FY2012.

United Negro College Fund Special Programs (UNCFSP)-NLM HBCU Access Project continued this year. The partnership works to incorporate NLM information resources into HBCU courses and other campus and community activities. During FY2012, four HBCUs were funded for community or campus outreach projects. Awardees included Hampton University, Johnson C. Smith University, Morehouse School of Medicine, and Shaw University. In June, the four principal investigators presented their project results to UNCFSP and NLM staffs. Following the meeting, an online training and a technical assistance workshop on developing outreach proposals were offered.

The **Prince George's Care Access Network (PG CAN)** is a health information awareness and resource initiative established to address high rates of mortality and high prevalence of chronic diseases in Prince George's County, Maryland. The initiative was developed through partnerships with the Prince George's Hospital Center (Dimensions Healthcare Systems), First Baptist Church of Glenarden, Health Action Forum of Prince George's

County and NLM. The faith-based project aims to increase awareness of the health disparity concerns in the county, improve quality of life and promote the elimination of health disparities. Using NLM's online resources and the MedlinePlus Magazine, the three participating churches developed programs to increase consumer health information awareness and disseminated information about diabetes, exercise, high blood pressure, nutrition, obesity and smoking. In addition, they conducted a variety of events, including: Healthy Sundays that focused on specific health topics; large and small health fairs where health information was disseminated and health screenings offered; presentations during Sunday School, church programs and to seniors ministry; and health-related bulletin boards. The churches will incorporate NLM resources into health related projects and link to NLM resources from their Web sites.

In FY2012, SIS staff expanded its outreach efforts working with **Black Greek Letter Organizations**. These organizations are involved actively in local outreach efforts to combat minority health disparities. Staff introduced members of Zeta Phi Beta Sorority, Delta Sigma Theta Sorority, Kappa Alpha Psi Fraternity, and The Links, a professional organization of women of color, to NLM resources to promote health literacy and health awareness.

Outreach to Minority Health Professionals

SIS continued its efforts to increase the awareness of and use of NLM online resources by minority health professionals. Hands on training, presentations and demonstrations were conducted at the 2012 NIH Residency Training programs of The Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), Fifth Annual National Conference on Health Disparities, Student National Medical Association Annual Medical Education Conference, National Medical Association, National Black Nurses Association, National Dental Association, Hispanic Dental Association and the Society of American Indian Dentist, as well as several regional conferences.

SIS hosted David Wharton and Dwane Sorrells, from the Choctaw Nation in Oklahoma, as Sacred Root Fellows for a weeklong visit to NLM in January. In August, SIS staff visited two Sacred Root fellowship sites in Oklahoma. Visits were made to Chickasaw Nation Medical Center in Ada, Oklahoma to meet with hospital administrators and staff of the Chickasaw Health Information Center (CHIC). In addition, staff visited with first year fellows from the Choctaw Nation. An overview of the health care system and facilities and a tour of several health care clinics and Choctaw Nation cultural sites were provided. Discussion of a project for year two of the fellowship centered on the development of a disaster plan for Choctaw Nation members.

Outreach to Librarians

NLM continues to support the training and education of minority librarians through funding of two graduate assistantships for students in the Knowledge River Program at the School of Library and Information Science, University of Arizona. These students work in the Health Sciences Library and are encouraged to pursue careers as health sciences librarians.

SIS supported programming at the 2nd National Joint Conference of Librarians of Color in Kansas City, Missouri and participated by providing a poster session on American Indian Health, "Our Stories."

Special Population Web Sites

In FY2012, SIS added digital stories collected previously with American Indian community members for American Indian Health Web Portal (AIH). In addition, digital storytelling workshops were held with the Montana Wyoming Tribal Leaders Council for a veterans project and the Shoshone Women's Society at Fort Washakie, Wyoming. Participants received training in digital storytelling methods and finished the training by completing production of their own digital story. Several of the stories were added to AIH. New topic areas under development for the portal are teen health, women's health, men's health and historical trauma.

Figure 3: Digital Storytelling

Digital storytelling workshop conducted by nDigiDreams at the Fond du Lac Human Services Health Center, Cloquet, Minnesota.

Work continued to transfer the Arctic Health portal, currently housed at the University of Alaska, Anchorage (UAA,) to NLM. The site is being redesigned with new images and templates and will be 508 compliant.
Women's Health Resources (WHR,) a collaboration with the Office of Research of Women's Health (ORWH), NIH, continued to expand research dissemination through the Web portal and social media. In FY2012, social media efforts grew significantly. Facebook (see http://www.facebook.com/WomensHealthResources) "likes" increased by 77% to 2,056 and Twitter (see

http://twitter.com/WomensHealthNIH) followers increased by 136% to 26,300.

SIS worked with ORWH in an outreach effort to increase the capacity of sex differences research by university researchers and increase female consumers' knowledge and use of research. Seven projects with an intended audience of students, faculty and health professionals, and five projects with the intended audience of consumers were awarded. These projects will serve as a pilot to determine if a program in this subject area would be well received by the community. Awardees included Howard University Stokes Medical Library, USET (United Southern and Eastern Tribes), Morehouse School of Medicine Library, Morgan State University Behavioral Health Sciences Department, Community Education Group (Washington, DC), TASSC (Technical Assistance for Social Services Collaborative) (Austin, TX,) University of Alaska Anchorage Medical Library, and Virginia Commonwealth University Tompkins McCaw Library.

SIS is a partner in the Refugee Health Information Network (RHIN) a national collaborative partnership of the Association of Refugee Health Coordinators, NLM and the Center for Public Service Communication (CPSC.) RHIN is committed to providing quality multilingual, multicultural health information resources for patients and those who provide care to resettled refugees and asylees. Members of the Association of Refugee Health Coordinators serve as subject matter experts for review of materials that are submitted for inclusion in the RHIN database. NLM commissioned the translation of documents related to HIV/AIDS, violence against women, and pregnancy into multiple formats (audio and video) and multiple languages for inclusion in RHIN.

Social Media

SIS continued its use of social media for promotion and outreach this year. The SIS social media team manages several Twitter feeds to promote and update users on resources and programs.   Twitter feeds include:
@NLM_SIS
@NLM_OSP
@NLM_DIMRC
@NLM4Caregivers
@AIDSInfo
@refugeehealth

There are now Facebook accounts for several SIS resources including NLM4Caregivers, Women's Health Resources and AIDS*info*.

The NLM4CareGivers pilot project is an outreach project using both Facebook (http://facebook.com/NLM_4Caregivers) and Twitter (@NLM_4Caregivers) to engage family caregivers and establish best practices for future outreach efforts. NLM_4Caregivers has changed its look and standardized its image across social media platforms. Outreach efforts included development of a new

educational piece on NLMs resources for caregivers. Also planned is programming for those who provide reference and research services to family caregivers around NLM and caregiving resources. On Facebook, NLM4Caregivers posts 30-60 times per month to its 504 followers. On Twitter, NLM_4Caregivers has increased its followers by 100% to 510.

K-12 Initiatives and Resources

NLM is in partnership with King Drew Medical Magnet High School in Los Angeles, Farrington High School, Honolulu, Hawaii and Kotzebue Middle/High School, Alaska to deliver a distance learning program via synchronous videoconferencing and collaboration technologies using Internet2. The program focuses on minorities while meeting school curriculum and community health needs. The project features a series of science topics selected by the high schools' teachers to match curriculum requirements. The program includes curriculum development, student evaluation, and two-way video teleclasses. The theme for FY2012 was Native American Health and Healing.

In March, SIS and Lister Hill Center staff made site visits to Farrington and King Drew High Schools to review and evaluate the distance learning program. During the visits, interviews with students, administrators, and teachers associated with the program were held.

Students from the St. John's High School, Johns Island, South Carolina visited NLM in April to present the results of their Teen Health Leadership project. SIS staff visited the school in May to meet with project staff from the high school and the Medical University of South Carolina Nursing School and Library to discuss progress made during the school year and make plans to develop a model curriculum based on the project so others might adopt and use it.

Figure 4: Students from St. John's High School

High School students from St. John's High School, Johns Island, SC, participating in the Teen Health Leadership Program and learning about NLM's information resources.

SIS coordinates the NLM Adopt-a-School Program, a partnership with the NLM Office of the Director. The program provides training in the use of NLM's consumer health databases, summer internship for students, enrichment of library resources including the donation of books and periodicals, guest lectures, and more. SIS coordinated summer internships for five students from Woodrow Wilson Senior High School in Washington, DC and Charles H. Flowers High School in Springdale, Maryland. SIS hosted two students from Woodrow Wilson High School who assisted the Outreach and Special Populations Branch with its social media efforts and Office of Clinical Toxicology with video tutorials on SIS resources. Other students were hosted by the Library Operations Division, Office of the NLM Director, and Lister Hill Center.

In March, SIS provided support for the 26th Association of Minority Health Professional Schools Annual Symposium in Long Beach, California. The meeting focused on local and regional pipeline programs that included enrichment activities to encourage and promote minority student interest in science and math. NLM introduced students to its online resources at an exhibit.

Tox Town was enhanced with several additions: US Southwest scene, drought location, and methane chemical. The new US Southwest scene introduces environmental hazards such as uranium tailings, abandoned mines, coal-fired power plants, dust storms and more. The drought scene introduces users to the causes and health effects of this indicator of the climate change. The Methane page of the Chemical List provides information sources and health effects of this greenhouse gas, which represent approximately 10 % of all greenhouse gas emissions in United States.

The K-12 team completed the Tox Town-based environmental health afterschool club curriculum for middle school students, and released as a set of PDF brochures, accessible via Tox Town Web site. The curriculum was promoted via outreach to a number of federal, national, and local organizations. The team also recruited four middle schools at diverse locations (Oakland, CA, Nashville, TN, Richmond, VA, and Garden City, MI) to participate in a pilot test of the curriculum. The schools are now in the process of implementing the curriculum. The K-12 team provides technical support for the implementation.

Environmental Health Student Portal is a Web site for middle school students and teachers, developed in FY2010. In FY2012, SIS completed a usability study of the portal, which produced valuable information for improving the site.

SIS completed and released GeneEd, a Web site that provides a single repository to genetics research, study guides, lesson plans, experiments and activities for teachers and students in grades 6-12. The concept for the site came from interest expressed by educators at a genetics conference and confirmed by a survey distributed

by the National Association of Biology Teachers. The Web site provides a single, safe source for teachers to use to introduce, teach, or supplement existing genetics course materials and to which to direct students. SIS led the development of the site, coordinating the input from NIH groups that specialize in genetics educational content and research. The groups included Genetics Home Reference NLM team, National Center for Biotechnology Information, National Human Genome Research Institute, and the Genetics and Rare Diseases Information Center, ORDR/NCATS.

In FY2012 members of the K-12 team participated as exhibitors and presenters in a number of conferences, including National Science Teacher Association annual conference, National Association of School Nurses annual conference, and annual convention American Middle Level Educators Association. An article describing using Tox Town to support science classroom argumentation was published in *Science Teacher*, the membership journal of the National Science Teacher Association, in the summer of 2012.

HIV/AIDS

The National Library of Medicine has continued its HIV/AIDS related outreach efforts to community-based organizations, patient advocacy groups, faith-based organizations, departments of health and libraries. The program, begun in 1994, provides support to design local programs for improving information access for HIV/AIDS patients, the affected community and 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; social media; or document access. In FY2012, NLM made awards to the following organizations: AIDS Educational Global Information System (AEGIS); Alliance of Border Collaboratives (ABC); Comunidades Unidas / Communities United; George Washington University; Hispanic Communications Network, LLC, Internet Sexuality Information Services, Inc.; Morgan State University; Pacific College of Oriental Medicine, LLC; Philadelphia FIGHT's AIDS Library and Critical Path Project; Red Ribbon Project; SisterLove, Incorporated; University of New Mexico; and Washington Area Consortium on HIV in Youth/Metro Teen AIDS.

SIS received set-aside funds to evaluate the AIDS Community Information Outreach Evaluation project from 1994-2012. This year the evaluation team from Columbia University developed the tool and gathered the data for a report on findings due next year.

NLM manages the AIDS*info*/*info*SIDA service on behalf of HHS. The steering group includes representation from NIH (NIAID, OAR, and NLM), FDA, CDC, and HRSA. The core components of AIDS*info* are the HHS Treatment Guidelines and HIV/AIDS-related clinical trials. These are updated on an ongoing basis. This year the following guidelines were updated:

- March 2012: *Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents*
- July 2012: *Recommendations for Use of Antiretroviral Drugs in Pregnant HIV-1-Infected Women for Maternal Health and Interventions to Reduce Perinatal HIV Transmission in the United States*

AIDS*info* updated the *Glossary of HIV/AIDS-Related Terms* in both English and Spanish in November 2011 and July 2012 respectively. The updated glossaries were released in both HTML and PDF. Subsequently, in September 2012, AIDS*info* updated the iPhone application to include 7th edition content.

In commemoration of World AIDS Day, December 1, 2011, the AIDS*info* and *info*SIDA Web sites were redesigned on the basis of extensive evaluation research, including analysis of the American Customer Satisfaction Index (ACSI) survey and usability studies conducted in both English and Spanish. As sister Web sites, both AIDS*info* and *info*SIDA were tailored to meet the needs of their primary audiences, and *info*SIDA received its own Web site domain—infoSIDA.nih.gov—to make it easier for users to bookmark pages and for search engines to index the site. Subsequently, in February 2012, a distinct mobile site domain for *info*SIDA was released: m.infoSIDA.nih.gov.

NLM SIS staff along with members from the AIDS*info* team helped staff the "Guidelines, Clinical Trials, and Information Resources" kiosk at the XIX International AIDS Conference held in Washington, DC in July 2012.

In September 2012, AIDS*info* redesigned its drug database. The database was updated to automatically display the FDA labels from DailyMed on the AIDS*info* Website, eliminating the need to manually update drug information for health professionals.

Research

SIS participated in a number of research activities in the fields of program evaluation, consumer health informatics, and science education. One such activity involved completing a project (conducted in FY2010 and FY2011) that analyzed published studies of community health information outreach efforts, identifying characteristics of successful health information outreach programs, as well as effective evaluation methods for capturing best practices. In FY2012, SIS completed the data analysis, prepared a manuscript that was accepted for publication in the *Journal of Medical Library Association*.

In the area of consumer health informatics, the Division continued working on a study of factors that affect patients' comprehension of consumer health materials in collaboration with colleagues from the University of Wisconsin, Madison. FY2012 efforts involved completing a paper on the types of comprehension errors, made by laypeople reading medical

documents, such as physicians' notes and descriptions of clinical trials which was published in the *Journal of Biomedical Informatics*.

SIS's third research area involves science education. The two science education projects, conducted in FY2012, continued from FY2011. The first project involved a study of team interactions between teachers, educational researchers, and Web developers co-designing educational activities for environmental health education. In FY2012, SIS received an invitation to prepare and submit an article about the study to the special Health and Environment issue of the *International Journal of Science Education*. The paper was completed and submitted to the

journal's editorial team in FY2012. The second project involved continuing a study of the impact of college major and biological knowledge on college students' views on commonly held incorrect health beliefs. The study is a collaboration between SIS, City College of NY and Georgetown University. The work conducted in FY2012 included data analysis and manuscript preparation. A preliminary version of the manuscript was accepted for presentation at the National Association of Research in Science Teaching conference. The journal manuscript is in the final stages of preparation for submission to the *Journal of Research in Science Teaching*.

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 an intramural research and development division of the US National Library of Medicine (NLM). LHNCBC seeks to improve access to high quality biomedical information for individuals around the world. It leads programs 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 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.

LHNCBC 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 members regularly publish their research results in the medical informatics, computer and information sciences, and engineering communities.

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 semi-annually to review LHNCBC's research projects and priorities. LHNCBC 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 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 NLM Personal Health Record (PHR) is a Web-based tool for consumers to keep track of their own and their families' health information. The goals of the PHR are to help consumers manage and understand their health care problems, to facilitate federal goals for clinical data interchange using national vocabulary standards, and to determine whether using personal health records can improve adherence to preventive care recommendations and improve consumer health.

Consumers can use the main PHR page to enter key health information, including medical conditions, surgeries, medications, allergies, and immunizations. They can also enter due dates for prescription refills and scheduling appointments. On the main PHR page consumers can also record questions they want to ask their doctor as well as medical contact information, and they can view educational material that pertains to their specific health information via information links to MedlinePlus and other trusted resources. On a separate page of the PHR, consumers can enter data for lab results, radiology reports, and other screening and diagnostic procedures. In addition, they can track measures of wellness including mood, diet, sleep, and exercise, as well as disease-specific parameters such as episodes of asthma or seizure frequency.

The PHR automatically assigns codes to the medications, observations, and problems as users enter them. These codes come from national vocabulary standards that are supported or developed by NLM, e.g., Logical Observation Identifiers Names and Codes (LOINC), RxNorm, and Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT). The strong use of vocabulary standards in the NLM PHR enables many automated features, such as personalized reminders about preventive care and healthy behaviors based on the specific data the consumer has entered.

In FY2012, LHNCBC researchers and developers continued to improve the capabilities of the PHR. On the technical side, we updated the Ruby on Rails platform to version 3.2. We updated the project infrastructure by switching to a new software version control system and organizing the software into multiple branches to better support a live system. We are well along on an upload feature to allow consumers to incorporate their medical documents (such as radiology or lab reports) in various formats and have developed functionality to track PHR usage statistically. We have also begun work on a feature to allow consumers to upload their health information from a structured Continuity of Care Document (CCD) – which meaningful use regulation requires all health care systems to provide to their patients. We implemented a new user interface that is more visually appealing than the old interface. To accommodate sightless individuals using

assistive technology (e.g. text-to-speech), we also created a completely parallel system that replaces graphical interactions with text-based navigation.

On the content side, we developed and implemented several new consumer-oriented panels including a pediatric developmental milestone tracker, Apgar record, and seizure activity log. Several more panels are in development including disease trackers for conditions such as sickle cell disease, migraine headaches, and chronic obstructive pulmonary disease. We also completely restructured the panel classification scheme to make it more intuitive and user-friendly. We reviewed and expanded the list of surgical procedures as well as the word and term synonyms. We began user testing and used user suggestions to redesign certain aspects of the PHR interface. In addition, we revised and expanded the dictionary of help text, which is available throughout the PHR.

This project addresses the long-standing NLM interest in electronic medical records systems and delivery of health care information to consumers and is closely aligned with the NLM strategic plan. It uses the nationally mandated vocabulary standards that NLM has supported, and it provides another consumer entry point to NLM's rich trove of patient-oriented data. We have continued negotiations with Suburban Hospital/Johns Hopkins Medicine for Suburban Hospital to host the PHR. Both parties have signed the Business Associate Agreement, and we are working on finalizing the software license and end user license agreements, both of which have been reviewed by the NIH Office of General Counsel. As part of these agreements, NLM would analyze de-identified data for research purposes. Early research would focus on PHR usability and usage patterns to guide the next round of development and research. Longer term, our goal is to evaluate the effect of using the PHR on consumer health behaviors.

Using Drug Databases to Assess Prescribing Practices and Continuity of Care

Medication history is an important part of patient assessment in emergency care. Studies have shown that a significant proportion of Emergency Department (ED) visits are related to adverse events of drugs. However, gathering such information from the patient is time-consuming, expensive, and, when patients are unconscious, it is infeasible. Studies have shown that patient-provided medication histories are incomplete. Suburban Hospital's ED employed Surescripts, a consortium of major Pharmacy Benefit Managers (PBM), to provide an electronic summary of a patient's full year prescription filling history. LHNCBC created a messaging interface engine, based on the open-source HL7 interface program called *Mirth*, which linked with Surescripts and delivered the prescription records for patients who had checked in for ED care. Before the system went live, Suburban Hospital collected both Surescripts data and

patient-provided history for quality assurance. Researchers obtained this information in a de-identified form, then compared the two sources of information. We found that Surescripts information, when available, significantly augmented the manual history, and covered a high proportion (88%) of a patient's current medications.

The concise prescription dispensing history report that we developed (based on the Surescripts data) is now routinely provided for patient care in the Suburban ED. Suburban Hospital ED clinicians have reported that the full year prescription history is also helpful in identifying potential problems of drug compliance, drug seeking behavior, and duplicative prescriptions.

We now have three years of Surescripts records for about 120,000 ED visits and are using this dataset to develop frequency statistics to pick the (most frequent) medications to display when a user types in a part of a medication's name (via an auto-complete algorithm) in other LHNCBC projects. We also have done preliminary analyses of the distribution of drug interaction messages and the frequency at which clinicians will be interrupted by a drug interaction messages under different choices of the severity of the interactions. A very few drugs are responsible for a very large share of such drug interaction alerts. These data also suggest that patients are often taking interacting drugs prescribed by two different practices, such that neither provider would likely get the warning about the interaction because both drugs would not be in the Electronic Medical Record of either provider. To pursue these questions further, we have obtained a database representing one year's worth of prescription information (25 million prescriptions) for a major US metropolitan area.

EMR Database Research and Natural Language System Development

We have used the MIMIC-II de-identified intensive care database under a restricted-use Memorandum of Understanding to conduct retrospective clinical studies on the significance of obesity and metabolic syndrome, interactions between feeding practices and blood transfusions in premature babies with necrotizing enterocolitis, and significance of vitamin levels in ICU mortality and post-discharge survival. *Critical Care* published our study that confirmed the "obesity paradox" (lower mortality odds after an ICU stay for overweight and obese patients compared to normal weight patients).

To facilitate the use of this database by researchers, LHNCBC staff mapped medications from multiple tables to RxNorm ingredients and radiology report headers to LOINC. We continue mapping nursing observations from MIMIC-II local codes to LOINC, RxNorm, and/or SNOMED CT as appropriate. In line with the NLM mission to facilitate access to health information resources, LHNCBC continues as a mirror site to provide access to PhysioNet, a 4.3TB database collection of physiologic waveform tracings gathered from health care

institutions world-wide by the MIT researchers who also developed MIMIC-II.

We developed natural language processing (NLP) techniques to extract important clinical variables from the narrative content of x-ray reports, discharge summaries, etc. in the MIMIC-II database. Even the best NLP techniques produce results with errors ranging from 5% to 10%., We have developed statistical methods that combine the large sets of NLP-extracted variables and the small sets of manually coded variables to improve the accuracy and tighten the confidence intervals for clinical studies that use NLP-derived variables. We have made the statistical methods, developed in R programming language, publicly available.

We continued evaluating our clinical NLP and information retrieval methods by participating in the 2012 Text REtrieval Conference (TREC), in which the task was to identify patient cohorts from a large set of de-identified clinical reports for comparative effectiveness studies. The LHNCBC team was the top-scoring team for the second year among 24 international participants from industry and academia.

Biomedical Imaging, Multimedia, and 3D Imaging

This research area has several objectives: build advanced imaging tools for biomedical research; create image-based tools for medical training and assessment; 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

We continued our collaboration with the American Society for Colposcopy and Cervical Pathology (ASCCP) in the operational use of our Teaching Tool, an image-based system to assess professional knowledge and skills in the field of colposcopy. More than 100 Obstetrics/Gynecology and Family Practice resident programs nationwide (at more than 95 universities and other premier institutions such as the Mayo Clinic) are using our Teaching Tool. To date, resident programs have used the Teaching Tool to administer more than 1,300 individual online exams of two types: Resident Online Exam (ROE) and the Colposcopy Mentorship Program (CMP). Planning is underway to administer a third (and more advanced) Colposcopy Resident Award exam through the Teaching Tool.

The National Cancer Institute (NCI) and their collaborators used another of our imaging programs, the Boundary Marking Tool, at a new site in Senegal (as well as existing sites at the University of Oklahoma Health Sciences Center, Costa Rica, Nigeria, the Netherlands, and Spain) to collect and annotate colposcopy images for biopsy studies and the creation of a worldwide database for cervical cancer research.

We also continued to use our Virtual Microscope (VM) to annotate histology images. We used expert-annotated images (from the University of Oklahoma and from Queens University, Belfast) to test our in-house developed MATLAB modules to analyze the geometry of epithelial segments in the images. The analysis includes finding rectangular regions that span the segments, segmenting the content of these regions, deriving features related to density of nuclei per unit area, and using these features to classify the segments as Normal, or as various grades of abnormality (CIN1, CIN2 and CIN3.) Work is in progress to compare these automatic classifications with "ground truth" from pathologists.

A new approach with potential great utility is to rapidly locate segments of epithelium tissue within large histology images of the uterine cervix. Our goal is to locate epithelium more accurately and one or two orders of magnitude faster than current methods. The method (1) uses compression information stored in the file to roughly separate tissue regions from background, then (2) carries out Graphical Processing Unit (GPU) processing to classify the tissue regions into epithelium and non-epithelium.

We incorporated new capability in our Multimedia Database Tool to retrieve and display histology images from the NCI ASCUS/LSIL (atypical squamous cells of undetermined significance / low-grade squamous intraepithelial lesion) Triage Study (ALTS). We also collaborated with academic researchers to develop interactive segmentation capabilities for very large images using GPUs. Developers successfully installed this capability in an in-house system equipped with two GPU processors, and used it for the segmentation of Gigabyte-sized histology images. Additional collaboration with academic groups included work toward developing high-fidelity image compression techniques for mobile platforms and work in biomedical case-based (text and image) information retrieval.

Content-Based Image Retrieval (CBIR)

CBIR is an active research area in the imaging research community because systems for image indexing, search, and retrieval all use many tools and techniques developed under this rubric. CBIR finds images (in repositories or the published literature) that are visually similar to a query image, as well as those that are semantically similar, or relevant to a query. For example, one chest x-ray might be visually similar to another, but CBIR uses "semantic similarity" to find a chest x-ray from another person who has the same lung disease.

Several practical systems and tools at LHNCBC rely on CBIR research. Our Open*i* system carries approximately 450,000 full-text publications that together include more than 1.2 million figures, including photographs, clinical images, charts, and other illustrations. The system automatically classifies the figures as "regular" images and graphical images, and then

sub-classifies"graphical" images as diagrams, statistical figures, flow charts and tables. It sub-classifies "regular" images into: x-ray, ultrasound, CT, MR, etc. We automated these classification steps by extracting more than 15 image features (color, texture, shape), then using those features in a Support Vector Machine (SVM)–based framework. We demonstrated the success of our classification methods in the international ImageCLEFmed competitions in 2011 and 2012, and then incorporated those methods in Open*i*.

We apply an additional step of extracting specific regions of interest (ROI) within images, based on the hypothesis that specific parts of an image may contain information that is more relevant to a query or concept, than would the entire image. Journal articles usually highlight these regions by "markers" or pointers – arrows, symbols – that our program must find. We have developed multiple approaches, e.g., edge or region-based segmentation, Markov Random Field (MRF) models to recognize arrow-type pointers, and a neural network-based technique to recognize asterisks. Our methods yield a precision of 85% and recall of 82%. More recent work in combining the detection of pointer boundaries and body (homogeneous pixel intensity area within the boundary) for pointer segmentation has improved the precision significantly over our previous edge-based segmentation – from an initial 25% precision to current 85% precision. Work is ongoing in improving performance for difficult cases in which pointers have exceptionally weak edges, or where the color of the pointer body is similar to that of the background.

The CBIR tasks of pointer identification and ROI extraction are combined with *textual* ROIs from figure captions or the body of the paper. Pairing visual ROIs with the corresponding textual mentions (e.g., "black straight arrow") enables the automatic indexing of the (visual) ROIs and the images of which they are a part. The tagged ROIs can then be used for image retrieval or building a visual ontology. We have developed a dynamic time warping (DTW) method to pair visual and textual ROIs. This pairing algorithm combines visual pointers with textual mentions by grouping recognized pointers by their visual characteristics first, and then searching for the best-matching pointer group with a text mention. Our experiments with ground truth text data shows that it successfully pairs over 96% of recognized true pointers with their textual mentions.

We also used CBIR to develop CervigramFinder, a research tool that automatically indexes and enables the retrieval of uterine cervix images (cervigrams) by shape, color, and texture features. Efficient searching by image features is a significant step toward locating records in large databases of cervigrams and patient data, such as the NCI's Guanacaste and ALTS databases, which contain a total of 100,000 cervigrams.

An LHNCBC project that is developing automated techniques to screen for tuberculosis and other pulmonary diseases is using CBIR to detect image features in chest x-rays. We have developed algorithms to automatically detect ribs, aorta, and other structures, and to segment lung areas. Our ongoing SVM method research aims to extract texture features to classify lungs as normal vs. abnormal.

We are also exploring use of distributed computing and GPUs for computer-intensive CBIR tasks, with a particular focus on image segmentation. Through our collaboration with Texas Tech University, we developed a method that uses GPU processing power for interactively tracing irregular object boundaries such as the separation between the epithelium and non-epithelial tissue in histology slides of the uterine cervix. We can then use the segmented regions to train classifiers to detect various stages of pre-cancer.

Interactive Publications

In this project, we investigate models for producing interactive publications that allow readers to actively manipulate images, tracings, and data in the publications – capabilities that we think will be demanded in the future. The project focuses on the standards, formats, and authoring and reading tools necessary for the creation and manipulation of such "interactive publications" that allow users to "interact with" video, audio, bitmapped images, interactive tables and graphs, and clinical DICOM images such as x-rays, CT, MRI, and ultrasound.

We have created tools for viewing and analyzing interactive publications (Panorama) and for authoring them (Forge). These tools are analogous to Adobe's Acrobat Reader and Professional for PDF viewing and editing documents. Panorama was one of nine semi-finalists out of 70 software tools for life sciences research that were entered in Elsevier's Grand Challenge contest two years ago. After we conducted a formal usability study, we recently enhanced Panorama by adding bar charts, and the capability to run natively on Mac OSX.

We also extended Panorama to provide Annotation Concepts. A Panorama user may click on text in an interactive publication which is sent to an NLM servlet (RIDeM, developed in-house) that identifies the corresponding UMLS concepts. The servlet returns an XML file to Panorama which parses it to provide the preferred UMLS term and semantic group. Annotation Concepts also provides linkouts for MedlinePlus, eMedline, Family Doctor, and other resources. Further work is ongoing to group concepts by semantic relationships, and investigators are exploring other grouping ideas.

In FY2012 we developed a Web-based version (Panorama Lite), using Adobe Flex, that eliminated the burden of having to download the Panorama software before using it. The only requirement to run this new version is to have Flash installed. Besides offering easy and intuitive usage, this client version has better line chart and graph support, and includes tables and subsets similar to the original Panorama. A feature unique to Panorama

Lite is a Map View that can present data for example, at the county, state, and/or country level, in a color-coded form.

Over the past year we have collaborated with two organizations to create interactive publications from their traditional static ones. The first was a publisher (ProQuest) and the second was a government agency (CDC's National Center for Health Statistics (CDC/NCHS)). We created two interactive papers for ProQuest from one of their open-source journals (*Sustainability: Science, Practice and Policy*). ProQuest announced the launch of these papers for public use in a press release.

For CDC/NCHS, we converted one of their key documents – *Health US In Brief* – to interactive form, and hosted it on our Web site. *In Brief* contains summary information on the health of the American people, including mortality and life expectancy, morbidity and risk factors such as cigarette smoking and overweight and obesity, access to and utilization of health care, health insurance coverage, supply of health care resources, and health expenditures. We presented this project to the NLM Board of Regents in September 2012, and received a favorable review. Planning is underway to select other CDC/NCHS documents for conversion to interactive form.

Screening of Chest X-rays for Tuberculosis in Rural Africa

In FY2012, LHNCBC continued its collaborative project with AMPATH (Academic Model Providing Access to Healthcare), an organization supported by USAID that runs the largest AIDS treatment program in the world. This project uses LHNCBC's imaging research and system development to fulfill NIH global health policy objectives. Our objective is to leverage in-house expertise in image processing to screen HIV-positive patients in rural Kenya for evidence of pulmonary tuberculosis (TB) in chest x-rays. Since chest radiography is important to the detection of TB and other pulmonary infections prevalent in HIV-positive patients, we have provided AMPATH with lightweight digital x-ray units readily transportable in rural areas. Their staff will take chest x-rays (CXR) of the population and screen them for the presence of disease. These x-ray units are already on site in the MOI University Hospital in Eldoret, Kenya and are being readied for deployment. The team completed design of vehicles to transport the x-ray units – one vehicle is being outfitted as a mobile x-ray truck.

Since the lack of sufficient radiological services in the area suggests the utility of automation to perform the screening, our in-house research effort focuses on developing software to automatically screen for disease in the CXR images. Our researchers are developing algorithms to automatically segment the lungs, detect and remove ribs, heart, aorta and other structures and then detect texture features characteristic of pulmonary disease. At present, the algorithms distinguish between two cases: abnormal vs. normal. After receiving IRB exemption, we obtained chest x-ray images to use as test and training sets:

400 from Montgomery County's TB Control Program, 850 from a source in India, and 8,200 from Indiana University. We also acquired an open-source Japanese set containing about 250 x-ray images.

Using these x-rays for training and testing, we have developed algorithms for detecting lungs and ribs focusing on region-based features such as log Gabor wavelets that exploit the orientation of anatomical structures. For lung segmentation, we have developed algorithms using region-based level sets and a novel graph-cut segmentation method, yielding an accuracy of about 95%. A robust identification of lung shape plays a role in detecting TB in CXR since many abnormalities (e.g., pleural effusion) exhibit deformation in lung shape. After extracting the lung fields, the algorithm measures various geometric features that discriminate between normal and effused cases. Ongoing work is in identifying the most successful geometric features.

In parallel, we are developing a binary SVM classifier that uses several features extracted from the x-rays as input, such as histograms of intensity, gradient magnitude and orientation, shape and curvature. Based on these input features, the SVM returns a confidence value, allowing an operator to inspect cases in which the classifier is uncertain. This initial classifier, showing an accuracy of about 80%, serves as our starting point for ongoing optimization.

Remote Virtual Dialog System (RVDS)

The Remote Virtual Dialog System (RVDS) will make the NLM "Dialogues in Science" series, currently only available in the NLM Visitor Center, available anywhere through the Internet. Support for the project is coming from stimulus funds made possible through the 2009 American Recovery and Reinvestment Act. The project involves the enhancement of programmatic capabilities of the virtual dialogue model to make it sustainable and to allow for expanded applications of the model. During FY2012, we completed the development of a voice-to-text conversion and recognition tool which is platform agnostic. Independent reviewers are providing feedback on an alpha version of the Web-based "Dialogues in Science" series. We are adding new interviews to the series and updating some of the current interviews.

Computational Photography Project for Pill Identification (C3PI)

Launched in September 2010, *Computational Photography Project for Pill Identification* (C3PI) intends to create an authoritative, comprehensive, public digital image inventory of the nation's commercial prescription solid dose medications. Ultimately our goal is to use these images to enable content-based information retrieval (CBIR) to promote patient drug-safety at the national level. Support for the project in FY2012 came from stimulus

funds made possible through the 2009 American Recovery and Reinvestment Act.

Initially working in partnership with the NLM Specialized Information Services Division and the US Veterans Administration to study content-based retrieval methods for medical image databases, researchers developed computer vision approaches for the automatic segmentation, measurement, and analysis of solid-dose medications from these pilot datasets including work on robust color classification tools to help identify prescription drugs. We are creating a collection of high resolution digital photographs of front and back surfaces of prescription tablets and capsules, 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).

In FY2012, we generated over 50,000 pictures of 1,500 samples of solid-dose pharmaceuticals from over 150 manufacturers and distributors. In addition, we are acquiring pictures from multiple cameras in a variety of lighting conditions to prepare a data collection for a broad effort or national computer vision/content-based information retrieval challenge (CV/CBIR) for the identification of medications. Staff also coordinated the deployment of a server-based repository and content management system to support distribution and curatorship of the image collection. In July 2012, NLM hosted a workshop of the Structured Product Labeling (SPL) Working Group – a collection of distributors, sponsors, and downstream users of SPL data. NLM staff presented a complete description of our image collection processes, the intended uses of the data, and the information outlets, including NLM's Pillbox, RxNav, and, in some cases, DailyMed.

In September 2012, LHNCBC hosted a workshop on the design of a CV/CBIR national challenge. Attendees included members of NIST, the director of the Face Recognition Grand Challenge (FRGC), and a representative of the FRGC academic winner (University of Houston).

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 testbed 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 in their original format or in PNG format to licensees over the Internet at no cost; and on DVD discs for a duplication fee. Almost 3,450 licensees in 64 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. More than 1,000

newspaper or magazine articles or radio programs have featured the Visible Human Project.

In FY2012, staff continued to maintain 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.

3D Informatics

The 3D Informatics Program (3DI) continued its research mission to address problems encountered in the world of three-dimensional and higher-dimensional, time-varying imaging. LHNCBC provides continuing support for image databases and continues to explore the growing need for image databases, including ongoing support for the National Online Volumetric Archive (NOVA), a collection 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. In anticipation of new sources of data from research partners contributing to the Insight Toolkit, the 3DI group is updating MIDAS software system and additional disk space. We are cultivating sources among research teams in confocal microscopy, and we are seeking collections derived from Visible Human Data including segmentations, annotations, and processed information. We continue to serve a broad community with these data, and seek to establish a leadership role through public data distribution.

Throughout FY2012, staff continued collaboration with the NCI's Laboratory for Cell Biology and with teams within LHNCBC to visualize and analyze complex 3D volume data generated through dual beam (ion-abrasion electron microscopy) and cryo-electron tomography. The work combines high performance computing with life sciences research, accelerating and empowering investigators in the detection and prevention of cancer and infectious diseases. The resulting visuals have enhanced the understanding and discoveries in the character of several immunological cells, cell structures and their interaction with pathological viruses including HIV.

In FY2012, OHPCC expanded our high resolution electron microscopy research to include processing data collected through transmission electron tomography. We are currently attempting to adapt research software that uses graphics processing units (GPUs) for high performance computing for sub-volume averaging and reconstruction. We are working to develop emerging methods into mature production software for the study of

protein structures on the surfaces of HIV and influenza virions. Additionally, LHNCBC staff helped supervise segmentation efforts for data from ion-abrasion electron microscopy to study stem-cell differentiation in murine myocytes. The unpublished results have raised more questions about the mechanisms for adult stem cell development and have suggested new studies in this critical research area.

The 3DI group also continued to investigate the use of rapid prototyping technologies in radiology. We analyzed the x-ray attenuation characteristics 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. We acquired and tested advanced software for managing 3D printing. A new set of models is under development including dosimetry models from CT scans of small animals. This work is conducted in partnership with the National Institute of Allergy and Infectious Diseases (NIAID).

Insight Tool Kit

The Insight Toolkit (ITK) is a public, open-source algorithm library for segmenting and registering high-dimensional biomedical image data. The current official software release – ITKv4.2.1 -- contains over 845,000 lines of open source code, making available a variety of image processing algorithms. ITK can be run on Windows, Macintosh, and Linux platforms, reaching across a broad scientific community that spans over 40 countries and more than 1,500 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 remains 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 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 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. Beyond the support of centers and software projects, the ITK effort has influenced end-user applications 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.

This year, LHNCBC and the ITK Project completed the release of ITK version 4 (ITKv4), ending the effort that has been supported by the American Reinvestment and Recovery Act. This program helped fund the efforts of groups including General Electric Global Research, the Mayo Clinic, Harvard University, Kitware, Inc., CoSmo Software, the University of Iowa, the University of Pennsylvania, Ohio State University, Old Dominion University, Carnegie Mellon University, Georgetown University, the University of North Carolina at Chapel Hill, and the University of Utah Scientific Computing and Imaging Institute. The research topics supported by these software development efforts include microscopy, digital histology, tumor micro-environments, zebrafish embryology, deconvolution methods for astronomy and astrophysics, image registration for neurosurgery, tumor volume measurement for lung cancer treatment, and video processing for security applications as well as healthcare.

Image and Text Indexing for Clinical Decision Support and Education

A picture is "worth a thousand words," and especially valuable for medical research and clinical practice. Scientists and the lay public can better understand complex biomedical concepts through visual means: e.g., radiographic images, photographs of organs, sketches, graphs or charts. This idea motivated us to develop the Open*i* system, which lets users search and retrieve medical citations enriched by relevant illustrations. Users may search by text queries as well as by example images.

In September 2010, the LHNCBC Board of Scientific Counselors reviewed our prototype multimedia search engine, which provided text-based search and image-feature based retrieval of other pre-indexed images visually simlar to an image in search results. In FY2012, we presented a more advanced pre-release version to the NLM Board of Regents. Both Boards recommended scaling the system to large collections and continuing research and development of high-accuracy image retrieval methods.

Since the reviews, we have: redesigned the system architecture and refactored the original code; improved the user interface; added functionality (such as filtering images based on their type, filtering journals by clinical specialty, and ranking papers by clinical content); acquired and processed a larger set of scientific publications; and incorporated distributed computing operations (Hadoop™ MapReduce) and enterprise level "bare-metal" virtualization.

The evolution of Open*i* has required the development of (1) new ways to represent images as text

strings, that could be indexed with traditional search engines such as Lucene (2) improved methods to automatically segment multi-paneled illustrations into single images, and to partition their captions to correspond to single images (3) improved methods to extract pointers (arrows, arrowheads, symbols) within images to identify regions of interest. In other work, we have begun to build a visual ontology and we have also developed methods for segmenting lung and brain tissues and extracting key features such as lung cysts, micronodules, and emphysema.from those images.

Open*i* is the first production-quality system of its kind in the biomedical domain, and it gives medical professionals and the public access to visual information from biomedical articles that are highly relevant to their questions, as well as a brief synopsis of the articles. Open*i* consistently ranked among the best in international competitions at ImageCLEF, Open*i* demonstrated the best retrieval results and placed third in image type classification in a 2012 medical image retrieval evaluation that attracted competitors from academia, industry and clinical settings.

Toward the end of FY2012 we released Open*i* to the public. It now provides access to over 1,200,000 figures extracted from more than 450,000 articles in the Open Access subset of PubMed Central. Usage grows by the week and now exceeds 11,000 unique visitors a day.

Turning The Pages (TTP)

The goal of the Turning The Pages (TTP) project is to provide the lay public a compelling experience of historically significant and normally inaccessible books in medicine and the life sciences. In this project, we build 3D models for books and develop animation techniques to allow users to touch and turn page images in a photorealistic manner on touch-sensitive monitors in kiosks at NLM, as well as "click and turn" in an online version. We have also built a 3D "scroll" model for the 1700 BC Edwin Smith medical papyrus which is "touched" (or clicked) and "rolled out." The online version of TTP is a popular Web site, attracting more than a million page views a month.

The kiosk now presents all eleven books; the Web version offers ten, and the iPad offers six. We are working to make all the books available in all three formats.

This year we added two more books to the iPad version, and we released Andrew Snape's *Anatomy of An Horse (sic)* in all three versions. Snape's comprehensive and beautifully illustrated book was originally published in London in 1683. We are currently creating the iPad and Web versions of Elizabeth Blackwell's *A Curious Herbal*, previously available only in the kiosk version. We have recreated all the graphics for Blackwell's book and updated the animation, and plan to launch the iPad and Web versions shortly. We have also added new features to the iPad version such as bookmarks and contextual zoom for curators' notes.

To reduce project costs, we have modified our production pipeline to capture the images of new books in-house rather than using commercial scanning companies. We also are investigating newer technologies, such as multi-touch monitors, for the next generation of the TTP kiosks.

Natural Language Processing and Text Mining

Medical Article Records System

NLM's flagship database, MEDLINE®, contains more than 20 million bibliographic records for articles from over 5,500 biomedical journals. To meet the challenge of producing these citations in an affordable way, researchers at LHNCBC develop automated techniques to extract bibliographic data (abstract, author names, affiliations, etc.) from both scanned paper journals and online journals.

While the bulk of citations come to NLM directly from publishers (in XML format), there are still approximately 820 journal titles that arrive in paper form. These papers are processed by the MARS production system, in operation for some years. MARS combines document scanning, optical character recognition (OCR), and rule-based and machine learning algorithms to extract citation data from paper copies of medical journals needed to complete bibliographic records for MEDLINE. Our algorithms extract this data in a pipeline process: segmenting page images into zones, assigning labels to the zones signifying its contents (title, author names, abstract, etc.), and pattern matching to identify these entities.

LHNCBC manages and continually improves the MARS system. For example, we are introducing three new features to improve MARS performance: (1) expansion of the MEDLINE character set, (2) capability for the Edit operators to correct errors made by the automated zoning process, and (3) a new user interface design for large screen monitors. The software implementation and integration test for the first two features have been completed, and are being deployed to the production system.

Citations that do come to NLM in electronic form from publishers often contain errors or have missing content. Missing items include: databank accession numbers (e.g., GenBank), NIH grant numbers, grant support categories, Investigator Names, and Commented-on Article information. The capture of Investigator Names can be especially difficult because some articles contain hundreds of such names and capturing the articles commented on by a paper requires operators to open and read other articles related to the one being processed. To automatically extract these fields from online articles, we have developed the Publisher Data Review System (PDRS) whose subsystems are based largely on machine learning algorithms such as Support Vector Machine.

PDRS was put in production in early FY2012, for open-access articles in NLM's PubMed Central. To extend automated data extraction to *all* online journals on

publisher's sites, including ones with restrictive copyrights, we are developing IMPPOA (*In-Memory Processing for Publisher Online Articles*), a system based on the PDRS platform and its machine learning algorithms, but implemented to process articles in RAM memory rather than downloading the articles to disc, we expect that this approach will eliminate publisher's concerns about copying articles into an external system disc. IMPOAA: (1) provides data missing from the XML citations sent in directly by publishers, (2) corrects errors in publisher data by extracting data from the articles on their sites and comparing these with the data sent to NLM, and (3) extracts data from articles for which publishers do not send in citations at all.

The systems outlined above rely on underlying research in image processing and lexical analysis which also enables the creation of new initiatives in which these techniques find application, such as the ACORN project.

Automatically Creating OldMedline Records for NLM

There has been a long-standing interest in expanding MEDLINE to include bibliographic records dating back to the late 19th century when Index Medicus was first developed, so we would have a complete electronic record of all citations in Index Medicus. These early citations appear in printed indexes published before 1960, and Library Operations (LO) has collected many of these with considerable manual effort. To automate this process, we have designed the ACORN system – combining scanning, image enhancement, OCR, image analysis, pattern matching, and related techniques to extract unique records from the printed indexes. These tasks are formidable considering the old typefaces and fonts (as well as a mix of different languages) in the indexes – producing a substantial degree of OCR errors. To overcome this problem in one of the indexes (Quarterly Cumulative Index Medicus, or QCIM), we developed a novel pattern matching technique that automatically finds and compares two versions of every citation from the subject and author listings, thereby minimizing the OCR errors encountered in each version. In addition, our system is designed to search MEDLINE as well as sources on the Web to avoid duplicating records that already exist elsewhere, and to further reduce OCR errors. Our system, designed to process indexes appearing both in paper form as well as on microfilm, has three main components: Quality Control, Processing, and Reconcile (for operator verification), which are currently being developed for demonstration in FY2013.

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 uses the MeSH headings from the PubMed® related articles which are precomputed by PubMed. Results from the two basic methods are combined into a ranked list of recommended indexing terms, incorporating aspects of MEDLINE indexing policy in the process.

The MTI system is in regular and increasing use by NLM indexers to index MEDLINE citations. MTI recommendations are available to them as an additional resource through the Data Creation and Maintenance System (DCMS). Because of the recent addition of subheading attachment recommendations, indexers now have the option of accepting MTI heading/subheading pairs in addition to unadorned headings. Our developers also created versions of MTI to assist in indexing the NLM History of Medicine book collection and for general cataloging. Due to its success with certain journals, MTI was designated as the first line indexer for 23 journals totaling 4,205 articles in FY2012. As a first line indexer (MTIFL), MTI indexing is still subject to the normal manual review process. The number of MTIFL journals will grow gradually and should prove to be a time and money saver for NLM. The Indexing Initiative (II) team and Library Operations (LO) are currently working together on developing ways to evaluate the performance of journals in the MTIFL program and to help identify future MTIFL journal candidates.

The II team worked closely with an NLM Associate Fellow whose ongoing project was designed to investigate the feasibility of automating the creation of functional annotations about genes, known as Gene Reference into Function (geneRIF). We have developed a prototype, the Gene Indexing Assistant (GIA), and integrated it into the Data Creation Management System used by the indexers for testing and evaluation.

The II team is also collaborating with LO to develop a new system designed to aid in indexing *Technical Bulletin* which is published by LO. The goal is to provide keywords for all past and future *Technical Bulletins* (1997 forward). The keywords will be automatically assigned by the program and manually reviewed prior to publication.

MetaMap is a critical component of the MTI system and used worldwide in bioinformatics research. Recent work has improved processing speed significantly, added XML (eXtensible Markup Language) output, implemented negation identification, and enabled users to supply their own acronyms/abbreviations list. MetaMap is available on Windows, Macintosh and Linux platforms. Users can build their own data sets with the MetaMap Data File Builder and access their local version of MetaMap via either an embedded Java API (Application Programming

Interface) or UIMA (Unstructured Information Management Architecture) wrapper. In FY2012, users downloaded approximately 1,700 copies of MetaMap, in the Java API and/or the UIMA Wrapper form. Of note, MetaMap is one of the NLM resources integrated into IBM's Watson system for healthcare applications.

Digital Preservation Research

The long-term preservation of documents in electronic form, both born-digital as well as those resulting from scanning paper, is a mandate for NLM as it is for other major libraries and archives. The goal of this LHNCBC project is to investigate and implement techniques for key preservation functions, including: automatically extracting metadata to enable future access to the documents, ingesting the documents and metadata into a storage system, and knowledge discovery from the archived material. To provide a platform for this research, we have built and deployed a *System for Preservation of Electronic Resources* (SPER). SPER builds on open-source systems and standards (e.g., DSpace, RDF) while incorporating inhouse-developed modules that implement key preservation functions.

We are focusing on two collections. One is a historic medico-legal collection of early 20th century court documents acquired from the FDA. NLM curators are using SPER to preserve the FDA documents, numbering 67,000 in total, and in 2012 they processed over 22,000 of these. These documents and their metadata are in a publicly accessible NLM Web site. The other collection, from NIAID, is a set of conference proceedings of the "US-Japan Cooperative Medical Science Program on Cholera" (CMSP), an international program conducted over a 50-year period from 1960 to 2011. Our CMSP document preservation activities include: (a) building a full repository for 2,800 research articles on cholera and 8,000 references on CMSP participants such as authors, panelists, attendees and Study Section reviewers, followed by (b) developing a portal for the public to search for research articles, institutes, authors and other participants from this dataset.

We have developed automated metadata extraction (AME) techniques to identify and extract three different types of metadata from the CMSP documents, namely: publication metadata with titles, authors and their affiliated institutions from research articles; investigator metadata with name, role, designation and affiliation of each person from the conference proceedings rosters; and Study Section metadata with names and affiliations of CMSP program reviewers from separate Study Section rosters. The AME processes include (a) layout analysis to recognize different types of information within a document set; (b) evaluation of the effectiveness of models such as Support Vector Machine and Hidden Markov Model for different metadata layouts; and (c) capture of relationships among various entities in the collection from the extracted metadata of different types. In 2012, we added two more

conference proceedings, for the years 2010 and 2011, to the CMSP repository. Furthermore, we used the extracted metadata to implement data analysis function for the CMSP document corpus to discover patterns and trends on factors such as important drugs, discoveries, investigators as well as international collaborations under the CMSP program over its 50 year span.

In addition, we are conducting research toward knowledge discovery from information preserved in this repository by (a) developing a domain-specific vocabulary, (b) generating RDF graphs or triples from the preserved information using this vocabulary and natural language processing techniques, and (c) building the corresponding knowledgebase. We are testing an external tool named LymbaGrid to develop the cholera vocabulary and generate the knowledgebase from the CMSP document corpus.

RIDeM/InfoBot

As part of the Clinical Information Systems effort, the RIDeM (Repository for Informed Decision Making) project seeks to automatically find and extract the best current knowledge in scientific publications. The knowledge is provided to several applications (Open*i* – a multimodal literature retrieval engine, Interactive Publications, and InfoBot) through RESTful Web services.

The related InfoBot project enables a clinical institution to automatically augment a patient's electronic medical record (EMR) with pertinent information from NLM and other information resources. The RIDeM API developed for InfoBot allows integrating patient-specific information (e.g., medications linked to formularies and images of pills, evidence-based search results for patient's complaints and symptoms, or MedlinePlus information for patient education) into an existing EMR system. For clinical settings that have no means to use the API, a Web-based interface allows information requests to be manually entered.

The InfoBot API integrated with the NIH Clinical Center's EMR system, CRIS, is in daily use since July 2009 through the *Evidence-Based Practice* tab in CRIS During the past year, the *Evidence-Based Practice* tab was accessed 28 times a day, on average, by over 890 returning NIH CC users.

In FY2012, we started extending RIDeM services to answer consumer health questions submitted to the NLM customer services that handle about 90,000 requests a year. We are developing a prototype Consumer Health Question Answering (CHQA) system to facilitate answering the requests. The prototype can classify the incoming requests as, for example, questions about health problems or requests to correct MEDLINE citations. Once the request type is recognized, CHQA provides information needed to answer the request. For MEDLINE quality assurance requests, the system automatically finds and retrieves the citation that set off the request, and then extracts information that helps answer the question, for

example, the citation publication status. The prototype is capable of understanding simple frequently asked questions (such as, "Is Parkinson's disease hereditary?") and finding sections of the NLM Genetic Home Reference articles that answer these questions.

De-identification Tools

De-identification enables research on clinical narrative reports. We are designing a de-identification system that will remove protected health identifiers from narrative clinical reports according to the provisions of the Privacy Rule that implementations portions of the Health Insurance and Accountability Act. 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.

We completed a version of the software system to be tested at the NIH Clinical Center. This version of the system is designed to de-identify clinical narrative text in the form of Health Level 7 (HL7) version 2. It is capable of utilizing information embedded in various HL7 fields as well as externally provided information such as name lists of the health care providers at NIH.

We designed the Visual Tagging Tool (VTT) - an editor for visualization and markup, which we use to produce gold standards against which to test the CTD system. Although it is designed specifically to facilitate and speed up manual tagging of identifiers that contain protected health information (PHI), we have made it publicly available to the greater NLP community for any kind of lexical tagging and text annotation. (http://lexsrv3.nlm.nih.gov/LexSysGroup/Projects/vtt/current/web/index.html)

We also are preparing a large corpus of clinical reports to serve the project as the gold standard based on VTT for de-identification. As of December 2012, the corpus comprises 20,994 clinical reports of 7,571 patients, where every piece of individually identifiable health information has been marked manually. We divided the set of annotated reports into a training set and a test set—the latter of which will be used to evaluate the overall de-identification performance of our system.

Librarian Infobutton Tailoring Environment (LITE)

Infobuttons (http://www.infobuttons.org) are context-aware links from one information system to another that anticipate users' information needs, take them to appropriate resources, and assist with retrieval of relevant information. To date, infobuttons are mostly found in clinical information systems (such as EHRs and PHRs) to provide clinicians and patients with access to literature and other resources that are relevant to the clinical data they are viewing. The NIH Clinical Center Laboratory for Clinical Informatics Development has worked with HL7 to develop an international standard to support the communication between clinical systems and knowledge resources. MedlinePlus Connect currently provides an HL7-compliant query capability.

In order to increase the usefulness of infobuttons, they are typically linked not to a specific resource, but instead to an "infobutton manager" that uses contextual information (such as the age and gender of the patient, the role of the user, and the clinical data being reviewed) to select from a large library of known resources those that seem most applicable to the situation. The infobutton manager customizes the links to those resources, using appropriate data from the context, and presents the user with the list of custom-selected, customized links. NIH is working with investigators at the University of Utah and the Veterans Administration to establish a freely available, HL7-compliant infobutton manager, known as "Open Infobutton" (http://www.openinfobutton.org) to be a national resource for EHR developers and users, providing all clinical systems users with the capability of integrating knowledge at the point of care.

Infobutton managers require knowledge bases to enable them to perform their customization work; Open Infobutton is no exception. The knowledge in these knowledge bases is very institution-specific, including the applications that might call the infobutton manager, the types of questions users might have, and the resources available for resolving those questions at the particular institution (local documents, site licenses, etc.). The Librarian Infobutton Tailoring Environment (LITE), is a user-friendly tool that can be used by an institution's medical librarians (or someone acting in that role) to provide Open Infobutton with the necessary knowledge for it to customize its responses to requests from that institution. The system is currently in alpha testing now in an installation at the University of Utah (http://lite.bmi.utah.edu). In 2012, we made LITE officially available to users of the OpenInfobutton system, with paper presentations and demonstrations at the American Medical Informatics Association.

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.

The Lexical Systems Group is nearing completion of a project to enhance the derivational-variants function of the lexical tools. Derivational variants are words related by a word-formation process like suffixation, prefixation or conversion (change of category). Earlier versions of the derivational variant system had only suffix rules and facts. These rules and facts are hand entered and curated. In order to add prefixation and conversion functionality to the system, the PDM team has developed a method to automatically extract candidate pairs of words that may be

derivationally related, which helps automate the creation of rules and facts for suffixation and conversion. The 2011 release of the Lexical Tools included around 4,500 derivational facts; the 2012 release had nearly 90,000 and the 2013 release will have over 121,000 derivational facts. These new derivational facts now include information on negative derivations and will be used to infer derivational rules.

During the year, our Web page had an average of 1,999 unique visitors per month. We had an average of 4,800 downloads per month in 2012. We provided support to 35 internal users, 14 US domestic users and 8 international users.

The 2013 release of the SPECIALIST Lexicon will contain over 469,992 records, representing over 857,502 forms, an increase of over 7,863 records from the 2012 release. Many of the new terms are derived from de-identified clinical records from our own De-identification project and from the MIMIC II database (described in another section).

Medical Ontology Research

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 EMR interoperability.

During FY2012, staff investigated the validation of value sets used in clinical quality measures as part of the Meaningful Use incentive program, in collaboration with the Office of the National Coordinator for Health Information Technology (ONC) and the Centers for Medicare & Medicaid Services (CMS), contributing to the creation of the Value Set Authority Center at NLM. We continue to explore the use of Semantic Web technologies for representing the UMLS and semantic predications, with application to literature-based discovery. Finally, we pursued our work on quality assurance in biomedical terminologies by investigating LOINC using description logic tools.

Research activities this year resulted in two journal articles, four papers in conference proceedings, one book chapter and five 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). We also co-chaired the scientific program committee of the Eighth International Conference on Data Integration in the Life Sciences (DILS 2012).

Semantic Knowledge Representation

The Semantic Knowledge Representation (SKR) project conducts basic research in symbolic natural language processing, based on the UMLS knowledge sources. A core resource is the SemRep program, which extracts semantic predications from text. SemRep was originally developed for biomedical research. Researchers are developing a general methodology for extending its domain to influenza epidemic preparedness, health promotion, and health effects of climate change.

The SKR project maintains a database of 60 million SemRep predications extracted from all MEDLINE citations that is available to the research community. This database supports the Semantic MEDLINE Web application, which integrates PubMed searching, SemRep predications, automatic summarization, and data visualization. The application helps users manage the results of PubMed searches by outputting an informative graph with links to the original MEDLINE citations and by providing convenient access to additional relevant knowledge resources, such as Entrez Gene, the Genetics Home Reference, and UMLS Metathesaurus. The Semantic MEDLINE technology was recently adapted for analyzing NIH grant applications, allowing NIH portfolio analysts to track emerging biomedical research trends and identify innovative research opportunities.

SKR efforts support innovative information management applications in biomedicine, as well as basic research. The project team is using semantic predications to find publications that support critical questions used during the creation of clinical practice guidelines (with support from NHLBI). Investigators are devoting significant effort to developing and applying the literature-based discovery paradigm using semantic predications. One such project is investigating the physiology of sleep and associated pathologies, such as declining sleep quality in aging, restless legs syndrome, and obstructive sleep apnea; another exploits predications and graph theory for automatic summarization of biomedical text. Further, the SKR team is collaborating with academic researchers in using semantic predications to help interpret the results of microarray experiments, to investigate advanced statistical methods for enhanced information management, to support formal models of knowledge representation, and to address the information needs of clinicians at point-of-care.

Information Resource Delivery for Researchers, Care Providers, and the Public

The LHNCBC performs extensive research in developing advanced computer technologies to facilitate the access, storage, and retrieval of biomedical and consumer health information.

ClinicalTrials.gov

ClinicalTrials.gov provides the public with comprehensive information about registered interventional and observational clinical research studies. ClinicalTrials.gov receives over 95 million page views per month and hosts approximately 900,000 unique visitors per month.

Data are submitted to ClinicalTrials.gov through a Web-based Protocol Registration System (PRS) by nearly 12,000 study sponsors including the US Federal government, pharmaceutical and device companies, academic, and international organizations. At the end of FY2012, the site had records for nearly 134,000 research studies, conducted in all 50 states and in over 180 countries. Approximately one-third of the studies are still open to recruitment. For the remaining two-thirds, the recruitment phase is over or the study has been completed. Over 7,400 of the closed studies display summary results tables describing primary and secondary outcomes, adverse events, and characteristics of the participants studied. In FY2012, new registrations were submitted at an average rate of 370 records per week, an increase of 6% from FY2011. The average rate of new results submissions was approximately 70 per week, an increase of 17% from FY2011. The continued growth in the use of ClinicalTrials.gov can be attributed to US laws as well as international recognition of the scientific and ethical importance of registration and results reporting. The combined registry and results database provides access to critical information about ongoing and completed clinical research for patients, healthcare providers, and policy decision makers.

NLM established ClinicalTrials.gov in FY2000 – to comply with requirements of the Food and Drug Administration Modernization Act of 1997, and to support NLM's mission of disseminating biomedical knowledge and advancing public health. Since that time, ClinicalTrials.gov has expanded significantly to support other international registration policies, such as the International Committee of Medical Journal Editors (ICMJE) policy requiring prospective trial registration as a condition of publication. In FY2012, ClinicalTrials.gov continued to expend significant effort on implementing and educating the public on the most recent law, Section 801 of the Food and Drug Administration Amendments Act of 2007 (FDAAA 801). NLM has been working with the NIH Office of the Director and other NIH Institutes and Centers as well as the Food and Drug Administration (FDA) on a Notice of Proposed Rulemaking (NPRM). When the NPRM is published, it will further elucidate the requirements of FDAAA 801 and request public comment on key implementation issues.

In FY2012, ClinicalTrials.gov staff focused on enhancing educational and outreach materials available to the public, as well as to sponsors and investigators affected by FDAAA 801. A key component of this strategy was the launch of a redesigned Web site with a new graphic design and reorganized written content, which was informed by user feedback, usability studies, and the need to better integrate educational materials about data submission and FDAAA 801. Another key component was targeted education and outreach on the results database and submission requirements. This included hosting two on-site workshops for personnel at academic research institutions and presenting at conferences, participating in

Webinars, and publishing in trade and peer-reviewed journals. Research projects in FY2012 investigated issues related to the ethical and scientific oversight of clinical research and the quality and timing of results reporting. Additionally, new tools were introduced in the PRS in FY2012 to support more efficient data entry and to expand resources available for uploading content available in other databases, such as NCI's Clinical Trial Registration Program. ClinicalTrials.gov also continued to provide technical advice and collaborate with other clinical study registries, professional organizations, funders, and regulators in working towards the development of global standards for trial registration and reporting to results databases. For example, a key activity was working with the European Medicines Agency (EMA) on developing a common set of data elements for results submission to both ClinicalTrials.gov and the EMA results database, which is being developed for possible release in 2013.

Genetics Home Reference (GHR)

Genetics Home Reference (GHR) is an online resource that offers 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 patients and their families, the GHR Web site currently includes user-friendly summaries of more than 1,900 genetics topics, including more than 800 genetic conditions, about 1,040 genes, 77 gene families, all the human chromosomes, and mitochondrial DNA. The Web site also includes a handbook called *Help Me Understand Genetics*, which provides an illustrated introduction to fundamental topics in human genetics including mutations, inheritance, genetic testing, gene therapy, and genomic research. In 2012, a new section on the Encyclopedia of DNA Elements (ENCODE) Project was added to the handbook.

Genetics Home Reference celebrated its ninth anniversary in 2012. In the past year, the project expanded its genetics content for consumers, adding 260 new summaries to the Web site. We intend to continue this rate of production in FY2013, covering additional Mendelian genetic disorders as well as more complex disorders. The team also plans to add more background information to *Help Me Understand Genetics*, including pages on next-generation DNA sequencing and the diagnosis and management of genetic disorders. In FY2012, the site averaged almost 28,000 visitors per day and about 27.6 million hits per month. 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.

Staff presented the Web site to several visiting groups, including health and science journalists as part of the Association of Health Care Journalists – NLM

Fellowship program. In addition, staff attended several major genetics conferences. At the annual meeting of the American Society of Human Genetics (ASHG), we presented a poster titled, "Genetics Home Reference Ten Years In: Where We Are Now."

Profiles in Science Digital Library

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

The content of *Profiles in Science* is created in collaboration with the History of Medicine Division of NLM, which processes and stores the physical collections. Several collections have been donated to NLM and contain published and unpublished materials, including manuscripts, diaries, laboratory notebooks, correspondence, photographs, poems, drawings and audiovisual resources. The *Profiles in Science* collections are consistently popular. The Web site averages over 60,000 unique visitors each month.

This year, the papers of pioneering surgeon Henry Swan were added to *Profiles in Science*. Staff added 402 transcripts of documents, to make handwritten items searchable and to provide alternatives to PDF format files. Staff also added 10 digital items to the 35 existing *Profiles in Science* collections. Currently, 142,180 image and 27,042 digital items are available on *Profiles in Science*. The Web site now features the archives of 33 prominent scientists and health advocates:

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 updating the *Profiles in Science* collections during FY2012, LHNCBC staff replaced the backbone of the architecture of the *Profiles in Science* Web site with the open-source Apache Solr enterprise search platform. We also continued to enhance documentation and migrate modules of the *Profiles in Science* digital library data creation software to eliminate dependence on unsupported software while improving reliability and security. We increased the visibility of the digitized items throughout the *Profiles in Science* Web site and search results through the use of preview images. We provided the research community insight into the digital library underlying the *Profiles in Science* Web sites through publication of "The *Profiles in Science* Digital Library: Behind the Scenes" in JCDL '12: Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries.

Evidence Based Medicine - PubMed for Handhelds

Developed and released in FY2003, PubMed for Handhelds facilitates evidenced-based medical practice with MEDLINE access at the point of care via smartphones, wireless tablet devices, netbooks or portable laptops. PubMed for Handhelds (PubMedHh) 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 abstracts allows a clinician to quickly read summaries at the point of need. A "consensus abstracts" element provides rapid review of multiple publications with smartphones at the point-of-care. This corresponds well with a recent review of PubMedHh server logs that showed that more than 90% of queries were clinical in nature. Randomized controlled trials using simulated clinical scenarios concluded recently at the Uniformed Services University, University of Botswana-University of Pennsylvania and the National Telehealth Center and Philippine General Hospital, Manila to evaluate the usefulness of abstracts in clinical decision making.

PubMed for Handhelds is available as an iOS (iPhone/iPad) app and an Android app. As of early December, the iOS app has been downloaded more than 25,000 times. For some search tools (PICO and askMEDLINE) queries from smartphone apps now account for 90% of queries.

Clinical Vocabulary Standards and Associated Tools

Multiple projects in this area continue to promote the development, enhancement, and adoption of clinical vocabulary standards. The CORE Problem List Subset of SNOMED CT is published in the UMLS as a specific content view. RxTerms facilitates the use of RxNorm as an interface for medication orders. Inter-terminology mapping promotes the use of standard terminologies by creating maps to administrative terminologies, which allows re-use of encoded clinical data. The Newborn Screening Guide combines terminology and electronic messaging systems

to facilitate care and research related to newborn screening. Another effort focuses on the development of a consumer-friendly medical problem and procedure terminology. LHNCBC continues to play an important role in the UMLS project in research related to the various UMLS knowledge sources and providing support in UMLS production and user support. The inter-terminology maps are also available through the UMLS.

During FY2012, we reviewed the codes contained in the National Quality Forum (NQF) eMeasures value sets, a crucial component of the Federal Meaningful Use regulations, to determine which codes were incorrect (did not exist as written or were obsolete), associated with incorrect terminology, or associated with an incorrect definition. We also began reviewing content of the NQF eMeasures for clinical validity and compatibility with clinical workflows.

The CORE Problem List Subset of SNOMED CT

SNOMED CT is a comprehensive, multi-lingual medical terminology for anatomic sites, organisms, chemicals, diagnoses, symptoms, findings, and other such concepts. The problem list is an essential part of the Electronic Health Record (EHR). The adoption of a common standard for the codes in this list prevents duplication of effort and promotes data interoperability. The 2014 US EHR certification requirements require the use of SNOMED CT to code the problem list.

Based on the analysis of the problem list vocabularies and their usage frequencies in seven large-scale US and overseas healthcare institutions, we identified a subset of the most frequently used problem list terms in SNOMED CT and we published it as the CORE (Clinical Observations Recording and Encoding) Problem List Subset of SNOMED CT. The CORE Subset can be a starter set for institutions that do not yet have a problem list vocabulary and this will save significant development effort and reduce variations between institutions. Existing problem list vocabularies can also be mapped to the CORE Subset to facilitate data interoperability.

Since its first publication in 2009, the CORE Subset has received considerable attention from the IHTSDO (International Health Terminology Standards Development Organization), the SNOMED CT user community, EHR software vendors and terminology researchers. It has been installed in various EHR products, and used as a focus for SNOMED CT-related research, mapping projects and quality assurance. The MedlinePlus Connect Project, which facilitates online linkage to patient education information, has mapped the CORE Subset to MedlinePlus health topics. In 2012, a clinical dataset from the Veterans Administration covering over 3 million patients was utilized to enrich the CORE Subset. It required the addition of a moderate number of new concepts (about 300) to cover the most frequently used terms in the new dataset, further lending support to the proposition that a relatively small number of SNOMED

CT concepts is sufficient to cover a high percentage of usage in most institutions. The CORE Problem List Subset is updated 4 times a year to synchronize with changes in SNOMED CT and the UMLS. The CORE Subset currently contains about 6,000 concepts.

Mapping between SNOMED CT and ICD codes

International Classification of Diseases (ICD) codes are required for public health reporting of population morbidity and mortality statistics. In the US, ICD-9-*CM* (the "Clinical Modification") is also used for reimbursement (soon ICD-10-*CM* will be required for this purpose). Because of this need, many existing EHR systems are still using ICD-based vocabularies to encode clinical data. However, ICD was not designed to capture information that is detailed enough to support clinical care. SNOMED CT is a much better clinical terminology and its use will be required to achieve "Meaningful Use" as defined in CMS regulations. To encourage the migration to SNOMED CT, and to enable EHRs to output ICD codes for administrative purposes, various maps between SNOMED CT and the ICD classifications have been developed. We published a SNOMED CT to ICD-10-*CM* rule-based map, covering 15,000 SNOMED CT concepts. This map allows users to encode patient problems in SNOMED CT terms, and then generate the appropriate ICD-10-*CM* codes in real-time for billing or other purposes. To demonstrate the use of the map, we developed the I-MAGIC (Interactive Map-Assisted Generation of ICD Codes) demo tool.

For an international project, in collaboration with the IHTSDO and the World Health Organization (WHO), we developed an analogous rule-based map between SNOMED CT and ICD-10 covering 19,000 SNOMED CT concepts. We adapted the I-MAGIC tool to showcase this map to ICD-10 as well. In a separate project, to help convert legacy ICD-9-*CM* encoded clinical data into SNOMED CT codes, we produced another map from ICD-9-*CM* to SNOMED CT.

We are currently planning to create maps between SNOMED CT and ICD-9-*CM*/ICD-10-*PCS* procedure codes, since SNOMED CT is also designated as the terminology standard for coding clinical procedures in Phase 2 of the "Meaningful Use" incentive program for electronic health records.

RxTerms

RxTerms is a free, user-friendly, and efficient drug interface terminology that links directly to RxNorm, the national terminology standard for clinical drugs. CMS used RxTerms in one of their pilot projects in the post-acute care environment. RxTerms is also used in the NLM PHR, and at least one EHR from a major medical institution in Boston.

RxTerms content and features are now being bundled with the RxNorm standard (the US standard drug

reference terminology that is required by meaningful use regulations). During FY2012, we aligned the data model of RxTerms and RxNorm by creating a new term type in RxNorm to cover the drug-route combination. We also worked with RxNorm on the federal vocabulary standard for identifying prescribed medications and ingredients, especially as needed for identifying drug allergies.

RxNav

Released in September 2004, RxNav was first developed as an interface to the RxNorm database and was primarily designed for displaying relations among drug entities. In addition to the browser, we created SOAP-based and RESTful application programming interfaces (APIs), to let users integrate RxNorm functions into their applications. Examples of such functions include mapping drug names to RxNorm, finding the ingredient(s) corresponding to a brand name, and obtaining the list of National Drug Codes (NDCs) for a given drug.

During FY2012, staff aligned RxNorm with the Anatomical Therapeutic Chemical (ATC) classification system developed by the World Health Organization (WHO) Collaborating Centre for Drug Statistics Methodology and widely used in Europe. Many RxNorm drugs are now linked to drug classes in ATC. Similarly, we also linked RxNorm drugs to pharmacologic actions from the Medical Subject Headings (MeSH), and enabled the remapping of obsolete drug codes to the current version of the drug codes.

A major effort this year was the development of RxMix, a graphical interface allowing users to create workflows – complex sequences of API functions – and to execute them on single values or on list of values, in batch. A typical use is the determination of the list of clinical drugs in RxNorm that have a given property in NDF-RT, for example the list of statin drugs, of penicillins. Ability to establish such list is critical for applications, such as allergy checking and clinical decision support.

We have integrated two other drug information sources with RxNav: RxTerms and NDF-RT, a resource that links drugs to their pharmacologic classes and properties, including indications, contra-indications and drug-drug interactions. Usage of RxNav, and the SOAP and RESTful application programming interfaces (APIs) for RxNorm, RxTerms and NDF-RT, received a combined total of about 50 million queries during FY2012 (a significant increase from 40 million queries last year). Users include clinical and academic institutions, as well as pharmacy management companies, health insurance companies, EHR vendors, and drug information providers. Developers of mobile apps have also started to integrate our APIs into their applications.

LOINC Standards for Identifying Clinical Observations and Orders

Federal Meaningful Use (MU) 2 electronic health records regulations require LOINC codes in lab result messages sent to ordering clinicians (within medical record systems, patient summaries and reports to public health). In FY2012, LHNCBC continued to work with the Regenstrief Institute, major laboratory companies, several NIH institutes, and other organizations to expand the size and breadth of the LOINC database. By the end of FY2012, LOINC had nearly 20,000 users in 148 countries and was translated into 11 languages and dialects. Users can pick any of these languages, search for words in the chosen language, and see the matching LOINC terms in that language plus English. To further expand LOINC globalization, we enabled language-specific Web pages in the LOINC Web browser (prompts and button labels in native language) – so far, the Web browser is completely understandable in two languages besides English (simplified Chinese and Italian).

We worked with Regenstrief and the LOINC Committee to create more than 5,500 new LOINC terms for both laboratory and clinical variables, and the LOINC database now contains nearly 70,000 terms. We released new terms for radiology (RAD), Neuro-QOL, OPTIMAL survey (American Physical Therapy Association), NEMSIS, the CARE long term care hospital (LTCH) survey and core terms for behavior for the Substance Abuse and Mental Health Services Administration (SAMHSA). During FY2012, we also edited existing molecular genetics terms to harmonize with Human Genome Organization (HUGO), Human Genome Variation Society (HGVS), and International System for Human Cytogenetic Nomenclature (ISCN) recommended nomenclature.

We worked with four of the eight largest international laboratory instrument vendors to help map or check the mapping of their internal instrument codes to LOINC codes, in order to facilitate electronic reporting of lab results. All eight such vendors now assert they provide LOINC codes for all the test codes their instruments can generate. We also worked with many smaller vendors to find (or create new) LOINC codes to describe the results of their test kits or instruments to fit their needs.

We continued to meet with other NIH organizations that are developing assessment instruments with the goal of closer alignment among NIH standard element development. We are collaborating with other NIH organizations (and Regenstrief Institute) to structure their assessment instruments and registry system values into the LOINC format and incorporate them into the LOINC database - a common framework that includes many kinds of clinical and research variables. We are serving on the Common Data Elements (CDE) Working Group to the trans-NIH BioMedical Informatics Coordinating (BMIC) Committee. We are working with the National Eye Institute (NEI) to restructure its packages

of assessment instruments for the National Ophthalmic Disease Genotyping Network (eyeGENE®), and with the NIH/NCATS Office of Rare Disease to revise their CDEs—and we plan to create corresponding LOINC codes. Staff is also working on a project with colleagues from multiple NIH components to develop CDEs for hemoglobinopathies using standard terminologies such as LOINC and SNOMED CT. We are also working with the National Institute for Neurological Diseases and Stroke (NINDS) on the NINDS CDEs and the Neuro-QOL measures.

Staff enhanced LOINC's support tools and databases, increased database field size to enable support for new LOINC numbers, and deleted 12 obsolete fields.

Newborn Screening Coding and Terminology Guide

LHNCBC has collaborated with the multiple federal, state, and other agencies to standardize all of the variables used in newborn screening (NBS) using national coding standards as required by Meaningful Use Stage 2. Our collaborators include the Health Resources and Services Administration (HRSA), the Centers for Disease Control and Prevention (CDC), the Association of Public Health Laboratories (APHL), the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), and the National Heart, Lung, and Blood Institute (NHLBI). We have created a comprehensive panel of LOINC terms for NBS and continue to create new LOINC terms as new conditions and tests come into play. We also periodically review and update existing codes based on user feedback.

We have mapped all of the NBS conditions to SNOMED CT, and have requested new codes from SNOMED CT as needed. All of these terms are intended for use in an HL7 NBS lab result message. During FY2012, we updated and released a new version of our guidance for the HL7 message implementation and our example HL7 NBS result message. Many state NBS laboratories across the country are adopting our codes and HL7 guidance, including Kentucky, Oregon (which is the regional laboratory for 5 additional states), Colorado, New York, Illinois, and Washington. We actively worked with many of these states during FY2012 and have continued to do so.

Critical congenital heart disease (CCHD) is the latest condition that was added to the NBS Recommended Uniform Screening Panel, and several states are beginning to implement CCHD screening. So we are developing standard codes for reporting CCHD screening results. As part of this CCHD effort, we are working with multiple groups, including Children's National Medical Center, HRSA and other federal agencies, state NBS programs and public health departments, hospitals, and pulse oximetry technology vendors. We represent NLM on the technical advisory team for the HRSA CCHD pilot program grantees, to help them with terminology and interoperability standards, and are on the steering committee for NewSTEPs, APHL's new Newborn Screening Technical assistance and Evaluation Program. LHNCBC is also working with our partners to standardize data collection and coding for short- and long-term follow-up, beginning with the laboratory tests for confirmation and diagnosis of conditions targeted by newborn screening, and we are working on standardizing NBS genetic testing result reporting.

Communication Infrastructure Research and Tools

LHNCBC 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

LHNCBC continues to investigate, review, and develop collaboration tools, research their application, and use the tools to support ongoing programs at the NLM. In our work with uncompressed high definition video over Internet Protocol (IP), we determined strengths and weaknesses of each of the three technologies (iHDTV, UltraGrid, and Conference XP) and we continue to overcome problems encountered in the delivery of uncompressed video due to differing platforms. We are monitoring the High Definition (HD) open-source work of Video Conferencing Tool (VIC) developers regarding H.264 compression. VIC is used by the AccessGrid, an open-source collaboration tool widely deployed in universities and research centers and used in the OHPCC Collaboratory for research work and to support NLM programs. Because Argonne National Laboratory ceased AccessGrid development and support of its open-source community, our team investigated newer, cloud collaboration tools, including proprietary ones that are standards compliant and that emulate the AccessGrid's pioneering model. Our aim is to identify technologies that might have sufficiently broad appeal to replace the AccessGrid platform. So far, the business models and licensing terms have proven more of a problem than the technologies.

The team published a comparison of the major compression/decompression (codecs) available in the OHPCC Collaboratory for High Performance Computing and Communications (Collab) last year, submitted a systematic review of uncompressed video technologies for publication, and drafted a report about the cloud technologies. In addition, the team has initiated review of video applications for mobile devices. The overall plan is to study and test these applications, then depending on the results, launch and test them in real settings.

Until recently, iHDTV was the only uncompressed video system sufficiently robust to use in a clinical trial but OHPCC staff have worked with the developers of UltraGrid at Masaryk University in Brno,

Czech Republic to integrate audio and other enhancements. This work led to several demonstrations of trans-Atlantic uncompressed videoconferencing at data rates of 1.5 gbps between NLM and Masaryk. Substantial work was done with the lead developer of ConferenceXP (CXP) at the University of Washington. The team was able to get the uncompressed as well as the compressed versions of the program to work this year and extensive testing was done with the National Center for Supercomputer Applications (NCSA). Although the technology works, there are latencies due to a variety of possible causes and the team has focused on UltraGrid and iHDTV, with a special interest in UltraGrid. OHPCC and Masaryk research groups share an interest in 3D HD videoconferencing, 4K video, varied forms of HD compression, and the use of dynamic circuit networks (DCN) to ensure quality of service. The team has implemented a 3D version of UltraGrid in the lab, but not over networks. The CXP tests with NCSA have led to some initial testing of their compressed/uncompressed Viper HD technology. The team continues to collaborate with the Rochester Institute of Technology (RIT) and the University of Puerto Rico Medical Campus to test open-source software for compressed HD videoconferencing based on the H.264 video standard and cloud technologies.

The installation of a 10 Gigabits per second (Gbps) network in the Collab has greatly facilitated collaboration with other institutions and our ability to test uncompressed video. The network capacity made the uncompressed trans-Atlantic videoconferencing demonstrations possible without using dynamic circuit network technology to reserve bandwidth. The team believes DCN technology should be further explored and attempts were made to establish an alliance with Northrup Grumman to test their software. The company has decided to keep development in-house and make it proprietary, however. 3D HD cameras have been acquired and components for 4K video have been purchased to further test the network infrastructure and continue work with Masaryk.

The team installed iHDTV systems at the Medical University of South Carolina (MUSC) to study uncompressed video's use as a diagnostic tool. We selected teledermatology as a research domain because previous research has shown it to be particularly difficult to use standard definition video to do remote dermatological exams. We are measuring diagnoses, clinician confidence, decisions to biopsy and physician and patient encounter satisfaction and are comparing telemedicine applications under the following conditions: 1) when patients are examined in-person, 2) when patients are examined using uncompressed high definition video, 3) when patients are examined using compressed high definition video using a standard employed by all major commercial videoconferencing manufacturers, and 4) when patient data (history and photos) are used to assess patients by typical store and forward methods. We chose to use iHDTV systems for the study because, at the time,

audio was not integrated into UltraGrid and ConferenceXP could only transmit compressed video. The study was delayed substantially by IRB complications, OMB clearance requirements, the death of the PI collaborator, and the subsequent incapacitating illness of her successor. We found a new MUSC PI and data collection has now begun. These problems necessitated an additional site visit to provide refresher training and a run through of the research protocols and methodology. Previous work at MUSC experimenting with video medical interpretation has been published.

Staff continued to work with SIS on a distance education outreach program for minority high school students and with the NIH Library to offer NCBI database and other bioinformatics training at a distance. In FY2012, staff delivered posters and presentations on the "virtual computer lab" methodology at the meetings of the American Telemedicine Association, the Medical Library Association, and the Research Centers at Minority Institutions annual symposium.

OHPCC Collaboratory for High Performance Computing and Communication

LHNCBC established the OHPCC Collaboratory for High Performance Computing and Communication (Collab) as a resource for researching, testing, and demonstrating imaging, collaboration, communications and networking technologies related to NLM's Next Generation Network initiatives. Staff use this infrastructure to test new technologies of interest to NLM and to conduct ongoing imaging, collaboration and distance learning research both within LHNCBC and outside 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 enables 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 the NLM 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, open-source technologies such as Conference-XP, iHDTV, UltraGrid, and the AccessGrid as well as cloud technologies and mobile apps.

Last year we acquired a 3D camcorder for the purpose of using it and/or dual non-3D HD camcorders to transmit 3D HD video in future videoconferencing research, and this year we acquired hardware to build systems capable of transmitting 4K video. We updated the Collab's streaming and Web servers and worked on replacing the Google appliance for searching the Web server with open-source software. Changes in streaming technology required extensive file conversions of

previously recorded video programs. Finally, staff published a retrospective review of the research generated as a result of NLM advance computing initiatives.

Disaster Information Management: Lost Person Finder

NLM's increasing interest in recent years in mitigating the effects of wide area disasters has resulted in information resources and tools from many parts of the library. In our Lost Person Finder (LPF) project, we address the problem of family reunification in the wake of a mass casualty event. LPF systems combine image capture, database and Web technologies, and address both hospital-based and community-wide disaster scenarios.

Web Site and Services

The heart of our system, now operational, is *People Locator* (PL), the main LPF Web site and its MySQL database. Developers extensively customized the open-source Sahana disaster management system to create, for example, a unified site to hold data from multiple disasters, thereby eliminating the need to build multiple Web sites and database instances for different disasters. PL can be searched by hospital counselors, relief workers, or the public. Searching or reporting can be done via computer or through mobile apps using Web services.

During 2012, we made PL more robust. It is now running on enterprise-level, load-balanced dual systems with failover, independent uptime monitoring, and an integrated SOLR indexing engine for fast record retrieval. We also developed a lighter-weight, more-portable version, *PLmobile* for special local use.

Mobile Apps

For hospital-based reporting, the triage process begins with *TriagePic*, a Windows application that hospital staff can use to quickly photograph arriving victims. These pictures, along with general health/triage status and minimal descriptive metadata (e.g., name, age range, gender) are packaged and sent by Web services to PL. In 2012, we moved TriagePic from a laptop-hosted system to tablets, to exploit their touch screens and front- and rear-facing cameras. Developers made the existing Windows 7 version more touch friendly and feature-rich, and deployed most recently on a Samsung 7 Slate. To provide a wide range of attractive platforms for hospital use, we are also developing Android and iPad versions.

In 2012, we further improved ReUnite, a smartphone app originally introduced during the Haiti earthquake for the public and aid workers, to better report and search for those missing during community-wide disasters. Our iOS5/6 version for the iPhone or iPod Touch is available from iTunes as a free download.

Deployments

From the beginning of the project, we have taken part in demonstrations and large-scale multi-institutional drills, as well as stood up events for international disasters, including the Christ Church Earthquake (February 2011), the Japanese Earthquake and Tsunami (March 2011), and the Joplin Tornado (May 2011), among others.

In FY2012 we also: (1) repurposed ReUnite to track patients as they were transferred from one hospital to another and this function was used by a hospital in Indianapolis (March); (2) supported Sahana Software Foundation for a demonstration to the Australian Red Cross (April); (3) served as a standby software expert for Hurricane Sandy and as a backstop-consultants for New York City shelter management software developers (there are commonalities between their software and ours) (October/November); and (4) demonstrated our software at the HALO 2012 counterterrorism conference in San Diego (October).

Face matching research

Our goal here is to enable users to find missing person records through automatic face recognition, a significant extension of our current method of searching by name or other text metadata. Our application has special challenges: unlike many other systems, our face matching needs to rely on a *single* photo of a person to identify her/his face in other images, and cannot therefore exploit traditional face recognition models that require large training sets. Researchers accomplished substantial work mainly in face localization, a key first step in face matching. Originally operating only on grayscale images, we extended the face localization subsystem to exploit human skin tones that serve as an effective clue for faces in images. With up to 95% accuracy achieved for face localization and detection, our system outscores most existing solutions, including commercial systems like FaceSDK by Luxand, Inc.

In addition, as a way to reduce the search space for face matching, researchers developed a near-duplicate image detector to identify images submitted for the same person. This detector successfully reduced the PL image collection by 33% by discarding such redundant information, and thereby speeding up face matching.

To support research and testing, we need annotated images. For this purpose, we have developed ground-truthing and annotation-judging tools. For instance, our ImageStats tool, originally used in-house for face annotations, was released to handle crowd-sourced annotations of public-domain face images by volunteer students (globally recruited) through the *Google Code-in* contest, through our collaboration with the Sahana Software Foundation.

Video Production, Retrieval, and Reuse Project

This development area encompasses four projects that contribute to the NLM Long Range Plan goal of promoting health literacy and increasing biomedical understanding.

NLM Media Assets Project

The NLM Media Assets Project provides the NLM with easy access to audio-video resources for improved biomedical communications. This includes:
- The NLM/APDB Tape Library Archive Project,
- The Hypervideo Personal Digital Library/ Digital Video Library (a computer aided search, retrieval and viewing database),
- The NLM/History of Medicine Exhibits Audiovisual Assets Management, and
- Archival management of the Visible Human Project film and digital image dataset.

The NLM Support Project provides NLM with the audio/video support and development needed to promote and augment NLM's operation. This includes:
- Support for the maintenance and operation of the NLM state-of-the-art auditorium, board room and conference rooms including video teleconferencing and NLM-wide Webcasting, and
- Ongoing production, post-production, and authoring services for the development of Internet video, interactive multimedia for large-screen and tablet devices and displays, and Blu-Ray DVD production.

The LHNCBC Research Support Project contributes to improving access to high quality biomedical imaging information. This project includes:
- The APDB/NCI collaboration on 3D visualization of molecular structures and functions in the discovery of disease and treatment,
- The Movement Disorders Database (a digital archive of movement disorders patients going through diagnostic routines, and development of interactive tools and applications which utilize video for clinical monitoring and diagnostics),
- The Profiles in Science video modules, and
- The Visible Human imaging and visualization research.

The LHNCBC Core Resources Project provides research into developing new technologies for disseminating biomedical information. This project includes:
- augmented reality modeling and applications,
- mobile device application development,
- ultra high definition imaging research,
- ongoing design and development of image-rich Web sites in support of biocommunications,
- audio/video/imaging archiving and asset management, and
- the LHNCBC Research Update Modules.

Interactive Mobile Applications

A number of LHNCBC 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 improvement of rendering times for videographics and 3D visuals and animations for DVD and other interactive multimedia productions.

Planning and development of interactive multimedia for the FY2012 NLM Exhibition "Native Voices: Native People's Concepts of Health and Illness" continued. APDB staff worked with the NLM Director, and staff from the Office of Communications and Public Liaison, Office of Health Information Programs Development, and History of Medicine Division to produce interactive video and videographic content for the exhibition Web site and the traveling exhibition program. Based on this outreach plan, APDB produced and developed an interactive iPad app featuring all video materials currently in the physical exhibition. Additional video interviews and segments were incorporated, as well as highly interactive search and retrieval functionality within the app. All video content was encoded in formats for distribution across multiple platforms including the iPad and mobile QR code applications featured throughout the onsite, Web, and traveling exhibition. Our focus on video compression codecs for small screen delivery, navigation, and search capabilities is an ongoing area of research related to the work of the exhibition as well as many other areas of NLM's information programs.

Digital Video Archive

APDB is now applying digital workflow management and file format standards (originally established for exhibition production support) to convert LHNCBC's large library of historical tape – which contains over four decades of NLM programs – into a viable digital repository accessible for future use. The Branch began testing use of the motion JPEG2000 for long-term archiving. Testing is in progress and results are being monitored. APDB expanded by more than 50 videos the extensive digital video library assembled for the NLM Director's exhibition interview database, which is part of APDB's ongoing effort in digitizing, organizing, and accessibly storing large-scale video libraries.

Biomolecular Visualization

APDB staff continued to collaborate with NCI's Laboratory for Cell Biology and with OHPCC to visualize and analyze complex 3D volume data generated through dual beam (ion-abrasion electron microscopy) and cryo-electron tomography. We applied 3D biological imaging technologies to data using advanced image segmentation methods and computational analysis to obtain an integrated molecular understanding of cellular architecture.

To facilitate a better understanding of the normal and pathological process of the subcellular environment, staff developed novel imaging techniques including linking correlative light microscopy and ion-abrasion scanning electron microscopy as a method of locating, identifying, and studying cellular or sub-cellular structures through the use of 3D reconstructions.

APDB produced 3D models, animations, and medical illustrations for various molecular content areas, including:

- HIV membrane fusion,
- HIV pathogenic distribution from T-cell to astrocyte,
- T-cell to T-cell HIV infection,
- mouse intestinal epithelium,
- dendritic cell/CD4+ T-cell interactions,
- role of GroEL in protein folding, and
- HIV-infected T-cell and neural stem cell junction.

The resulting visuals have enhanced the understanding and discoveries in the character of several immunological cells, cell structures, and their interaction with pathological viruses including HIV.

Computing Resources Projects

The Computing Resources (CR) Team has a variety of core projects that build, administer, support, and maintain an integrated and secure infrastructure to facilitate LHNCBC's research and development (R&D) activities. The integrated secure infrastructure contains network, security, and facility management, and system administration support for a large number of individual workstations and shared servers.

The network management includes the planning, implementation, testing, deployment and operation of high-speed networks over Internet and Internet-2. One core project implements the 10-gigabit network, and studies many advanced communication protocols to support LHNCBC collaboration activities and research projects. Another core project implements a network monitoring system that displays network usages in real time. The network management team also participates in the study of Trusted Internet Connection (TIC) consolidation and evaluates the impacts to the NIH and NLM.

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 IT working environment. One core project studies and implements a unified patch management to improve LHNCBC's overall security measures. Another core project implements the automated security audit system that ensures all system at LHNCBC comply with policies. The security management team also studies and evaluates the network performance impact of Web anti-virus software, and coordinates annual penetration testing to ensure network security. The facility management team deploys new IT equipment and servers, including power acquisition, network planning, cabling

connection, and space allocation in the central computer room as well as at co-location facilities. Another core project studies, designs, and implements an enterprise console management system that enables LHNCBC to remotely manage large numbers of servers.

The system administration team provides LHNCBC-wide IT services such as DNS, NIS, data backup, printing, and remote access to ensure an efficient business operation. Core projects include Federal Information Security Management Act (FISMA) compliance facilitation and support and centralized network storage to support Continuity of Operation (COOP) requirements. Other projects include a centralized ticketing system for better customer support and an enterprise secure remote access system to meet emergency requirements like pandemic flu. Additionally, the system administration team supports shared computing resources such as security audit, system buildup, and security certification.

Training and Education at LHNCBC

LHNCBC is a major contributor to the training of future scientists and provides training for individuals at many stages in their careers. Our Informatics Training Program (ITP), ranging from a few months to two years or more, is available for visiting scientists and students. Each fellow is matched with a mentor from the research staff and participates actively in LHNCBC research projects.

During FY2012, 49 participants from 16 states and six countries received training and conducted research in a wide range of disciplines: 3-D image processing, biomedical ontology research, biomedical terminology research, content-based information retrieval, de-identification of medical records, evidence-based medicine systems, image, text and document processing research, information retrieval research, literature-based discovery research, natural language processing research, personal health record research, pill identification research, research into collaboration tools, semantic Web research and systems for disaster management, research in question answering systems, research on large biomedical data sets. The program emphasizes its focus on diversity through participation in programs for minority students, including the Hispanic Association of Colleges and Universities and the National Association for Equal Opportunity in Higher Education summer internship programs.

The ITP also sponsors a Clinical Informatics Postdoctoral Fellowship Program, funded by LHNCBC, to attract young physicians to NIH to pursue research in informatics. This program is run jointly with the NIH Clinical Center to bring postdoctoral fellows to labs throughout NIH. LHNCBC 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. We also host a two-month NLM Rotation Program which provides

trainees from NLM-funded Medical Informatics programs an opportunity to learn about NLM programs and current LHNCBC research. The rotation includes a series of lectures showcasing research conducted at NLM and provides an opportunity for trainees to work closely with established scientists and fellows from other NLM-funded programs.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

David Lipman, MD
Director

The National Center for Biotechnology Information (NCBI) was established as a division of the National Library of Medicine in November 1988 by Public Law 100-607. 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 24 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 major role in the increased use of bioinformatics. Recently, next-generation sequencing technologies have been a source of large volumes of sequence data. NCBI meets the challenge of collection, organization, storage, analysis, and dissemination of scientific data by designing, developing, and providing the public with the tools, databases, and technologies that will enable genetic discoveries of the 21st century.

NCBI has a multidisciplinary staff of 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 molecular biology and genetics information and associating it with related information in the biomedical literature.
- Performs research into advanced methods of computer-based information processing for analyzing the structure and function of biologically important molecules and compounds.
- Facilitates the use of databases and software by researchers and healthcare personnel.
- Coordinates efforts to gather and disseminate biotechnology information worldwide.

Molecular Biology Information Resources

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

GenBank

The primary source 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. Thousands of sequence records are submitted prior to journal publication. Records submitted to NCBI's international collaborators—EMBL (European Molecular Biology Laboratory) in the UK and DDBJ (DNA Data Bank of Japan)—are shared through an automated system of daily updates. Other cooperative arrangements, such as those with the US Patent and Trademark Office for sequences from issued patents, ensure that the collection contains all available relevant data.

GenBank is divided into separate divisions based on taxonomy and sequence data collection methods. Eleven taxonomy divisions contain sequences for over 380,000 species. There are numerous high-throughput sequencing divisions. The Whole Genome Shotgun (WGS) division includes contigs (overlapping reads) from WGS projects. Annotations are permitted in WGS assemblies, and records are updated as sequencing progresses and new assemblies are computed.

The fastest growing division in GenBank is TSA (Transcriptome Shotgun Assembly), which contains shotgun assemblies of primary (mRNA) sequences deposited in dbEST or the Short-Read Archive (SRA).

The amount of data submitted to GenBank continues to grow. GenBank's two major divisions, WGS

and non-WGS, combined contain over 240 million sequence records and over 450 billion basepairs. The traditional nucleotide sequences division increased to 156 million records in FY2012, from 141 million records in FY2011. The WGS division grew to over 84 million records, up from 65 million in FY2012. Five updated releases of GenBank were made public over the past year.

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. GenBank now contains more than 1,200 complete genomes from bacteria and archaea, with 20 percent of these deposited during the past year. The number of eukaryote genomes with significant coverage and assembly continues to increase as well, with over 460 WGS assemblies now available.

In order to simplify access to, and improve the quality of, the enormous amounts 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 resources within and outside NCBI, including biological databases, genomes, 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 housed at the NCBI.

NCBI has developed various tools for GenBank data submission. The BankIt submission tool allows the author to submit sequences within an online form and validates submissions by flagging errors before the sequences are deposited. Sequin is a stand-alone tool for updating and submitting large groups of sequences to the database. The latest version of Sequin, 12.21, was released in July 2012. New features include additional Submission Wizards for TSA (transcriptome shotgun assembly), ITS (internal transcribed space), microsatellite, and mitochondrial D-loop and control region sequences. A Sequence Method Page was added with information on sequencing technology and assembly methods.

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. NCBI's genomic information databases include: BioProject, dbSNP,

RefSeq, CCDS, dbGaP, Gene, Probe, UniGene, HomoloGene, dbVar, Genome Reference Consortium, GEO, and Epigenomics. 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 RefSeq collection contains 17,977,767 proteins for 18,512 organisms, representing a 37 percent increase in the number of proteins and a 14 percent increase in the number of organisms in the last year. There are a number of sub-branches for the RefSeq project with curatorial support provided by different sections for microbial genomes, viral genomes, plant and fungal genomes, higher eukaryotes, and so on. The Fungal section of RefSeq realized significant growth in the number of organisms (59 percent), followed by the Invertebrates (22 percent).

In the past year, over 56,000 new or updated RefSeq records were made public for the 44 taxa (nine invertebrates, 32 vertebrates, and three plants) supported by the (primarily) transcript-based, curation-supported, process flow. The count reflects the pool of RefSeq records that are considered well-supported and are available for updates by RefSeq curation staff. RefSeq curation activities resulted in content modification for over 50,000 Gene records and manual review of over 16,000 RefSeq records. In addition, annotation of sequence records was expanded to include reports of primary transcript evidence that supports the long-range exon combination represented by the RefSeq record. The RefSeq curation group provides significant support to NCBI's eukaryotic whole genome annotation process flow by setting up the annotation run and defining input data requirements, reviewing intermediate QA results, providing QA testing for new data loaded to NCBIs genome browser (Map Viewer) prior to public release, and more. RefSeq curation staff also serve as a point of contact for research communities and provide education about NCBI resources (particularly genomics resources), facilitate data submissions, and respond to community requests for curatorial improvements for transcript and protein representation of their organisms of interest.

The Consensus Coding Sequence (CCDS) database identifies a core set of consistently annotated, high-quality human and mouse protein coding regions. Both human and mouse datasets were updated in the past year, yielding current totals of 27,511 CCDS IDs for human and 23,027 CCDS IDs for mouse. The CCDS project is supported by curatorial review so that annotation updates to proteins that have a CCDS ID are done in a

synchronized manner that retains consistency in different international genome browsers. Updates are reviewed and must be agreed upon by members of an international collaboration. Public annotation is provided to explain the evidence for the curation decision to make a change. Over the last fiscal year, 267 existing CCDS entries were subject to collaborative review and 170 explanatory public notes were added to the database. During this time, the CCDS collaboration also worked on reconciling annotation differences for human protein-coding genes that are not yet represented in the CCDS database.

NCBI's 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. Gene is heavily used and manages information for more than 10 million genes from almost 10,000 taxa. About 80,000 full records are retrieved each day from about 22,000 sessions.

This year, much software development was focused on improving representation of phenotype information to support the Genetic Testing Registry (GTR) and other medical genetics projects. Links were added between Gene, MedGen and GTR, and phenotype definitions were standardized across all resources.

The representation of human genes as RefSeqGene records continued to grow, from 4,405 at the beginning of the fiscal year, to 4,849 at the end. As part of the LRG collaboration, RefSeqGene records were also revised to include more splice variants.

Activity within the Locus Reference Genomic (LRG) collaboration also increased. Currently about 685 RefSeqGenes have been assigned LRG accessions, with 234 now public. As a corollary of establishing RefSeqGene, collaborations with the dbSNP and dbVar groups continued in order to establish sequence annotation of clinically important variants, and to release the information in a more timely fashion. Several suggestions about the human reference sequence were provided to the Genome Reference Consortium, based on the detailed analysis of genes being targets for RefSeqGene.

To support the medical genetics community for which RefSeqGene was implemented, key tools have been developed to process human variation data according to Human Genome Variation Society (HGVS) standards. These tools are a critical component of the infrastructure to analyze human variation. Variation Reporter, for instance, allows users to upload variants and receive a report of known variants and functional consequences.

Extensive testing, quality assurance, and documentation are essential to the release of data in Gene, Map Viewer, and other resources that depend on Gene to establish accurate computational links to gene definitions. Gene staff is heavily involved in annotation of genes on genomic sequences via the genome annotation pipeline. In FY2012, further improvements were made to the genome annotation pipeline process, which now supports more than 50 taxa. Because of improvements to the data flow,

key genomes such as human and mouse were annotated several times, making the representation in Gene more timely and complete.

The Map Viewer is NCBI's 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, and 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. Users are able to zoom into progressively greater levels of detail, down to sequence data for areas of interest. In FY2012, there was a major effort to improve the Map Viewer ideogram. The Genome Decoration Page that allows users to place their own annotation data on a genome also is being updated to use the new ideogram. The Evidence Viewer is a Map Viewer feature that provides graphical biological evidence supporting a particular gene model.

The NCBI Remap tool allows users to remap coordinates of genes, SNPs, and other markers from one genome assembly to another. The tool accepts a variety of genome annotation formats and uses the alignments of one assembly to project annotation features on the other assembly. This year, the Remap service added an API so remaps may be implemented programmatically. The Clinical Remap tool provides coordinate remapping between genome assemblies and reference standard RefSeqGene records.

The Genome Reference Consortium (GRC) is an international collaboration that aims to update and improve 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. NCBI also generates the final assembly after collaboration and quality assurance. In FY2012, zebrafish was added to GRC. Also this year, GRC implemented an automated process to identify changed alignments along with a Web interface that allows review by curators. GRC made extensive updates to the GRCh37 release of the human genome by generating multiple representations of complex regions and by releasing fixes limited to specific regions. GRC is working with EBI and Sanger to analyze changes needed for GRCh38.

UniGene is highly useful as a resource for cDNA clone selection and rudimentary expression profiling by many communities of scientists. Interactive usage of UniGene remains steady with 20,000 Web hits daily, 6,000 of which involve reviewing clusters. While UniGene has expanded to include 114 organisms, most of the interactive traffic is for human and mouse entries.

The HomoloGene database of homologous genes acts as a complement to the Gene and UniGene resources. HomoloGene covers 20 model animal and plant genomes and provides a reliable and comprehensive database of

gene homologs, offering statistics on inter-species sequence and protein domain conservation. The latest HomoloGene release, 66, included the addition of genes and sequences for the Rhesus monkey as well as updated annotation for human, chimpanzee, cow, dog, rat, chicken, zebrafish, fruit fly, yeast, Arabidopsis, and rice.

The database of Short Genetic Variations (dbSNP) is a comprehensive catalog of common genetic variation. dbSNP contains over 178 million submissions of human genome data that have been processed and reduced to a non-redundant set of 52 million refSNP clusters. Over 100 other organisms are represented in the SNP database, with 124 million submissions curated to 81 million refSNP clusters. In FY2012, NCBI's SNP Website was redesigned with a new interface similar to other resources in NCBI's Entrez search and retrieval system.

UniSTS is a comprehensive database of sequence tagged-sites derived from STS-based maps and other experiments. Usage of the UniSTS resource remains steady, at about 2,500 hits per day.

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

UniVec, NCBI's non-redundant database of vector sequences, is used in conjunction with the VecScreen tool to screen nucleotide sequences for contamination with foreign sequences introduced during the cloning or sequencing process. The UniVec database was updated to build 7.0 during the year, and the number of sequences represented in the database increased by 16 percent. The newly added sequences include: 96 complete vector sequences and 157 adaptor, primer and multiple-cloning site sequences. The new additions include sequences that were sources of vector contamination in submitted GenBank sequence. This update to the UniVec database will enable searches run using NCBI's VecScreen tool to detect more of the foreign sequences added by various cloning procedures.

Comparative Genome Data

The Genome database provides information for a wide variety of organisms including eukaryotes, Archea, viruses, and bacteria. In FY2012, the Genome site was completely redesigned and reorganized. The new data model is based on organism information, including its genome structure, available assemblies and annotations, and related genome-scale projects such as transcriptome sequencing, epigenetic studies, and variation analysis. Search results and record views are now discovery-oriented with a column on the Web display that provides options for analysis tools and easy access to related data.

Information about genomes is represented in several highly interlinked NCBI databases: BioProject, BioSample, Genome, and Assembly.

The BioProject database, formerly named Genome Project, provides access to large-scale biomolecular projects that include a variety of data types, such as single-cell organism genomes, metagenomes, transcriptome sequencing, genotyping, variation and epigenomics. The resource provides new tracking of several data elements, including more precise information about a project's scope, material, and objectives.

The new Assembly database, released in January 2012, provides statistics, update history, and links to sequences for prokaryotic and eukaryotic genome assemblies. The database tracks changes by providing an accession and version number to each assembly. It also contains the placement of each scaffold in the assembly along with the name and sequence accession and version for each chromosome and scaffold. In addition, Assembly organizes and provides assembly metadata such as assembly names and synonyms and statistical reports. Assemblies may also be linked through the Genome database main page or from a genome record.

Specialized Databases and Tools

The Genetic Testing Registry (GTR) was developed as a centralized resource for test providers to submit information voluntarily about genetic tests for inherited and somatic genetic variations. Released to the public in February 2012, GTR includes detailed information about available genetic tests, including their purpose, methodology, validity, evidence of the test's usefulness, as well as laboratory contacts and credentials. GTR also provides access to information from other relevant resources, such as GeneReviews. The initial release of GTR included content from the GeneTests, an NIH-supported resource; in June 2012, a submission site for new tests was implemented.

The Influenza Virus Resource is a comprehensive collection of flu sequences. Samples collected from around the globe include viruses obtained from birds, pigs, humans, and other species. Links are provided to other flu resources containing sequences, publications, and general flu virus information. Nearly 44,000 new influenza virus sequences were added to the Influenza Sequence Database in FY2012. About 27,500 were from the NIAID Influenza Genome Sequencing Project, the NIAID-funded Centers of Excellence for Influenza Research and Surveillance, the Centers for Disease Control and Prevention, and about 40 other institutions worldwide.

The Sequence Read Archive (SRA) was developed to handle large amounts of data generated from massively parallel sequencing experiments. The repository accepts various formats of high-throughput sequencing data, including that produced by platforms from Illumina, Complete Genomics, Applied Biosystems, and others. In addition to raw sequence data, SRA stores alignment

information in the form of read placements on reference sequences. SRA is NIH's primary archive of high-throughput sequencing data and cooperates with the International Nucleotide Sequence Database Collaboration (INSDC) for data exchange.

The Clone Database integrates information about genomic clones and libraries, including sequence data, genomic position, and distributor information. CloneDB provides descriptions, sources, and detailed statistics on available genomic libraries for a large number of organisms. In FY2012, CloneDB was updated with an Entrez search capability and new Web page. Also, algorithm updates were performed to reduce overlapping clone placements, and new clone placement reports were developed and made available on the FTP site. Eight species received updated clone placements, including multiple human and mouse assemblies as a result of patch releases; a total of 10 organisms received updated libraries and end sequences. Data for all organisms with fingerprint maps at the British Columbia Genome Sciences Centre were loaded into CloneDB and are displayed on the individual genomic clone record pages under the newly indexed term, "Has fingerprint."

The Gene Expression Omnibus, or GEO, is a public repository for large-scale functional genomic data generated by microarray- and sequencing-based technologies. GEO now contains over one million records. In FY2012, GEO processed over 200,000 new records, representing 25 percent growth over last year. Several GEO resource improvements were made in FY2012, including release of a new Web application called GEO2R. This tool greatly expands the utility of the GEO database by enabling inexperienced users to perform sophisticated statistical analyses of GEO data to help identify and visualize differential gene expression. Additionally, a new pipeline was established to broker GEO data to NCBI's BioProject database, and more proactive approaches have been implemented to acquire citation information for GEO data so that a greater amount of data can be appropriately linked with corresponding PubMed entries. Also, more granular Entrez indexing was applied to GEO, as well as a "Similar Studies" link, helping users locate data relevant to their interests.

NCBI's BioSample database contains descriptions of biological source materials used in experimental assays. Currently, the database contains over 1.6 million records that were either imported from SRA or dbGaP, or submitted directly using the newly released BioSample submission portal. The submission portal supports sample attribute packages in compliance with Genomic Standards Consortium standards as well as newly developed pathogen reporting standards. Additionally, several collections of commercial samples resulting from collaborations with American Type Culture Collection, Coriell, and the Leibniz Institute (DMSZ) repositories are public, including authenticated human cell lines and bacterial pathogen collections that will be incorporated into developing Pathogen Detection resources.

The NCBI Epigenomics database serves as a comprehensive resource for whole-genome epigenetic datasets selected from the GEO database. Epigenetics is the study of stable and heritable changes in gene expression that occur independently of the primary DNA sequence. The Epigenomics database contains over 4,000 studies, over 1,200 samples, and over 1,100 data tracks from five well-studied species. Data can be browsed by experiment or sample and a "Compare Samples" beta tool is available.

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. Over 28,500 taxa are represented in the database. 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 NCBI Genome Workbench is an integrated application for visualization and analysis of sequence data. It is designed to provide a flexible platform for development of new analytic and visualization techniques. Multiple releases were provided to the public in FY2012, supplying users with a number of new features, including default view functionality, expanded generic table and graphical sequence views, scale bar for trees with distances, integrated BAM files loader, support for new file formats, MacOS-X 64-bit support, and many others.

The Genome Workbench application was downloaded more than 33,000 times in FY2012, and average monthly usage doubled. Much of the development and visualization features of Genome Workbench transfer directly to its Web-based counterpart, Sequence Viewer (the graphical display for nucleotide and protein sequences). Sequence Viewer is designed to be an embeddable component to complement other information-rich views.

Variation Data Resources

In FY2012, NCBI introduced a preview version of a new database called ClinVar. The database facilitates access to and communication about the relationships asserted between human variation and observed health status. ClinVar data is comprised of reports of variants found in patient samples, assertions made regarding their clinical significance, information about the reporter, and other supporting data. ClinVar also acts as a sister site to the Genetic Testing Registry by archiving reports of the relationships between human variations and phenotypes along with supporting evidence.

The database of genomic structural variation (dbVar) contains data on variant DNA less than or equal to one basepair in size. Structural variation data submissions are accepted from whole genome comparative studies and locus- and gene-specific data from quantitative studies. The database, therefore, is organized by study. In FY2012, a dbVar Genome Browser was released, offering the ability to browse multiple studies at once and search for any feature annotated on the genome.

The 1000 Genomes Project is an international consortium that aims to build a comprehensive map of human genetic variation. Started in 2008, the project has collected data from 2,600 individuals originating from 26 populations around the world. NCBI has been making data from the project available since 2008, and in FY2012 collaborated with the Project to make the data available through Amazon Web Services' cloud. Cloud access allows any researcher to access and analyze the data at a fraction of the cost it would take for their institution to acquire the internet bandwidth, data storage and analytical computing capacity needed for the huge 1000 Genomes datasets.

NCBI developed a 1000 Genomes Browser in FY2012 that shows variants, genotypes, and supporting sequence read alignments from the 1000 Genomes Project. Users may perform searches using gene names, dbSNP accession numbers, or chromosomal positions. The graphic within the browser is based on NCBI's Sequence Viewer graphical tool. A table of genotypes organized by the 1000 Genomes Project is shown beneath the graphic view where users can drill down to show individual-level genotypes.

Variation Reporter, a new NCBI tool, takes a set of locations in a human genome assembly and identifies known human variations at those positions. The variations are collected from the NCBI refSNP data set. Variation reporter also accepts a variety of genome annotation formats. The results provide the location of variants along with clinical information, minor allele frequency, links to literature, and functional consequences.

A new Variation Web site was unveiled this year as a gateway to a variety of NCBI variation resources, including 1000 Genomes Browser, dbSNP, dbVar, ClinVar, dbSNP, the Variation Reporter, Clinical Remap, and the Phenotype Genotype Integrator.

Chemical Information

PubChem is organized as three linked databases: PubChem Substance, PubChem Compound, and PubChem BioAssay. Together, they form a complete resource for information on 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.

Now in its eighth year, PubChem reached a number of significant milestones in 2012, including: over 100 million contributed substance records; over 200 million biological screening results, and over 200 data

contributors. PubChem also routinely received more than one million requests per day from more than 100 thousand users. Through a series of contributions, including a high-profile one from IBM, PubChem now offers links to patent information for six million chemical structures, with over 59 million chemical-patent links covering US, European, and World Intellectual Property Organization patents spanning the years 1790-2012.

PubChem released a number of new search and data interfaces in the past year. PubChem now provides a RESTful Web interface (PUG REST) to aspects of PubChem data and services that simplifies requirements to externally access PubChem information. A PubChem Classification Browser now allows PubChem data to be explored using annotation to hierarchical ontologies and classification systems (such as MeSH, ChEBI, KEGG BRITE, and GO). A new layer to PubChem, called PubChem3D, allows one to use computed 3-D descriptions of chemical structures to interrelate, search, subset, visualize, analyze, and download data. Through integration with DrugBank, the Hazardous Substances Data Bank, and other databases, the annotation of chemical records is now dramatically expanded, including pharmacology, drug target information, toxicology, safety, and handling information.

PubChem made a number of technological changes in the last year. It is now Hypertext Transfer Protocol Secure (HTTPS) compatible. Heavily used pages, such as PubChem Compound Summary pages, were optimized to download faster, require less data to be transferred to users (limiting bandwidth requirements), provide greater navigation control, and to have an updated look and feel. In addition, indexing of PubChem Summary pages by search engines such as Google, Bing, and Yahoo is now improved.

PubChem continues to serve as a public repository of bioassay results for small molecules and RNAi, and continues to exchange bioassay data with the ChEMBL database at the European Bioinformatics Institute (EBI). The bioactivity data volume in PubChem grew significantly in 2012; the BioAssay database now contains over 600 thousand bioassay records and more than 200 million bioactivity test results. To extend the BioAssay database as an annotation resource for other NCBI molecular databases, inter-linking between NCBI's Protein, Gene, CDD databases and the PubChem BioAssay database via assay targets has been completed. A summary view and a detailed bioactivity data view is now provided from Protein and Gene records. Similarly, a bioactivity link is provided for the CDD Summary page, which shows a summary of the bioactivity data available for protein members belonging to a protein sub-family. There is an ongoing effort to enhance the PubChem deposition interface, in particular the bioassay submission portal, to ease bioactivity data exchange.

Protein Information

The Protein Clusters database contains Reference Sequence (RefSeq) proteins from the complete genomes of prokaryotes, plasmids, and organelles. The proteins are clustered and annotated based on sequence similarity and function, then used as a basis for genome-wide comparison.

Molecular Modeling Database (MMDB)

NCBI's Molecular Modeling DataBase (MMDB) mirrors biopolymer structures submitted to the Protein Data Bank (PDB). MMDB is augmented with annotation of structural domains and explicit links to relevant literature. MMDB is a source of protein and nucleotide sequences, chemicals tracked by PubChem, and novel conserved domains in the CDD, as well as structural neighbors computed by the VAST algorithm on compact structural domains. MMDB is processed and released on a weekly basis, following weekly updates of the PDB archive.

In FY2012, MMDB started presenting "merged" structure records for the first time. Due to limitations of the data exchange format, PDB needs to break up records describing large structural complexes (such as the structures of ribosomes) into several "split" data files. During processing of the structure data, those "split" records are now merged back together and presented as intended by the original data depositors.

In FY2012, methodology was developed to simplify and enhance the presentation of similar structures and structure neighbors in Entrez. Instead of focusing on single 3D domains, similarity of whole biological complexes will be presented wherever possible. This has been accomplished by utilizing the existing structure neighboring data computed by VAST. By focusing on similarities between whole functional complexes, lists of structure "neighbors" will be simpler and shorter, while at the same time highlighting the conservation of molecular interactions on top of biopolymer structure and sequence. A revised neighboring service will be made available to the public in FY2013.

Protein structures in MMDB provide annotations for protein sequences in NCBI's Entrez databases, for sequence results in the NCBI BLAST service, and for sequence variation tracked by the NCBI SNP database. To this end, the number of protein sequences directly linked to 3D structure was significantly increased in FY2012, employing sequence identifiers provided by the Protein Data Bank. Also, work has been initiated on the CBLAST service to provide links to 3D structure data from results pages of COBALT, NCBI's multiple sequence alignment platform.

Conserved Domain Database (CDD)

Nine releases of CDD were processed and published in FY2012 (versions 3.00 to 3.08). These included a mirror of Pfam version 26, TIGRFAM releases 11 and 12, and 2,013 new or updated NCBI-curated models. CDD curators processed 3,259 unique models in the in-house tracking database.

In October 2011, NCBI announced the release of a new version of the Conserved Domain Architecture Retrieval Tool (CDART), which was re-engineered for improved performance and better ranking and filtering of results. Also in October, a new version of the CDTree/Cn3D software bundle was released.

In November 2011, an updated version of the CD-Search tool was released that adds nucleotide sequences as queries, in addition to previous searching by protein sequences and sequence identifiers. Nucleotide queries are translated in all possible reading frames and combined results for these translations are presented by default. Conserved domain searches are now being run by default for nucleotide sequences shorter than 10,000 base pairs that are submitted to blastx, the translating BLAST search that scans all six translations of the submitted query against protein databases.

With the release of CDD v3.07 in August 2012, high confidence associations between proteins and conserved domain models are also calculated and provided for domain models imported from external sources, which has increased the specificity and accuracy of domain annotation provided by CDD.

In FY2012, the CDD group engaged in a concerted effort to increase coverage of protein 3D structures with conserved domain models by characterizing new protein domain superfamilies that are not yet modeled by CDD's source databases such as Pfam or TIGRFAMs. This is ongoing work, as the Protein Structure Initiative continues to deliver representative 3D structures for previously uncharacterized protein and protein domain families. CDD was presented at the 5th Conference of the International Society for Biocuration.

NCBI Biosystems Resource

NCBI's BioSystems database was developed as a complement to other databases, such as those that show the components and products of biologic pathways along with corresponding annotations and links to the literature. The BioSystems database was developed to: serve as a centralized repository of data; connect biosystem records with related literature, molecular, and chemical data via NCBI's Entrez system; and facilitate computation of biosystems data. BioSystems includes several types of records, including pathways, structural complexes, and functional sets, and is designed to accommodate other types of information as it becomes available. Data has been collected from a number of source databases.

During FY2012, the BioSystems database incorporated the Gene Ontology (GO) project, a controlled vocabulary of terms for gene product characteristics and gene product annotation data. GO records are linked to associated genes or proteins. LIPID Metabolites and Pathways Strategy (LIPID MAPS) data were also added to BioSystems. LIPID MAPS is an effort to identify and quantitate the lipid species in mammalian cells.

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 possible to rapidly scan huge sequence databases for similar sequences and to evaluate the resulting matches statistically. 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 and expanded for effectiveness and ease of use. BLAST contains 15 specialized pages corresponding to NCBI databases. During FY2012, a prokaryotic 16S ribosomal RNA database was added to the pull-down list on the mail Nucleotide BLAST service. The 16S BLAST database is useful for identifying or establishing the taxonomic affinities of unknown bacterial 16S sequences such as those from environmental or organismal samples. A new Domain Enhanced Lookup Time Accelerated BLAST (DELTA-BLAST) search algorithm was developed for searches that require more sensitivity than a standard protein-protein BLAST search. DELTA-BLAST identifies conserved domains and uses this information to construct and perform a search against the protein database. A new Microbial BLAST service was added to the BLAST suite as well. The redesign allows selection of taxonomic categories through an auto-complete mechanism; multiple taxonomic groups can be added or deleted, and nucleotide search sets have been divided into Complete and Draft genomes. Also, the BLAST+ Help Manual located on the NCBI Bookshelf was extensively revised and updated.

Stand-alone BLAST provides a platform for users wanting to perform searches locally. Version 2.2.27 of stand-alone BLAST was the most recent release this year. The Netblast (Blastcl3) client that provided batch search access to the Web BLAST service was discontinued and replaced by a remote option in BLAST+, the command line application. Support for the OLD_BLAST URL parameter was discontinued as well in order to facilitate enhancements to a new BLAST report page that is coming in 2013.

Integration of Clinical, Genetic, and Environmental Databases

The database of Genotypes and Phenotypes, dbGaP, was originally built to house, display and distribute data produced in Genome Wide Association Studies (GWAS). However, the primary mission of dbGaP has expanded beyond GWAS to represent evolving technologies and data types. The data that make up dbGaP studies can typically be categorized into three main classes: 1) phenotype data, consisting of clinical, anthropomorphic, demographic and exposure variables collected from a few hundred to

thousands of individuals; 2) molecular data produced using SNP chip arrays, expression arrays, epigenetic assays and/or sequencing of RNA or DNA; and 3) Study Documents. Study Documents are protocols or data collection instruments that describe how data or samples were collected and/or processed. The system has proven that it is readily adaptable to house data for many different types of studies that include human phenotype information.

Study Submissions

dbGaP has released 292 studies since 2007, 66 of which were released in FY2012. Some of the data in dbGaP is public, such as summary-level data, while some can only be obtained through an authorized access system (namely, individual-level phenotype/genotype data and some high-throughput sequencing data).

Linking Studies to Other NCBI Resources

All dbGaP studies are cross-referenced in the BioProjects database at the study administrative and data levels. Currently, disease traits for each dbGaP study are linked to one or more terms in Medical Subject Headings (MeSH). dbGAP staff is actively working on establishing links to MedGen records using the MeSH to MedGen links. All samples submitted to dbGaP are loaded and linked to the Entrez BioSample db. Links to PubMed articles identified by the submitter are also made for each dbGaP study.

Cumulative Summary Counts

Collectively, the studies released to date include measurements for a total of 528,718 research participants. The database includes:
- Over 134,596 individual-level phenotype and exposure variables.
- 3,288 documents, which allow users to explore the studies through natural language descriptions, collection forms, or scientific protocols, and also provide models for future research.
- Tens of trillions of individual genotypes, i.e., single-nucleotide measurements of participant DNA sequence. These potentially reveal systematic and heritable genetic differences between affected and unaffected individuals.
- 3,552 files of phenotypic association results have been submitted, and each file contains at least 100,000 p-values. The p-values are test statistics to determine the association of the tested phenotype trait with many locations (SNPs) in the human genome.

Authorized Access System Download Activity

The dbGaP authorized access system is the NCBI portal where Principal Investigators (PI) request access for individual-level data submitted to dbGaP. By the end of FY2012, 3,295 research projects from 2,014 Primary

Investigators had been created in the approval system. Each description of a proposed research activity is prepared by a PI as context and justification for their access to individual-level data. A project may include multiple study/consent group datasets, and a Data Access Request (DAR) is created by the system for each dataset.

Data Usability: Tools and Software Development

During 2012, there was significant progress in the development of the dbGaP submission portal, and release is imminent. The portal will reduce manual processing of incoming data files. Significant development effort will be devoted to the construction of the data tables and displays for Expression Quantitative Trait Loci (eQTL) results. This new class of data can be thought of as molecular phenotype association data, testing the association of SNPs to gene expression.

The new Phenotype-Genotype Integrator (PhenGenI) is a service that integrates genome-wide association study (GWAS) catalog data from NIH's National Human Genome Research Institute with molecular and literature databases from NCBI. PhenGenI takes chromosome location, gene, SNP, or phenotype as input and provides annotated tables of SNPs, genes, association results, a dynamic sequence viewer, and gene expression data.

Entrez Retrieval System

The Entrez search, retrieval, and indexing system was originally developed for searching nucleotide and protein sequence databases and related MEDLINE citations, but has since expanded to become the indexing and search foundation for all of NCBI's major resources. With Entrez, users quickly and easily search gigabytes of sequence and literature data. A key feature of the system is the concept of "neighboring," which automatically identifies references or sequences that are related to a user's research. The ability to traverse the literature and the molecular sequences via "neighbors" and links provides an efficient and intuitive way of accessing data. Entrez currently supports and integrates 38 databases, including the sequence databases, chemicals, and the biomedical literature.

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 are provided to the full text of articles, when available, via NCBI's PubMed Central database or the originating journal. PubMed, which serves as the foundation of NCBI's bibliographic information system, contains over

21 million citations from more than 36,772 journals, some dating back to the 1800s.

PubMed is continually updated and enhanced for better functionality and more precise search results. During FY2012, a Filters sidebar with user-selectable terms replaced the PubMed Limits page, making features more visible on the search page. PubMed's author search algorithm was improved to show similar author citations according to a ranking method. The ranking algorithm takes into account metadata for citations, such as co-authors, journal, and affiliation, with the goal of disambiguating common author names. The "Send to" menu was updated to include a "Citation manager" selection to provide a convenient way to import citations into reference manager software. Tools such as PubMed's "Results by Year" and PMC's "PMC images search" were added to the discovery column on all PubMed pages. The "Results by Year" tool displays a graph for search results with 10,000 or more citations. The PMC Images search shows thumbnail images from PMC that relate to the performed search. PubMed Mobile has replaced the PubMed Text version; users are automatically redirected to PubMed Mobile.

My NCBI is an Entrez feature that allows users to store searches and results, with the option of automatically updating searches and receiving results via email. Searches can be done directly on the homepage with a "Search NCBI Databases" link along with an "NCBI Site Preferences" link that allows users to set up display preferences. The PubMed Abstract display was updated with a "Save items" portlet to provide an easy way to add citations to a My NCBI collection. This feature also provides a drop-down menu of saved collections. The My NCBI My Bibliography Award View display, a tool to assist eRA Commons users to comply with NIH Public Access policy, was enhanced.

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 users from a PubMed citation directly to the full text of an article available through their library subscription program.

In FY2012, the number of organizations participating in LinkOut increased to close to 3,600, representing a six percent growth rate over the prior year. LinkOut participants include 2,790 libraries, over 510 full-text providers, and 300 providers of non-bibliographic resources, such as biological and chemical databases. Participation in Outside Tool, a service linking to external resources, increased to over 960 institutions. LinkOut users can now link to 99 million Entrez records, including links to the full text of 64 percent of PubMed records from over 10,100 journals. Usage of LinkOut resources reached over 39 million hits per month, and about 1.5 million hits per weekday.

A number of internal processes of LinkOut have been improved. LinkOut indexing sped up significantly after adopting a new approach to batch unique queries and their results into a dictionary file, which replaced direct processing of each query in the past. LinkOut DTD was revised to handle large files more efficiently and provide additional LinkOut entities to help providers construct linking rules easily. Additional filters were created to help users identify free resources from a specific provider easily. The Library Submission utility has also been enhanced as libraries are now able to select electronic publication date, specify multiple date ranges, and set a longer delay and retention period for their holdings.

In the coming year, focus will be on the utility for full-text and non-bibliographic providers. The goals are to easily correct errors due to redundant submission, allow providers to send links using a utility rather than XML files, and to provide a clear picture of existing link patterns for better quality control.

PubMed Central

PubMed Central (PMC) archives, indexes, and provides free and unrestricted access to full-text articles from life science journals. This repository integrates with the PubMed biomedical literature database of indexed citations and abstracts.

Use of PMC continues to increase in concert with the growth of available articles. On a typical weekday, PMC has up to 750,000 unique users retrieving over 1.4 million articles. As of October 2012, there were over 2.5 million articles in PMC.

PMC now has approximately 1,450 participating journals that deposit either all of their content or the subset of NIH-funded articles they publish. For publication year 2011, PMC received close to 80,000 NIH-funded articles and more than 100,000 additional peer-reviewed papers from participating journals. About 44 percent of the NIH-funded articles came from the journals that deposit final published articles directly into PMC; the remainder came as author manuscripts via the NIH Manuscript Submission System.

In August 2012, the National Information Standards Organization (NISO) announced the publication of a new American National Standard, JATS: Journal Article Tag Suite. This new national standard is based on the NLM Document Type Definitions (DTDs) developed and used by NCBI and PubMed Central. The full NLM Journal Archiving and Interchange Tag Suite reached version 3.0 before the NISO standardization process.

PubMed Health

The PubMed Health collection of systematic reviews grew from 5,500 in FY2011 to over 21,000 during the year; the collection includes systematic reviews of clinical effectiveness published in the last ten years. PubMed Health now has close to comprehensive coverage of reliable systematic reviews, combining reviews from leading health technology assessment agencies (such as the US Agency for Healthcare Research and Quality) and the Cochrane Collaboration with the systematic reviews incorporated in the Database of Reviews of Effects (DARE) from the Centre for Reviews and Dissemination (CRD). The DARE database of reviews was the principal addition to the systematic review collection this year.

New systematic reviews are being added to PubMed Health at an annual rate of close to 3,000, with hundreds of updated systematic reviews added as well. In FY2012, approximately 175 new full-text reviews were added.

Also added this year were more than 1,000 new summaries or articles for consumers and clinicians based on systematic reviews. Over a third of this new content came from a new PubMed Health information partner, the National Cancer Institute (NCI), which is providing evidence-based summaries from its Physician Data Query (PDQ) service. PDQ covers more than 100 of the most common cancers, generally with versions for patients and health professionals.

A key priority for PubMed Health in the last half of the year was the development of topic pages, which incorporate a new interface. The topic pages gather together clinical effectiveness research resources on a topic, clustered around background information. This design directs greater focus towards CER content, while enabling flexibility in the selection of background information. The new system will also highlight articles of general importance that cut across specific topics.

Bookshelf

The NCBI Bookshelf gives users access to the full text of over 600 textbooks and documents in life sciences, medicine, and healthcare. In addition to textbooks from commercial publishers, the Bookshelf includes tutorials and help documents authored by NCBI, NLM, and NIH staff. In FY2012, the Bookshelf added 323 new titles and 373 chapters to existing books.

Bookshelf made a number of improvements to the search feature: namely, making it easier to get to the result of interest; the ability to sort by relevance, alphabetically by title, or publication date; and a search for images run in parallel with the book search and highlighted in the images portlet.

The NLM LitArch Open Access Subset was also released in FY2012. The OA subset allows for the sharing of XML, PDF files, and images for a subset of titles that are either in the public domain or whose licenses and participatory agreements permit this.

E-Utilities

Entrez Programming Utilities (E-Utilities) are a set of server-side programs that use a fixed URL syntax that translates a standard set of input parameters into the values necessary for software components to search for and

retrieve data. All E-Utility programs are described in the Entrez Programming Utilities Help Manual.

E-Utility DTDs were updated in early FY2012 to include new Entrez databases including SRA, BioSample, and BioProject. E-Utilties for EFetch and E Summary (version 2) were released this year with an enhanced XML output for Entrez document summaries.

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 databases and software applications.

The research conducted by CBB has strengthened NCBI applications and databases by providing innovative algorithms and approaches (e.g., BLAST, VAST, CDD, and text mining) 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. CBB also performs research in natural language processing and text mining, with several of the results being used to improve the interactions between Web users and the NCBI Web pages.

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

The Board of Scientific Counselors (BSC), 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 BSC's thirty-eighth meeting was held in April 2012.

Bioinformatics Training and Support

Outreach and Education

NCBI's outreach and public services component is an essential activity to ensure that the user community is aware of all NCBI services and is trained to make effective use of those services. The audience for NCBI databases is very broad. The resources are used not only by molecular biologists and health professionals, but by students, educators, librarians, and science writers, as well as the general public. Garnering feedback from the user community is vital in order to provide services that meet their actual research needs and anticipate future requirements.

The public services division provides user support via e-mail and telephone, by staffing conference exhibits, and through training materials and seminars based on NCBI resources. 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 tutorial materials. NCBI staff published ten articles on NCBI resources in the 2012 Database Issue of *Nucleic Acids Research*. In addition, several resource articles were published in other scientific journals regarding dbVar, DELTA-BLAST, CCDS, Primer-BLAST, and CDD.

NCBI's social media presence continues to increase as usage of NCBI's Facebook, Twitter, and YouTube sites increased in FY2012. The amount of educational materials published online has also increased and improved. NCBI added ten tutorial videos to its YouTube site this year including:

- PubMed: The Filters Sidebar
- GEO2R: Analyze Geo Data
- GTR: Homepage and Basic Search Functions
- GTR: Locate a Test in Under Three Minutes

- My Bibliography
- Genome Workbench: Loading a Genome
- BLAST Results: Expect Values, Part 1 and Part 2
- Highlight Sequence Features
- Need the Full Text Article?
- PubMed Advanced Search Builder

The *NCBI News* online newsletter provides updates on new and improved resources, as well as a featured resource in each issue. Three issues of the NCBI News were published this year. NCBI also provides 17 "Announce" e-mail lists that give users the opportunity to receive information on new and updated services and resources from NCBI. Twenty-five RSS Web feeds are available for updates and announcements on various NCBI resources.

Training

The NCBI Education Program provides Discovery Workshops and Webinars aimed at various types of users, from beginners to seasoned researchers. Training for NCBI resources are provided for a wide audience.

NCBI Discovery Workshops take place three times a year on the NIH campus. These two-day workshops, taught by service desk staff, incorporate seminar and hands-on instruction and focus on a variety of NCBI topics: Sequences, Genomes, and Maps; Proteins, Domains, and Structures; NCBI BLAST Services; and Human Variation and Disease Genes. Additional workshops are provided at other locations with assistance by the National Network of Libraries of Medicine (NN/LM).

Various workshops were provided at meetings such as the American Society of Human Genetics (ASHG) annual conference for biomedical researchers and genetic professionals.

EXTRAMURAL PROGRAMS

Valerie Florance, PhD
Associate Director

The Extramural Programs Division (EP) administers extramural grant programs for NLM as authorized by the Medical Library Assistance Act (MLAA) and Public Health Service Act. EP's first grant awards were issued in 1965. Funds are expended as grants-in-aid to the extramural community in support of the Library's mission. Review and award procedures conform to NIH policies.

EP awards several categories of grants, all of which pertain to biomedical informatics and the management and dissemination of biomedical knowledge. Biomedical informatics research applies computer and information sciences to improve the access, storage, retrieval, management, dissemination and use of biomedical information. Applications are received through 'parent' NIH funding announcements or through special funding opportunity announcements issued by EP. Each year, NLM makes new and/or continuing awards in these five grant categories: Research Projects, Resources, Career Development, Research Career Training, and Small Business Research & Development.

Overview of FY2012

NLM's EP FY2012 base budget for grant awards was $44,281,000, the same amount as in FY2011. One hundred thirty-seven new and continuing awards were made with NLM's appropriated funds. An additional $3.3 million in co-funding was received for 11 awards from other NIH and Department of Health and Human Services (DHHS) sources. FY2012 funds were awarded to 35 organizations in 32 states (**Figure 5**).

Figure 5: Location of NLM FY2012 Extramural Awards
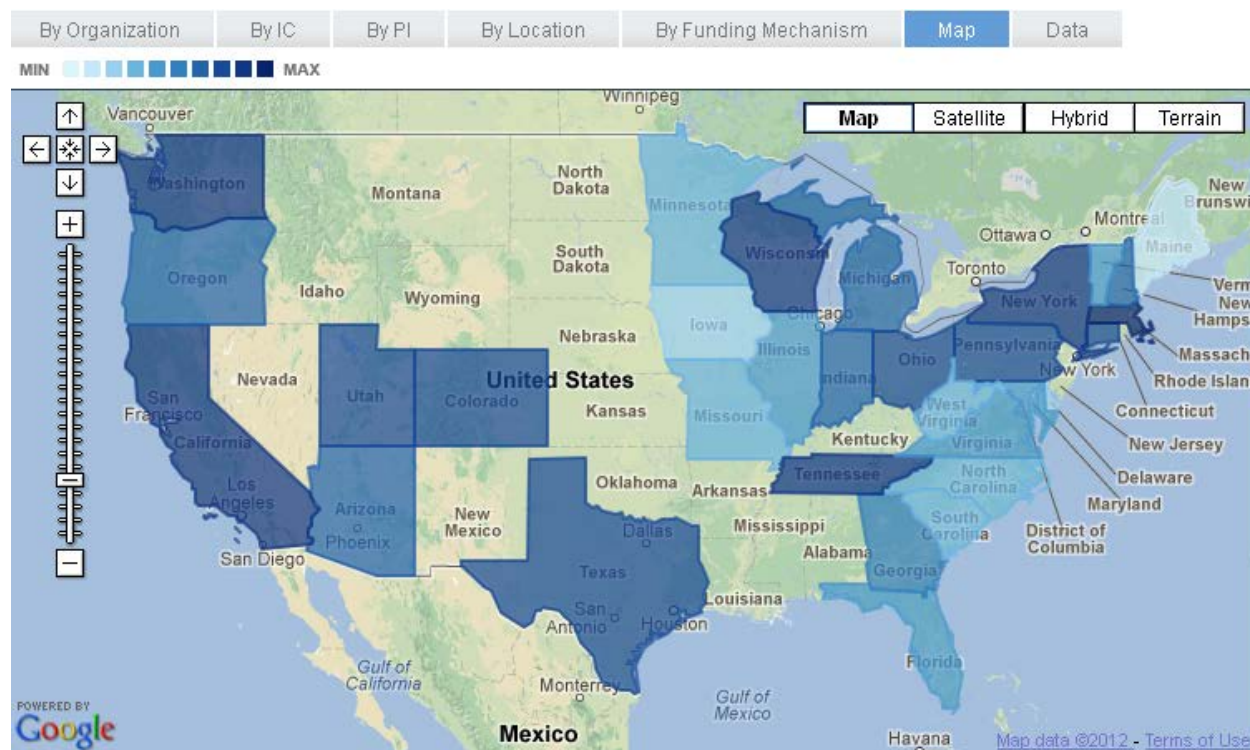
List of States with NLM Funding: Arizona, California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Illinois, Indiana, Iowa, Maine, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Hampshire, New York, North Carolina, Ohio, Oregon, Pennsylvania, South Carolina, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, and Wisconsin.

Table 6 and **Figure 6** show the way NLM's grant funds are distributed across the different types of grants mechanisms; 'other research' is the category used to represent Resource support, including NLM's Scholarly Works and Applied Informatics grants. **Appendix 5** provides a list of new projects funded in FY2012.

Table 6: FY2012 EP Grants Budget

Grants Mechanisms	$ Amount
Research Project Grants	$26,798,342
Research Project Grants (SBIR/STTR)	$768,445
Research Centers (Special)	$2,473,127
Other Research (Career)	$851,751
Other Research (Other)	$3,094,341
Training	$10,097,994
Contracts, LRP, IAAs	$197,000
Total:	**$44,281,000**

Figure 6: FY2012 EP Grants Budget

**FY2012 EP Grants Budget
$44,281**

Other Research 7%
Research Center 6%
Small Business 2%
Career 2%
R&D Contracts 0.4%
Training 23%
Research Project Grants 61%
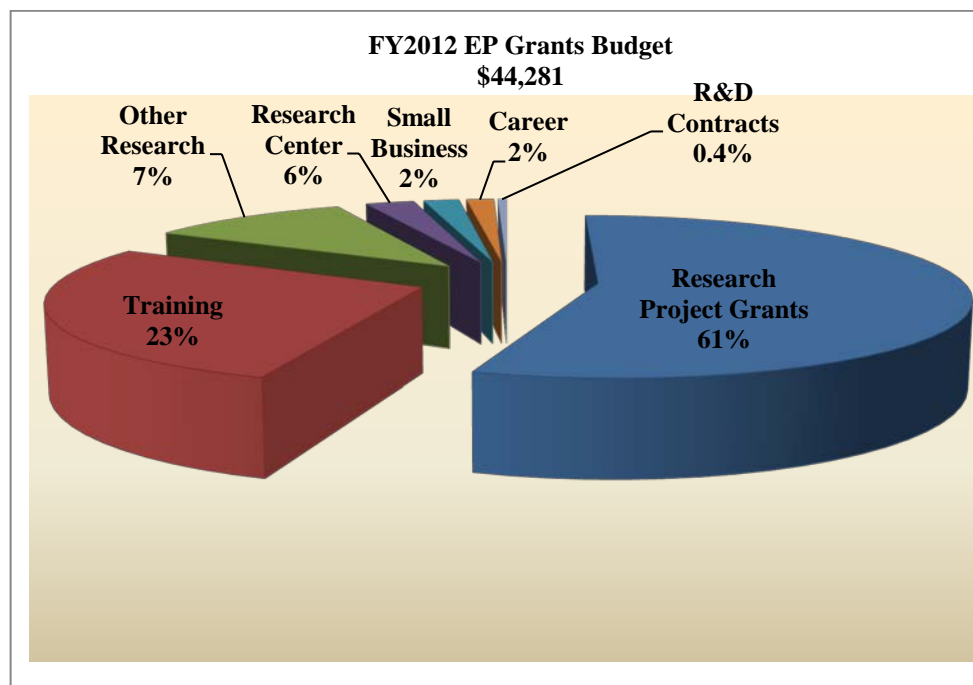
In 2011, EP began reporting grant expenditures using six Investment Area codes that reflect the topic scope of NLM's funded research. The Investment Areas are: Health Care, Bioinformatics, Translational Informatics, Public Health Informatics, Basic Informatics and User Sciences. **Figure 7** uses EP's Investment Areas to show NLM's FY2012 research grant expenditure patterns. The year-one cost per grant in the six areas varies from $199,228 (Basic Informa-tics) to $331,489 (User Sciences). Clinical Informatics remains the area most heavily supported, representing about 50 percent of NLM's research investments.

Figure 7: Research Grant Investment Areas, $ 7.6 million, 32 awards
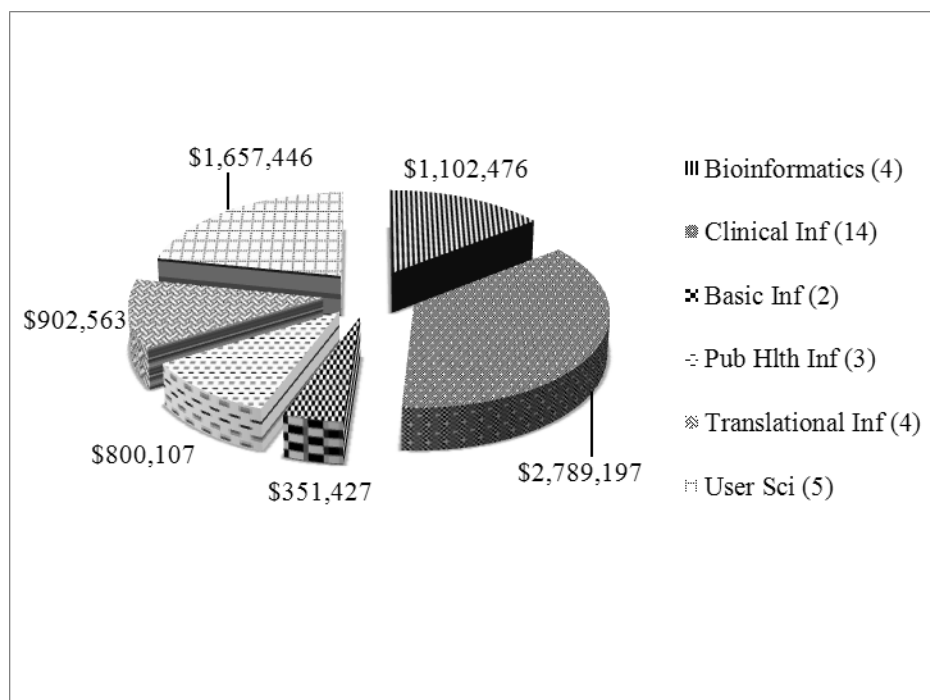
For all ARRA awards made in FY2009 and FY2010, expenditures must be completed by September 30, 2013. Consequently, the role of EP staff during FY2012 focused on monitoring and documenting the completion of NLM's portfolio of ARRA grants. Program officers requested annual progress reports from all ARRA grants that received multi-year funding, and grants management staff monitored spending patterns on a sample of the largest ARRA awards. A summary of NLM's ARRA funded projects can be found at http://www.nlm.nih.gov/recovery/.

NLM launched a new grant program in FY2012, the NLM Administrative Supplement for Informationist Services. This unique collaborative funding program provides research grantees of other Institutes with supplemental funds from NLM to add an in-context information specialist, usually a librarian, to their project team. With co-funding from two of the participating Institutes, eight awards were made, investing nearly $250,000 to support 15 librarian informationists on research projects funded by the following six Institutes: Aging (NIA), Cancer (NCI), Dental (NIDCR), Deafness (NIDCD), Eye (NEI) and NLM.

In a major administrative change, NIH Common Fund developed new models for award and management of its four high-risk/high-reward grant programs. Beginning in FY2013, applications received in the four programs will be assigned to Institutes and Centers for management and funding will be a mix of Common Fund and Institute funds across the life of each grant. The four programs are: NIH

Director's Pioneer Award, NIH Director's New Innovator Award, NIH Director's Early Independence Award, and the NIH Director's Transformative Research Award. In the future NLM might administer or co-fund grants in any of these programs. For FY2012, this change resulted in management by NLM of a second NIH Director's Pioneer Award grant, fully funded by the NIH Common Fund.

Outcomes and Impact of NLM Research Grants

A project by NLM Associate Fellow Susan Roy in early FY2012 provided a longitudinal view of the publication and citation patterns of research funded by NLM grants. Ms. Roy's study found that between FY2008 and FY2011, NLM grantees produced 2,723 book and article publications. Her study also highlighted important publishing differences between the clinical and bioinformatics research communities. For example, clinical informatics grants produced about 9 publications per grant, compared to 13 per grant for bioinformatics. The lag time to article publication was about 1.5 years from grant award date for clinical informatics and just around one year for bioinformatics.

In FY2012, NLM grantees published more than 300 articles in more than 120 different journals. The top four journals in which NLM grantees publish are JAMIA (29 articles); Journal of Biomedical Informatics (15 articles); PLoS One (10 articles); Bioinformatics (8 articles). An additional six journals had six articles apiece

that cited NLM support. Citations are a measure used at NIH to assess the near and longer term impact of grantee research. Citations to articles published in 2012 are few. However, the 266 NLM grantee articles published in 2011 have been cited 499 times. More than 6000 citations in SCOPUS cite articles that acknowledge NLM support.

Looking at articles published by NLM grantees whose grants have been active for the past three years, the most cited articles in clinical, translational and bioinformatics have different characteristics (**Table 7**). Clinical informatics articles typically have fewer authors per article and fewer grants acknowledged; the opposite is true for bioinformatics articles.

Table 7: Example Citation Patterns in Different Areas of Informatics

NLM Grantee	Journal	Article Title	Year Published	# Citations Since Published*	# Citations Recent 2 yrs*	Area
Schneeweiss, S and 3 others	Circulation	Cardiovascular Outcomes and mortality in patients using…	12/2009	99	80	Clinical
Moore, JH and 6 others	Genome Biology	Missing Heritability and Strategies for Finding the underlying causes of ….	06/2010	179	140	Translational
Salzberg, SL and 8 others	Nature Biotechnology	Transcript Assembly and Quantification by RNA Seq…	05/2010	372	258	Bioinformatics

(*does not include self-citations)

Other notable outcomes of FY2012 NLM extramural funding:

- NIH Director's Pioneer Awardee Dr. Markus Covert (Stanford University) published *"A Whole-Cell Computational Model Predicts Phenotype from Genotype"* the first complete computerized simulation of an organism, the cover story for Cell Vol. 150, issue 2, July 20, 2012, pp. 389-401. Dr. Covert's research is co-funded by the NIH Common Fund and NLM.
- Integrating Biology and the Bedside (i2b2), the National Center for Biomedical Computing funded and administered by NLM, has over 300 registered users of its toolbox. More than 84 organizations in the US and abroad have adopted the i2b2 tools to support clinical research using data from electronic health records.
- NLM grantee Dr. Joshua Denny (Vanderbilt University) received the AMIA New Investigator Award for translational informatics supported in part by his NLM research grant.

- NLM grantee Dr. Peter Embi (Ohio State University) received an AMIA Distinguished Paper Award for his NLM-supported clinical research informatics research on clinical trial alerts in electronic health record systems.

Success Rates of Grant Applicants

Success rates are computed by dividing the number of awards by the number of applications reviewed in a fiscal year. **Table 8** shows success rates in 2010, 2011 and 2012 for NLM's core grant programs for applications funded with appropriated funds. NLM's success rates continued to drop for most areas in 2012 due to an increase in the number of applications faced by a flat budget. Efforts to maintain the Research grant success rate at 15 percent or better were not successful. Success rates for all NIH Institutes and Centers are posted at http://report.nih.gov/success_rates/index.aspx.

Table 8: Success Rates - Core NLM Grant Programs FY2010 through FY2012

Core Grant Programs	Activity Code	2010	2011	2012
Research	R01	20%	18%	14%
	R21	27%	9%	11%
Career	K22	N/A	N/A	50%
	K99	100%	25%	26%
Resource	G13	20%	9%	8%

Research Grants

Extramural research support is provided through grant programs that fund investigator-initiated research. EP's research grants, funded with appropriated funds, support both basic and applied informatics projects involving the application of computer and information science approaches in clinical medicine, translational science, public health and basic biomedical research.

Research Grant Program

EP receives R01 research grant applications on three deadlines each year through the NIH parent announcement, NLM's Express grant program and various multi-institute NIH initiatives in which NLM participates. NLM's research investment areas include clinical informatics, public health informatics, bioinformatics, translational bioinformatics, consumer health informatics and information sciences.
- 16 awarded R01 applications (15 in FY2011)

Exploratory/Developmental Grants

EP receives R21 exploratory/developmental grant applications through the NIH parent announcement and various multi-institute NIH initiatives in which NLM participates. This program supports high-risk/high-reward projects, proof of concept and work in new interdisciplinary areas.
- 4 awarded R21 applications (3 in FY2011)

Conference Grants

Support for conferences and workshops (R13) is provided through the NIH parent announcement. NLM restricts its participation to small awards for scientific meetings in focused areas of biomedical informatics and bioinformatics. Applicants must obtain approval from EP program staff before they can apply.
- 2 awarded R13 application (1 in FY2011)

Small Business (SBIR/STTR) Research Grants

By law, all grant-issuing agencies set aside a portion of available research funds for Small Business Innovation Research (SBIR) grants. The 2011 reauthorization legislation instituted increases to the incremental percentage of the set-aside devoted to small business or technology transfer for each of the next six years, beginning in FY2012. This year's increments were 2.6% for small business and .35% for technology transfer. NLM's SBIR/STTR interests for FY2012 focused on modeling tools for climate and environmental effects on human health, new technologies for integrating patient reported outcome data from EMR for development of clinical trials and tools to improve EHR usability. In FY2012, EP received applications through the NIH parent announcement, and through a trans-NIH SBIR FOA titled "Small Business Alzheimer's Disease Research".
- 5 awarded small business applications (5 in FY2011)

NIH Director's Pioneer Award

Since FY2009, NLM has co-funded with the Common Fund an NIH Director's Pioneer award to Dr. Markus Covert at Stanford University for "A Gene-Complete Computational Model of Yeast". In FY2012, NLM became the administrative home of another NIH Director's Pioneer award, this one to Dr. Ram Samudrala at the University of Washington for "Novel Paradigms for Drug Discovery: Computational Multi-target Screening."

Administrative Supplement for Informationist Services:

NLM, in conjunction with seven other NIH Institutes and Centers, issued a funding opportunity announcement (FOA) for administrative supplements to provide funds to supported research and center grants (R01, P01, P30, P50 and U01) in order to enhance the storage, organization, management and use of electronic research data through the involvement of informationists. Nineteen applications were received and reviewed.
- 8 awarded informationist supplements

Resource Grants

Resource Grants use appropriated funds to support dissemination and management of health-related information. These grants are not research grants, and are reviewed with appropriate criteria. The G08 Resource grants support the development and deployment of knowledge management tools, resources, and services that address unmet needs for a broad audience. The G13 Scholarly Works grants support the preparation of scholarly manuscripts in health sciences, history of medicine and public health policy areas.

Information Resource Grants to Reduce Health Disparities (G08)

In March 2012, NLM reissued its request for applications (RFA) entitled Information Resource Grants to Reduce Health Disparities (RFA-LM-12-001) with an application deadline of May 12, 2012. The focus of the RFA was projects that bring useful, usable health information to health disparity populations and their health care providers through the use of computer and information science. Forty three applications were received in FY2012. The applications will be reviewed and awards will be made in FY2013.

Grants for Scholarly Works (G13)

NLM alone, among the NIH Institutes and Centers, provides grant funds to support the preparation of scholarly manuscripts. The first grants awarded by NLM in 1965 were for history of medicine projects. The Scholarly Works program continues to play a key role in important areas of biomedical scholarship, particularly in the history of science and medicine. Although the number of awards

remained the same as 2011, the success rate fell slightly due to a small increase in applications.

- 5 awarded Scholarly Works applications (5 in FY2011)

Training and Career Awards

NLM remains the principal US source of support for research training in biomedical informatics. EP provides both institutional training support and individual career transition support.

NLM's University-based Biomedical Informatics Research Training Programs (T15)

Five-year institutional training grants support pre-doctoral, post-doctoral, and short-term informatics research trainees at university-based programs across the country (see **Figures 8 and 9**). Fourteen new five year awards were made in July 2012, a reduction from the eighteen programs supported in the past. Two of the awards were made to new programs, at University of California San Diego and Ohio State University. Six existing programs (numbers 1,2,7,9,10 and 18 in Figure 4) did not receive new awards, but will offer one additional year of training support for early pre-doctoral trainees.

Collectively, the 14 continuing programs emphasize training in health care informatics (12 programs); bioinformatics and computational biology (12 programs); clinical research translational informatics (12 programs); and public health informatics (10 programs). The National Institute of Dental and Craniofacial Research (NIDCR) expanded support for trainees in dental informatics from two to nine trainees. They are located at the University of Pittsburgh, Oregon Health Sciences University and the University of Wisconsin-Madison.

Figure 8: NLM's University Based Programs

1. **University of California Irvine** (Irvine, CA)

2. **University of California Los Angeles** (Los Angeles, CA)

3. **University of California San Diego** (San Diego, CA)

4. **Stanford University** (Stanford, CA)

5. **University of Colorado** Anschutz Medical Campus (Aurora, CO)

6. **Yale University** (New Haven, CT)

7. **Regenstrief/Indiana University** (Indianapolis, IN)

8. **Harvard University** (Medical School) (Boston, MA)

9. **Johns Hopkins University** (Baltimore, MD)

10. **University of Missouri-Columbia** (Columbia, MO)

11. **Columbia University Medical Center** (New York, NY)

12. **Ohio State University** (Columbus, OH)

13. **Oregon Health & Science University** (Portland, OR)

14. **University of Pittsburgh at Pittsburgh** (Pittsburgh, PA)

15. **Vanderbilt University** (Nashville, TN)

16. **Rice University** (Houston, TX)

17. **University of Utah** (Salt Lake City, UT)

18. **University of Virginia** (Charlottesville, VA)

19. **University of Washington** (Seattle, WA)

20. **University of Wisconsin Madison** (Madison, WI)

Figure 9: FY2012 NLM Trainee Slots at Universities

FY2012 NLM Trainee Slots at Universities
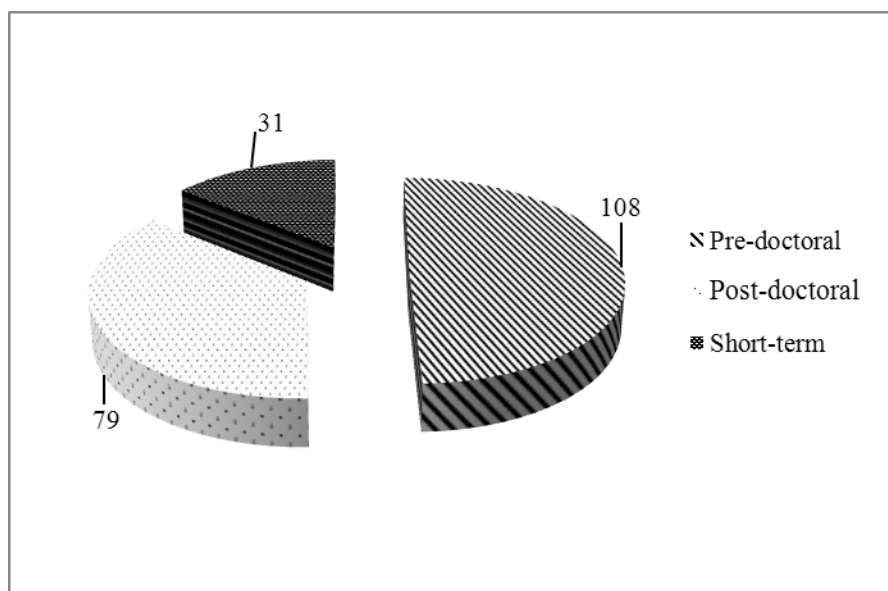
The annual NLM Informatics Training Conference took place in Madison, WI on June 26-27, 2012. Approximately 260 people attended, including directors, faculty, staff, and trainees from all 18 existing NLM training programs; training directors of two new programs funded July 1, 2012; faculty and trainees from the Veterans Administration informatics training sites; NLM staff; and invited guests. Trainee research projects were presented in plenary, parallel or open-mic sessions by 60 informatics trainees. An additional 42 trainees presented research-related posters at the meeting. Two Webinars were offered for attendees by NLM staff, with focus on career transition awards and training program fiscal administration. Attendees voted for best speaker, best poster and best open-mic talk. The award winners were:

- Best Open-Mic: Rimma Pivarov, Columbia University, "Prediction of Kidney Disease Outcomes Using Structured and Narrative Longitudinal Data".
- Best Poster Day 1: Daniel Gadala-Maria, Yale University, "Pipeline for Analysis of Immunoglobin Sequences on the Repertoire Scale".
- Best Poster Day 2: Perry Evans, Yale University, "Evaluating Melanoma Whole Exome Sequences Suggests New Driver Genes".
- Best Presentation Day 1: Krystal Haerian, Columbia University, "Pharmacovigilance and Drug Repurposing Model and Application: a Case Study".

- Best Presentation Day 2: There were two winners: Christine Peterson, Rice University, 'Building Inferring Metabolic Networks Using the Bayesian Graphical Lasso" *and* Timothy Imler, Regenstrief Institute, "Building the Next Generation of Medical Education: Case Based Online Interactive Learning from Drawing Board to the Conference Room"

Details of the agenda and a photo gallery are available at http://www.nlm.nih.gov/ep/trainingconf2012.html.

K99/R00 Pathway to Independence Awards

In January 2006, NIH announced a new career transition program, the NIH Pathway to Independence award (K99/R00), which combines a two-year mentored period with a three-year un-mentored research period. Although applications to this program are not restricted to NLM's informatics trainees, they are preferred applicants.
- 5 awarded K99/R00 applications (3 in FY2011)

K22 Independent Career Development Award

In FY2010, NLM re-launched its K22 early career award program. Several years of experience showed that a significant number of NLM's trainees, particularly those with MD degrees, were not applying for K99/R00 awards, so this program was reinstated to meet their needs.
- 1 awarded K22 application (0 in FY2011)

Early Recruitment to Informatics Research Careers

Summer Research Experience Program (R25)

Building on a similar successful ARRA program, NIH launched a new NIH Summer Research Experience Program in 2011, which provides a high quality research experience for high school and college students during the summer academic break. NLM limits participation in this program to its University-based informatics training programs (T15). Three awards were made in FY2011, to University of California Irvine, Vanderbilt University and Oregon Health Sciences University. A total of 24 summer research trainee slots were supported in FY2012.

Short Term Trainee Positions (STTP) to Enhance Diversity

Since 2007, NLM's Short-Term Trainee Program has focused on increasing diversity in the field of biomedical informatics by providing short term training experiences for under-represented minorities, and disabled or disadvantaged individuals. With the new awards, this program grew from nine slots to 31 slots offered at 10 different sites. Typically, STTP trainees are appointed for 3 month periods during the summer.

Pan-NIH Projects and Interagency Collaborations

NCBC

The National Centers for Biomedical Computing (NCBC) are cooperative agreement awards funded under the NIH Common Fund. As a result of an open competition for the second five-year period, seven NCBC centers were funded in 2010. NLM continues to administer one NCBC center, "Informatics Integrating the Bench and Bedside (i2b2)," based at Harvard University's Brigham and Women's Hospital, which received four years of additional funding in 2010. Each year, the NIH Common Fund provides a decreasing share of the funds for i2b2. NLM's share of support for i2b2 will reach 100 percent of the funding for i2b2 in FY2013. The National Heart, Lung and Blood Institute also provides co-funding support for i2b2. NLM program officers have scientific advisory roles in three NCBC centers: i2b2 at Harvard University, SIMBIOS at Stanford University and iDASH at University of California San Diego. I2b2 is the first center grant NLM has funded, and is one of EP's most productive grantees in terms of publications (23 articles cited this grant in FY2012) and software sharing.

Multi-institute Grant Programs

NLM participates in two types of multi-institute grant programs: general and topical. General programs such as the Academic Research Enhancement Award (AREA) grants, diversity and reentry supplements are fundamental components of NLM's overall grant program. NLM also selectively participates in topic-focused multi-institute funding announcements. The active multi-institute programs NLM participates in are listed in Appendix 1.

The applications for multi-institute programs are reviewed by the NIH Center for Scientific Review (CSR). Those that receive fundable priority scores are considered for awards alongside grants reviewed by NLM's study section. 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 within NIH

In FY2012, NLM provided co-funding support to an NIH Director's Pioneer Award, an STTR award administered by NHGRI, an International Research Training Award issued by FIC and the iDASH NCBC administered by NHLBI. These collaborative funding arrangements represent co-funding from NLM to other Institutes in the amount of $683,949. NLM received co-funding from the NIH Director's Office, NHLBI and NIDCR in the amount of $1,558,980.

Interagency Agreements and Funding

NLM grantee Dr. John Brownstein received $40,000 in supplemental grant funding support from the Office of the Assistant Secretary for Preparedness and Response (ASPR), DHHS, for a special health–related initiative relating to his HealthMap platform and data sources, aiding development of a tool to provide epidemiologic/disease outbreak information for hurricane impacted areas, resulting in improved situational awareness and timely analysis of disease threats.

Along with several other NIH Institutes, NLM provides support to the Protein Sequence Databank (PDB) at Rutgers University. NLM provided $197,000 in FY2012. The PDB, which is administered by the National Science Foundation (NSF), is the single world-wide repository for the processing and distribution of 3-D biological macromolecular structure data.

NSF and seven NIH Institutes, including NLM, issued a joint funding solicitation for research projects entitled "Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA). Hundreds of applications were reviewed by NSF study sections; reviewers used both NSF and NIH criteria for applications with a health or biomedical research focus. NLM chose two projects from this initiative to transfer to NIH for funding consideration in FY2013.

Extramural Programs Web Site
http://www.nlm.nih.gov/ep/index.html

In FY2012, there were 76,759 visitors to the EP Web site, an increase of 25 percent over the previous year. The EP Web site continued to have a substantial interest from

international visitors who represented almost 22 percent of total visits. The most frequently viewed pages were those describing the grant programs, EP Home page, information pages such as the NLM 2012 Funding Plan and Grant Deadlines, Frequently Asked Questions, and a page of links to NIH-produced tutorials for preparing grant applications.

According to the American Customer Satisfaction Index (ACSI) survey results, the overall satisfaction score for the EP Web site was 75 out of a possible 100, which is similar to the satisfaction ratings for the NLM main site.

The EP Web site was enhanced by the addition of a new Grantee Spotlight section. Seven grantee spotlights were posted during FY2012.

Outreach Activities: At the 2012 NLM Informatics Training Conference, EP program officers presented a Webinar, "Introduction to Career Development Awards and New Investigator R01 Grants", for trainees in or nearing the final year in their fellowships. Program staff organized a series of Informatics Lectures by NLM grantees. Dr. Jason Moore presented on "Machine Learning Approaches to the Genetic Analysis of Complex Traits," on November 2, 2011, at Natcher Conference Center, NIH. Dr. Michael Wagner presented on "Decision-theoretic Model of Disease Surveillance and Control and a Prototype Implementation for the Disease Influenza," on April 18, 2012, at the NLM Visitor Center. Dr. John Hurdle presented on "Nutritional Informatics: Integrating Real-time Dietary Patterns into the EHR," on June 6, 2012, at the same venue. These lectures were well attended by NLM and NIH staff.

Grants Management Office

NLM grants staff awarded $44,128,000 of extramural grants and agreements in FY2012 (**Appendix 3**). Grants management staff initiated a review of NLM's ARRA grants due to close in FY2013, with strong emphasis on the final Federal Financial Report submission and the final reconciliation of funds in the Payment Management System (PMS), DHHS and the NIH Office of Financial Management. This will be an ongoing review from FY2012 through FY2014 until all ARRA grants are fully closed.

Scientific Review Office

Grant Review Activities: Overall, 257 applications were reviewed for which NLM was the primary assignment. Of those, 216 were reviewed by NLM. The remaining 42 applications were reviewed by CSR including SBIR/STTR grant applications.

Biomedical Library and Informatics Review Committee (BLIRC): NLM's standing study section, BLIRC, evaluates grant applications assigned to NLM for possible funding. BLIRC met three times in FY2012 and reviewed 152 applications (as compared to 118 in 2011). The Committee (**Appendix 8**) reviews applications for most biomedical informatics and bioinformatics research applications, knowledge management/applied informatics, career support, and fellowships.

Special Emphasis Panels (SEPs): Four Special Emphasis Panels were held during FY2012 compared to 8 in FY2011. 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 direct conflict of interest exists between the application and a member of the BLIRC or when the number of applications received is simply too large for BLIRC to handle. Overall, NLM's SEPs reviewed a total of 64 applications during FY2012, compared to 180 in FY2011.

Review Concurrence: Concurrence with the results of initial review, called second-level review, is performed by the Board of Regents (BOR). The BOR Extramural Programs Subcommittee conducts early concurrence reviews electronically on the most fundable research grants, reviews program concepts and special initiatives such as the Training Grant RFA, and provides advice on selected applications. In FY2012, the subcommittee held two early concurrence panels; 22 grants were voted on and approved. The BOR conducts an en bloc vote for all applications assigned to NLM as primary or secondary Institute. For FY2012, a total of 4,220 NLM grant applications were included in the en bloc votes (428 primary and 3,792 dual assignments).

Appendix 1: RFA/PA Actions in FY2012

Table 9: RFA/PA Actions in FY2012

Announcement	Title	Expiration
PAR-11-208	NLM Express Research Grants in Biomedical Informatics and Bioinformatics (R01)	May 8, 2014
PAR-11-260	Research Project Grants (NIH Parent R01)	September 8, 2014
PA-11-261	Exploratory/Developmental Grants for Biomedical Informatics and Bioinformatics (NIH Parent R21)	September 8, 2014
PA-12-212	NIH Support for Conferences and Scientific Meetings (Parent R13/U13)	September 8, 2014
PA-12-006	Academic Research Enhancement Award (AREA) (Parent R15)	January 8, 2015
RFA-LM-12-001	NLM Information Resource Grants to Reduce Health Disparities (G08)	May 23, 2012
PA-12-158	NLM Administrative Supplements for Informationist Services in NIH-funded Research Projects (Admin Supp)	June 6, 2012
PAR-11-084	NLM Grants for Scholarly Works in Biomedicine and Health (G13)	February 24, 2012 (Reissued as PAR-13-014)
PAR-10-195	NLM Independent Career Development Award for Biomedical Informatics (K22)	May 8, 2013
PA 11-197	NIH Pathway to Independence Award (K99/R00)	September 8, 2014

Table 10: Multi-Institute Active Announcements in which NLM Participates

Announcement	Title	Expiration
PA-12-149	Research Supplements to Promote Diversity in Health-Related Research (Admin Supp)	September 30, 2015
PA-12-150	Research Supplements to Promote Reentry into Biomedical and Behavioral Research Careers (Admin Supp)	September 30, 2015
PAR-10-133/135	Understanding and Promoting Health Literacy Research Grants (R01) (R21)	May 8, 2013
PAR-10-136/137	Behavioral and Social Science Research on Understanding and Reducing Health Disparities Research Grants (R01) (R21)	May 12, 2013
	Advancing Novel Science in Women's	January 8, 2013

Announcement	Title	Expiration
PAS-10-226	Health Research (ANSWHR) Exploratory/Developmental Research Grants (R21)	
PAR-10-235	Climate Change and Health: Assessing and Modeling Population Vulnerability to Climate Change (R21)	May 25, 2012
RFA-LM-12-002 **RFA-EY-13-001**	NIH Basic Behavioral and Social Sciences Opportunity Network (OppNET) Grants	December 18, 2012 November 1, 2012
PA-12-088	Small Business Innovation Research (SBIR) Grant (R43/R44)	January 8, 2013
PA-12-089	Small Business Technology Transfer Grants (STTR) (R41/R42)	January 8, 2013
PAR-09-220	Innovations in Biomedical Computational Science and technology Initiative (SBIR [R43/R44])	September 8, 2013
PA-12-196	Innovative Health Information Technology for Broad Adoption by Healthcare Systems and Consumers (SBIR)(R44)	January 8, 2015
RFA-OD-12-003	Small Business Alzheimer's Disease Research (SBIR [R43/R44])	May 1, 2012

Appendix 2: FY2012 NLM New Grants Awarded for New Projects

Research Grants (R01)

ABERNETHY, NEIL FRANKLIN
1 R01 LM011180-01A1
UNIVERSITY OF WASHINGTON
Improving Network Analysis and Visualization for Infectious Disease Control

ARCHER, KELLIE J
1 R01 LM011169-01
VIRGINIA COMMONWEALTH UNIVERSITY
Informatic tools for Predicting an Ordinal Response for High-Dimensional Data

BUI, ALEX
1 R01 LM011333-01
UNIVERSITY OF CALIFORNIA LOS ANGELES
RUMI: A Patient Portal for Retrieving Understandable Medical Information

CARRELL, DAVID S
1 R01 LM011366-01
GROUP HEALTH COOPERATIVE
Scalable and Robust Clinical Text De-Identification Tools

CHEN, ELIZABETH S
1 R01 LM011364-01
UNIVERSITY OF VERMONT & ST AGRIC COLLEGE
Leveraging the EHR to Collect and Analyze Social, Behavioral & Familial Factors

GANZ, AURA
1 R01 LM011100-01A1
UNIVERSITY OF MASSACHUSETTS AMHERST
Mass Casualty Management System (DIORAMA-II)

GONZALEZ, GRACIELA HERNANDEZ
1 R01 LM011176-01
ARIZONA STATE UNIVERSITY-TEMPE CAMPUS
Mining Social Network Postings for Mentions of Potential Adverse Drug Reactions

KOYUTURK, MEHMET
1 R01 LM011247-01A1
CASE WESTERN RESERVE UNIVERSITY
Enhancing Genome-Wide Association Studies via Integrative Network Analysis

LAI, ALBERT M
1 R01 LM011116-01A1
OHIO STATE UNIVERSITY
An Information Fusion Approach to Longitudinal Health Records

MANDL, KENNETH D
1 R01 LM011185-01
CHILDREN'S HOSPITAL BOSTON
Active Patient Participation in a Disease Registry for Comparative Effectiveness

OCHS, MICHAEL F
1 R01 LM011000-01A1
JOHNS HOPKINS UNIVERSITY
Modeling Transcriptional Reprogramming by Markov Chain Monte Carlo Sampling

SHEN, LI
1 R01 LM011360-01
INDIANA UNIV-PURDUE UNIV AT INDIANAPOLIS
Bioinformatics Strategies for Multidimensional Brain Imaging Genetics

STATNIKOV, ALEXANDER
1 R01 LM011179-01A1
NEW YORK UNIVERSITY SCHOOL OF MEDICINE
Methods for Accurate and Efficient Discovery of Local Pathways.

ZENG, QING
1 R01 LM011334-01
UNIVERSITY OF UTAH
Assist Patients with Medication Decisions

ZHAO, ZHONGMING
1 R01 LM011177-01
VANDERBILT UNIVERSITY
Mapping the Genetic Architecture of Complex Disease via RNA-seq and GWAS Data

NIH Pathway to Independence Award (K99)

NAHM, MEREDITH
1 K99 LM011128-01A1
DUKE UNIVERSITY
Pioneering Health Information Quality (K99)

JIANG, XIAOQIAN
1 K99 LM011392-01
UNIVERSITY OF CALIFORNIA SAN DIEGO
Protection of Records: Privacy (PReP) Technology for Medical Research

JONNALAGADDA, SIDDHARTHA REDDY
1 K99 LM011389-01A1
MAYO CLINIC
Improving the Efficiency and Efficacy in Authoring
Essential Clinical FAQs

NHO, KWANGSIK
1 K99 LM011384-01
INDIANA UNIV-PURDUE UNIV AT INDIANAPOLIS
Integration of Bio-, Medical, and Imaging Informatics for
Complex Diseases

PENNATHUR, PRIYADARSHINI R
1 K99 LM011309-01
UNIVERSITY OF IOWA
Technology, Cognitive Work, and Patient Safety: A
Information-Oriented Model

Career Transition Award (K22)

UNERTL, KIM M
1 K22 LM011382-01
VANDERBILT UNIVERSITY
Developing New Approaches to Chronic Disease Care:
Collaborative Team Workflow

Conference Grants (R13)

BASHSHUR, RASHID L
1 R13 LM011263-01
UNIVERSITY OF MICHIGAN AT ANN ARBOR
Best Practices in Telemedicine Symposium-Workshop

UZUNER, OZLEM
1 R13 LM011411-01
STATE UNIVERSITY OF NEW YORK AT ALBANY
Challenges in Natural Language Processing for Clinical
Narratives

Exploratory/Developmental Grants (R21)

BLUM, JAMES MARLOW
1 R21 LM011026-01A1
UNIVERSITY OF MICHIGAN AT ANN ARBOR
Exploring the Feasibility of Computational Markers to
Predict Atrial Fibrillation

BURD, RANDALL S
1 R21 LM011320-01A1
CHILDREN'S RESEARCH INSTITUTE
A Paper-Digital Interface for Time-Critical Information
Management

EDWARDS, JOHN R
1 R21 LM011199-01
WASHINGTON UNIVERSITY
A Machine Learning Approach for Fine-Scale Genome
Wide DNA Methylation Analysis

JACQUEZ, GEOFFREY M
1 R21 LM011132-01A1
BIOMEDWARE
Exploratory Evaluation of Homomorphic Cryptography
for Confidentiality Protection

**Grants for Scholarly Works in Biomedicine
and Health (G13)**

GREENE, JEREMY A
1 G13 LM010890-01A1
JOHNS HOPKINS UNIVERSITY
When is a Medicine Good Enough? The Problem of
Therapeutic Equivalence, 1959-2009

MORABIA, ALFREDO
1 G13 LM010884-01A1
QUEENS COLLEGE
History of Epidemiology: Evolution and Contribution to
Medicine and Public Health

LANG, RAYMOND PAUL
1 G13 LM011224-01
LEONARD CHESHIRE DISABILITY
Global Institutions and the Politics of International
Disability

NATHANSON, CONSTANCE ALLEN
1 G13 LM010931-01A1
COLUMBIA UNIVERSITY HEALTH SCIENCES
Disaster and the Reinvention of Public Health in France

WILLIAMS, JOHN A
1 G13 LM010912-01A1
UNIVERSITY OF MICHIGAN AT ANN ARBOR
Implementation of a Pancreas Knowledgebase

**Small Business Innovation Research (SBIR) and Small
Business Technology Transfer (STTR) Awards (R41-
R44)**

BOCK, ROBERT
1 R43 LM011404-01
LUMENETEC, LLC
The Climate Change and Health Gateway (CCHG) for
Enhancing Biomedical, Health,

MAISEL, JAMES
1 R43 LM011165-01A1
ZYDOC MEDICAL TRANSCRIPTION, LLC
Applying NLP to Free Text as an EHR Data Capture
Method to Improve EHR Usability

KIM, YOUNG-SEOK
1 R43 LM011412-01
TACTUS TECHNOLOGIES, INC.
Mobile Cadaver Lab: An Innovative Platform to
Supplement Medical Education for Mo

SEIFERT, GREGORY JOHN
1 R43 LM011646-01
ADVANCED MEDICAL ELECTRONICS
CORPORATION
Tools for Coordination Among Caregivers of Alzheimers
Disease Patients

WEISS, JOHN CHRISTOPHER
1 R43 LM011408-01
DYNAMIC CLINICAL SYSTEMS, INC.
Research Platform Integrating Patient Reported and
Clinical Outcomes Data Sources

High Priority, Short Term Project Award (R56)

ZHENG, WENJIN JIM
1 R56 LM010680-01A1
MEDICAL UNIVERSITY OF SOUTH CAROLINA

Bridging Genomics and Medicine by Ontology
Fingerprints

**University Biomedical Informatics Research Training
Programs (T15)**

PAYNE, PHILIP RICHARD ORRIN
1 T15 LM011270-01
OHIO STATE UNIVERSITY
The OSU Clinical and Translational Research Informatics
Training Program (CTRIP)

OHNO-MACHADO, LUCILA
1 T15 LM011271-01
UNIVERSITY OF CALIFORNIA SAN DIEGO
San Diego Biomedical Informatics Education & Research
(SABER)

Appendix 3: FY2012 Operating Budget by NIH Mechanism and NIH Activity Code

Table 11: FY2012 Operating Budget Request by NIH Mechanism Groupings

NIH Grouping	No.	Total Amount
Research Project Grants (R01, R03, R21, R00, RL1, U01)	78	$26,798,342
SBIR/STTR (R41, R42, R43, R44)	5	$768,445
Other Research - Research Careers (K99, K22)	8	$851,751
Other Research - Biomedical Research Support (P41)	2	$1,285,652
Other Research - Other (G08, G13, R13, R25, D43)	26	$1,808,689
Training - Institutional (T15)	14	$10,097,994
R&D Contracts (L30, L40, N01, Y03)	1	$197,000
National Centers for Biomedical Computing Award (U54)	1	$2,473,127
EP Grant budget excluding TAPS and Operations	**135**	**$44,281,000**

Table 12: FY2012 Operating Budget Request by NIH Activity Code

NIH Activity Code	No.	Total Amount
D43: **International Training Grants in Epidemiology (cofund)**	-	$235,995
DP1: **Pioneer Award**	-	$396,000
G08: **Knowledge Management & Applied Informatics; Planning Grant for IAIMS**	7	$790,172
G13: **Scholarly Works in Biomedicine and Health**	12	$529,144
K99: **Pathway to Independence**	7	$698,960
K22: **NLM Independent Career Development Award for Biomedical Informatics**	1	$152,791
L30: **Extramural Loan Repayment Program**	-	$ -
L40: **Extramural Loan Repayment Program**	-	$ -
N01: **NN/LM Contracts**	-	$ -
P41: **Biomedical Resource Grant**	2	$1,285,652
R00: **Pathway to Independence**	9	$2,126,622
R01: **Research Project Grants**	61	$22,854,139
R03: **Small Project Grants**	1	$64,827
R13: **Conference Grants**	4	$71,436
R15: **Academic Research Enhancement Award (AREA)**	-	$ -
R21: **Exploratory/Developmental Grants**	7	$1,356,754
R25: **Education Projects**	3	$181,942
R41: **Small Business Technology Transfer (STTR)**	-	$ -
R42: **Small Business Technology Transfer (STTR)**	-	$89,000
R43: **Small Business Innovation Research (SBIR)**	5	$679,445
R44: **Small Business Innovation Research (SBIR)**	-	$ -
RL1: **Linked Research Project Grant**	-	$ -
T15: **University Biomedical Informatics Research Training Programs**	14	$10,097,994
U01: **Cooperative Agreement**	-	$ -
U54: **NCBC Roadmap Center**	1	$2,473,127
Y03: **Inter-Agency Agreement**	1	$197,000
EP budget excluding TAPS and Operations	**135**	**$44,281,000**

OFFICE OF COMPUTER AND COMMUNICATIONS SYSTEMS

Ivor D'Souza
Director

The Office of Computer and Communications Systems (OCCS) provides efficient, cost-effective computing and networking services, application development, and technical advice and collaboration in information sciences. OCCS provides some NLM services directly, but also indirectly supports the IT platform used by the NLM's research and management programs.

OCCS provides the NLM's backbone computer networking capacities, and assists other NLM components in local area networking; operates and maintains the NLM Computer Centers; develops software; and provides extensive customer support. OCCS helps to coordinate, integrate and standardize the vast array of computer services available throughout all of the organizations comprising the NLM. The Division also serves as a technological resource for other parts of the NLM and for other Federal organizations with biomedical, statistical and administrative computing needs.

Brief discussions of this year's activities in the following subject matter areas will be presented:
- Controlled Medical Vocabularies
- Consumer and Public Health
- IT Infrastructure Services
- Medical Literature Support and Document Delivery Services
- Outreach and Customer Services

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. The 2012AA edition of the UMLS Metathesaurus includes 160 source vocabularies in 21 languages. The Metathesaurus contains over 2.7 million concepts, a 12 percent increase from last year, with nearly 10.8 million names for those concepts. In the past two release cycles (2012AB and 2012AA), more than 38 English language and 20 translated sources were successfully updated or added.

OCCS developed additional features for the Metathesaurus release environment to produce an Active Subset of the UMLS for distribution to users. This subset is targeted at users who are interested in using the UMLS to access standard terminologies that are actively being maintained.

The Division implemented a Web service to support the needs of the Office of the National Coordinator for Health Information Technology (ONC), in the construction of Value Sets. Additionally, OCCS executed a feature enhancement of the Appendix pages to externalize and enable easy updating. This will avoid having to redeploy the entire suite when there is an update.

During each calendar year, the UMLS team produces two releases of UMLS for download and UMLS Terminology Services (UTS) for online access. Each release updates many existing sources within UMLS. Each source update goes through a rigorous editing cycle to ensure its content quality prior to publication.

Value Set Authority Center (VSAC) Project

Value Set Authority Center (VSAC) is a new project sponsored by and initiated with requirements from the Office of the National Coordinator for Health Information Technology (ONC) and the Centers for Medicare & Medicare Services (CMS). VSAC is a repository for storage, search, and retrieval, and will provide an authoring environment for creating purpose-specific subsets of standard terminologies known as value sets. Value sets are used in the specification of quality measures for healthcare, Health Level Seven, Inc. (HL7) messaging guides, patient reported outcomes measures, and artifacts, important for meaningful use of electronic health records (EHRs). Value Sets require an official authority center where value sets can be authored, curated and widely accessed in standard ways.

OCCS developed a public-facing Web application to allow users to authenticate, obtain documentation or self-help materials, browse and search for value sets, and to download or create report of value sets. The development involved a Web 2.0 front-end user interface accessing data from the backend purely through a data service. This data service is used to provide the Sharing Value Sets (SVS) RESTful (Representational State Transfer) Application Programming Interface (API) to retrieve Value Sets.

Project collaborators include ONC, CMS, the National Quality Forum (NQF), the Agency for Healthcare Research and Quality (AHRQ) United States Health Information Knowledgebase (USHIK) team, Mayo Clinic's Common Terminology Services 2 (CTS2) group, MITRE's Cypress Team, and a myriad of developers and vendors of electronic medical record systems.

OCCS demonstrated the beta release via Web meeting to our collaborators in September of 2012. The terminologies loaded in VSAC included:
- Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT);
- Logical Observation Identifiers, Names and Codes (LOINC);
- Current Procedural Terminology (CPT)/Current Dental Terminology (CDT);

- International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM);
- International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM);
- International Classification of Disease, 10th Revision, Procedure Classification System (ICD-10-PCS);
- RxNorm; Medical Subject Headings (MeSH);
- Clinical Vaccine Names (CVX);
- Source of Payment (SOP);
- Health Service Location (HSLOC);
- Race & Ethnicity.

VSAC establishes NLM as the authority center for value sets and increases our collaboration with many agencies and vendors through the promotion of standards such as SVS, CTS2, and REST.

International Health Services Terminology Standards Development Organization (IHTSDO) Project

The International Health Services Terminology Standards Development Organisation (IHTSDO) is a not-for-profit association that develops and promotes the use of Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) to support safe and effective health information exchange. The NLM is the US Member of the IHTSDO and distributes SNOMED CT.

During FY2012, OCCS worked with the Lister Hill Center, Library Operations and international collaborators to produce, document, validate, and publish:

- The first phase of SNOMED CT to International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) mappings, within which over 8,500 SNOMED CT concepts were individually mapped yielding a complete set of over 27,000 mappings;
- The first phase of SNOMED CT to International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) mappings for use within the United States, within which over 6,600 SNOMED CT concepts were individually mapped yielding a complete set of over 24, 000 mappings;
- The March 2012 edition of US Extension in Release Format version 1 (RF1), later followed by an draft Release Format version 2 (RF2) Snapshot edition;
- The September 2012 edition of US Extension in Full RF2, an RF1 compatibility package and an RF1 release from the published data using IHTSDO tooling. With this release, US Extension to SNOMED CT is fully compliant with the release structure of the international edition itself.

In addition, we developed and deployed:

- A US Extension to SNOMED CT modeling environment based on the IHTSDO "workbench" used to model the international edition of SNOMED CT. Based on the international toolkit and software archetypes, this series of projects is specifically tuned to support modeling of the US Extension and contains customizations to support US-specific use cases as well as integrations with the content request system.
- The US SNOMED CT Content Request System (USCRS), which is now the official gateway for SNOMED CT users within the US to suggest content, changes to the US Extension to SNOMED CT or the international SNOMED CT. The deployment contains feature enhancements, such as email notifications, configurable user settings, user management tools, and workbench integrations, within-application user documentation, training Webcasts, and demonstrated scalability with over 8,700 requests. Efforts to extend the platform to support international reuse are underway and should enable tight integration between the US Extension and SNOMED CT.

We also validated and published additional Kaiser Convergent Medical Terminology (CMT) subsets (e.g. Ear, Nose and Throat (ENT)/Gastro-Intestinal (GI)/ Infectious Disease (ID) Problem List Subset) and US Extension subsets (e.g. - Route of Administration Subset).

RxNorm Project

RxNorm, produced by NLM, is a standardized nomenclature for clinical drugs. We released eight versions of the RxNorm Editing System this year. These releases provide the editors of RxNorm data with additional tools and functionalities to perform their tasks and create new data elements for inclusion into RxNorm data releases. We implemented the ability to create and edit ingredient atoms for FDA Structured Product Labels (MTHSPL) ingredient data, including inert ingredients and their synonyms, which support allergy interaction data requested by the Department of Health and Human Services. We optimized modules for speed. We expanded string statuses for greater granularity in defining data, which improves the ability to edit and QA data, and added support for additional data items to encompass the RxTerms dataset.

In order for RxNorm to provide regular, dependable, and accurate representation of available drugs in the United States, we provide regular releases of the RxNorm datasets. We completed concurrent releases with the Department of Veteran Affairs (VA) National Drug Reference Terminology (NDF-RT) and the UMLS datasets. We created: 12 monthly regular and prescribable releases (24 releases), 52 weekly regular and prescribable releases (104 weekly releases); and streamlined and optimized weekly release processes.

Our goal this year was to improve RxNorm process performance by at least 10 percent. We were able to surpass this reducing our weekly release downtime by 60 percent and reducing the resynchronization process with Unified Medical Language System (UMLS) from over 12 days to eight days. We also reduced the monthly processing by 50 percent.

These performance enhancements were achieved during a year in which we also filled hundreds of requests

for system changes to accommodate new data functionality for editors.

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. Accomplishments included:
- The cutover activities for Descriptors and Qualifiers for 2013.
- A new version of the Global Citation Management System (GCMS) Extensible Markup Language (XML).
- Loading the Online Mendelian Inheritance in Man (OMIM) Terminology and the Genetics Home Reference (GHR) Terminology into the M2000 database.
- The MeSH SCR Cutover (a switch between 2011 and 2012 MeSH).
- The new MARC file for 2012 MeSH version.
- The new MeSH and Pharmacological Actions XML files, based on the 2012 MeSH.
- A new version of GCMS XML.
- The Post Authorization Module conversion of the satellite MeSH Applications.

We completed the 2012 cutover for all active MeSH Translation Management System (MTMS) languages and continued to provide operational support for all MTMS activities. The GCMS activities completed this year included the continued work on the interface changes of the MeSH Manual Tasks System with Other DCMS Elements Manual Tasks System. We also completed the production implementation for the 2012 Year End Processing (YEP) and the archiving of 2012 MeSH YEP Data. The YEP reports now include Supplemental Chemical Records (SCR) and Quality Control (QC) reports.

Consumer and Public Health

MedlinePlus and MedlinePlus Mobile

MedlinePlus released three versions this year, in June, November and December. These releases included an auto-complete feature, which will provide users with search term suggestions from the MedlinePlus dictionary and will help to improve user experience. We added new features and enhanced the display.

We released the MedlinePlus Web Service Application Programming Interface (API) in June, which makes MedlinePlus data more easily accessible from other applications. We also expanded the MedlinePlus Health Topics Web Service (API), offering users more flexibility in how they can implement solutions and leverage more MedlinePlus data for their own applications.

To ensure accessibility of our online products by all individuals, including those with disabilities, we are reviewing all of our applications and content to ensure accessibility. NLM products using Adobe Flash have been a particular concern.
- We identified that the MedlinePlus Input System Flash modules were not accessible, and we completed the framework for converting these for accessibility. We converted all MedlinePlus Body Map pages from Flash to JavaScript in order to make them Section 508 compliant. This also enables them to display effectively on Apple iOS-based devices.
- Since MedlinePlus' news video provider, HealthDay, announced they would be changing the video format they deliver to us (which will increase the resolution and usability), we selected JWPlayer as our new video player software. It supports both HTML5 and Flash on cross-browser platforms. In determining the best way to handle MP4 video with captioning, we decided to adapt the MP4 3GPP format (i.e. MP4 with embedded captioning). This format works well with JWPlayer on all browsers and, most importantly, it supports iOS devices such as iPad and iPhone.
- We updated the MedinePlus Search Cloud Widget Flash component, by adding keyboard navigation to it. This enhancement will allow users to navigate the Flash widget by using only a keyboard. We released the new version of MedlinePlus Search Cloud Widget in September of 2012.
- We ensured that all of our enhancements and modifications for both MedlinePlus and MedlinePlus Mobile are accessible and following Section 508 standards.

MedlinePlus saw over 745 million page views and over 213 million unique visitors. Additionally, MedlinePlus and Spanish MedlinePlus were both ranked in the top 10 for federal Web sites and both received a satisfaction score of 87 out of a possible 100 on the ACSI E-Government Satisfaction Index, a survey that is one of the most comprehensive and representative reflections of the citizen experience with federal government Web sites. There were over 100 production requests implemented this year.

"MedlinePlus Mobile: Consumer Health Information On-the-Go" article was submitted to IEEE's *IT Professional* magazine and appeared in the May/June issue of *IT Pro* magazine.

In order to improve the usability of the MedlinePlus Mobile Web site, a usability test hosted by General Services Administration's (GSA) First Fridays Program was conducted in February. We identified and resolved the top three usability issues. We released a new version of MedlinePlus Mobile in September, which included an enhancement to the global search results page by indicating MedlinePlus page type in the results. This will allow users to identify if the search results are a topic, drug, or an encyclopedia. We added a search box which lets users perform a search from any page. We also

created a new Drug A-Z index page, to simplify navigation.

MedlinePlus Connect

MedlinePlus Connect offers an Application Programming Interface (API) to provide external entities access to NLM health information services. Particularly, it offers to match the Uniform Resource Locator (URL) of MedlinePlus Topics pages to the International Classification of Diseases book 9 (ICD9)/Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) codes they supply. MedlinePlus Connect now also serves up MedlinePlus Web pages related to medications in the same manner by using RxNorm Concept Unique Identifiers (RxCUI) and National Drug Codes (NDC). We enhanced medication responses by integrating the RXCUI mapping file from The American Society of Health-System Pharmacists (ASHP) into the MedlinePlus Connect process. We also added the drug links to the MedlinePlus Connect Web service response, which helps users access the MedlinePlus content directly. We implemented additional security features in order to access MedlinePlus Connect mapping files. We set up a reporting feature to improve responses. We also implemented Spanish Support for Drug Information.

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 access to FDA-approved labels of medications.

The DailyMed batch processing system was migrated from a Solaris to a Linux virtual machine which resulted in cost savings, less physical hardware, ability to share resources with others, and easier upgrading and maintainence. The full zip downloads were split, and we added daily, weekly, and monthly update zip downloads to optimize SPL (Structured Product Labeling) downloads using DailyMed FTP (File Transfer Protocol) services. We also split the zip file into multiple sections, i.e. Human Rx, Human OTC, Animal, Homeopathic, and Miscellaneous (bulk ingredients, vaccines, etc.). We implemented Web services for DailyMed, which will save users and developers up to one hour of downloading time. We enhanced the search capabilities and updated the DailyMed Web pages to enhance view of Structured Product Labels (SPL) images.

DailyMed recognized more than eight million unique visitors, a 40 percent increase from last year, and served over 85 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, video, and narration appropriate for older Americans with low vision and/or low hearing, etc. Content includes general background information, open-captioned videos, quizzes and frequently asked questions. There are currently 60 topics available in NIHSeniorHealth. We redesigned the Web site this year and added functionality.

NIHSeniorHealth received over 8 million site hits and over 604,000 unique visitors, of whom over 26 percent were international visitors.

NLM Digital Repository Project

The NLM Digital Repository Project supports collection and preservation of a wide variety of digital objects including manuscripts, pamphlets, monographs, images, movies, audio, and other items. The repository includes digitized representations of physical items, as well as born digital objects. OCCS provides system architecture and software development resources to assist in the implementation and maintenance of the NLM Digital Repository.

During FY2012, OCCS supported the successful completion of the Medical Heritage Library project. OCCS added a new collection and doubled the number of books and videos preserved in the repository. All the films have Section 508-compliant captions and transcripts, and are fully searchable with the unique OCCS-developed Video Search software. We added 41 digitized historical films in a new collection named *Tropical Disease Motion Pictures*. OCCS optimized the video conversion process, generated video derivative files in several access formats, and ingested all video files and metadata into the repository. In June, OCCS demonstrated the Video Search software at the Institute of Electrical and Electronics Engineers (IEEE) Joint Conference on Digital Libraries (JCDL).

For the Medical Heritage Library (MHL) Project, NLM was obligated to provide approximately 5,700 digitized books to be made widely available in NLM's Digital Repository and the Internet Archive (IA), in order to meet the conditions of a grant from the Alfred P. Sloan Foundation to the Open Knowledge Commons (OKC), and then NLM. To meet this objective, OCCS continued to develop and refine the software "pipeline" that has enabled hundreds of MHL books to be processed each month. OCCS ingested 5,078 MHL books and 950,000 MHL pages this year. These items were ingested into the *Medicine in the Americas* collection, which now contains 6,443 books and 1,237,303 pages. We created a presence for NLM on the Internet Archive. We designed an approach for providing NLM-digitized materials to IA and implemented software to automate this approach. In FY2012, we transmitted 6,453 books to IA, completing

NLM's commitment to the MHL project. We continue to scan, process, and ingest new materials into the Digital Repository and send them to IA, at the rate of hundreds of books per month.

To improve operational response, OCCS developed a Digital Repository Monitoring Tool. This tool monitors the availability of each major component of the system on each virtual server, making it more efficient to ensure that the system is operational. Four new versions were released, which included adding support for ingesting, accessing and navigating a set of multiple volumes in an efficient manner (unique in digital library interfaces), and monitoring.

OCCS began ingesting a wider variety of content into the Digital Repository including MHL books scanned from microfilm by an external vendor, the "Heckler" report (Report of the Secretary's Task Force on Black & Minority Health), and NLM's new *Hidden Treasure* publication. In total, there was one new collection, 41 new videos, over 5,000 new books, and approximately 950,000 page images added to the repository. We used approximately 5.7 million files in ingesting new books and videos. In addition, *Digital Collections* recognized more than 76,000 unique visitors and over 1,060,000 page views.

Digitization Scan QA system

OCCS developed a Web-based quality assurance system for the NLM Digitization Program to automate, improve and streamline the quality assurance processes. We made enhancements to allow processing of multi-volume books. We added new alert messages to enable better error diagnosis. We have undertaken a review to streamline the process. This year, we processed 1,566 items with 337,947 image files.

Health Services Research Projects in Progress (HSRProj)

HSRProj contains descriptions of research in progress funded by federal and private grants and contracts for use by policy makers, managers, clinicians and other decision makers. It provides access to information about health services research in progress before results are available in a published form.

In FY2012, we updated the HSRProj Web site with over 1,500 new projects, which increased the total number of projects available for searching to 24,561. Additionally, we relocated 1,202 records with "Final Date" between 1/2007 & 12/2007 to the archive file.

PHPartners & HSR Info Central Input System

The Public Health Partners Web site (PHPartners.org) assists the public health workforce find and use information effectively to improve and protect the public's health. This is a joint project among US government agencies (e.g. the Centers for Disease Control and Prevention, Agency for Healthcare Research 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 Library of Medicine, National Network of Libraries of Medicine).

OCCS added new content, security enhancements, and new capabilities to the HSR Info Central Input System which included completing a new home page Web site. The new page has a new look with a HSRIC graphic banner and URL link graphics designed by the NIH Medical Arts department. We presented the new home page at the annual Health Research Meeting held in Orlando Florida in June, and we received very positive feedback. We added two new functions to the system. One allows the user to easily review suggestions submitted for PHPartners' Tutorial site. The second is a delete feature for the site, "Review Questions and Comment" form. Both features will assist with user efficiency.

This year, our accomplishments included the addition of new content to the Healthy People 2020 site pages, and the creation of the Leading Health Indicators (LHI) page. We added new capabilities to the Healthy People 2020 Structured Evidence Queries page, as well as new content to the Healthy People 2020 site pages. We continue to enhance our security measures in order to keep up with the latest cyber threats.

Comparative Effectiveness Resource (CER) Page

OCCS created a CER Web page. The CER page provides specialized searches of published research (PubMed), and research still in progress (HSRProj & ClinicalTrials.gov), to help inform investigations of comparative effectiveness.

IT Infrastructure Services—IT Security Initiatives

NLM continued to accomplished new security projects with preventive and active measures against new and emerging cyber-attacks that have the potential to impair NLM's ability to provide public information services. This year we engaged to meet the OMB Continuous Monitoring initiative and further strengthen our security posture.

HHS Trusted Internet Connection (TIC)

The NLM participated in the planning for the HHS Trusted Internet Connection (TIC) network implementation that will take place in FY2013. The OMB has mandated that all federal agencies reduce the number of connections to the Internet and pass traffic through TIC approved security centers. HHS is providing its own TIC network, being implemented and operated via CIT/NIH networks and HHS security facilities. NLM obtained approval to pass "unrestricted" data traffic through its own established Internet and Internet2 connections, and only "restricted" traffic needs to pass through the HHS TIC network. This will allow NLM to continue to meet its mission in the most efficient and cost effective manner with respect to communications and network services.

HHS CSIRC Security Device Installation Project

In order to achieve agency-level IT security goals for security event monitoring, detection, analysis, eradication and recovery with the Department of Health and Human Services (HHS) and NIH, the NLM has worked continuously with the HHS Computer Security Incident Response Center (CSIRC) and NIH Incident Response Team (IRT) and has successfully installed a high-speed intrusion detection system, a network monitoring system, and utilized the HHS-wide security information and event management system. As a result, NLM is now able to work more closely with the HHS and NIH security resources to promptly receive security alerts, improve speed and effectiveness of incident response, and better protect NLM IT from ongoing cyber threats.

The HHS requested that all qualified Trusted Internet Connection Access Providers (TICAP) meet department-level security controls. The HHS CSIRC team joined the NLM network and security teams to perform a review and attestation of the required security controls at NLM. After the completion of the Trusted Internet Connection Controls Attestation, NLM met all the security requirements and no remediation is required in order to be authorized as the Trusted Internet Connection Access Provider (TICAP) and continue to provide Internet I and II connections.

Web Site and Host Security

NLM must provide secure and trustworthy services for the general public by preventing application-level compromise from cyber-attacks that may damage our system and data integrity. We established a standard operating procedure that requires any Web site or host that will be open to the public complete both authenticated system and application scans to preemptively ensure there are no known security issues. In addition, new releases of existing Web sites or services must also go through the same process. The new procedure will further assure that no known vulnerabilities are accessible from outside of NLM.

NIH Security Projects Workgroup (SPWG)

The NLM Security workgroup participated with the NIH SPWG on the HHS/CSIRC ArcSight implementation, network segmentation standards, and Network Access Control (NAC) tool evaluations and implementation. In addition, NLM also presented desktop patching best practices to the group. The work promotes NLM's security program and collaboration with other NIH ICs and meets requirements from the OMB Continuous Monitoring initiative and other NIH-wide security compliance initiatives.

Tenable Security Center

In order to support the HHS Computer Security Incident Response Center and enhance NLM's security event analytics for detection and alerting of relevant security incidents in real-time, the NLM security team worked with all NLM divisions to fully adopt the new version of Tenable Security Center for our Vulnerability Assessment and Remediation program. In addition, the NIH Incident Response Team is able to share NLM's current vulnerability scan and network asset data with the NIH and view the NLM's Vulnerability Assessment and Remediation program, meeting requirements in the Office of Management and Budget's Continuous Monitoring initiative.

National Library of Medicine (NLM) Network Penetration Test

In order to identify any security blinds spots that may cause risks to our digital assets, NLM completed an annual independent internal penetration test. The overall objective of the assessment was to conduct an internal network vulnerability assessment to identify exploitable vulnerabilities on the internal network that expose NLM to risk. This work helps NLM to continuously advance the NLM security posture and proactively identify any security weaknesses in desktops, servers, network and security devices, and applications. In addition, NLM system administrators and application team leads remediated all reported vulnerabilities and learned positive practices from the reports and debriefs.

Desktop Security

For desktop security, NLM's automated patch management program applied over 170,000 patches on commodity desktops fixing known vulnerabilities to software.

More than 1,000 desktop computers were updated with over 20 million signatures. This is a 42 percent increase from FY2011 and demonstrates widespread and continuous threats across cyberspace. Windows PCs were kept safe by deploying 154 security patches and over 20 million antivirus signatures to each of over 900 Windows PCs. This activity was crucial in achieving excellent results in the annual penetration test where not a single Windows PC was breached by the penetration testers.

Network Server Security

All Windows, Linux and Solaris server vulnerabilities were mitigated as discovered. All servers were patched during this period with the latest patches available from Microsoft, Red Hat and Oracle/Sun. Also, we have completed remediation of security vulnerabilities discovered during the CIT penetration test and the NLM/CIT penetration test.

Implementing HSPD-12 to Improve Access Control to NLM Systems

Desktop and network security improved with the implementation of two-factor authentication requirements consistent with presidential initiative HSPD-12. In most cases the two factors are a physical card and knowledge of a PIN number. Two-factor authentication was first required for login to NIH enterprise applications, such as ITAS (the NIH Integrated Time and Attendance System), but is also now applied to login to Windows PCs. This authentication method makes it considerably harder to logon to a system with someone else's credentials. Most users are using their Smartcard (AKA PIV card – Personal Identity Verification) to log into their desktop systems and to enterprise systems like ITAS, and the remainder will join them by the end of the calendar year.

Mandatory Annual NIH Information Security Awareness Training

The Office of Personnel Management (OPM) requires annual security awareness for all staff who use IT resources. NIH officially announced the 2012 Annual Information Security Awareness Refresher in April 2012. NLM, with over 1,500 users, reached the targeted goal: to be one of the first three ICs to accomplish 100% completion each year before the NIH deadline. Additional role-based training for staff with significant security responsibilities also met the requirements for additional role-based training "every three years" in compliance with the HHS requirement. This training promotes awareness of the threats facing the NLM and best practices to avoid any security adversaries.

IT Infrastructure Services—NLM High Speed Communication Network

OCCS continually monitors and appropriately refreshes the NLM communications network to make sure that it stays in front of NLM requirements for increased demands for bandwidth, reliability and security.

Public Internet connectivity services to NLM continued to be provided through a newly awarded contract with Level 3. Internet connectivity was upgraded from 2 Gbps (Gigabits per second) to 3 Gbps, with an ability to quickly upgrade to 10 Gbps if needed. The circuit connects NLM to Level 3's Internet point-of-presence in McLean, VA. CIT and NLM have a peering arrangement where, in the event the primary NLM Level 3 circuit fails, NLM Internet services will be provided by automatically failing over to use the CIT/NIH Internet connections with Verizon and Time Warner. This failover capability is tested once per month to ensure high availability.

Internet2 is an important network resource for the connection of NLM and the research community. NLM/Internet2 connectivity is now provided by two 10

Gpbs links to the high-speed backbone network via the Mid Atlantic Exchange (MAX) at the University of Maryland. One connection is dedicated to handling traffic via the Commercial Peering Service (CPS) of MAX and another 10 Gbps link gives NLM an additional 10 Gbps connection to support large transfers of genome datasets and other biomedical information. These circuits also serve to provide redundancy in the event that one fails.

The effective useable bandwidth of the NIH Consolidated Collocation Site (NCCS) offsite computer center to Internet and Internet2 connections was increased by 100% from 1Gbps to 2 Gbps. The current bandwidth limiting factor is the CIT connection between their perimeter switches at the NCCS which is limited to 2 Gbps. They can increase this if required. Currently the NCCS/Internet2 link is 1 Gbps (over a 2.5G channel) and the NCCS/Internet link is 1 Gbps.

In FY2012, OCCS installed air-blown fiber optic backbone cabling for the NCBI and LHC networks in Building 38A. This extension of the existing NLM air-blown fiber optic backbone tube system facilitates and reduces the cost of installation of additional fiber optic cabling as new requirements arise.

OCCS increased the bandwidth and reliability of the private network that connects NLM off-site components located in Democracy 1 & 2, and Rockledge, to Building 38A by deploying Coarse Wave Division Multiplexing (CWDM) technology over leased fiber optic cables. This increased the useable bandwidth from 1 Gbps to 3 Gbps between Building 38A and Democracy 2, and from 1 Gbps to 2 Gbps between Democracy 2 to Democracy 1.

Implementing IPv6

To meet the US Office of Management and Budget mandate, OCCS continued the implementation of the Internet Protocol version 6 (IPv6) that runs in parallel with the older IPv4 protocol. In June 2012, NLM successfully participated in the World IPv6 Launch by permanently turning on IPv6 for the NLM Home Page, DOCLINE, LocatorPlus, DailyMed, and NIH Senior Health Web sites. Since then, IPv6 has been implemented on 136 of NLM's 208 major Web sites or applications. OCCS is an active leader in NIH's IPv6 implementation by providing support and contributing to the NIH IPv6 working group.

IT Infrastructure Services—Green Computing and Energy-Saving Initiatives

Server Virtualization

During FY2012, OCCS increased the use of virtualization and green computing of our server infrastructure by 11%. The number of virtualized servers has increased from 429 to 471. Virtualization is the foundation for our internal cloud strategy. This will enable rapid deployment of information systems, boost computing resource utilization,

and increase the energy efficiency of NLM's mission critical applications.

The virtual infrastructure allows for more efficient use of physical server hardware and data center consolidation. The number of physical servers has been reduced from 185 to 152, a 14 % reduction. The smaller number of physical servers will reduce energy costs, cooling costs, and our datacenter footprint. Power consumption has been reduced from 93 Kilowatts (kW) to 66 kW and OCCS server rack space usage went down from 878 to 695 rack space units occupied in the NLM Data center. The following charts show the progress on virtualization from FY2010 to FY2012, and corresponding power and rack space usage during this same period.

Figure 10: OCCS Server Virtualization, FY2010 to FY2012
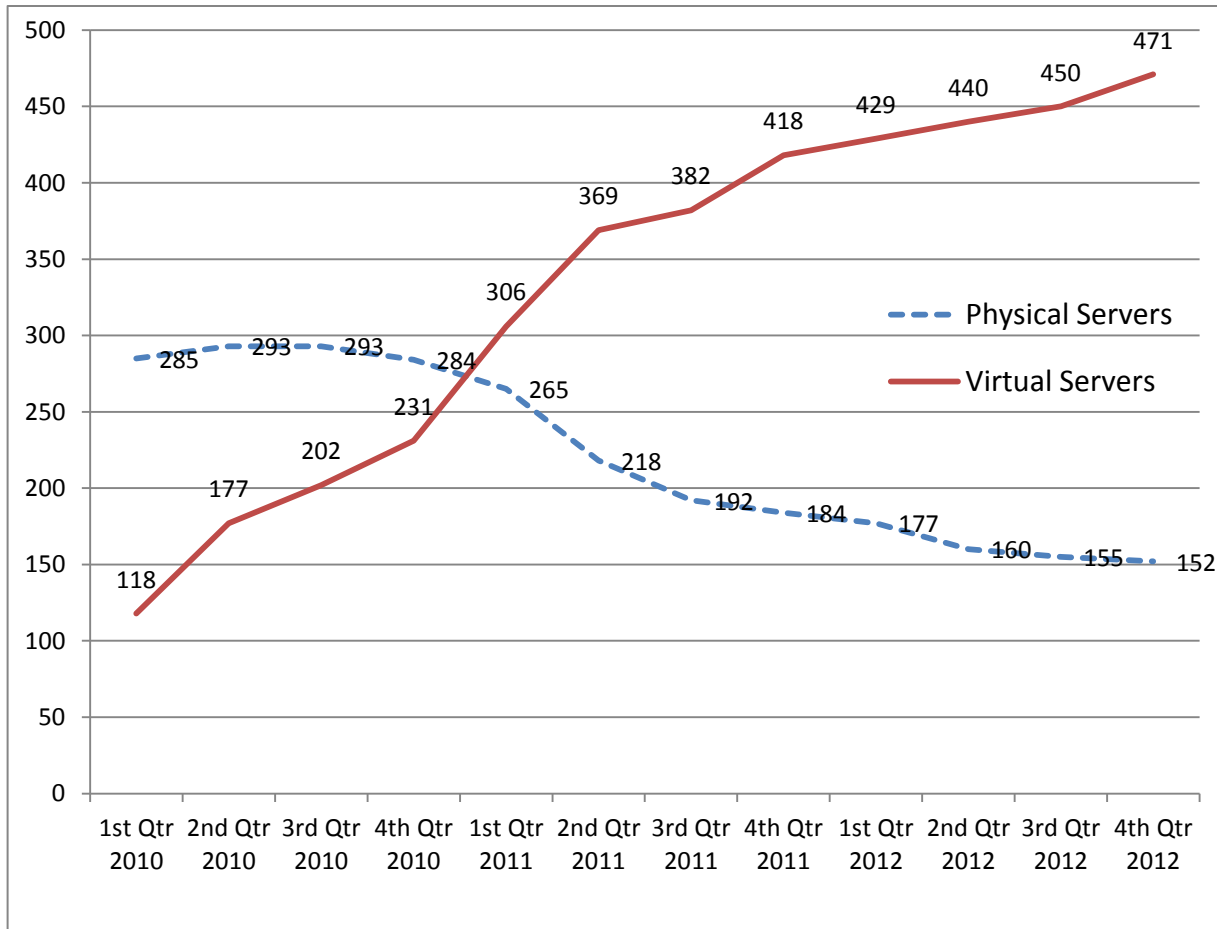
The chart below shows the power consumption reduction (FY2010-FY2012) corresponding with server virtualization:

Figure 11: OCCS Server Power Consumption FY2010-FY2012
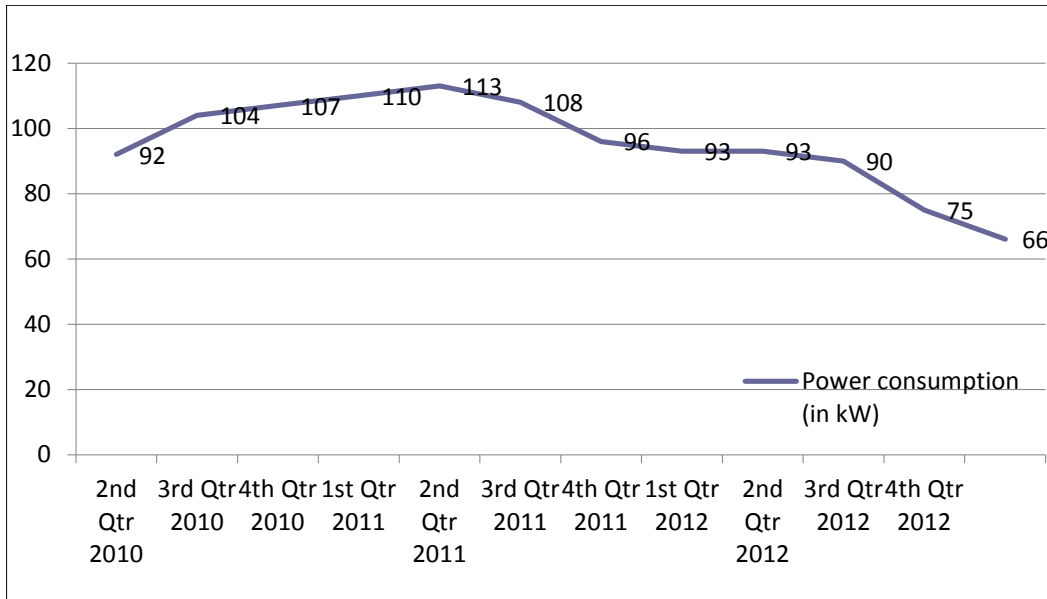
The chart below shows the corresponding reduction of rack space from FY2010 to FY2012 corresponding with server virtualization:

Figure 12: OCCS Rack Space Use FY2010-FY2012
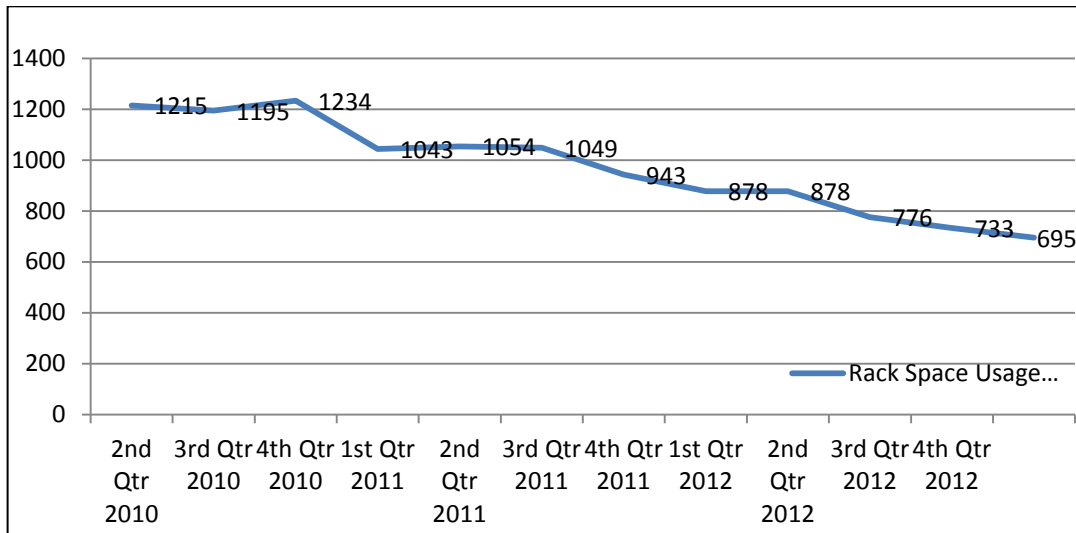
VMware Virtualization Infrastructure

We increased the capacity of virtualization environment (CPU and memory) at NLM and NCCS site by 25% and 35% respectively by upgrading older servers with newer generation of server hardware. Also, we upgraded the virtualization environment from VMware vSphere 4.1 to vSphere 5.0.

NLM Data Center Redesign

In collaboration with other NLM teams, the OCCS's Facilities Management Section (FMS) conducted a redesign of the legacy NLM Data center that will save over $200,000 a year on energy costs. By using industry-best practices, the redesign lowered the ratio of total facility power, which includes cooling, lighting and other costs, in contrast with the power consumed by the Information Technology gear alone. Known as the Power Use Efficiency (PUE) metric, this ratio focuses attention on power needs that are incidental to IT gear, such as cooling and lighting. The redesign led to Data center PUE dropping from 1.71 (2010) to 1.62 (2011), and in 2012, the PUE was reduced to 1.56. The redesign also produced a 33 percent increase of total Uninterrupted Power Supply (UPS) capacity, and a 150 percent increase in cooling capacity, allowing for future growth.

The redesign had several components, and both mechanical and electrical infrastructures were improved leveraging emerging technologies. For example, mixing of hot and cold air within the NLM Data Center was contributing to energy inefficiencies. A series of in-row air barriers were installed to block air flow across empty cabinet spaces and further isolated hot and cold aisle air flow. An exhaust duct system was installed throughout the Data Center to become the hot air plenum. To further minimize hot and cold air mixing, we also introduced perforated ceiling tiles above hot aisles to let hot air pass more quickly into the return plenum.

Also part of the redesign, we installed nine energy-efficient Stultz Computer Room Air Handler (CRAH) units, replacing five older Liebert CRAHs. These new CRAHs could be strategically aligned to serve the hot/cold aisle architecture. Coupled with separation of hot and cold air streams, the cooling systems are also operated more efficiently using higher chilled water supply and return temperatures. A new cooling control system contributes further energy savings by automatically adjusting the CRAH cooling speeds based on the heat load at the rack level.

In recognition of these ongoing infrastructure efficiency improvement efforts, the team was awarded the 2011 Health and Human Services (HHS) Green Champions Electronic Stewardship Award in April 2012.

IT Infrastructure Services—Desktop and Mobile Infrastructure Initiatives

The Desktop Services Section team was responsible for several projects and processes that included upgrading NLM computers from Microsoft Office 2007 to Microsoft Office 2010, implementing a Mobile Device Management process, establishing an IT Asset Loaner Program, development and implementation of a new Incident Management process, and evaluation and testing for deploying the 64-bit Windows 7 operating system among others.

Increased Savings and Improved Management of Mobile Devices

On November 9, 2011, the President signed Executive Order 13589–Promoting Efficient Spending. This Executive Order requires federal departments to submit plans to OMB for reducing the combined costs associated with travel, conference spending, employee IT devices, printing, executive fleet efficiencies and promotional items by not less than 20% below FY 2010 levels by the end of FY 2013. In an effort to fulfill this Executive Order, the Order Fulfillment team was tasked with streamlining NLM's accounts with various cellular carriers. The team was able to reduce wireless device cost by 38 percent from $25,077 to $15,220. This was accomplished by implementing a mobile device management process that increased the accountability of wireless devices and promoted more efficient use and sharing of devices.

Raised Service Levels by Improving Management of Incidents and Requests

The IT Service Center (ITSC) serves the NLM community through the processing and management of incidents and service requests and serves as the communication hub for broadcast messages across the Library. The ITSC is currently staffed by a three-person team of ITIL v3 certified Help Desk professionals. During FY2012, The staff managed over 26,000 e-mail messages and approximately 12,000 incident and service requests. During this same time, a new Incident Management logging and tracking tool, known internally as the Information Technology Service Management System (ITSMS), was introduced to support tracking and logging incidents. The ITSC team participated in the design, customization, and testing of the new system. Several OCCS teams collaborated to design the workflow of incident and service request routing within the ITSMS tool. The OCCS Facilities Management Section also participated and successfully adopted the (ITSMS) Incident Management workflow to promote consistent incident and service request response, and to ensure all parties are well informed. The ITSC staff will continue to spearhead efforts to solidify processes related to incident management to include communications during

catastrophic events. The education and training of OCCS staff continues to be at the forefront of these efforts.

Achieved Cost Savings by Conducting Training In-House

The Information Technology Service Center (ITSC) Team Lead provided Microsoft Office 2010 training to OCCS-managed program areas. The training was well attended by NLM staff and the trainer received high praise for sessions. This training was offered in-house and saved thousands of dollars.

Achieved Cost Savings through the Use of an IT Loaner Program

The OCCS Store IT Asset Loaner Program was developed and enhanced to meet the need of allowing organizations to use IT equipment for a short timeframe for testing, evaluation of equipment before committing to purchase, and while a primary computer is being rebuilt. The program offers an effective way to minimize lost productivity; it also provides cost savings and an efficient use of mobile devices and computer equipment.

Improved Performance From Efficient Use of Virtual Computing

Servers used for deploying patches and managing antivirus software on all Windows personal computers (PC) were transitioned to run as virtual servers on NLM enterprise hosts. The new antivirus software was transitioned to run on the virtual servers as they are more powerful, yet also save energy. The antivirus software version was upgraded to provide support for Windows 7 PCs and offers more thorough protection than previous versions thereby benefiting all 900+ client PCs, once fully implemented. Similarly, a secondary patching server was also upgraded with new software and virtual hardware, and offers more reliable performance and improved speed to ensure that patches can be delivered during non-business hours so as not to interfere with employee productivity.

IT Infrastructure Services—Web Tools

Web Analytics

NLM uses the WebTrends software package to track the number of pages served over time by the sites being managed and to provide detailed analysis of trends in site usage, audience composition, and other matters. This year, we deployed the page tagging data collection method on additional and larger NLM sites, such as MedlinePlus. This method is a Web analytics industry standard that gives richer and more accurate usage data. NLM will be using this method not only to improve insights on how online visitors navigate the NLM Web sites but also to comply with the Federal Digital Strategy guidelines released earlier this year.

Search Engine

We released the Spanish MedlinePlus Web service and the revised MedlinePlus Web service in June. The revision added links associated with Health Topic pages to the output of the MedlinePlus Web service and implemented field searching. This will allow users to narrow their queries to only the parts of the health topic record they are interested in searching.

Medical Literature Support and Document Delivery Services

Data Creation and Maintenance System (DCMS)

The major event for DCMS each year is the Baseline extraction, which is a re-release of all DCMS citations that follows the MeSH Year End Processing (YEP). The team also implemented the Gene Indexing Assistant (GIA), which will decrease the Gene indexing process time, and provide a potential cost savings to NLM, upwards of $80,000. We developed a new DCMS user interface function to support the GIA.

We changed the DCMS program to automatically index Commentaries. The Indexing Section is predicting a $350,000 saving in Contract Indexing costs per year starting with this fiscal year, which will be a five percent reduction from the $7.5 million spent last fiscal year. We have automatically indexed 19,233 comments, saving approximately $192,000 from the Indexing costs for this fiscal year.

We maximized the performance of the DCMS XML Extractor by running parallel queries, which has given us a considerable performance gain.

As part of the annual baseline extraction of Meeting Abstracts, Lister Hill Center's (LHC) Medical Text Indexer (MTI) provided more than 133,000 records loaded and re-released to the Gateway system, a 10 % increase over the prior year. The 2012 MEDLINE®/ PubMed® baseline database contains over 20 million records, a 9 % increase for data exported over last year.

OLDMEDINE records have out-of-date MeSH terms. This project's goal is to map OLDMEDINE terms to current MeSH terms. We switch records from a status of OLDMEDLINE to MEDLINE once all terms on any OLDMEDINE record are mapped. To date, over 1.6 million of the over 1.9 million records have been completed and switched to MEDLINE.

DOCLINE

DOCLINE® is the National Library of Medicine's automated interlibrary loan (ILL) request routing and referral system. The purpose of the system is to provide efficient document delivery service among libraries in the National Network of Libraries of Medicine (NN/LM). ILL requests processed by the National Library of Medicine for

other libraries are also entered, sent to Relais, and tracked through this system.

Loansome Doc enables individual PubMed® users to order documents found in MEDLINE from libraries. It is available to users in the US and internation-ally. A user can order articles from a list of citations retrieved from PubMed by sending requests to a library for the full-text documents.

We released several new versions this year, including upgrades to application security and to accessibility in order to pass acceptance tests for Section 508 standards, for both Loansome Doc and DOCLINE. We added Odyssey as a delivery method. We modified the interface to alert borrowers more clearly about copyright law. Moreover, we added a yearly National Print Retention report. PubMed versioned articles can now be ordered. We made modifications to speed up the Loansome Doc request status display. We added support for Internet Explorer 9 (IE9).

We reengineered DOCLINE to handle the subset of requests to NLM that are not entered into DOCLINE by the requestor.

Last fiscal year, Interlibrary Loan requests were manually entered into Siebel as Service Requests and then manually entered into DOCLINE. This year, Interlibrary Loan requests came directly into Siebel through the new ILL form. The standard format allowed much faster entry into DOCLINE.

Voyager Integrated Library System (ILS)

We completed the Year-End Processing (YEP) of the Voyager database in November of 2011. There were changes made to improve the integrity of data authority headings, imprint data, series statement, special producers' information, and other fields. We created baseline files of all records in both MARC and XML formats.

The OCCS Voyager team is working on a project to synchronize the NLM authority file with the Library of Congress' authority file. TSD catalogers have previously done this complex process manually.

Print Journal Donation Tracking System

In 2009, NLM launched a Journal Donation System that helped DOCLINE libraries determine if NLM can use any volumes from the print journals they plan to discard. This year, enhancements were introduced that now also make it possible for non-DOCLINE libraries to donate, and the user interface was improved to meet Section 508 accessibility standards. Since its launch, a total of 8,230 gifts have been added to the NLM collection using this system.

Literature Selection Technical Review Committee (LSTRC)

The Literature Selection Technical Review Committee (LSTRC) application records current and historical data concerning committee review of journal titles in the health and biomedical sciences. We implemented an application enhancement to support administrators in selecting the review date for journals and assigning journals to LSTRC committee members for review.

A MEDLINE review application form is required for all journals initially reviewed by the Literature Selection Technical Review Committee for inclusion in MEDLINE. Completed applications are accessible in LSTRC administrative.

LSTRC committee members use the MEDLINE reviewer Web form. This year we added additional search categories and new data fields to capture additional evaluation criteria. We also revised existing data fields to reflect changes in publication trends. We addressed Web accessibility for all pages.

E-Resources

The Databases, Resources & APIs database provides one place to easily find all NLM electronic resources including datasets, biomedical databases, application-programming interfaces (APIs), Web-based and mobile applications, and hyperlinks to informational tools. The administration system allows MEDLARS Management Section (MMS) staff to manage the growing number of resources NLM offers and keep the database up-to-date.

OCCS created the administration system and added tasks for administrators. We made several enhance-ments to the ability of administrators to add and edit subject terms. We also added Search terms. We have also made all pages Web accessible. We updated the public site to enable searching by category, subcategory, and subject term.

Images from the History of Medicine (IHM)

The Images from the History of Medicine system provides access to the nearly 70,000 images in the prints and photograph collection of NLM's History of Medicine Division (HMD). The collection includes portraits, pictures of institutions, caricatures, genre scenes and graphic art in a variety of media, illustrating the social and historical aspects of medicine. New images and new bibliographic data were loaded into Luna several times this year. This year IHM recognized more than 360,000 unique visitors and over 18 million page views to its site.

HMD Directory

The History of Medicine Division (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. We have made every effort to make certain that all pages are accessible by people of all abilities and disabilities.

We converted the Emergency Preparedness and Response mobile Web site into an iOS application for HMD. This application was tested and accepted by HMD and released to the public as "LibraryFloods" on March 27.

Serials Extract File (SEF)

The Serials Extract File is a critical internal product that serves as a bridge between other applications. Among numerous upgrades and fixes, we added XML citation provider's data to SEF. This is a new process where the data is provided by National Center for Biotechnology Information (NCBI), for weekly update in the SEF. We modified the Serials Viewer and Manage Serials Form, to make this data available to Library Operations (LO) users. We updated all serials records from Voyager (ILS) data in January. Processing was created and data was added for new SEF data field "Serials Record Flag." This field is used by the LO staff to indicate which serial e-resources are known to be available at no charge. We updated approximately 6,500 serial records from Voyager. We also added new data fields of Medical Text Indexer First Line (MTIFL) and MTIFL Timestamp. MTIFL is software that analyzes journal article titles and abstracts to suggest MeSH terms, which are then revised by an indexer. Certain journals perform better with MTI suggestions than others. We generated the 2012 List of Serials Indexed (LSI) publication in XML format for NLM File Transfer Protocol (FTP) server.

NLM Technical Bulletin Index

Staff in BSD/MMS (Bibliographic Services Division/ MEDLARS Management Section) will use the list to create a controlled vocabulary for searching the NLM Technical Bulletin. This new vocabulary will improve the search experience for public users who access the NLM Technical Bulletin and will greatly reduce NLM staff time necessary to add index terms to individual articles in TeamSite.

Outreach and Customer Services

OCCS staff continued to support the NLM Outreach Program through volunteering and providing technical support to over 30 outreach events, some of which are described briefly below. In addition, OCCS provided active support for several NLM exhibition initiatives.

Native Voices Exhibition

OCCS enhanced the *Native Voices* Web site and implemented an iPad application. The Web site and iPad application provide information about concepts of health and medicine among contemporary American Indians, Alaska Natives, and Native Hawaiians. They explore the connection between wellness, illness, and cultural life through a combination of interviews with Native people, artwork, objects, and interactive media. They honor the Native tradition of oral history and establish a unique collection of information.

Native Voices Web site visitors can now explore all of the *Native Voices* video interviews also available at the exhibit kiosk. Six new videos and enhancements that show the creation of the healing totem and its journey to NLM were also added, along with a video of the Opening Ceremonies. The iPad application provides an easy way for users to watch an overview of the exhibition or individual story that demonstrates how the Native people of today in the United States enhance their wellness through both traditional and Western healing practices.

Web Exhibit

Web Exhibit activities included the continuation of the redesign for Consumer Health: Outreach Projects Database (OPD). We created several User Interface Designs to be able to better understand the presentation of content and create reports more easily. We held discussions with the stakeholders to come up with ways to avoid redundancy between data entries across related applications such as Outreach Activity Reporting Form (OARF), Consumer Health, and Web Exhibits. We also incorporated the Search Functionality to be much more efficient and accurate.

Exhibit Asset Manager

We created the Exhibit Asset Manager database for the History of Medicine Division (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. We added a new Tribal Search feature this year. We also created several registrar reports.

OCCS and Other NLM Volunteers Staff the NLM Booth at PowWows and Other Conferences and Outreach Events

Outreach conferences and exhibits increase awareness of the NLM and its many goals. During FY2012, at least 26 conferences and 12 Native American PowWow events had an NLM presence—events attended by over 240,000 visitors. The NLM Outreach Program has been successful at getting the news to those who otherwise may not learn about all that is offered through the many Web sites hosted and supported at the NLM.

Customer Service Support System (Siebel)

Siebel software allows NLM to manage customer-facing communications. The NLM Customer Service Team monitors and responds to questions from medical libraries and the public that arrive into Siebel from the NLM "Contact Us" page. We added new users and projects. There were upgrades to the Siebel package, servers, and database environments.

Mobile Web Site—App of Apps

This project strives to make it easier for users of all mobile platforms to find mobile-friendly applications and resources developed by the NLM. OCCS developed two proofs of concepts that were accepted and which underwent comment and review by the NLM mobile users. The first was written using jQuery Mobile and the second using Sencha Touch. Based on feedback gathered from the two prototypes, the committee selected a final look, feel, and feature set. The selected technology is JQuery Mobile, due to its accessibility on a broad range of devices. We launched the App of Apps mobile Web site this year.

NIH Combined Federal Campaign (CFC) Web Site

As the NLM took a turn in FY2012 leading the NIH-wide CFC campaign, OCCS contributed to the creation and maintenance of the NIH CFC Web site. We also created a photo contest viewing and voting application in support of the NIH IC Directors' Challenge. It was released from October 31 through November 4 and had over 4,000 participants from the NIH community. Throughout the course of the CFC, we supported the advisory committee by updating content, publishing photos and adding new features to the pages as needed.

ADMINISTRATION

Todd D. Danielson
Associate Director for Administrative Management

Table 13: Financial Resources and Allocations, FY2012
(Dollars in Thousands)

Budget Allocation

Program Area	Amount
Extramural Programs	$62,411
Intramural Programs	$260,656
Library Operations	$75,001
Computer and Communications Systems	$30,429
Lister Hill National Center for Biomedical Communications	$47,274
National Center for Biotechnology Information	$88,299
Specialized Information Services	$19,653
Research Management and Support	$14,255
Total Appropriation	$337,322
Plus: Reimbursements	$46,283
Total Resources	**$383,605**

Personnel

In October 2011, **Jeffrey Reznick, PhD**, was selected as the Chief of the History of Medicine Division (HMD), replacing Dr. Elizabeth Fee, who was named Chief Historian at NLM. Dr. Reznick has served as senior curator at the National Museum of Health and Medicine, Executive Director at the Orthotic and Prosthetic Assistance Fund, and also assistant director of Emory University's Institute for Comparative and International Studies. From November 2009 to March 2011, he served as the deputy chief of HMD, and since April, was the Acting Chief. As a social and cultural historian of medicine and war, he is author of two books as well as numerous book reviews, journal articles, and entries in reference works. He received his MA and PhD from Emory University and his BA from the University of Rochester.

In October 2011, **Tomasz Adamusiak, MD, PhD,** joined the Cognitive Science Branch (CGSB) at the Lister Hill National Center for Biomedical Communications (LHNCBC). Dr. Adamusiak received his medical and doctoral degrees from Medical University of Lodz, Poland in 2006 and 2008 respectively. His research interest is in development of biomedical ontologies. Dr. Adamusiak is working on quality assurance in LOINC using description logics for checking its consistency. He is mentored by Olivier Bodenreider, MD, PhD, branch chief, CGSB.

In December 2011, **Swapna Abhyankar, MD,** was appointed to a staff scientist position with the Computer Science Branch (CSB) at the Lister Hill National Center for Biomedical Communications (LHNCBC) after completing her clinical informatics postdoctoral fellowship at the Lister Hill Center, begun in February 2010. She has helped develop the NLM Personal Health Record system, and served as the clinical informatics lead on the development of a standard set of codes for the Newborn Screening Project. In addition, she has been active in normalizing and using the data in the rich and extensive MIMIC II database to answer clinical questions. Dr. Abhyankar received her BS degree with distinction from Stanford University in 1995 and her MD from the University of Michigan in 2002, and spent several years in private pediatric practice with a large group in Maryland as well as with Holy Cross Hospital. She will be working with Clement McDonald, MD, Lister Hill Center Director.

In January 2012, **Rainer Winnenburg, PhD,** joined the Cognitive Science Branch at the Lister Hill National Center for Biomedical Communications (LHNCBC). Dr. Winnenburg received his doctorate degree in October 2011, in bioinformatics from the Technische Universitat Dresden, Germany. His research interest is in clinical natural language processing. He is mentored by Olivier Bodenreider, MD, PhD, Branch Chief, CGSB.

In January 2012, **Heather Dobbins, PhD,** was appointed to a staff scientist position with the Cognitive Science Branch at the Lister Hill National Center for Biomedical Communications (LHNCBC). In this capacity, she oversees the submission and quality review process for the results database at ClinicalTrials.gov. Prior to joining the staff, she was the lead results analyst for the ClinicalTrials.gov task on the NLM Clinical Information Services contract and has specialized knowledge in the reporting of results in a structured results database. Dr. Dobbins earned a BS in physics and a BS in biology at the University of Maryland. She completed her PhD in neuroscience at the University of Maryland at Baltimore in the Department of Neuroscience, in the School of Medicine, and subsequently held a dual appointment as a Postdoctoral Research Associate and Howard Hughes Medical Institute Teaching and Learning Fellow in the Physics Department at the University of Maryland. Dr. Dobbins will be working with Deborah Zarin, MD, director of ClinicalTrials.gov and Rebecca Williams, PharmD, MPH, assistant director of ClinicalTrials.gov.

In March 2012, **Elizabeth Workman, PhD,** joined the Cognitive Science Branch at the Lister Hill National Center for Biomedical Communications (LHNCBC). She earned a PhD in biomedical informatics from the University of Utah in December of 2011. She will use dynamic text summarization of SemRep semantic predications, a process developed in her dissertational work, in the field of serendipitous discovery. In this pursuit, she will analyze the interactions of information seeking behavior, literature-based discovery, and dynamic

text summarization methodology. Her mentor is Thomas Rindflesch, PhD, CgSB.

In April 2012, **Raymond Francis Sarmiento, MD,** joined the Office of High Performance Computing and Communications in the *Lister Hill National Center for Biomedical Communications (LHNCBC)* as a postdoctoral fellow. Dr. Sarmiento earned his medical degree in 2008 from the University of the Philippines, College of Medicine in Manila. After graduation, he joined the National Telehealth Center as the coordinator for research and development. He also became the lead project coordinator for the Medical Informatics Unit of the University of the Philippines College of Medicine in the development of an EMR system for the Philippine Dermatological Society. His current research includes: developing a controlled medical vocabulary for several Philippine languages; optimizing askMEDLINE for use in low-resource clinical environments; using short messaging service in developing a medication reminder application for tuberculosis patients; and evaluating the impact of telemedicine systems on clinical process flow and patient outcomes. He began his fellowship at the Lister Hill Center in April 2012. His mentor is Dr. Paul Fontelo.

In July 2012, **Loren Frant** was appointed Deputy Chief of the Public Services Division (PSD), Library Operations. Ms. Frant came to NLM as an Associate Fellow in 2004. Following her associate year, Ms. Frant accepted a position as a systems librarian in PSD's Reference and Web Services Section, where she led a team of librarians delivering Web site redesigns and database improvements. She served as the technical lead for MedlinePlus and was then appointed head of the Health Information Products Unit in 2008, a position she held until selected as deputy chief. As the head of this unit, Loren led all MedlinePlus strategic decisions and directed the operations for a suite of important products and services, including MedlinePlus, MedlinePlus en español, MedlinePlus Connect, and MedlinePlus Web services. Ms. Frant is a skilled spokesperson who has presented on MedlinePlus and MedlinePlus Connect on numerous occasions to the Board of Regents, the NLM Online Users Group at the Medical Library Association, MLA, to other NIH Institutes, the Institute of Medicine, the HHS Web Council, and two Health Data Initiative Forums hosted by HHS.

In August 2012, **Kenneth Koyle** was appointed deputy chief of the History of Medicine Division in Library Operations. In this capacity, he shares responsibility for directing and managing all activities of the Division. Mr. Koyle is a retired Army officer with more than 25 years of service. From 2010 until July 2012, he served as Deputy Chief of the US Army Medical Department's (AMEDD) Center of History and Heritage at Fort Sam Houston,

Texas. He was executive officer of the Center and the sole active duty historian in the Army Medical Department, responsible with the chief of the Center for supervision of an 18-person staff of history, archives, and museum personnel and administration of a $2.1M annual budget, as well as historical research in support of AMEDD and the Office of the Surgeon General. Prior to his tenure at AMEDD, Mr. Koyle was a medical history fellow at the Uniformed Services University of the Health Sciences, where he received his master's degree in history. He holds a second master's degree in adult education from Penn State University.

In July 2012, **Dan Gerendasy, PhD,** joined NLM as the new chief of international programs in the Office of Health Information Program Development. Dr. Gerendasy came to NLM from the NIH Center for Scientific Review (CSR), where for the last eight years his responsibilities as a Scientific Review Officer included leading reviews of major international and global health research projects and training programs. His award-winning efforts in international and global health are wide ranging and include service as CSR's representative to the NIH International Representatives Forum, frequent briefings of foreign delegations on the peer review process at NIH, service on an Institute of Medicine working group on global health, coordination for the Director of CSR of international visits and visitors, and leadership of NIH committees and work groups focusing on global health to name but a few. Dr. Gerendasy is particularly interested in scientific capacity building in the developing countries. He received a BA in biology from Grinnell College, an MS in biochemistry from Northwestern University, and a PhD in molecular biology from the University of Arizona. His NIH-funded research program focused on molecular mechanisms of changes in synaptic function. Dr. Gerendasy spent 12 years at the Scripps Research Institute starting as a postdoctoral fellow and leaving as Assistant Professor before moving to the SUNY Health Science Center at Stony Brook, and then to NIH.

In April 2012, **Szilárd Vajda, PhD,** began his fellowship at the Lister Hill National Center for Biomedical Communications (LHNCBC). Dr. Vajda holds a PhD in computer science from Henri Poincaré University, Nancy, France. His expertise is in pattern recognition and machine learning applied to document image analysis and handwriting recognition. He has developed mobile applications for the Android platform. He comes to NLM with a broad-based research experience from: the Robotics Research Institute at the Technical University of Dortmund, Germany; Furukawa Electric Institute of Technology in Budapest, Hungary; Loria Research Center in Nancy, France; and the Indian Statistical Institute in Kolkata, India. Dr. Vajda is developing face recognition techniques for the Lost Person Finder project in CEB with Dr. Sameer Antani and Dr. George Thoma.

NLM Associate Fellows Program for 2012 - 2013

Diana Almader-Douglas received her MA in information resources and library science from the University of Arizona, Tucson, in 2011. Ms. Almader-Douglas was awarded a Knowledge River Scholarship, which focuses on educating information professionals on issues pertaining to Latino and Native American populations. While completing her degree, she was employed as a graduate assistant at the Arizona Health Sciences Library, where she provided document delivery and reference services. During her program, she also worked as a student affiliate at the Southern Arizona VA Health Care System's medical library, where she provided reference services and assisted with collection development projects. Ms. Almader-Douglas earned a Master of Arts degree in counseling from Webster University in 1998. As a licensed professional counselor, she assisted state-funded behavioral health recipients throughout her career. Ms. Almader-Douglas' undergraduate degree is from Northern Arizona University, Flagstaff, Arizona, where she majored in psychology.

Karen Gutzman received her MS degree in library science in 2012 from the University of North Texas with an emphasis in Health Informatics. While completing her degree, she worked as a research assistant with the National Children's Study, which is the largest study of children's health in the US She was a recipient of the Medical Library Association scholarship and she completed an internship with the University of Houston's digital library. Her background includes teaching biology and English in Liptovsky Mikulas, Slovakia and receiving a Master of Arts degree in educational leadership from Luther Seminary in St. Paul, MN. Ms. Gutzman received her undergraduate degree in biological sciences from South Dakota State University.

RoseMary Hedberg received her MLIS degree in 2012 from the School of Information at the University of South Florida, Tampa. While completing her degree, she worked as a graduate research assistant at the All Children's Hospital Medical Library and volunteered as an intern at the Moffitt Research Center Biomedical Library. She served as the student chapter president for the American Society for Information Science and Technology and was selected as the school's representative for the 2012 American Library Association Student to Staff program. Prior to completing her graduate studies, she worked at a sleep diagnostic and research center as a patient quality assurance coordinator. Ms. Hedberg received her undergraduate degree in psychology from the University of South Florida.

Kevin Read received his MLIS and MAS from the University of British Columbia (UBC), in Vancouver, Canada. While completing his degrees, Mr. Read worked as a student librarian at the Biomedical Branch and Science and Engineering Branch Libraries performing reference and instructional services. Mr. Read's most recent work has been collaborating with the pediatric palliative network, PedPalNET, at the British Columbia Children's and Women's Health Centre building, a strategy to provide health information to patients' families, and working with the British Columbia College of Physicians and Surgeons Library as a consultant to improve their Finding Medical Evidence training workshops. He has also worked with the Canadian National Aboriginal Health Organization and the UBC Centre for Healthcare Management to help improve their understanding and use of social media tools to promote health information services. Mr. Read received his BA with honors in Islamic art and architecture from the University of Victoria in Victoria, Canada.

Retirements

In June 2012, **Sheldon Kotzin** retired after nearly 44 years of federal service at NLM. Mr. Kotzin earned a Master of Library Science degree from Indiana University in 1968 and following graduation, came to the NLM as a Library Associate. He subsequently served as head, Catalog Maintenance Unit in the Technical Services Division, Head of the Collection Access Section (then Loan and Stack) in the Public Services Division, and coordinator of the National Network of Libraries of Medicine (then the Regional Medical Library Network). Mr. Kotzin became chief of the Bibliographic Services Division in 1981 and was appointed associate director for library operations in 2006. Since 1998, he had served as executive editor of MEDLINE and administrator of the Literature Selection Technical Review Committee, the body that reviews and recommends journals for indexing in MEDLINE. He also served as NLM's representative to the International Committee of Medical Journal Editors, a group of clinical journal editors who establish standards for submission of journal articles and comment on ethical principles related to publication in biomedical journals. The Medical Library Association elected him a fellow in 2007. In May, Mr. Kotzin received NLM's Frank Bradway Rogers Award and a Board resolution congratulating and thanking him for his outstanding career. At his June 14 farewell, he was presented with many tributes including a statement from Rep. Chris Van Hollen in the Congressional Record, May 30, 2012, that concludes "Mr. Speaker, I am pleased to represent Sheldon Kotzin in the U.S. Congress and ask my colleagues to join me in extending to him our gratitude and appreciation for his years of service to our country and his contributions to international scientific medical communication."

In July 2012, **Nadgy P. Roey** retired from her position as program advisor/ethics coordinator for the National Library of Medicine ending 32 years of federal government service, 23 of which was with the National Library of Medicine. Ms. Roey began her Federal career

in 1980 at the Office of Personnel Management serving in various roles. She joined the National Institutes of Health (NIH) Office of the Director in September 1986 as a personnel management specialist. In October of 1989 she joined the National Library of Medicine and served as a Supervisory personnel management specialist. In 1992, she became the National Library of Medicine's (NLM) personnel officer, a role she maintained until NIH centralized the human resources function. In 2002, she became a program advisor to the NLM Executive Officer and continued in her role as Ethics Coordinator. Her superior leadership and expert guidance managing the National Library of Medicine's Ethics Program was recognized by a 2005 NIH Merit Award. In 2008, she was once again recognized with an NIH Director's Award for her leadership and skill in the daily administration and coordination of an outstanding ethics program.

Departures

In May 2012, **Phillip D. Osborne** accepted a position as the director of the National Oceanic and Atmospheric Administration's National Capital Acquisition Division. Mr. Osborne served as the director of acquisitions for the National Library of Medicine since April 2, 2006. During his six years, Mr. Osborne's efforts have contributed immeasurably to NLM's successful mission accomplishments, and he capably steered the Library through numerous significant changes in HHS and NIH contracting policy and procedures. His contributions included the expansion of the acquisition office to cover the contracting needs of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the award of a large number of contracts using funding from the American Recovery and Reinvestment Act (ARRA), and stabilization of the NLM Acquisition Office's workforce.

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 recipient of the 2012 award was **Dr. E. Michael Gertz f**or his important contribution to the molecular characterization of two previously unrecognized immune diseases.

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 2012 award were **Dr. Marilu Hoeppner** in recognition of substantial contributions for the successful planning and coordination of efforts that have been critical for the rapid growth and success of NCBI's Electronic Bookshelf project, and **Mr. Sheldon Kotzin** for leading Library Operations in continually adopting process

improvements to provide biomedical information as effectively as possible to our users.

The NLM Director's Honor Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to **Dr. Michael Sappol** in recognition of his editorship and project management of the book, *Hidden Treasure: The National Library of Medicine*, and the special contributions of this book to the NLM's 175th anniversary year.

The Phillip C. Coleman Award was presented to **Angela B. Ruffin, PhD**, for being there when others need her and for helping NLM's staff fulfill its potential. The EEO Special Achievement Award was presented to **Laura L. Bartlett** for her outstanding efforts supervising and mentoring several diverse summer school students as well as working with stay-in-school and other student interns and teachers, and **Patricia R. Carson** for her continuing leadership and successful support of NLM's "Mentoring in Medicine" and "Adopt a School" programs.

The NIH Merit Award was presented to two individuals and one group. The individual recipients were **Ione C. Auston** for her vision, leadership, and invaluable contributions to NLM's activities in support of health technology assessment, effectiveness research, and the evolving field of comparative effectiveness research, and **Dianne P. Babski** for exemplary contributions that increase the visibility and use of NLM free data resources in support of the US government open data initiative. The NIH Merit Group Award was presented to the **Open1Group** members **Sameer K. Antani, PhD,** and **Dina Demner-Fushman, MD, PhD**, for leading the research, development, and launch of Open*i*, a novel system for searching the open access literature and retrieving relevant images and text.

The NIH Director's Award was presented to two NLM staff members as individual awards and eleven NLM staff as part of NIH group awards. Individual awards were presented to **Alan R. Aronson, PhD** in recognition of sustained excellence and outstanding leadership in the development of the Medical Text Indexer, which increases the efficiency of the indexing process, and **Ivor L. D'Souza** for leading significant expansion of NLM's computing and communications capacity, while simultaneously improving energy efficiency, containing costs, and directing development of innovative applications. The National Institute of Mental Health, National Database for Autism Research Team Group Award included **Michael F. Huerta, PhD**, for exceptional work in developing the National Database for Autism Research (NDAR). The Office of the Director, Genetic Testing Registry Policy Team Group Award included **Brandi L. Kattman, Jennifer M. Lee, PhD, David J. Lipman, MD, Donna R. Maglott, PhD, Adriana J. Malheiro, James M. Ostell, PhD, Wendy S. Rubinstein,**

MD, PhD, and **Janet B. Coleman** (Contractor), in recognition of exceptional dedication and outstanding performance in establishing the public policy foundation for the Genetic Testing Registry. Another Office of the Director, NCRR/NCATS/ORIP Reorganization Group Award included **Valerie Florance, PhD**, for the extraordinary vision, leadership, coordination, and cooperation in realigning NIH programs to advance translation at NIH. Additionally, The Office of the Director, Common Fund Global Health Leadership Team Group Award included **Dan D. Gerendasy, PhD**, for outstanding service in the coordination of the Common Fund Global Health Initiatives.

NLM Staff were recognized for their outstanding efforts by outside organizations and other HHS components as follows.

The Association of Academic Health Sciences Libraries (AAHSL) presented the Cornerstone Award to **Donald A.B. Lindberg, MD**, for a notable or important contribution made during the most recent four years that has assisted AAHSL in achieving its mission or had significant impact on the profession of academic health sciences librarianship. NLM Staff were elected to fellowships **James T. Case, DVM., PhD,** and **Suresh Srinivasan** in the American College of Medical Informatics in recognition of significant and sustained contributions to the field of medical informatics; and **Eugene V. Koonin, PhD,** was elected a fellowship in the American Academy of Microbiology in recognition of an exemplary career in basic and applied research, teaching, clinical and public health, industry or government service.

The HHS*innovates* Award "Secretary's Pick" was awarded to **Dwight M. Clarke, Wei Ma, Ivor L. D'Souza, Laura Lee (NIH/CC), Chandra S. Kola,** (Contractor), **Cindy Notobartolo** (Suburban Hospital), and **David Y. Zhang** (Contractor), for the Electronic Patient Tracking in Disasters designed to help hospitals manage high volumes of incoming patients in disaster situations. The HHS Green Champions Electronic Stewardship Award was presented to **Ivor L. D'Souza, Mehryar Ebrahimi, James W. Fleshman, PhD, Dimpal Patel, Don R. Preuss, Michael W. Simpson, James I. Snowden, Brian E. Szamborski,** and Contractors **Steven C. Geppi, Malee Kaolawanich (NIH/OD), Victor G. Previll, Salim S. Saah, Shreenath Shetty, and Bill Watts** in recognition of the Upgrade to the NLM Data Center for transforming a legacy, 30 year old data center into an energy-efficient data center, which has resulted in an annual energy savings of over $200,000.

Table 14: FY2012 Full-Time Equivalents (Actual)

Program Area	Count
Office of the Director	8
Office of Health Information Programs Development	5
Office of Communication and Public Liaison	10
Office of Administration	58
Office of Computer and Communications Systems	47
Extramural Programs	18
Lister Hill National Center for Biomedical Communications	67
National Center for Biotechnology Information	279
Specialized Information Services	44
Library Operations	264
TOTAL FTEs	**800**

Director's Education Fund

Established in 1998 by the NLM Director, Donald A.B. Lindberg, MD, the Director's Employee Education Fund continued to provide opportunities to the multicultural and diverse NLM workforce to enhance and advance their careers and reach their fullest potential.

The NLM Diversity Council continued the coordination and management of the Director's Education Fund and, in 2012, the fund supported 41 employees in taking 67 courses. The majority of NLM staff who took advantage of this opportunity were enrolled in the University of Maryland (34), followed by Strayer University (7) and Catholic University and Capella University (4 respectively). Other institutions attended included: American University, Johns Hopkins University, Notre Dame College, Syracuse University, Kent State University, Walden University, Montgomery College, Columbia Union College, and Washington Adventist University. Course disciplines included: Health IT Standards and Systems Interoperability; Special Topics in Librarianship; Health Care Management; Public Programs; Doctoral Dissertation Seminars; Information Access Services; Marketing Management; Principles and Issues of Information Systems; Health Policy and Management; Image Processing; and History of Medicine among others. The Diversity Council continues to publicize the availability of the fund to NLM employees.

NLM 2012 Health Exposition

The NLM Diversity Council sponsored the fifth annual *Healthy Lifestyles for You and Your Family Expo* on April 26, 2012. NLM's educational family-oriented health expo focusing on healthy living styles for both adults and children was attended by over 1000 children and their parents.   Activities included Mad Science, innovative, hands-on *Cells* and *Taste   and   Smell* workshops   for children to encourage scientific literacy in children and *Treat   Your   Body   Right*, presented   by   the   Emmy

award-winning FoodPlay Productions, featuring live theatre and interactive media to promote family healthy eating and exercise. NLM sponsored the following events: *Children's Book Hospital*, about creating books, *Totally Toxic*, about environmental awareness, an interactive Internet video with Hawaii, and a Scavenger Hunt of

Native Voices: Native Peoples' Concepts of Health and Illness, currently on exhibition at NLM. The NLM Healthy Lifestyles expo has received positive recognition from parents, other NIH Institutes and the public in promoting healthy lifestyles and scientific education to children in their formative years.

Appendix 4: Regional Medical Libraries

1. **MIDDLE ATLANTIC REGION**
Middle Atlantic Region
University of Pittsburgh
Health Sciences Library System
200 Scaife Hall, 3550 Terrace Street
Pittsburgh, Pennsylvania 15261
Phone: (412) 648-2065 Fax: (412) 624-1515
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
Phone: (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
Phone: (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
Phone: (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
Phone: (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
Phone: (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
222 Maple Avenue
Shrewsbury, MA 01545
Phone: (508) 856-5979 Fax: (508) 856-5977
States served: CT, MA, ME, NH, RI, VT
*URL:*http://nnlm.gov/ner

Appendix 5: 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.

Chairperson

Tanji, Virginia, MSLS, MED
Director
Health Science Library
University of Hawaii at Manoa
School of Medicine
Honolulu, HI 96813-5534

Members

Evens, Ronald G., MD
Mallinckrodt Professor of Radiology
Washington University School of Medicine
Saint Louis, MO 63110

Fleming, David A., MD, MA, FACP
Professor and Chairman
Department of Internal Medicine
Director, MU Center for Health Ethics
University of Missouri School of Medicine
Columbia, MO 65212

Gottlieb, Katherine, MBA, DPS (h.c.)
President and CEO
Southcentral Foundation
Anchorage, AK 99508

Lewis, Henry, PharmD
President
Florida Memorial University
Miami Gardens, FL 33054

MacKay, Trudy, PhD
Distinguished University Professor of Genetics
Department of Genetics
North Carolina State University
Raleigh, NC 87695-7614

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 84112

Roskies, Ralph Z., PhD
Professor of Physics, University of Pittsburgh
Scientific Director, Pittsburgh Supercomputing
 Center
Pittsburgh, PA 15213

Ryan, Mary L., MLS, MPH
UMSAS Library Director/Professor
University of Arkansas for Medical Sciences
 Library
Little Rock, AK 72205-7186

Scutchfield, F. Douglas, MD
Peter P. Bosomworth Professor of Health
 Services Research and Policy
University of Kentucky College of Public Health
Lexington, KY 40536-0003

Appendix 6: Board of Scientific Counselors, Lister Hill Center for Biomedical Communications

The Board of Scientific Counselors (BSC) provides advice on NLM's intramural research and development programs for the Lister Hill Center for Biomedical Communications.

Chairperson

Ash, Joan S., PhD
Professor and Vice-Chair
Department of Medical Informatics and Clinical
 Epidemiology
Oregon Health & Science University
Portland, OR 97201

Members

Bakken, Suzanne, DNSc, RN, FAAN
Professor
Departments of Nursing and Biomedical
 Informatics
Columbia University
New York, NY 10032

Chapman, Wendy W., PhD
Associate Professor
Division of Biomedical Informatics
University of California, San Diego
La Jolla, CA 92093-0505

Chueh, Henry C., MD
Director and Chief
Laboratory of Computer Science
Division of Biomedical Informatics
Massachusetts General Hospital
Boston, MA 02114

Hammond, William E., PhD
Director
Duke Center for Health Informatics
Duke University
Durham, NC 27705

Horii, Steven C., MD
Professor of Radiology and Clinical Director
Medical Informatics Group
Department of Radiology
Hospital of the University of Pennsylvania
Philadelphia, PA 19104

Mandl, Kenneth D., MD
Associate Professor
Harvard Medical School
Director, Intelligent Health Laboratory
Children's Hospital Informatics Program
Children's Hospital Boston
Boston, MA 02115

Were, Martin C., MD
Assistant Professor of Medicine
Department of General Internal Medicine and
 Geriatrics
Indiana University School of Medicine
Indianapolis, IN 46202-3012

Appendix 7: 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.

<u>Chairperson</u>

Levine, Arthur S., MD
Senior Vice Chancellor for Health Sciences
Dean, School of Medicine
University of Pittsburgh
Pittsburgh, PA 15261

<u>Members</u>

Allewell, Norma M., PhD
Dean
Department of Chemistry and Biochemistry
University of Maryland
College Park, MD 20742

Babbitt, Patricia C., PhD
Professor
Department of Bioengineering & Therapeutic
 Sciences
University of California, San Francisco
San Francisco, CA 94158-2330

Benham, Craig J., PhD
Professor
Department of Biomedical Engineering
UC Davis Genome Center
University of California, Davis
Davis, CA 95616

Lee, Christopher J., PhD
Professor
Department of Chemistry & Biochemistry
Molecular Biology Institute
University of California, Los Angeles
Los Angeles, CA 90095-1570

Lynch, Michael R., PhD
Distinguished Professor
Department of Biology
Indiana University
Bloomington, In 47405

Seidman, Christine E., MD
T.W. Smith Professor of Medicine and Genetics
Harvard Medical School
Boston, MA 02115

Weng, Zhiping, PhD
Professor and Director
Program in Bioinformatics and Integrative
 Biology
University of Massachusetts Medical School
Worcester, MA 01605

Appendix 8: Biomedical Library and Informatics Review Committee

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

Chairperson
Cooper, Gregory F., MD, PhD
Professor
Department of Biomedical Informatics
University of Pittsburgh
Pittsburgh, PA 15260

Members

Alpi, Kristine M., MLS, MPH
Director & Adjunct Assistant Professor
Department of Population Heath &
 Pathobiology
Veterinary Medicine Library
North Carolina State University
Raleigh, NC 27607

Cardozo, Timothy J., MD, PhD
Associate Professor
Department of Biochemistry and Molecular
 Pharmacology
New York University School of Medicine
New York, NY 10016

Gonzalez, Graciela H., PhD
Assistant Professor
Department of Biomedical Informatics
Arizona State University
Scottsdale, AZ 85259

Gorman, Paul N., MD
Associate Professor
Medical Informatics & Clinical Epidemiology
Oregon Health & Science University
Portland, OR 97239

Hurdle, John F., MD, PhD
Associate Professor
Department of Biomedical Informatics
University of Utah School of Medicine
Salt Lake City, UT 84112

Kahn, Michael G., MD, PhD
Associate Professor
Department of Pediatrics
University of Colorado
The Children's Hospital
Aurora, CO 80045
Kalet, Ira J., PhD

Professor Emeritus
Radiation Oncology Department
Department of Biomedical Informatics and
 Medical Education
University of Washington
Seattle, WA 98195-6043

Kochi, Julia K., MILS
Director, Digital Library & Collections
Library & Center for Knowledge Management
University of California, San Francisco
San Francisco, CA 94143-0840

Lussier, Yves A., MD
Professor of Medicine and Engineering
Clinical Research Information Officer
Assistant Vice President of Health Affairs
The University of Illinois in Chicago
Chicago, IL 60637

Moore, Jason H., PhD
Professor of Genetics
Geisel School of Medicine
Dartmouth College
Lebanon, NH 03756

Murphy, Beverly, MLS
Assistant Director
Marketing and Publishing
Webmaster, DUMCL Online
Duke University Medical Center Library
Durham, NC 27710

Pestian, John P., PhD
Professor, Pediatrics
Director, Computational Medicine Center
Cincinnati Children's Hospital Medical Center
University of Cincinnati
Cincinnati, OH 45229

Shatkay, Hagit, PhD
Associate Professor
Department of Computer and Information
 Sciences
University of Delaware
Newark, DE 19716

Sittig, Dean F., PhD
Professor

UT Memorial Hermann Center for Healthcare
 Quality and Safety
The University of Texas School of Health
 Informatics Sciences at Houston
Houston, TX 77030

Srinivasan, Padmini, PhD
Professor
Computer Science
The University of Iowa
Iowa City, IA 52242-1479

Starren, Justin, MD, PhD
Chief, Division of Biomedical Informatics
Department of Preventive Medicine
Director, Northwestern University Biomedical
 Informatics Center

Feinberg School of Medicine, Northwestern
 University
Chicago, Il 60611

Subramaniam, Shankar, PhD
Professor and Chair
Department of Bioengineering
University of California, San Diego
La Jolla, CA 92093-04112

Veinot, Tiffany, CE, PhD
Assistant Professor
School of Information
University of Michigan
Ann Arbor, MI 48109-1285

Appendix 9:   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.

<u>Chairperson</u>

Christopher, Mary M., DVM, PhD
Professor of Pathology
Departments of Pathology, Microbiology &
 Immunology
School of Veterinary Medicine
University of California
Davis, CA 95606

<u>Members</u>

Cabello, Felipe C., MD
Professor
Dept. of Pathology, Microbiology & Immunology
New York Medical College
Valhalla, NY 10595

Copeland, Robert L., PhD
 Associate Professor
Department of Pharmacology
Howard University
Washington, DC 20059

Courtney, Karen L., PhD
Assistant Professor
School of Health Information Science
Victoria, BC, Canada, V8W 3P5

Crummett, Courtney, MS, MLS
Bioinformatics Librarian
Massachusetts Institute of Technology
Science Library
Cambridge, MA 02139

Gwinn, Marta, MD
Senior Consultant
Office of Public Health Economics
Centers for Disease Control and Prevention
Atlanta, GA 30333

Jackson, Gretchen P., MD, PhD
Assistant Professor
Surgery and Biomedical Informatics
Vanderbilt Children's Hospital
Nashville, TN 37232

Moreno, Carlos A., MD
Professor and Chairman

Department of Family and Community Medicine
University of Texas Health Science Center,
 Houston
Houston, TX 77030

Ogunyemi, Omolola, PhD
Director
Center for Biomedical Informatics
Charles Drew University of Medicine and
 Science
Lynwood, CA 90262

Pascoe, John M., PhD
Professor
Department of Pediatrics
Wright State University
Dayton, OH 45404

Phillips, William R., MD
Clinical Professor
Department of Family Medicine
University of Washington
Seattle, WA 98195

Philpott, Caroline C., MD
Chief, Genetics and Metabolism Section
Liver Diseases Branch, NIDDK
National Institutes of Health
Bethesda, MD 20892-1800

Smith, Paul D., MD
Associate Professor
Department of Family Medicine
University of Wisconsin Medical School
Madison, WI 53715-1896

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 52242-1098

Zhang, Ge, MD, PhD
Assistant Professor
Department of Biomedical Engineering
University of Akron
Akron, OH 44325-0302

Appendix 10: 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.

Chairperson

Ward, Gary E., PhD
Professor, Department of Microbiology & Molecular
 Genetics
University of Vermont
Burlington, VT 05405

Members

Anderson, Ivy, MLS
Director, Collection Development & Management
Program
California Digital Library
Oakland, CA 94612

Bird, Christopher J., BA
Solicitor, Legal Department
Wellcome Trust
London, England NW12BE
United Kingdom

Blanton, Ronald E., MD
Professor of Medicine
Center for Global Health and Diseases
Case Western Reserve University
Cleveland, OH 44106-7286

Bourne, Philip E., PhD
Professor
Department of Pharmacology
San Diego Supercomputing Center
University of California, San Diego
La Jolla, CA 92093-0505

Colamarino, Sophia A., PhD
Consulting Associate Professor
Department of Psychiatry and Behavioral
 Sciences
Stanford University Medical School
San Francisco, CA 94123

Courant, Paul N., PhD
University Librarian and Dean of Libraries
University of Michigan Library
University of Michigan
Ann Arbor, MI 48109-1205

Fassler, Jan S., PhD
Professor
Department of Biology
University of Iowa
Iowa City, IA 52242

Henderson, Cynthia L., MLS
Executive Director
Louis Stokes Health Science Library
Howard University
Washington, DC 20059

Kann, Maricel G., PhD
Assistant Professor
Department of Biological Sciences
University of Maryland
Baltimore, MD 21250

Meglio, Delores, MS
Vice President, Public Relations
Knovel Corporation
New York, NY 10017

Rossner, Mike, PhD
Executive Director
The Rockefeller University Press
New York, NY 10021

Tanner, R. Michael, PhD
Vice President and Chief Academic Officer
Association of Public and Land-Grant
 Universities
Washington, DC 20005-4722

Thibodeau, Patricia L., MLS, MBA
Associate Dean for Library Services & Archives
Medical Center Library
Duke University
Durham, NC 27710

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 78229-3900

Appendix 11: Organizational Acronyms and Initialisms Used in this Report

Acronym	Meaning of Acronym
AAHSL	Association of Academic Health Sciences Libraries
AABB	Non-profit association formerly known as American Association of Blood Banks
AAPA	American Academy of Physicians Assistants
ABC	Advanced Biomedical Tele-Collaboration Test Bed)
ACLA	American Clinical Laboratory Association
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
AHILA	Association for Health Information and Libraries in Africa
AHRQ	Agency for Healthcare Research and Quality
AIDSinfo	Acquired Immune Deficiency Syndromeinfo (database)
ALTBIB	Alternatives to Animal Testing
AIH	American Indian Health (Web portal)
AME	Automated Metadata Extraction
AMIA	American Medical Informatics Association
AMPA	American Medical Publishers Association
AMPATH	Academic Model Providing Access to Healthcare
AMWA	American Medical Women's Association
APDB	Audiovisual Program Development Branch
API	Applied Programming Interface
APIRE	American Psychiatric Institute for Research and Education
ARRA	American Recovery and Reinvestment Act
ASCCP	American Society for Cervical Pathology and Colposcopy
ASHG	American Society of Human Genetics
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

Acronym	Meaning of Acronym
BMT	Boundary Marking Tool
BN	Brand Name
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
CARe	Candidate Gene Association Resource project
CAS	Collection Access Section
CBB	Computational Biology Branch
CBIR	Content-Based Image Retrieval
CBRN	chemical, biological, radiological and nuclear (incidents)
CCB	Configuration Control Board
CCDS	Consensus CoDing Sequence
CCHIT	Commission for Healthcare Information Technology
CCR	Central Contractor Registration
CCRIS	Chemical Carcinogenesis Research Information System
CDART	Conserved Domain Architecture Retrieval Tool
CDD	Conserved Domain Database
cDNA	Complementary DNA
CEB	Communications Engineering Branch
CEL	Affymetrix Cell intensity (file)
CgSB	Cognitive Science Branch
CHEBI	Chemical Entities of Biological Interest
ChEMBL	Computational Chemical Biology Group database
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
CMAX	Collaborative Multi-Agency eXercise (BHEPP disaster drill)
CMS	Centers for Medicare and Medicaid Services
CMT	Convergent Medical Terminology
COOP	(NIH Pandemic Flu) Continuity of Operations Plan

Acronym	Meaning of Acronym	Acronym	Meaning of Acronym
CORE	Clinical Observations Recording and Encoding	DLXS	Digital Library Extension Service
CoreBio	Core Bioinformatics Facility	DMSZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen (German Collection of Microorganisms and Cell Cultures)
CounterACT	Countermeasures Against Chemical Threats		
CPS	Commercial Peering Service	DNA	Deoxyribonucleic Acid
CPSC	Center for Public Service Communication	DOT	Department of Transportation
CPT	Current Procedural Terminology	DPR	Digital Preservation Research
CRAC	Computer Room Air Conditioner	DRAGON	Dynamic Resource Allocation in GMPLS Optical Networks
CRAH	Computer Room Air Handler		
CRD	Centre for Reviews and Dissemination (England)	DRESWG	Digital Repository Evaluation and Selection Working Group
CRI	Clinical Research Informatics	DRIG	Digital Repository Implementation Group
CRISP	Computer Retrieval of Information on Scientific Projects	DTD	Document Type Definition
		DVTS	Digital Video Transport System
CSB	Computer Science Branch	EAI	Emergency Access Initiative
CSI	Commission on Systemic Interoperability	EBI	European Bioinformatics Institute
CSIRC	Computer Security Incident Response Center (HHS)	EBP	Evidence-Based Practice
		ECHO	European Community Humanitarian Office
CSR	Center for Scientific Review	Educollab	Educational Collaborators
CT	Computer Tomography	EEO	Equal Employment Opportunity
CTD	Clinical Text De-identification	EFTS	Electronic Funds Transfer Service
CTD	Comparative Toxicogenomics Database	EHR	Electronic Health Record
CTS	Communications Technology Satellite	EMBL	European Molecular Biology Laboratory
CTSA	(NIH Roadmap) Clinical Translational Science Award Centers	EMR	Electronic Medical Record
		EMS	Emergency Medical Services
CUIs	Concept Unique Identifiers	EnHIP	Environmental Health Information Partnership
CWDM	Coarse Wave Division Multiplexing		
DAC	Data Access Committees	EnHIOP	Environmental Health Information Outreach Program
DAR	Data Access Request		
DARE	Database of Reviews of Effects	EP	Extramural Programs
DART/ETIC	Developmental and Reproductive Toxicology/Environmental Teratology Information	EPA	Environmental Protection Agency
		eQTL	expression quantitative trait loci
		eRA	Electronic Research Administration
DBA	Data Base Administrator	ERG	Emergency Response Guidebook
dbEST	Database of Expressed Sequence Tags Center	ESI	Early Stage Investigators
		EST	Expressed Sequence Tag
dbGaP	Database of Genotypes and Phenotypes	ETIC	Environmental Teratology Information Center
dbMHC	Database for the Major Histocompatibility Complex		
		eTK	Electronic Thorndike and Kibre
dbRBC	Database of Red Blood Cells	EUREKA	Exceptional, Unconventional Research Enabling Knowledge Acceleration
dbSNP	Database of Single Nucleotide Polymorphism		
		E-Utilities	Entrez Programming Utilities
DCMS	Data Creation and Maintenance System	eVK	Electronic Voights and Kurtz
DDBJ	DNA Data Bank of Japan	FAES	Foundation for Advanced Education in the Sciences
DDD	Drug Delivery Devices		
DDoS	Distributed Denial of Service (attack)	FDA	Food and Drug Admnistration
DEAS	Division of Extramural Administrative Support	FDCC	Federal Desktop Core Configuration
		FHA	Federal Health Architecture
DELTA-BLAST	Domain Enhanced Lookup Time Accelerated BLAST	FIC	Fogarty International Center
		FISMA	Federal Information Security Management Act
DHHS	Department of Health and Human Services		
DHS	Department of Homeland Security	FMS	Facilities Management Section
DICOM	Digital Imaging and Communications in Medicine	FNLM	Friends of the National Library of Medicine
		FTE	Full Time Employee
DIMRC	Disaster Information Management Research Center	FTP	File Transfer Protocol
		GAIN	Genetic Association Information Network
DIRLINE	Directory of Information Resources Online	Gbps	Gigabits per Second
		GCMS	Global Citation Management System

Acronym	Meaning of Acronym	Acronym	Meaning of Acronym
GDP	Genome Decoration Page	ICCVAM	Interagency Coordinating Committee on The Validation of Alternative Methods
GDS	GEO DataSet	ICD	International Classification of Diseases
GEO	Gene Expression Omnibus (database)	ICMJE	International Committee of Medical Journal Editors
GENSAT	Gene Expression Nervous System Atlas		
geneRIF	Gene Reference Into Function	ICs	Institutes and Centers (of NIH)
GENE-TOX	Genetic Toxicology	ICT	Information and Communication Technologies
GHR	Genetics Home Reference		
GIA	Gene Indexing Assistant	IDE	Integrated Development Environment
GIS	Geographic Information System	IDS	Intrusion Detection System
GO	Grand Opportunity grant	IE8	Internet Explorer 8
GO-ESP	Grand Opportunities-Exon Sequencing Project	IEB	Information Engineering Branch
		IGS	Intergenic Spacer
GMAC	Grants Management Advisory Committee	IGSTK	Image Guided Surgery Toolkit
GPS	Global Position System	IHTSDO	International Health Terminology Standards Development Organization
GPU	Graphics Processing Unit		
GRC	Genome Reference Consortium	IHM	Images from the History of Medicine
GRMS	Global Records Management System	ILL	Interlibrary Loan
GSA	General Services Administration	ILS	Integrated Library System
GSS	Genome Survey Sequences	IMPAC	Information Management Planning Analysis And Coordination
GTR	Genetic Testing Registry		
GUI	Graphic User Interface	InCHIs	IUPAC International Identifiers
GWAS	Genome Wide Association Studies	INDSC	International Nucleotide Sequence Database Collaboration (formerly DDBJ/EMBL/GenBank)
HapMap	Haplotype Map		
HAVnet	Haptic Audio Video Network for Education Technology		
		infoSIDA	infoSíndrome de Inmunodeficiencia Adquirida (database)
HBCU	Historically Black Colleges and Universities		
		IP	Interactive Publications
HD	High Definition	IPv6	Next Generation Internet, Version 6
HGVS	Human Genome Variation Society	IRB	Institutional Review Board
HHS	Health and Human Services	IRC	In-Row Coolers
HIPAA	Health Insurance Portability and Accounting Act	IRIS	Integrated Risk Information System
		IRMA	Image Retrieval for Medical Applications
HITSP	Healthcare Information Technology Standards Panel	ISO	International Organization for Standardization
HLA	Human Leukocyte Antigen		
HL7	Health Leven Seven, Inc.	ISTO	Image Storage and Transmission Optimization
HMD	History of Medicine Division		
HSDB	Hazardous Substances Data Bank	IT	Information Technology
HPCC	High Performance Computing and Communications	ITP	Informatics Training Program
		ITER	International Toxicity Estimates for Risk
HPV	Human Papillomavirus	ITK	Insight Toolkit
HRSA	Health Resources and Services Administration	ITP	Informatics Training Program
		ITS	Internal Transcribed Space
HSRIC	HRS (Health Services Research) Information Central	ITSC	Information Technology Service Center
		ITSMS	Information Technology Service Management System
HRSInfo	Health Services Research Information		
HSRProj	Health Services Research Projects	IUPAC	International Union of Pure and Applied Chemistry
HSRR	Health Services and Sciences Research Resources		
		JDBC	Java Database Connectivity
HSTAT	Health Services and Technology Assessment Text	JDI	Journal Descriptor Indexing
		JDMS	Journal Descriptor Maintenance System
HTTP	Hypertext Transfer Protocol	JRE	Java Runtime Environment
HuGENet	Human Genome Epidemiology Network	KEGG	Kyoto Encyclopedia of Genes and Genomes
I3	Image Indexing Initiative		
IAIMS	Integrated Advanced Information Management Systems	KSS	Knowledge Source Server (data)
		LactMed	Drugs and Lactation (database)
IBIS	Inferred Biomolecular Interactions Server	LAN	Local Area Network
ICC	Incident Command Center	LC	Library of Congress

Acronym	Meaning of Acronym	Acronym	Meaning of Acronym
LHI	Leading Health Indicators (HHS)	MTI	Medical Text Indexer
LHNCBC	Lister Hill National Center for Biomedical Communications	MTMS	MeSH Translation Management System
		NAC	Network Access Control
LID	Laboratory for Informatics Development	NA-MIC	National Alliance of Medical Image Computing
LIPID MAPS	Lipid Metabolites and Pathways Strategy	NAML	Network of African Medical Libraries
LITE	Librarian Infobutton Tailoring Environment	NAS	National Academy of Sciences
LJI	List of Journals Indexed	NASA	National Aeronautics and Space Administration
LO	Library Operations		
LOINC	Logical Observations Identifiers, Names, Codes	NCBC	National Center for Biomedical Computing
LPF	Lost Person Finder	NCBI	National Center for Biotechnology Information
LRG	Locus Reference Genomic		
LRP	Long Range Plan (NLM)	NCCS	NIHI Consolidated Collocation Site
LSD	Lysosomal Storage Disorders	NCHS	National Center for Health Statistics
LSI	List of Serials Indexed	NCI	National Cancer Institute
LSTRC	Literature Selection Technical Review Committee	NCRR	National Center for Research Resources
		NCVHS	National Committee on Vital and Health Statistics
LVG	Lexical Variant Generator		
LWS	Lifecycle Work Station	NDC	National Data Codes
MARC	Machine- Readable Cataloging	NDF-RT	National Drug File – Reference Terminology
MARG	Medical Article Records Groundtruth		
MARS	Medical Article Records System	N*e*HC	National *e*-Health Collaborative
MAX	Mid Atlantic Exchange, U. of Maryland	NEI	National Eye Institute
MCI	Mass Casualty Incident	NGI	Next Generation Internet
MDoT	MEDLINE Database on Tap	NHANES	National Health and Nutrition Examination Surveys
MDT	Multimedia Database Tool		
MEDLARS	Medical Literature Analysis and Retrieval System	NHGRI	National Human Genome Research Institute
MEDLINE	MEDLARS Online	NHIN	National Health Information Network
MegaBLAST	Basic Local Alignment Search Tool	NHLBI	National Heart. Lung, and Blood Institute
MEME	Metathesaurus Editing and Maintenance Environment	NIA	National Institute on Aging
		NIAID	National Institute of Allergy and Infectious Diseases
MEO	Medical Education and Outreach		
MEPI	Medical Education Partnership Initiative (Africa)	NIBIB	National Institute of Biomedical Imaging and Bioengineering
MeSH	Medical Subject Headings	NICHD	National Institute of Child Health and Human Development
MHC	Major Histocompatibility Complex		
MHL	Medical Heritage Library	NICHSR	National Information Center on Health Services Research and Health Care Technology
MID	Manuscript Identifiers		
MICAD	Molecular Imaging and Contract Database		
MIM	Mentoring In Medicine	NIDCD	National Institute on Deafness and other Communication Disorders
MIM	Multilateral Initiative on Malaria		
MIMCom	MIM Communications Working Group	NIDCR	National Institute of Dental and Cranio-facial Research
MIN	Multiple Ingredient (term type), RxNorm		
MIRS	Medical Information Retrieval System	NIDDK	National Institute of Diabetes, Digestive, and Kidney Diseases
MLA	Medical Library Association		
MLAA	Medical Library Assistance Act	NIEHS	National Institute of Environmental Health Sciences
MLB	Medical Language Branch (database server)		
		NIGMS	National Institute of General Medical Sciences
MLP	Molecular Libraries Program (at NIH)		
MMDB	Molecular Modeling DataBase	NIH	National Institutes of Health
MMS	MEDLARS Management Section	NIHMS	NIH Manuscript Submission
MMTx	MetMap Technology Transfer	NIH PI	NIH Pathways to Independence Award
MOR	Medical Ontology Research	NIMH	National Institute of Mental Health
MOU	Memorandum of Understanding	NIMHD	National Institute on Minority Health and Health Disparities
mRNA	Messenger Ribonucleic Acid		
MS	Mass Spectrometry	NINDS	National Institute of Neurological Disorders and Stoke
MTHSPL	Metathesaurus Structured Product Labels		

Acronym	Meaning of Acronym	Acronym	Meaning of Acronym
NIOSH	National Institute for Occupational Safety and Health	PAHO	Pan American Health Organization
NIST	National Institute of Standards and Technology	PAL	Potential Abstract Labels (tool)
		PBM	Pharmacy Benefit Manager
NLM	National Library of Medicine	PCA	Personal Computer Advisory Committee
NLM LitArch	NLM Literature Archive	PCR	Polymerase Chain Reaction
NLP	National Language Processing System	PDA	Personal Digital Assistant
NN/LM	National Network of Libraries of Medicine	PDRS	Publisher Data Review System
		PDB	Protein Data Bank
NNMC	National Naval Medical Center	PDF	Portable Document Format
NNO	National Network Office	PDL	Personal Digital Library
NOAA	National Oceanic and Atmospheric Administration	PDM	Patient Data Management
		PDQ	Physician Data Query
NOSC	Network Operations and Security Center	PFIF	Person Finder Interchange Format
NOVA	National Online Volumetric Archive	PheGenI	Phenotype-Genotype Integrator
NPL	National Priorities List (as for Superfund)	PHLIP	Public Health Law Information Project
NQF	National Quality Forum	PHII	Public Health Informatics Institute
NRCBL	National Reference Center for Bioethics Literature	PHP	Public Health Partners
		PHR	Personal Health Record
NSF	National Science Foundation	PHS	Public Health Service
NTCC	National Online Training Center and Clearinghouse	PI	Pathway to Independence award
		PI	Principal Investigator
OA	Open Access	PII	Personally Identifiable Information
OACF	Onsite Alternate Computing Facility	PICO	Patient/Population, Intervention, Comparison, and Outcome
OAM	Office of Administrative Management		
OARF	Outreach Activity Reporting System	PID	Pathway Interaction Database (NCI)
OCCS	Office of Computers and Communications Systems	PIN	Precise Ingredient (term type), RxNorm
		PIV	Personal Identify Verification
OCHD	Coordinating Committee on Outreach, Consumer Health and Health Disparities	PL	Person Locator
		PLAWARe	Programmable Layered Architecture With Artistic Rendering
OCIO	Office of the Chief Information Officer (NIH)		
		PMC	PubMed Central
OCPL	Office of Communication and Public Liaison	PMCI	PubMed Central International
		PMC ID	PubMed Central Identification (number)
OCR	Optical Character Recognition	PRS	Protocol Registration System
OD	Office of the Director	PSD	Public Services Division
ODIMRC	Office of the Disaster Information Management Research Center	PubMedHh	PubMed for Handhelds
		PUE	Power Use Efficiency
OERC	Outreach Evaluation Resource Center	PUG	PubChem Power User Gateway
ORF	Original Release Format	QA	Quality Assurance
OHIPD	Office of Health Information Programs Development	QCIM	Quarterly Cumulative Index Medicus
		RAC	Real Application Clusters
OMB	Office of Management and Budget	RCDC	Research Condition and Disease Categorization
OMIA	Online Inheritance in Animals (database)		
OMIM	Online Mendelian Inheritance in Man (database)	RCSB	Research Collaboratory for Structural Bioinformatics
OMSSA	Open Mass Spectrometry Search Algorithm	RDMS	Rare Disease Maintenance System
		RefSeq	Reference Sequence (database)
ONC	Office of National Coordinator (for Health Information Technology)	RefSNP	Reference SNP (database)
		RELACIGER	Red Latinoamericana de Centros de Información en Gestión del Riesgo de Desastres (Latin American Network of Risk Management Centers)
OPASI	Office of Portfolio Analysis and Strategic Initiatives		
OPD	OLDMEDLINE Serials Application		
OPD	Outreach Products Database	REMM	Radiation Even Medical Management
OSA	Optical Society of America	RF2	Release Format version 2
ORWH	Office of Research on Women's Health	RFA	Request for Applications
OSIRIS	Open Source Independent Review and Interpretation System	RFID	Radio Frequency Identification
		RFP	Request for Proposals

Acronym	Meaning of Acronym	Acronym	Meaning of Acronym
RHIN	Refugee Health Information Network	STEM	Science, Technology, Engineering and Math
RIDeM	Repository for Informed Decision Making	STR	Short Tandem Repeat
RML	Regional Medical Library	STTP	Short-Term Trainee Program
RNA	Ribonucleic Acid	STTR	Small Business Technology Transfer Research
RNAi	RNA Interference		
RPS-BLAST	Reversed Position Specific BLAST	STS	Sequence Tagged Site
RQS	Request Submission and Tracking System	SVM	Support Vector Machine
RRF	Rich Release Format	TA	Title Abbreviation(s)
RSS	Really Simple Syndication	TBL	The bottom line
RTECS	Registry of Toxic Effects of Chemical Substances	TDI	3D Informatics (Group)
		TEHIP	Toxicology and Environmental Health Information Program
RTLS	Real Time Location System		
RVDS	Remote Virtual Dialogue System	TERA	Toxicology Excellence for Risk Assessment
RWJF	Robert Wood Johnson Foundation		
SAB	Source Abbreviations	TIC	Trusted Internet Connection
SBIR	Small Business Innovation Research	TICAP	Trusted Internet Connection Access Partners
SCID	Severe Combined Immunodeficiency		
SCR	(MeSH) Supplemental Chemical Records	TIE	Telemedicine Information Exchange
SDK	Software Development Kit	TIFF	Tagged Image File Format
SEF	Serials Extract File	TKMT	Traditional Korean Medical Terms
SEIM	Security Event and Information Management System	TILE	Text to Image Linking Engine
		TIOP	Toxicology Information Outreach Project
SEO	Search Engine Optimization	TOXLINE	Toxicology Information Online
SEPs	Special Emphasis Panels	TOXNET	Toxicology Data Network
SEQ	Structured Evidence Queries	TPA	Third Party Annotation (database)
SIDA	Swedish International Development Agency	TREF	Terminology Representation and Exchange Format
SIG	Special Interest Group		
SII	Scalable Information Infrastructure	TRI	The Toxics Release Inventory
SIS	Specialized Information Services	TSA	Transcriptome Shotgun Assembly
SKR	Semantic Knowledge Representation	TSD	Technical Services Division
SMART	Scalable Medical Alert and Response Technology	TT	Teaching Tool
		TTP	Turning the Pages
SNOMEDCT	Systematized Nomenclature of Medicine Clinical Terms	UCUM	Unified Codes for Units of Measure
		UID	Unique Identifier (PubMed)
SO	Signing Official	UIMA	Unstructured Information Management Architecture
SOAP	Simple Object Oriented Protocol (formerly Simple Object Access Protocol)		
		UKPMC	United Kingdom PubMed Central
SOC	(HHS) Secretary's Operation Center	UMLS	Unified Medical Language System
SPER	System for the Preservation of Electronic Resources	UMLSKS	UMLS Knowledge Source Server
		UN	United Nations
SPIN	Shared Pathology Informatics Network	UPS	Uninterrupted Power Supply
SPIRS	Spine Pathology Image Retrieval System	USHIK	United States Health Information Knowledgebase (team)
SPL	Structured Product Labels (FDA)		
SPWG	Special Projects Workgroup (NIH)	UTS	UMLS Terminology Services
SRA	Short Read Archive	VSAC	Value Set Authority Center
STB	Systems Technology Branch	WGS	Whole Genome Shotgun
		YEP	Year End Processing

Cover:

Spirit of Eagles

Chholing Taha (Cree Indian) 2010

The Artist's Interpretation:

The four colors of man at the top are expressed in the form of pictograph thunderbirds or eagles, honoring the wisdom of all the Elders from all nations. The blue flowers in the border illustrate the many plant medicines Creator has left for man to heal the community, family and person. The light blue thunderbirds/eagles represent the positive influence the Spirit of the Ancestors (wisdom from experience) has from infancy to adulthood. The group of wigwams at the bottom speaks of the importance of community commitment and collaboration, the foundation for holding all people upright for healthy living.

National Library of Medicine

OFFICE OF THE DIRECTOR [1]
Dr. Donald A. B. Lindberg
DIRECTOR
Betsy Humphreys
DEPUTY DIRECTOR
Dr. Milton Corn
DEPUTY DIRECTOR
for Research & Education

BOARD OF REGENTS

OFFICE OF ADMINISTRATION
Todd Danielson

OFFICE OF COMMUNICATIONS & PUBLIC LIAISON [2]
Kathleen Cravedi

OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT [3]
Dr. Michael F. Huerta

OFFICE OF COMPUTER & COMMUNICATIONS SYSTEMS
Ivor D'Souza
- MEDICAL LANGUAGE BRANCH — **Suresh Srinivasan**
- SYSTEMS TECHNOLOGY BRANCH — **Ivor D'Souza (Acting)**
- APPLICATIONS BRANCH — **Wei Ma**

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 — **Janice Kelly**
- OFFICE OF THE DISASTER INFORMATION MANAGEMENT RESEARCH CENTER — **Stacey Arnesen**

DIVISION OF EXTRAMURAL PROGRAMS
Dr. Valerie Florance
- BIOMEDICAL LIBRARY & INFORMATICS REVIEW COMMITTEE

DIVISION OF LIBRARY OPERATIONS
Sheldon Kotzin
- National Network of Libraries of Medicine — **Dr. Angela Ruffin**
- Medical Subject Headings Section — **Dr. Stuart Nelson**
- National Information Center on Health Services Research & Health Care Technology — **Lisa Lang**
- PUBLIC SERVICES DIVISION — **Martha Fishel**
- BIBLIOGRAPHIC SERVICES DIVISION — **David Gillikin**
- TECHNICAL SERVICES DIVISION — **Jennifer Marill**
- HISTORY OF MEDICINE DIVISION — **Dr. Jeffrey S. Reznick**
- LITERATURE SELECTION TECHNICAL REVIEW COMMITTEE

LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS
Dr. Clement J. McDonald
- CLINICAL TRIALS — **Dr. Deborah Zarin**
- COMMUNICATIONS ENGINEERING BRANCH — **Dr. George Thoma**
- COMPUTER SCIENCE BRANCH — **Dr. George Thoma (Acting)**
- AUDIOVISUAL PROGRAM DEVELOPMENT BRANCH — **Anne Altemus (Acting)**
- COGNITIVE SCIENCE BRANCH — **Dr. Olivier Bodenreider**
- OFFICE OF HIGH PERFORMANCE COMPUTING & COMMUNICATIONS — **Dr. Michael J. Ackerman**
- LHNCBC 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**
- NCBI BOARD OF SCIENTIFIC COUNSELORS
- PUBMED CENTRAL NATIONAL ADVISORY COMMITTEE

1. Includes:
Assistant Director for Policy and Legislative Development – Jerry Sheehan
Assistant Director for High Performance Computing and Communications – (Vacant)
Assistant Director for Health Services Research Information – Dr. Michael J. Ackerman
Assistant Director for Applied Informatics – Lisa Lang

2. Includes Public Liaison Officer – Melanie Modlin

3. Includes International Programs

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

FY2012

Chart 1: NLM Organization Chart